

# Mode testing, critical bandwidth and excess mass

Jose Ameijeiras–Alonso

and

Rosa M. Crujeiras

and

Alberto Rodríguez–Casal\*

Department of Statistics, Mathematical Analysis and Optimization

Universidade de Santiago de Compostela

April 26, 2017

## Abstract

The identification of peaks or maxima in probability densities, by mode testing or bump hunting, has become an important problem in applied fields. This task has been approached in the statistical literature from different perspectives, with the proposal of testing procedures which are based on kernel density estimators or on the quantification of excess mass. However, none of the existing proposals provides a satisfactory performance in practice. In this work, a new procedure which combines the previous approaches (smoothing and excess mass) is presented and compared with the existing methods, showing a superior behaviour. A real data example on philatelic data is also included for illustration purposes.

*Keywords:* Bootstrap calibration; multimodality; testing procedure; philately.

---

\*The authors gratefully acknowledge the support of Projects MTM2016–76969–P (Spanish State Research Agency, AEI) and MTM2013–41383–P (Spanish Ministry of Economy, Industry and Competitiveness), both co-funded by the European Regional Development Fund (ERDF), IAP network from Belgian Science Policy. Work of J. Ameijeiras-Alonso has been supported by the predoctoral grant BES–2014–071006 from the Spanish Ministry of Economy, Industry and Competitiveness.

# 1 Introduction

Simple distribution models, such as the Gaussian density, may fail to capture the stochastic underlying structure driving certain mechanism in applied sciences. Complex measurements in geology, neurology, economics, ecology or astronomy exhibit some characteristics that cannot be reflected by unimodal densities. In addition, the identification of the (unknown) number of *peaks* or modes is quite common in these fields. Some examples include the study of the percentage of silica in chondrite meteors (Good and Gaskins, 1980), the analysis of the macaques neurons when performing an attention-demanding task (Mitchell et al., 2007), the distribution of household incomes of the United Kingdom (Marron and Schmitz, 1992), the study of the body-size in endangered fishes (Olden et al., 2007) or the analysis of the velocity at which galaxies are moving away from ours (Roeder, 1990). In all these examples, identifying the number (and location) of local maxima of the density function (i.e. modes) is important *per se*, or as a previous step for applying other procedures.

An illustrative example which has been extensively considered in mode testing literature can be found in philately (the study of stamps and postal history and other related items). Research in this field has been motivated by the use of stamps for investment purposes. Before the 1940 decade, stamps images were printed on a variety of paper types and, in general, with a lack of quality control in manufactured paper, which led to important fluctuations in paper thickness, being thin stamps more likely to be produced than thick ones. Given that the price of any stamp depends on its scarcity, the thickness of the paper is crucial for determining its value. However, there is not a standard rule for classifying stamps according to their thickness (not being available such a classification in catalogues), becoming this problem even harder in stamp issues printed on a mixture of paper types with

possible differences in their thickness. A stamp issue with this problematic is the Hidalgo catalogue, from the nineteenth century, and this particular example has been shown in several references in the literature as a paradigm of the problem of determining the number of modes/groups. This example will be revisited in this work, recalling previous analysis and comparing results with the ones provided by the new testing procedure presented in this paper.

A formal hypothesis test for a null hypothesis of a certain number of modes can be stated as follows. Let  $f$  be the density function of a real random variable  $X$  and denote by  $j$  the number of modes. For  $k \in \mathbb{Z}^+$ , the testing problem on the number of modes can be formulated as:

$$H_0 : j = k \quad \text{vs.} \quad H_a : j > k. \quad (1)$$

There have been quite a few proposals in the statistical literature for solving (1) and the different techniques can be classified in two groups: a first group of tests based on or using a critical bandwidth, introduced by Silverman (1981), further studied by Hall and York (2001) and also used by Fisher and Marron (2001); and a second group of tests based on the *excess mass*, such as those ones proposed by Hartigan and Hartigan (1985), Müller and Sawitzki (1991) and Cheng and Hall (1998). These methods are briefly revised and compared in this paper, where a new proposal gathering strength from both areas is also introduced, outperforming the existing procedures, in testing unimodality and more general hypotheses.

Apart from the formal testing procedures, and as a complementary tool for them, a first step when confronting the problem of identifying modes in a data distribution is the exploration of a nonparametric estimator of the underlying probability density, which can be done by kernel methods. Classical kernel density estimation (Wand and Jones, 1995, Ch. 2) allows for the reconstruction of the data density structure without imposing parametric

restrictions (only subjected to mild regularity assumptions) but at the expense of choosing an appropriate bandwidth parameter, which controls the degree of smoothing. A direct observation of a kernel estimator may lead to inaccurate or even wrong conclusions about the mode density structure. This can be easily noticed from the plots shown in Figure 1, where the kernel density estimator for the stamp dataset with different bandwidths is shown. Based on this estimator, from an exploratory perspective, there are several alternatives for identifying modes such as the SiZer map (Chaudhuri and Marron, 1999), the mode tree and the random forest (Minnotte and Scott, 1993; Minnotte et al., 1998). Although these tools are helpful in supporting the results of formal testing procedures on the number and location of modes, apart from giving some insight on the global mode structure, the interpretation of the outputs from these procedures requires an *expert eye*.

This paper presents a new testing procedure combining the use of a critical bandwidth and an excess mass statistic, which can be applied to solve (1) in a general setting. In Section 2, a review on mode testing methods is presented, considering both tests based on critical bandwidth and on excess mass, jointly with the new proposal. A simulation study comparing all the procedures, in terms of empirical size and power, is included in Section 3. Section 4 is devoted to data analysis, revising the stamp dataset and presenting new results. Some final comments and discussion are given in Section 5. Details on the simulated models, a modification of the proposal when the compact support of the variable is known, a more flexible testing scenario and the computation of the proposed test are provided in the Appendix.

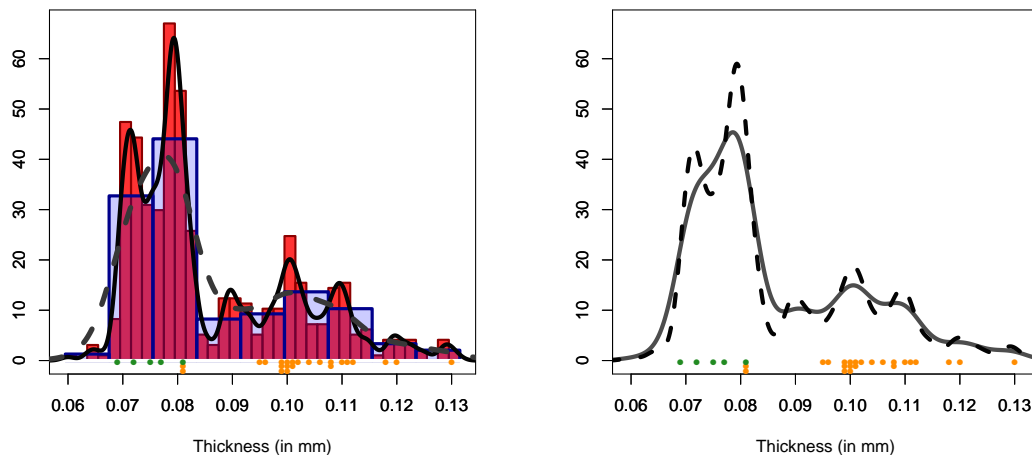


Figure 1: Sample of 485 stamps from the 1872 Hidalgo Issue of Mexico. Points: stamps watermarked with *LA+-F* (green) and *Papel sellado* (orange). Kernel density estimators with gaussian kernel and different bandwidths; left panel:  $h = 0.003910$  (rule of thumb -solid line-) and  $h = 0.001205$  (plug-in rule -dashed line-, see Wand and Jones, 1995, Ch. 3); right panel: critical bandwidths  $h_4 = 0.002831$  (solid line) and  $h_7 = 0.001487$  (dashed line). Left: histograms with different bin widths (0.002 -red- and 0.008 -blue-).

## 2 A review on multimodality tests

Different proposals for multimodality tests will be briefly revised in this section. Section 2.1 includes a review on the methods using the critical bandwidth, and excess mass approaches are detailed in Section 2.2. A new proposal, borrowing strength from both alternatives, is presented in Section 2.3: an excess mass statistic will be calibrated from a modified nonparametric kernel density estimator using a critical bandwidth.

## 2.1 Tests based on the critical bandwidth

For a certain number of modes  $k \in \mathbb{Z}^+$ , the critical bandwidth (Silverman, 1981) is the smallest bandwidth such that the kernel density estimator has at most  $k$  modes:

$$h_k = \min\{h : \hat{f}_h \text{ has at most } k \text{ modes}\},$$

where  $\hat{f}_h$  denotes the kernel density estimator, computed from a random sample  $\mathcal{X} = (X_1, \dots, X_n)$ , with kernel  $K$  and bandwidth  $h$ :

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right). \quad (2)$$

Silverman (1981) proposed to use the critical bandwidth with the gaussian kernel as a statistic to test  $H_0 : j \leq k$  vs.  $H_a : j > k$ , being its use justified by the fact that, with a gaussian kernel, the number of modes of  $\hat{f}_h$  is a nonincreasing function of  $h$ . Hence,  $H_0$  is rejected for large values of  $h_k$ , whose distribution is approached using bootstrap procedures. Specifically, the proposed methodology consists in generating  $B$  samples  $\mathcal{X}^{*b} = (X_1^{*b}, \dots, X_n^{*b})$  with  $b = 1, \dots, B$ , from  $\hat{f}_{h_k}$ . By computing the critical bandwidth,  $h_k^{*b}$ , from each sample  $\mathcal{X}^{*b}$ , given a significance level  $\alpha$ , the null hypothesis is rejected if  $\mathbb{P}(h_k^* \leq h_k | \mathcal{X}) \geq 1 - \alpha$ . When  $k = 1$ , the testing problem tackled by Silverman (1981) coincides with (1). However, for a general  $k$ , the null hypothesis in (1) is more restrictive than the one considered by Silverman (1981), but asymptotic consistency of the test is only derived for  $j = k$  (for a deeper insight see Hall and York, 2001, or Appendix C).

Hall and York (2001) proved that the previous bootstrap algorithm does not provide a consistent approximation of the test statistic distribution under the null hypothesis and suggested a way for accurate calibration when  $k = 1$ . Given a closed interval  $I$  where the null hypothesis is tested ( $f$  has a single mode in  $I$ ), if both the support of  $f$  and the interval  $I$  are unbounded then properties of  $h_1$  (critical bandwidth when  $k = 1$ ) are generally

determined by extreme values in the sample, not by the modes of  $f$ . To avoid this issue, the testing problem (1) is reformulated as follows:

$$H_0 : j = 1 \text{ in the interior of a given closed interval } I \text{ and no local minimum in } I, \quad (3)$$

and the critical bandwidth is redefined accordingly as:

$$h_{\text{HY}} = \min\{h : \hat{f}_h \text{ has exactly one mode in } I\}. \quad (4)$$

An issue that should be kept in mind in the computation of this critical bandwidth is that even if  $K$  is the gaussian kernel, the number of modes of  $\hat{f}_h$  inside  $I$  is not necessarily a monotone function of  $h$ . But under relatively general conditions (see Hall and York, 2001), the probability that number of modes is monotone in  $h$  converges to 1 for such a kernel. Hall and York (2001) proposed using  $h_{\text{HY}}$  as a statistic to test (3). Similarly to Silverman (1981) the null distribution of  $h_{\text{HY}}$  is approximated by bootstrap, generating bootstrap samples from  $\hat{f}_{\text{HY}}$ .

Unfortunately, the critical bandwidths for the bootstrap samples  $h_{\text{HY}}^{*b}$ , are smaller than  $h_{\text{HY}}$ , so for an  $\alpha$ -level test, a correction factor  $\lambda_\alpha$  to compute the p-value  $\mathbb{P}(h_{\text{HY}}^* \leq \lambda_\alpha h_{\text{HY}} | \mathcal{X}) \geq 1 - \alpha$  must be considered. Two different methods were suggested for computing this factor  $\lambda_\alpha$ , the first one based on a polynomial approximation and a second one using Monte Carlo techniques considering a simple unimodal distribution.

The previous proposal could be extended, as mentioned by Hall and York (2001), to test that  $f$  has exactly  $k$  modes in  $I$ , against the alternative that it has  $(k + 1)$  or more modes there, extending the critical bandwidth in (4) for  $k$  modes, namely  $h_{\text{HY},k}$ . Nevertheless, in this scenario, the bootstrap test cannot be directly calibrated under the hypothesis that  $f$  has  $k$  modes and  $(k - 1)$  antimodes, since it depends on the  $(2k - 2)$  unknowns  $(c_i/c_1)$ , where  $c_i = f^{1/5}(t_i)/|f''(t_i)|^{2/5}$  (assuming  $f''(t_i) \neq 0$  for all  $i$ ), and  $t_i$  being the ordered turning points of  $f$  in  $I$  with  $i = 1, \dots, (2k - 1)$ ; which notably complicates the computations.

Finally, it should be also commented that the use of the critical bandwidth for testing (1) is not limited to its use as a test statistic. Consider a Cramér–von Mises test statistic:

$$T = n \int_{-\infty}^{\infty} [F_n(x) - F_0(x)]^2 dF_0(x) = \sum_{i=1}^n \left( F_0(X_{(i)}) - \frac{2i-1}{2n} \right)^2 + \frac{1}{12n}. \quad (5)$$

where  $F_0$  is a given continuous distribution function,  $\{X_{(1)} \leq \dots \leq X_{(n)}\}$  denotes the ordered sample and  $F_n$  is the empirical distribution function. Fisher and Marron (2001) proposed the use of (5) for solving the general problem of testing  $k$  modes ( $H_0 : j \leq k$ ) by taking  $F_0(x) = \hat{F}_{h_k}(x) = \int_{-\infty}^x \hat{f}_{h_k}(t) dt$  and derived the statistic:

$$T_k = \sum_{i=1}^n \left( \hat{F}_{h_k}(X_{(i)}) - \frac{2i-1}{2n} \right)^2 + \frac{1}{12n}, \quad (6)$$

where the null hypothesis is rejected for large values of  $T_k$ . To approximate the distribution of the test statistic (6) under the null hypothesis, a bootstrap procedure is also proposed. It will be seen in Section 3 that the behaviour of the Fisher and Marron (2001) proposal is far from satisfactory.

## 2.2 Tests based on excess mass

Müller and Sawitzki (1991) confront the testing problem (1), employing a different perspective, under the following premise: for  $k \in \mathbb{Z}^+$ , a mode is present where an excess of probability mass is concentrated. Specifically, given a continuous real density function  $f$  and a constant  $\lambda$ , the excess mass is defined as:

$$E(f, \lambda) = \int_{C_\lambda} f(x) dx - \lambda ||C_\lambda||,$$

where  $C_\lambda = \{x : f(x) \geq \lambda\}$  and  $||C_\lambda||$  denotes the measure of  $C_\lambda$ . If  $f$  has  $k$  modes, the excess mass can be defined as:

$$E(f, \lambda) = \sup_{C_1, \dots, C_k} \left\{ \sum_{m=1}^k \mathbb{P}(C_m) - \lambda ||C_m|| \right\},$$



where the supremum is taken over all families  $\{C_m : m = 1, \dots, k\}$  of pairwise disjoint connected sets over the support of  $f$ . Under the assumption that  $f$  has  $k$  modes,  $E(f, \lambda)$  can be empirically estimated just replacing  $\mathbb{P}(C_m)$  by  $\mathbb{P}_n(C_m) = (1/n) \sum_{i=1}^n \mathcal{I}(X_i \in C_m)$ , being  $\mathcal{I}$  the indicator function. The difference  $D_{n,k+1}(\lambda) = E_{n,k+1}(\mathbb{P}_n, \lambda) - E_{n,k}(\mathbb{P}_n, \lambda)$  measures the plausibility of the null hypothesis, that is, large values of  $D_{n,k+1}(\lambda)$  would indicate that  $H_0$  is false. Using these differences, Müller and Sawitzki (1991) proposed the following test statistic:

$$\Delta_{n,k+1} = \max_{\lambda} \{D_{n,k+1}(\lambda)\}, \quad (7)$$

rejecting the null hypothesis that  $f$  has  $k$  modes for large values of  $\Delta_{n,k+1}$ . Müller and Sawitzki (1991) showed that this statistic is an extension of the *dip* test introduced by Hartigan and Hartigan (1985), just valid for the unimodal case, since both quantities (dip and excess mass) coincide up to a factor, for the unimodality case. In addition, the proposal of Müller and Sawitzki (1991) for testing unimodality is the same as that one of Hartigan and Hartigan (1985) and considers a Monte Carlo calibration, generating resamples from the uniform distribution.

In view of the extremely conservative behaviour of the calibration of the previous proposals (see Section 3 for results), Cheng and Hall (1998) designed a calibration procedure based on the following result: for large samples and under the hypothesis that  $f$  is unimodal, the distribution of  $\Delta_{n,2}$  is independent of unknowns except for a factor  $c = (f^3(x_0)/|f''(x_0)|)^{1/5}$ , where  $x_0$  denotes the unique mode of  $f$ . Using this fact, for the case  $k = 1$ , Cheng and Hall (1998) approximated the distribution of  $\Delta_{n,2}$  employing the values of  $\Delta_{n,2}^*$  obtained from the samples generated from a parametric calibration distribution  $\Psi(\cdot, \kappa)$ , being  $\kappa$  a certain parameter. Depending on the value of  $d = c^{-5}$ , different parametric distributions were suggested by the authors: a normal ( $d = 2\pi$ ), a beta ( $d < 2\pi$ ) or a rescaled Student

$t$  ( $d > 2\pi$ ). For estimating  $d$ , if  $\hat{x}_0$  denotes the largest mode of  $\hat{f}_h$ , then

$$\hat{d} = |\hat{f}_{h'}''(\hat{x}_0)|/\hat{f}_h^3(\hat{x}_0),$$

is used, where  $\hat{f}''$  and  $\hat{f}$  are kernel estimators with a gaussian kernel and  $h'$  and  $h$  are their respective asymptotically optimal global bandwidths, replacing the unknown quantities for the ones associated with a  $N(0, \hat{\sigma}^2)$ , being  $\hat{\sigma}^2$  the sample variance.

The methodology proposed by Cheng and Hall (1998) consists in generating samples from  $\Psi(\cdot, \hat{\kappa})$ , where  $\hat{\kappa}$  and the distribution family are chosen using  $\hat{d}$ . The excess mass statistic given in (7) when  $k = 1$ , that is  $\Delta_{n,2}^*$ , is computed from the resamples and, for a given significance level  $\alpha$ , the null hypothesis is rejected if  $\mathbb{P}(\Delta_{n,2}^* \leq \Delta_{n,2}|\mathcal{X}) \geq 1 - \alpha$ .

## 2.3 A new proposal

The previous tests show some limitations for practical applications: first, just the proposals of Silverman (1981) and Fisher and Marron (2001) allow to test (1) for  $k > 1$ . Despite the efforts of Cheng and Hall (1998) and Hall and York (2001) for providing good calibration algorithms, it will be shown in Section 3 that the behaviour of all the proposals is far from satisfactory. Specifically, the test presented by Silverman (1981) is very conservative in general (although sometimes can show the opposite behaviour) and the proposal of Fisher and Marron (2001) does not have a good level accuracy. The new method proposed in this work overcomes these drawbacks by considering an excess mass statistic, as the one proposed by Müller and Sawitzki (1991) with bootstrap calibration. Unlike Cheng and Hall (1998), a completely data-driven procedure will be designed, using the critical bandwidth under  $H_0 : j = k$ ,  $k \in \mathbb{Z}^+$ .

**The proposal, in a nutshell.** Consider then the testing problem (1) and take the excess mass statistic given in (7), under the null hypothesis. Given  $\mathcal{X}$ , generate  $B$  resamples

$\mathcal{X}^{*b}$  ( $b = 1, \dots, B$ ) of size  $n$  from (a modified version of)  $\hat{f}_{h_k}$ . For a significance level  $\alpha$ , the null hypothesis will be rejected if  $\mathbb{P}(\Delta_{n,k+1}^* \leq \Delta_{n,k+1} | \mathcal{X}) \geq 1 - \alpha$ , where  $\Delta_{n,k+1}^*$  is the excess mass statistic obtained from the generated samples. It should be also noted that the procedure can be easily adapted to handle Hall and York (2001) scenario: to test the null hypothesis that  $f$  has almost  $k$  modes in the interior of a given closed interval  $I$ , if  $I$  is known, use (a modified version of)  $\hat{f}_{h_{HY,k}}$  to generate the samples. From this brief description, two questions arise: How is this *modified* version of  $\hat{f}_{h_k}$  constructed? Does the procedure guarantee a correct calibration of the test? In fact, the construction of the modified kernel density estimator, namely the calibration function and subsequently denoted by  $g$ , is intended to ensure the correct calibration, under some regularity conditions.

**Regularity conditions** (RC1) The density function  $f$  has continuous derivative. (RC2) There exist  $t_1$  and  $t_2$ , such that  $f$  is monotone in  $(-\infty, t_1)$  and in  $(t_2, \infty)$ . (RC3) The only points satisfying  $\{x : f'(x) = 0 \text{ and } f(x) \neq 0\}$  are the modes and antimodes of  $f$ , denoted as  $x_i$ , with  $i = 1, \dots, (2j - 1)$ ; and  $f''(x_i) \neq 0$ . (RC4)  $f''$  exists and is Hölder continuous in a neighbourhood of each  $x_i$ .

Under these premises, Cheng and Hall (1998) indicated that the distribution of the test statistic (7) is independent from the underlying density  $f$ , except for the values in the modes and antimodes of  $d_i = |f''(x_i)|/f^3(x_i)$  with  $i = 1, \dots, (2j - 1)$ . It can be seen, assuming that  $f$  has  $k$  modes, that the distribution of  $\Delta_{n,k+1}$  can be approximated by  $\Delta_{n,k+1}^*$ , just with bootstrap values obtained from samples coming from a calibration distribution with  $k$  modes. This calibration distribution must satisfy that the estimated values  $\hat{d}_i$  converge in probability to  $d_i$ , as  $n \rightarrow \infty$ , for  $i = 1, \dots, (2k - 1)$ . As mentioned before, it is for that reason that Cheng and Hall (1998) proposed using a parametric family depending on the value of  $d$  when  $k = 1$ . Two issues appear related with their calibration procedure. First,

it is not an easy task to construct a parametric family having the desired values of  $d_i$ , for  $i = 1, \dots, (2k - 1)$ , when  $k > 1$ . In addition, the second-order limiting properties of the test depend on the form of the density function. Then, a better behaviour is expected if the calibration function is “more similar” to the real density function. Our method deals with these two issues to get a test having a good performance in the finite-sample case and allowing to solve the general problem of testing  $k$  modes.

In order to estimate the location of modes and antimodes, as well as the value of the density in these points under  $H_0$ , the kernel density estimator (2) with critical bandwidth  $\hat{f}_{h_k}$  is a good candidate. In addition, if  $f$  is also a bounded density with bounded support  $[a, b]^1$ , twice-differentiable in  $(a, b)$  and  $\lim_{x \rightarrow a^+} f'(x) > 0$ ,  $\lim_{x \rightarrow b^-} f'(x) < 0$ , the result in Mammen et al. (1992) holds and the critical bandwidth is of order  $n^{-1/5}$ , being then optimal for estimating  $f$ . Unfortunately, a suitable bandwidth for estimating  $f$  cannot be used for effectively estimating  $f''$ , being, the optimal bandwidth for estimating the second derivative of the density is of order  $n^{-1/9}$ . For this reason, the kernel density estimator with critical bandwidth  $h_k$  cannot be used as calibration function  $g$ , which needs to provide a good estimator of  $d_i$ . Under the conditions of Mammen et al. (1992), an appropriate estimator can be calculated using the ratio  $\hat{d}_i = |\hat{f}_{h_{PI}}''(\hat{x}_i)|/\hat{f}_{h_k}^3(\hat{x}_i)$ , where  $h_{PI}$  is a plug-in bandwidth. In this work, the plug-in rule for the second derivative will be obtained deriving the asymptotic mean integrated squared error and replacing  $f$  in its expression using a two-step procedure (see, for example, Wand and Jones, 1995, Ch. 3).

The calibration function  $g$  will be constructed by modifying  $\hat{f}_{h_k}$  in a neighbourhood of  $\{x : \hat{f}_{h_k}'(x) = 0\}$ , being such values a finite collection (see Silverman, 1981), having positive estimated density. The expression of this calibration function  $g$  is given in (8) and explained below. Depending on the nature of these points, two modifications in their neighbourhood

---

<sup>1</sup>When  $[a, b]$  is known, the critical bandwidth proposed by Hall and York (2001) can be used.

will be done. If the point  $\hat{x}_i$  is a mode or an antimode of  $\hat{f}_{h_k}$  then the modification will preserve its location and its estimated density value, and its second derivative will satisfy  $g''(\hat{x}_i) = \hat{f}_{h_{\text{PI}}}''(\hat{x}_i)^2$ . For  $i = 1, \dots, (2k - 1)$ , this will be the role of the function  $J$ , in the neighbourhood  $(\mathbf{r}_i, \mathbf{s}_i)$  of  $\hat{x}_i$  for guaranteeing  $|g''(\hat{x}_i)|/g^3(\hat{x}_i) = \hat{d}_i$  and satisfying (RC4). The second modification will remove the  $t$  saddle points of  $\hat{f}_{h_k}$ , denoted as  $\zeta_p$ , with  $p = \{1, \dots, t\}$ . In order to ensure that  $g$  satisfies (RC3), in a neighbourhood  $(z_{(2p-1)}, z_{(2p)})$  of  $\zeta_p$ , a function  $L$  will be introduced. Finally, as  $\hat{f}_{h_k}$  verifies (RC1) and (RC2), and the modifications of the functions  $J$  and  $L$  will be done in bounded neighbourhoods and preserving condition (RC1),  $g$  will satisfy also these two conditions. Then, the calibration function  $g$  will be constructed as follows:

$$g(x; h_k, h_{\text{PI}}, \boldsymbol{\varsigma}) = \begin{cases} J(x; \hat{x}_i, h_k, h_{\text{PI}}, \varsigma_i) & \text{if } x \in (\mathbf{r}_i, \mathbf{s}_i) \text{ for some } i \in \{1, \dots, (2k - 1)\}, \\ L(x; z_{(2p-1)}, z_{(2p)}, h_k) & \text{if } x \in (z_{(2p-1)}, z_{(2p)}) \text{ for some } p \in \{1, \dots, t\}, \\ & \text{and } \zeta_p \notin (\mathbf{r}_i, \mathbf{s}_i) \text{ for any } i \in \{1, \dots, (2k - 1)\}, \\ \hat{f}_{h_k}(x) & \text{otherwise,} \end{cases} \quad (8)$$

where  $\boldsymbol{\varsigma}$  has  $k$  components  $\varsigma_i \in (0, 1/2)$ , with  $i = 1, \dots, k$ , determining at which height of the kernel density estimation the modification is done. Values of  $\varsigma_i$  close to 0 imply a modification in an “small” neighbourhood around the mode or antimode. Note that a little abuse of notation was made as  $g$  will depend on the function  $\hat{f}_{h_k}$  (not only on  $h_k$ ) and on the values  $\hat{f}_{h_{\text{PI}}}''(\hat{x}_i)$ , for  $i = 1, \dots, (2k - 1)$ . An example of the effect of  $g$  can be seen in Figure 2 and it will be further described. Before defining functions  $J$  and  $L$ , to ensure that

---

<sup>2</sup>Note that, although asymptotically the sign of  $\hat{f}_{h_{\text{PI}}}''(\hat{x}_i)$  is always correct, in the finite-sample case, it may not be negative in the modes or positive in the antimodes. In that case, an abuse of notation will be done, denoting as  $h_{\text{PI}}$  to the critical or other plug-in bandwidth in order to guarantee that the sign of this second derivative remains correct.

$g$  has continuous derivative, a link function  $l$  must be introduced:

$$\begin{aligned} l(x; u, v, a_0, a_1, b_0, b_1) &= \frac{a_0 - a_1}{2} \left( 1 + 2 \left( \frac{x-u}{v-u} \right)^3 - 3 \left( \frac{x-u}{v-u} \right)^2 \right) \exp \left( \frac{2(x-u)b_0}{a_0 - a_1} \right) + \\ &+ \frac{a_0 - a_1}{2} \left( 2 \left( \frac{x-u}{v-u} \right)^3 - 3 \left( \frac{x-u}{v-u} \right)^2 \right) \exp \left( \frac{2(v-x)b_1}{a_0 - a_1} \right) + \frac{a_0 + a_1}{2}, \end{aligned} \quad (9)$$

where  $a_0 \neq a_1$  and  $v > u$ . Two issues must be noticed in this function. First, it allows a smooth connection between two functions, being  $u$  and  $v$  the starting and ending points where the link function is used,  $a_0$  and  $a_1$  the values of the connected functions on these points and  $b_0$  and  $b_1$  their first derivative values. Second, if the signs of  $b_0$ ,  $b_1$  and  $(a_1 - a_0)$  are the same, then the first derivative of  $l$  will not be equal to 0 for any point inside  $[u, v]$ .

The form of  $J$  is given in equation (10) and its construction guarantees that  $\hat{x}_i$  is the unique point in which the derivative is equal to 0 in the neighbourhood where it is defined. The construction of  $J$  is achieved with the  $\mathcal{K}$  function defined bellow and properly linked with the link function (9) to preserve (RC1). The  $\mathcal{K}$  function is defined as follows

$$\mathcal{K}(x; \hat{x}_i, \mathbf{p}_i, \mathbf{q}_i, \eta_i) = \mathbf{p}_i \left( 1 + \delta_i \left( \frac{x - \hat{x}_i}{\eta_i} \right)^2 \right)^{\eta_i^2 \frac{\delta_i \cdot \mathbf{q}_i}{2\mathbf{p}_i}},$$

being  $\delta_i$  a value indicating if  $\hat{x}_i$  is a mode ( $\delta_i = -1$ ) or an antinode ( $\delta_i = 1$ ). The value  $\eta_i$  will be defined later and it will depend on  $\varsigma_i$ . The second derivative of this function exists and is Hölder continuous in  $(\hat{x}_i - \eta_i/2, \hat{x}_i + \eta_i/2)$ . The following equalities are also satisfied:  $\mathcal{K}(\hat{x}_i; \hat{x}_i, \mathbf{p}_i, \mathbf{q}_i, \eta_i) = \mathbf{p}_i$  and  $\mathcal{K}''(\hat{x}_i; \hat{x}_i, \mathbf{p}_i, \mathbf{q}_i, \eta_i) = \mathbf{q}_i$ . Then, denoting as  $\boldsymbol{\rho}_i = (\hat{x}_i, \hat{f}_{h_k}(\hat{x}_i), \hat{f}_{h_{\text{PI}}}''(\hat{x}_i))$ , the  $J$  function can be defined as follows

$$J(x; \hat{x}_i, h_k, h_{\text{PI}}, \varsigma_i) = \begin{cases} l(x; \mathbf{r}_i, \mathbf{v}_i, \hat{f}_{h_k}(\mathbf{r}_i), \mathcal{K}(\mathbf{v}_i; \boldsymbol{\rho}_i, \eta_i), \hat{f}_{h_k}'(\mathbf{r}_i), \mathcal{K}'(\mathbf{v}_i; \boldsymbol{\rho}_i, \eta_i)) & \text{if } x \in (\mathbf{r}_i, \mathbf{v}_i), \\ \mathcal{K}(x; \boldsymbol{\rho}_i, \eta_i) & \text{if } x \in [\mathbf{v}_i, \mathbf{w}_i], \\ l(x; \mathbf{w}_i, \mathbf{s}_i, \mathcal{K}(\mathbf{w}_i; \boldsymbol{\rho}_i, \eta_i), \hat{f}_{h_k}(\mathbf{s}_i), \mathcal{K}'(\mathbf{w}_i; \boldsymbol{\rho}_i, \eta_i), \hat{f}_{h_k}'(\mathbf{s}_i)) & \text{if } x \in (\mathbf{w}_i, \mathbf{s}_i), \end{cases} \quad (10)$$

being  $\mathbf{v}_i = \hat{x}_i - \eta_i/2$  and  $\mathbf{w}_i = \hat{x}_i + \eta_i/2$ . As it was mentioned, the function  $J$  described in (10) (and hence also the calibration function  $g$ ) depends on the constant  $\varsigma_i \in (0, 1/2)$ . Ordering the modes and denoting as  $\hat{x}_0 = -\infty$  and  $\hat{x}_{(2k)} = \infty$ , that is  $-\infty = \hat{x}_0 < \hat{x}_1 < \dots < \hat{x}_{2k-1} < \hat{x}_{2k} = \infty$ , the remaining unknowns values in (10) will be obtained as follows. First, it is necessary to decide at which height  $\vartheta_i$  the modification in  $\hat{f}_{h_k}$  is done. For values of  $\varsigma_i$  close to 0,  $\vartheta_i$  will be close to  $\hat{f}_{h_k}(\hat{x}_i)$ ; while for values close to 0.5,  $\vartheta_i$  will be in the middle point between  $\hat{f}_{h_k}(\hat{x}_i)$  and the highest (or lowest if  $\hat{x}_i$  is an antimode) value of  $\hat{f}_{h_k}$  in the two closest modes or antimodes ( $\hat{x}_{i-1}$  and  $\hat{x}_{i+1}$ ). Second, once the height is decided,  $\mathbf{r}_i$  and  $\mathbf{s}_i$  will be the left and the right closest points to  $\hat{x}_i$  at which the density estimation is equal to  $\vartheta_i$ . Third, in order to link correctly the  $\mathcal{K}$  function, it is necessary to define  $\eta_i$  ensuring that  $\mathcal{K}(\hat{x}_i \pm \eta_i/2; \boldsymbol{\rho}_i, \eta_i)$  will be higher (lower in the antimodes) than  $\vartheta_i$ . With this objective,  $\eta_i$  is chosen in such a way that  $\mathcal{K}(\hat{x}_i \pm \eta_i/2; \boldsymbol{\rho}_i, \eta_i)$  is near  $\hat{f}_{h_k}(\hat{x}_i)$  and as close as possible to the middle point between  $\vartheta_i$  and  $\hat{f}_{h_k}(\hat{x}_i)$ . Also, the value  $\eta_i$  will ensure that the neighbourhood  $[\mathbf{v}_i, \mathbf{w}_i]$  in which  $\mathcal{K}$  is defined is inside  $(\mathbf{r}_i, \mathbf{s}_i)$ . Finally,  $\hat{f}_{h_k}'$  must be different to 0 in the four points  $(\mathbf{r}_i, \mathbf{v}_i, \mathbf{w}_i$  and  $\mathbf{s}_i)$  where the two link functions are employed. An example of the modifications achieved by the  $J$  function in the modes and antimodes of  $\hat{f}_{h_k}$  is shown in Figure 2 and the complete characterization is provided bellow

$$\begin{aligned}
\vartheta_i &= \hat{f}_{h_k}(\hat{x}_i) + \delta_i \cdot \varsigma_i \cdot \min \left( |\hat{f}_{h_k}(\hat{x}_i) - \hat{f}_{h_k}(\hat{x}_{i-1})|, |\hat{f}_{h_k}(\hat{x}_i) - \hat{f}_{h_k}(\hat{x}_{i+1})| \right), \\
\mathbf{r}_i &= \inf \{x : x > \hat{x}_{i-1}, \delta_i \cdot \hat{f}_{h_k}(x) \leq \delta_i \cdot \vartheta_i \text{ and } \hat{f}_{h_k}'(x) \neq 0\}, \\
\mathbf{s}_i &= \sup \{x : x < \hat{x}_{i+1}, \delta_i \cdot \hat{f}_{h_k}(x) \leq \delta_i \cdot \vartheta_i \text{ and } \hat{f}_{h_k}'(x) \neq 0\}, \\
\eta_i &= \sup \{ \gamma : \gamma \in (0, \min(\hat{x}_i - \mathbf{r}_i, \mathbf{s}_i - \hat{x}_i)), \delta_i \mathcal{K}(\hat{x}_i + \gamma/2; \boldsymbol{\rho}_i, \gamma) \leq \delta_i (\hat{f}_{h_k}(\hat{x}_i) + \vartheta_i)/2 \\
&\quad \text{and } \hat{f}_{h_k}'(\hat{x}_i \pm \gamma/2) \neq 0 \}.
\end{aligned} \tag{11}$$

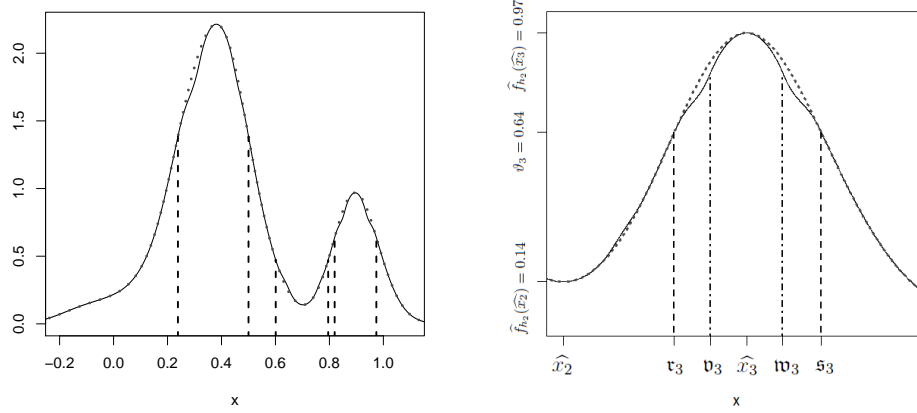


Figure 2: Sample of  $n = 1000$  observations from model M16 (described in Appendix A). Dotted grey line:  $\hat{f}_{h_2}$ . Solid line:  $g(\cdot; h_k, h_{PI}, (0.4, 0.4, 0.4))$ . Dashed line: support of  $J(\cdot; \hat{x}_i, h_2, h_{PI}, 0.4)$ , with  $i = 1, 2, 3$ . Dot-dashed line: support of  $\mathcal{K}(\cdot; \boldsymbol{\rho}_3, \eta_3)$ . Left: in the support  $(-0.2, 1.1)$ . Right: in a neighbourhood of the mode  $\hat{x}_3$ .

In order to proceed with the modification achieved with the  $L$  function, assume that this estimator has  $t$  saddle points  $\zeta_p$ , with  $p = 1, \dots, t$ . Define as  $\xi = \min\{|x - y| : x, y \in (\zeta_1, \dots, \zeta_t) \cup (\mathbf{r}_1, \mathbf{s}_1, \dots, \mathbf{r}_{2k-1}, \mathbf{s}_{2k-1})\}$ . Then, if  $\zeta_p$  is not inside the interval where the  $J$  functions are defined, the neighbourhood used to remove the stationary and turning points will be delimited by  $z_{(2p-1)} = \zeta_p - \varpi\xi$  and  $z_{(2p)} = \zeta_p + \varpi\xi$ , with  $\varpi \in (0, 1/4)$ . In the simulation study, the value of  $\varpi$  will be taken close enough to 0 to avoid an impact in the value of the integral associated to  $g$ . Once these points are calculated, the saddle points can be removed from  $g$  with the link function by taking  $L$  equal to

$$L(x; z_{(2p-1)}, z_{(2p)}, h_k) = l(x; z_{(2p-1)}, z_{(2p)}, \hat{f}_{h_k}(z_{(2p-1)}), \hat{f}_{h_k}(z_{(2p)}), \hat{f}'_{h_k}(z_{(2p-1)}), \hat{f}'_{h_k}(z_{(2p)})). \quad (12)$$

To construct the calibration function, first, values of  $\varsigma_i \in (0, 1/2)$ , for  $i \in \{1, \dots, (2k -$



1)} must be fixed. Then, using the  $J$  function (10) with the values given in (11) and the  $L$  function (12), the function  $g$  defined in (8) satisfies the specified regularity conditions and  $|g''(\hat{x}_i; h_k, h_{\text{PI}}, \boldsymbol{\varsigma})|/g^3(\hat{x}_i; h_k, h_{\text{PI}}, \boldsymbol{\varsigma})$  converges in probability to  $d_i$ . With this modification the calibration function also preserves the structure of the data under the hypothesis that  $f$  has  $k$  modes. However, this calibration function  $g$  may not be a density function since

$$q(\boldsymbol{\varsigma}) = \int_{-\infty}^{\infty} g(x; h_k, h_{\text{PI}}, \boldsymbol{\varsigma}) dx \quad (13)$$

may not be equal to 1. To guarantee that  $g$  is indeed a density, a possible approach consists in proceeding with a search of values for  $\boldsymbol{\varsigma}$  such that  $q(\boldsymbol{\varsigma})$  is equal to 1, which is guaranteed since

$$\lim_{\varsigma_i \rightarrow 0^+, \forall i \in (1, \dots, 2k-1)} q(\boldsymbol{\varsigma}) = \int_{-\infty}^{\infty} \hat{f}_{h_k}(x) dx = 1.$$

For convenience, in the simulation study, the employed approach will be followed using  $\varsigma_i$  close enough to 0 ( $\forall i \in \{1, \dots, 2k-1\}$ ) in order to avoid an impact on the integral value.

To conclude this section it is important to mention the necessary condition of the bounded support in order to ensure the results of Mammen et al. (1992). Although, in general, it will not have an important effect as it can be seen in Section 3, in case of a known support,  $h_{\text{HY},k}$  can be used to create the calibration function. An alternative approach for known support is given in Appendix B. Once a conclusion about the number of modes is made, when both the estimation of the modes and antimodes is needed, this known support plays an special role. From the results of Romano (1988) and Mammen et al. (1992), it can be proved that  $\hat{f}_{h_{\text{HY},k}}$  provides a good estimation of their localization. Also, it should be reminded that, unlike Silverman (1981) and Fisher and Marron (2001) proposals, the one presented in this paper considers  $H_0 : j = k$  instead of  $H_0 : j \leq k$ . A deeper insight is presented in Appendix C.

### 3 Simulation study

The aim of the following simulation study is to compare the different proposals presented in Section 2. Samples of size  $n = 50$ ,  $n = 200$  and  $n = 1000$  ( $n = 100$  instead of  $n = 1000$  in power studies) were drawn from twenty five different distributions, ten of them unimodal (models M1–M10), ten bimodal (models M11–M20) and five trimodal (models M21–M25) (see Appendix A). For each choice of sampling distribution and sample size, 500 realizations of the sample were generated. Conditionally on each of those samples, for testing purposes, 500 resamples of size  $n$  were drawn from the population. Tables 1–5 report the percentages of rejections for significance levels  $\alpha = 0.01$ ,  $\alpha = 0.05$  and  $\alpha = 0.10$  under different scenarios: testing unimodality vs. multimodality (Tables 1 and 2); testing bimodality against more than two modes (Table 4) and power analysis (respectively Tables 3 and 5). The procedures considered include the proposals by Silverman (1981) (SI), Fisher and Marron (2001) (FM), Hall and York (2001) (HY), Hartigan and Hartigan (1985) (HH), Cheng and Hall (1998) (CH) and the new proposal (NP) in this paper. Note that for testing  $H_0 : j = 2$ , only SI, FM and NP can be compared. For the critical bandwidth test HY, the two proposed methods for computing  $\lambda_\alpha$  have been tried, with very similar results. The ones reported in this section correspond to a polynomial approximation for  $\lambda_\alpha$ .  $I = [0, 1]$  is used both for HY and for NP, when the interval containing the modes is assumed to be known (Table 6). Further computational details are included in Appendix E.

**Testing unimodality vs. multimodality.** From the results reported in Tables 1 and 2, it can be concluded that SI is quite conservative: even for high sample sizes, the percentage of rejections is below the significance level, and quite close to 0 even for  $\alpha = 0.10$ . An exception to this conservative behaviour can be found in models M5 (with outliers in tails) and M6 (with a *flat tail*), finding percentages of rejections between 0.03 and 0.052

for  $n = 200$  or  $n = 1000$ , when  $\alpha = 0.10$ . Regarding FM, a systematic behaviour cannot be concluded: the percentage of rejections is above the significance level for models M1, M5, M7, M9 or M10, but it can be also below the true level for M2, M4 or M8.

The behaviour of HY is quite good when using  $I = [0, 1]$  for the different distributions and large sample sizes. For  $n = 1000$ , the percentage of rejections is quite close to  $\alpha$ , except for model M5 (for  $\alpha = 0.05$ , below level) and for models M6 and M7 (for  $\alpha = 0.10$ , above level). However, the percentage of rejections is usually below the significance level for small sample sizes. Exceptions to this general pattern are found for model M1 ( $n = 200$ ), M3 ( $n = 50$ ) and M10 ( $n = 200$ ), where percentage of rejections is close to  $\alpha$  and models M3 ( $n = 200$ ), M6 ( $n = 200$ ) and M7 with percentages above  $\alpha$ . Nevertheless, it should be kept in mind that the support where unimodality is tested must be known. Similarly to SI, the results obtained with HH are quite conservative. For instance, for  $n = 1000$ , even taking  $\alpha = 0.10$ , the percentage of rejections is always below 0.002.

Calibration seems correct in simple models for CH, although slightly conservative in some cases, such as for models M4 ( $n = 1000$ ), M5 ( $n = 200$ ) and M8 ( $n = 200$ ). As expected, the parametric calibration distributions do not capture, for example, the skewness and this affects the second-order properties in more complex models. This effect is reflected in the asymmetric M3, M7 and M10 ( $n = 1000$ ), or model M9, where the percentage of rejections is below  $\alpha$ , and for M6 where is considerably higher than the significance level.

Finally, regarding the new proposal NP, it can be concluded that the calibration is quite satisfactory, even for complicated models, with a slightly conservative performance for M3 ( $n = 200$ ), M4 ( $n = 1000$ ) or M7 ( $n = 200$ ), being this effect more clear for model M9. The only scenario where the percentage of rejections is above  $\alpha$  is for M6 with  $n = 200$ , but this behaviour is corrected when increasing the sample size. Although the performance is better for higher sample sizes ( $n = 200$  or  $n = 1000$ ), in some cases, such as M9 or M16 (in

the bimodal case), it can be seen that even for  $n = 1000$ , a percentage of rejections close to  $\alpha$  is hard to get. In this difficult cases, the knowledge of the support can be used for obtaining better results as it was reported in Table 6, where the percentage of rejections is close to  $\alpha$  for the sample sizes  $n = 200$  and  $n = 1000$ .

Regarding power behaviour (and just commenting on the three methods which exhibit a correct calibration), results are reported in Table 3: none of the proposals is clearly more powerful. For instance, for M13, HY clearly detects the appearance of the second small mode, whereas the other approaches do not succeed in doing so. For M11, M12, M14 and M15 ( $n = 50$ ), CH presents the highest empirical power and HY shows the lowest one.

**Assessing bimodality.** For testing  $H_0 : j = 2$ , Hall and York (2001) prove that, even knowing the density support, SI cannot be consistently calibrated by a bootstrap procedure, similar to the one used for the unimodality vs. multimodality test. Hence, in some models such as M15 or M16 ( $n = 50$ ), the percentage of rejections is close to  $\alpha$ , whereas the conservative behaviour observed in the unimodality test, is just perceived in M14 and M19, for large sample sizes. Also, in this case, there is a model where the percentage of rejections is considerably higher than the significance level, M20 ( $n = 200$  and  $n = 1000$ ), being this bimodal scenario similar to the conservative M19, just generating some outliers. FM presents again an erratic behaviour: for M17 (except for  $n = 200$ ), M18 or M19, the percentage of rejections is below  $\alpha$ , whereas the opposite happens for M11, M12, M15, M16 or M20 (except for  $n = 50$ ).

For testing bimodality, NP presents good results. The percentage of rejections is close to the significance level, except for M12 ( $n = 200$ ) and M13 ( $n = 200$ ), slightly below  $\alpha$ , and M11 ( $n = 50$ ), M15 ( $n = 50, n = 200$ ), M16 and M19 ( $n = 50$ ), slightly above  $\alpha$ . For  $n = 1000$ , all the results are good except for M16, but the calibration problem is corrected (as seen in Table 6) applying NP with known support, taking for that purpose  $I = [0, 1]$ .

So, just the new proposal presents a correct calibration. Hence, power results reported in Table 5 are only judged for the new proposal: power increases with sample size, detecting that all the alternative distributions do not satisfy the null hypothesis, except for M21 ( $n = 50$ ) and M24 ( $n = 50$ ), but with small sample sizes.

## 4 Real data analysis

As explained in the Introduction, the value of stamps depends on its scarcity, and thickness is determinant in this sense. However, in general, the designation of thick, medium or thin stamps is relative and can only refer to a particular stamp issue. Otherwise, making uniform categories for all stamp issues may lead to inaccurate classifications. In addition, there is not such a differentiation between groups available in stamps catalogs, leaving this classification to a personal subjective judgment. The importance of establishing an objective criterion specially appears in stamp issues printed on a mixture of paper types, with possible differences in their thickness.

A stamp issue where the problem of determining the number of different groups of stamps appears is in the 1872 Hidalgo issue. First, for this particular issue, and in general in the Mexican ones, it is known that the handmade paper presents a high variability in the thickness of the paper. Second, since of scarcity of ordinary white wove paper, other types of paper were used to produce the Hidalgo issue. A small quantity of “vertically laid” paper, a fiscal type of white wove paper denominated *Papel Sellado* (some of them were watermarked vertically), other type of white wove, the *La Croix-Freres* of France (some of them with a watermark of *LA+-F*) and also another unwatermarked white wove paper might also have been used. It is estimated than the watermark of *Papel Sellado* can appear in between 6 and 18 stamps in each sheet of 100. For the *La Croix-Freres*

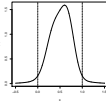
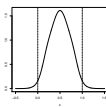
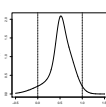
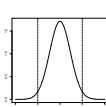
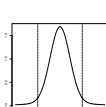
		$\alpha$	0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10
M1			SI			FM			HY		
		$n = 50$	0(0)	0(0)	0.002(0.004)	0.010(0.009)	0.076(0.023)	0.178(0.034)	0(0)	0.022(0.013)	0.050(0.019)
		$n = 200$	0(0)	0.002(0.004)	0.020(0.012)	0.056(0.020)	0.162(0.032)	0.262(0.039)	0.002(0.004)	0.046(0.018)	0.090(0.025)
		$n = 1000$	0(0)	0(0)	0.020(0.012)	0.036(0.016)	0.126(0.029)	0.210(0.036)	0.002(0.004)	0.052(0.019)	0.096(0.026)
			HH			CH			NP		
		$n = 50$	0(0)	0.006(0.007)	0.022(0.013)	0.022(0.013)	0.072(0.023)	0.140(0.030)	0.010(0.009)	0.064(0.021)	0.120(0.028)
		$n = 200$	0(0)	0.002(0.004)	0.002(0.004)	0.014(0.010)	0.058(0.020)	0.122(0.029)	0.010(0.009)	0.044(0.018)	0.120(0.028)
		$n = 1000$	0(0)	0(0)	0(0)	0.006(0.007)	0.048(0.019)	0.104(0.027)	0.008(0.008)	0.052(0.019)	0.108(0.027)
			SI			FM			HY		
M2		$n = 50$	0(0)	0(0)	0.012(0.010)	0.004(0.006)	0.040(0.017)	0.076(0.023)	0(0)	0.024(0.013)	0.068(0.022)
		$n = 200$	0(0)	0(0)	0.006(0.007)	0(0)	0.006(0.007)	0.056(0.020)	0(0)	0.030(0.015)	0.082(0.024)
		$n = 1000$	0(0)	0(0)	0.014(0.010)	0(0)	0.014(0.010)	0.040(0.017)	0.004(0.006)	0.038(0.017)	0.080(0.024)
			HH			CH			NP		
		$n = 50$	0(0)	0.012(0.010)	0.028(0.014)	0.046(0.018)	0.100(0.026)	0.140(0.030)	0.016(0.011)	0.070(0.022)	0.122(0.029)
		$n = 200$	0(0)	0.002(0.004)	0.004(0.006)	0.020(0.012)	0.074(0.023)	0.164(0.032)	0.004(0.006)	0.050(0.019)	0.114(0.028)
		$n = 1000$	0(0)	0(0)	0(0)	0.008(0.008)	0.032(0.015)	0.092(0.025)	0.006(0.007)	0.030(0.015)	0.082(0.024)
			SI			FM			HY		
		$n = 50$	0(0)	0(0)	0.002(0.004)	0.026(0.014)	0.112(0.028)	0.222(0.036)	0.008(0.008)	0.066(0.022)	0.108(0.027)
M3		$n = 200$	0(0)	0(0)	0.004(0.006)	0.014(0.010)	0.072(0.023)	0.146(0.031)	0.030(0.015)	0.088(0.025)	0.146(0.031)
		$n = 1000$	0(0)	0(0)	0.006(0.007)	0.002(0.004)	0.050(0.019)	0.128(0.029)	0.018(0.012)	0.070(0.022)	0.120(0.028)
			HH			CH			NP		
		$n = 50$	0(0)	0(0)	0.004(0.006)	0.002(0.004)	0.032(0.015)	0.056(0.020)	0.004(0.006)	0.042(0.018)	0.078(0.024)
		$n = 200$	0(0)	0(0)	0.002(0.004)	0.002(0.004)	0.004(0.006)	0.030(0.015)	0.002(0.004)	0.022(0.013)	0.054(0.020)
		$n = 1000$	0(0)	0(0)	0(0)	0.002(0.004)	0.012(0.010)	0.032(0.015)	0.006(0.007)	0.032(0.015)	0.082(0.024)
			SI			FM			HY		
		$n = 50$	0(0)	0(0)	0.002(0.004)	0.002(0.004)	0.018(0.012)	0.060(0.021)	0(0)	0.020(0.012)	0.050(0.019)
		$n = 200$	0(0)	0(0)	0.004(0.006)	0(0)	0.012(0.010)	0.044(0.018)	0.004(0.006)	0.026(0.014)	0.074(0.023)
M4		$n = 1000$	0(0)	0.004(0.006)	0.014(0.010)	0(0)	0.010(0.009)	0.046(0.018)	0.008(0.008)	0.052(0.019)	0.090(0.025)
			HH			CH			NP		
		$n = 50$	0(0)	0.004(0.006)	0.018(0.012)	0.016(0.011)	0.064(0.021)	0.118(0.028)	0.014(0.010)	0.050(0.019)	0.102(0.027)
		$n = 200$	0(0)	0(0)	0(0)	0.008(0.008)	0.032(0.015)	0.082(0.024)	0.004(0.006)	0.030(0.015)	0.080(0.024)
		$n = 1000$	0(0)	0(0)	0(0)	0.004(0.006)	0.034(0.016)	0.066(0.022)	0(0)	0.028(0.014)	0.066(0.022)
			SI			FM			HY		
		$n = 50$	0(0)	0(0)	0.004(0.006)	0.186(0.034)	0.366(0.042)	0.494(0.044)	0(0)	0.006(0.007)	0.038(0.017)
		$n = 200$	0(0)	0(0)	0.032(0.015)	0.268(0.039)	0.500(0.044)	0.612(0.043)	0.002(0.004)	0.030(0.015)	0.074(0.023)
		$n = 1000$	0(0)	0.004(0.006)	0.030(0.015)	0.210(0.036)	0.380(0.043)	0.504(0.044)	0.006(0.007)	0.028(0.014)	0.080(0.024)
M5			HH			CH			NP		
		$n = 50$	0(0)	0(0)	0.006(0.007)	0.004(0.006)	0.052(0.019)	0.084(0.024)	0.006(0.007)	0.062(0.021)	0.106(0.027)
		$n = 200$	0(0)	0(0)	0(0)	0.010(0.009)	0.034(0.016)	0.064(0.021)	0.012(0.010)	0.050(0.019)	0.092(0.025)
		$n = 1000$	0(0)	0(0)	0(0)	0.006(0.007)	0.022(0.013)	0.082(0.024)	0.006(0.007)	0.052(0.019)	0.106(0.027)

Table 1: Percentages of rejections for testing  $H_0 : j = 1$ , with 500 simulations (1.96 times their estimated standard deviation in parenthesis) and  $B = 500$  bootstrap samples.

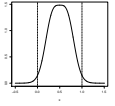
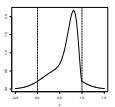
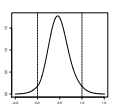
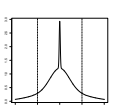
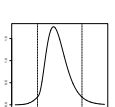
		$\alpha$	0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10
M6			SI			FM			HY		
		$n = 50$	0(0)	0(0)	0.012(0.010)	0.004(0.006)	0.040(0.017)	0.082(0.024)	0.002(0.004)	0.022(0.013)	0.074(0.023)
		$n = 200$	0(0)	0.012(0.010)	0.052(0.019)	0.010(0.009)	0.064(0.021)	0.122(0.029)	0.012(0.010)	0.110(0.027)	0.196(0.035)
		$n = 1000$	0.002(0.004)	0.020(0.012)	0.048(0.019)	0.008(0.008)	0.042(0.018)	0.100(0.026)	0.048(0.019)	0.118(0.028)	0.216(0.036)
			HH			CH			NP		
		$n = 50$	0(0)	0.006(0.007)	0.012(0.010)	0.028(0.014)	0.092(0.025)	0.168(0.033)	0.008(0.008)	0.050(0.019)	0.112(0.028)
		$n = 200$	0(0)	0.008(0.008)	0.012(0.010)	0.050(0.019)	0.136(0.030)	0.236(0.037)	0.018(0.012)	0.088(0.025)	0.160(0.032)
		$n = 1000$	0(0)	0.002(0.004)	0.002(0.004)	0.038(0.017)	0.112(0.028)	0.202(0.035)	0.016(0.011)	0.046(0.018)	0.116(0.028)
			SI			FM			HY		
M7		$n = 50$	0(0)	0(0)	0(0)	0.072(0.023)	0.246(0.038)	0.378(0.043)	0.012(0.010)	0.072(0.023)	0.146(0.031)
		$n = 200$	0(0)	0.002(0.004)	0.006(0.007)	0.064(0.021)	0.210(0.036)	0.368(0.042)	0.016(0.011)	0.078(0.024)	0.144(0.031)
		$n = 1000$	0(0)	0(0)	0(0)	0.060(0.021)	0.214(0.036)	0.346(0.042)	0.004(0.006)	0.072(0.023)	0.134(0.030)
			HH			CH			NP		
		$n = 50$	0(0)	0(0)	0.010(0.009)	0.008(0.008)	0.026(0.014)	0.082(0.024)	0.006(0.007)	0.032(0.015)	0.084(0.024)
		$n = 200$	0(0)	0(0)	0(0)	0(0)	0.014(0.010)	0.042(0.018)	0.002(0.004)	0.028(0.014)	0.070(0.022)
		$n = 1000$	0(0)	0(0)	0(0)	0(0)	0.012(0.010)	0.036(0.016)	0.004(0.006)	0.042(0.018)	0.094(0.026)
			SI			FM			HY		
		$n = 50$	0(0)	0(0)	0(0)	0.006(0.007)	0.024(0.013)	0.062(0.021)	0(0)	0.012(0.010)	0.046(0.018)
M8		$n = 200$	0(0)	0(0)	0.002(0.004)	0.002(0.004)	0.022(0.013)	0.064(0.021)	0(0)	0.024(0.013)	0.054(0.020)
		$n = 1000$	0(0)	0(0)	0(0)	0.002(0.004)	0.018(0.012)	0.048(0.019)	0.010(0.009)	0.054(0.020)	0.092(0.025)
			HH			CH			NP		
		$n = 50$	0(0)	0(0)	0.006(0.007)	0.006(0.007)	0.034(0.016)	0.078(0.024)	0.006(0.007)	0.032(0.015)	0.076(0.023)
		$n = 200$	0(0)	0(0)	0(0)	0.004(0.006)	0.026(0.014)	0.066(0.022)	0.006(0.007)	0.028(0.014)	0.088(0.025)
		$n = 1000$	0(0)	0(0)	0.002(0.004)	0.016(0.011)	0.038(0.017)	0.082(0.024)	0.014(0.010)	0.044(0.018)	0.088(0.025)
			SI			FM			HY		
		$n = 50$	0(0)	0(0)	0(0)	0.018(0.012)	0.084(0.024)	0.198(0.035)	0(0)	0.006(0.007)	0.014(0.010)
		$n = 200$	0(0)	0(0)	0(0)	0.048(0.019)	0.182(0.034)	0.328(0.041)	0.002(0.004)	0.022(0.013)	0.060(0.021)
M9		$n = 1000$	0(0)	0(0)	0(0)	0.014(0.010)	0.160(0.032)	0.318(0.041)	0.012(0.010)	0.048(0.019)	0.086(0.025)
			HH			CH			NP		
		$n = 50$	0(0)	0(0)	0(0)	0.002(0.004)	0.016(0.011)	0.032(0.015)	0.004(0.006)	0.026(0.014)	0.068(0.022)
		$n = 200$	0(0)	0(0)	0(0)	0.002(0.004)	0.018(0.012)	0.042(0.018)	0.010(0.009)	0.046(0.018)	0.084(0.024)
		$n = 1000$	0(0)	0(0)	0(0)	0.002(0.004)	0.010(0.009)	0.014(0.010)	0.004(0.006)	0.020(0.012)	0.062(0.021)
			SI			FM			HY		
		$n = 50$	0(0)	0(0)	0(0)	0.016(0.011)	0.054(0.020)	0.116(0.028)	0(0)	0.014(0.010)	0.050(0.019)
		$n = 200$	0(0)	0(0)	0.006(0.007)	0.018(0.012)	0.092(0.025)	0.182(0.034)	0.006(0.007)	0.038(0.017)	0.086(0.025)
		$n = 1000$	0(0)	0(0)	0.008(0.008)	0.022(0.013)	0.094(0.026)	0.168(0.033)	0.010(0.009)	0.050(0.019)	0.096(0.026)
M10			HH			CH			NP		
		$n = 50$	0(0)	0.002(0.004)	0.008(0.008)	0.004(0.006)	0.046(0.018)	0.086(0.025)	0.012(0.010)	0.044(0.018)	0.094(0.026)
		$n = 200$	0(0)	0(0)	0(0)	0.014(0.010)	0.042(0.018)	0.078(0.024)	0.010(0.009)	0.062(0.021)	0.094(0.026)
		$n = 1000$	0(0)	0(0)	0(0)	0.008(0.008)	0.028(0.014)	0.074(0.023)	0.008(0.008)	0.040(0.017)	0.104(0.027)

Table 2: Percentages of rejections for testing  $H_0 : j = 1$ , with 500 simulations (1.96 times their estimated standard deviation in parenthesis) and  $B = 500$  bootstrap samples.

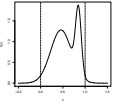
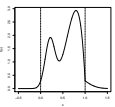
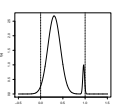
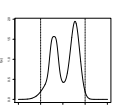
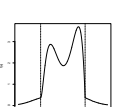
		$\alpha$	0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10
M11			SI			FM			HY		
		$n = 50$	0(0)	0.008(0.008)	0.100(0.026)	0.084(0.024)	0.250(0.038)	0.394(0.043)	0.008(0.008)	0.168(0.033)	0.330(0.041)
		$n = 100$	0.002(0.004)	0.154(0.032)	0.368(0.042)	0.374(0.042)	0.640(0.042)	0.738(0.039)	0.168(0.033)	0.502(0.044)	0.630(0.042)
		$n = 200$	0.094(0.026)	0.304(0.040)	0.500(0.044)	0.600(0.043)	0.788(0.036)	0.860(0.030)	0.378(0.043)	0.604(0.043)	0.714(0.040)
			HH			CH			NP		
		$n = 50$	0.012(0.010)	0.088(0.025)	0.156(0.032)	0.196(0.035)	0.370(0.042)	0.494(0.044)	0.090(0.025)	0.238(0.037)	0.376(0.042)
		$n = 100$	0.060(0.021)	0.182(0.034)	0.274(0.039)	0.442(0.044)	0.630(0.042)	0.722(0.039)	0.228(0.037)	0.418(0.043)	0.542(0.044)
		$n = 200$	0.106(0.027)	0.238(0.037)	0.356(0.042)	0.584(0.043)	0.768(0.037)	0.824(0.033)	0.328(0.041)	0.506(0.044)	0.600(0.043)
M12			SI			FM			HY		
		$n = 50$	0(0)	0.006(0.007)	0.058(0.020)	0.332(0.041)	0.584(0.043)	0.716(0.040)	0(0)	0.174(0.033)	0.422(0.043)
		$n = 100$	0.006(0.007)	0.186(0.034)	0.496(0.044)	0.662(0.041)	0.838(0.032)	0.896(0.027)	0.224(0.037)	0.642(0.042)	0.802(0.035)
		$n = 200$	0.026(0.014)	0.486(0.044)	0.804(0.035)	0.914(0.025)	0.966(0.016)	0.986(0.010)	0.584(0.043)	0.912(0.025)	0.950(0.019)
			HH			CH			NP		
		$n = 50$	0.042(0.018)	0.118(0.028)	0.202(0.035)	0.268(0.039)	0.480(0.044)	0.594(0.043)	0.100(0.026)	0.266(0.039)	0.400(0.043)
		$n = 100$	0.158(0.032)	0.346(0.042)	0.460(0.044)	0.636(0.042)	0.788(0.036)	0.844(0.032)	0.346(0.042)	0.578(0.043)	0.680(0.041)
		$n = 200$	0.262(0.039)	0.490(0.044)	0.622(0.043)	0.818(0.034)	0.926(0.023)	0.956(0.018)	0.524(0.044)	0.752(0.038)	0.844(0.032)
M13			SI			FM			HY		
		$n = 50$	0(0)	0(0)	0(0)	0.064(0.021)	0.492(0.044)	0.776(0.037)	0.634(0.042)	0.858(0.031)	0.892(0.027)
		$n = 100$	0(0)	0.272(0.039)	0.956(0.018)	0.914(0.025)	0.998(0.004)	1(0)	0.994(0.007)	1(0)	1(0)
		$n = 200$	0.898(0.027)	1(0)	1(0)	1(0)	1(0)	1(0)	1(0)	1(0)	1(0)
			HH			CH			NP		
		$n = 50$	0(0)	0.004(0.006)	0.014(0.010)	0.014(0.010)	0.052(0.019)	0.096(0.026)	0.016(0.011)	0.058(0.020)	0.112(0.028)
		$n = 100$	0(0)	0(0)	0(0)	0(0)	0.032(0.015)	0.062(0.021)	0.008(0.008)	0.050(0.019)	0.102(0.027)
		$n = 200$	0(0)	0(0)	0.002(0.004)	0.006(0.007)	0.048(0.019)	0.366(0.042)	0.016(0.011)	0.276(0.039)	0.758(0.038)
M14			SI			FM			HY		
		$n = 50$	0.002(0.004)	0.044(0.018)	0.352(0.042)	0.662(0.041)	0.864(0.030)	0.922(0.024)	0.050(0.019)	0.576(0.043)	0.830(0.033)
		$n = 100$	0.234(0.037)	0.922(0.024)	0.990(0.009)	0.990(0.009)	1(0)	1(0)	0.920(0.024)	1(0)	1(0)
		$n = 200$	0.866(0.030)	1(0)	1(0)	1(0)	1(0)	1(0)	1(0)	1(0)	1(0)
			HH			CH			NP		
		$n = 50$	0.226(0.037)	0.514(0.044)	0.646(0.042)	0.718(0.039)	0.860(0.030)	0.924(0.023)	0.460(0.044)	0.716(0.040)	0.806(0.035)
		$n = 100$	0.784(0.036)	0.946(0.020)	0.978(0.013)	0.994(0.007)	0.996(0.006)	1(0)	0.950(0.019)	0.992(0.008)	0.994(0.007)
		$n = 200$	1(0)	1(0)	1(0)	1(0)	1(0)	1(0)	1(0)	1(0)	1(0)
M15			SI			FM			HY		
		$n = 50$	0(0)	0.004(0.006)	0.038(0.017)	0.042(0.018)	0.118(0.028)	0.228(0.037)	0.014(0.010)	0.128(0.029)	0.268(0.039)
		$n = 100$	0(0)	0.034(0.016)	0.136(0.030)	0.118(0.028)	0.334(0.041)	0.472(0.044)	0.066(0.022)	0.342(0.042)	0.500(0.044)
		$n = 200$	0.008(0.008)	0.176(0.033)	0.370(0.042)	0.266(0.039)	0.524(0.044)	0.636(0.042)	0.298(0.040)	0.576(0.043)	0.736(0.039)
			HH			CH			NP		
		$n = 50$	0.010(0.009)	0.046(0.018)	0.072(0.023)	0.098(0.026)	0.242(0.038)	0.374(0.042)	0.054(0.020)	0.156(0.032)	0.274(0.039)
		$n = 100$	0.014(0.010)	0.070(0.022)	0.150(0.031)	0.232(0.037)	0.424(0.043)	0.542(0.044)	0.124(0.029)	0.288(0.040)	0.400(0.043)
		$n = 200$	0.026(0.014)	0.104(0.027)	0.194(0.035)	0.364(0.042)	0.582(0.043)	0.690(0.041)	0.192(0.035)	0.400(0.043)	0.548(0.044)

Table 3: Percentages of rejections for testing  $H_0 : j = 1$ , with 500 simulations (1.96 times their estimated standard deviation in parenthesis) and  $B = 500$  bootstrap samples.



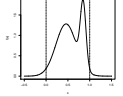
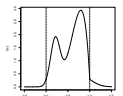
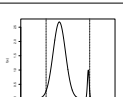
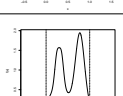
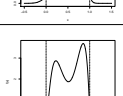
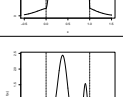
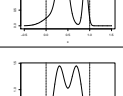
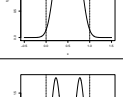
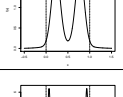
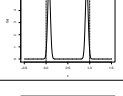
		$\alpha$	0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10
			SI			FM			NP		
M11		$n = 50$	0(0)	0.012(0.010)	0.058(0.020)	0.016(0.011)	0.076(0.023)	0.152(0.031)	0.028(0.014)	0.088(0.025)	0.178(0.034)
		$n = 200$	0.006(0.007)	0.026(0.014)	0.062(0.021)	0.092(0.025)	0.276(0.039)	0.392(0.043)	0.014(0.010)	0.056(0.020)	0.108(0.027)
		$n = 1000$	0.010(0.009)	0.028(0.014)	0.066(0.022)	0.080(0.024)	0.218(0.036)	0.316(0.041)	0.006(0.007)	0.046(0.018)	0.084(0.024)
M12		$n = 50$	0(0)	0.004(0.006)	0.034(0.016)	0.030(0.015)	0.100(0.026)	0.178(0.034)	0.004(0.006)	0.034(0.016)	0.074(0.023)
		$n = 200$	0.006(0.007)	0.022(0.013)	0.038(0.017)	0.038(0.017)	0.134(0.030)	0.184(0.034)	0.002(0.004)	0.030(0.015)	0.076(0.023)
		$n = 1000$	0.014(0.010)	0.022(0.013)	0.030(0.015)	0.052(0.019)	0.094(0.026)	0.168(0.033)	0.008(0.008)	0.046(0.018)	0.082(0.024)
M13		$n = 50$	0(0)	0.002(0.004)	0.026(0.014)	0.006(0.007)	0.044(0.018)	0.102(0.027)	0.006(0.007)	0.052(0.019)	0.118(0.028)
		$n = 200$	0(0)	0.004(0.006)	0.022(0.013)	0(0)	0.024(0.013)	0.056(0.020)	0.006(0.007)	0.028(0.014)	0.070(0.022)
		$n = 1000$	0.002(0.004)	0.012(0.010)	0.044(0.018)	0.004(0.006)	0.036(0.016)	0.072(0.023)	0.010(0.009)	0.044(0.018)	0.088(0.025)
M14		$n = 50$	0(0)	0.010(0.009)	0.022(0.013)	0.020(0.012)	0.072(0.023)	0.132(0.030)	0.008(0.008)	0.034(0.016)	0.074(0.023)
		$n = 200$	0(0)	0(0)	0.006(0.007)	0.018(0.012)	0.088(0.025)	0.152(0.031)	0.004(0.006)	0.034(0.016)	0.088(0.025)
		$n = 1000$	0.002(0.004)	0.006(0.007)	0.008(0.008)	0.024(0.013)	0.058(0.020)	0.114(0.028)	0.008(0.008)	0.056(0.020)	0.092(0.025)
M15		$n = 50$	0(0)	0.014(0.010)	0.048(0.019)	0.008(0.008)	0.072(0.023)	0.136(0.030)	0.012(0.010)	0.078(0.024)	0.158(0.032)
		$n = 200$	0.004(0.006)	0.040(0.017)	0.090(0.025)	0.032(0.015)	0.128(0.029)	0.214(0.036)	0.024(0.013)	0.106(0.027)	0.200(0.035)
		$n = 1000$	0.004(0.006)	0.040(0.017)	0.088(0.025)	0.062(0.021)	0.154(0.032)	0.224(0.037)	0.014(0.010)	0.048(0.019)	0.104(0.027)
M16		$n = 50$	0.008(0.008)	0.048(0.019)	0.080(0.024)	0.092(0.025)	0.210(0.036)	0.284(0.040)	0.006(0.007)	0.064(0.021)	0.112(0.028)
		$n = 200$	0.014(0.010)	0.032(0.015)	0.036(0.016)	0.058(0.020)	0.128(0.029)	0.186(0.034)	0.016(0.011)	0.098(0.026)	0.200(0.035)
		$n = 1000$	0.006(0.007)	0.012(0.010)	0.014(0.010)	0.038(0.017)	0.082(0.024)	0.156(0.032)	0.010(0.009)	0.074(0.023)	0.150(0.031)
M17		$n = 50$	0(0)	0.014(0.010)	0.036(0.016)	0.002(0.004)	0.028(0.014)	0.060(0.021)	0.012(0.010)	0.060(0.021)	0.136(0.030)
		$n = 200$	0(0)	0.028(0.014)	0.068(0.022)	0.008(0.008)	0.046(0.018)	0.096(0.026)	0.008(0.008)	0.070(0.022)	0.106(0.027)
		$n = 1000$	0.002(0.004)	0.004(0.006)	0.026(0.014)	0.002(0.004)	0.026(0.014)	0.048(0.019)	0.008(0.008)	0.038(0.017)	0.074(0.023)
M18		$n = 50$	0(0)	0(0)	0.002(0.004)	0(0)	0.008(0.008)	0.038(0.017)	0.004(0.006)	0.054(0.020)	0.108(0.027)
		$n = 200$	0(0)	0.008(0.008)	0.020(0.012)	0.004(0.006)	0.032(0.015)	0.050(0.019)	0.006(0.007)	0.048(0.019)	0.108(0.027)
		$n = 1000$	0(0)	0.008(0.008)	0.028(0.014)	0.004(0.006)	0.028(0.014)	0.066(0.022)	0.002(0.004)	0.034(0.016)	0.080(0.024)
M19		$n = 50$	0(0)	0(0)	0(0)	0(0)	0.004(0.006)	0.034(0.016)	0.024(0.013)	0.070(0.022)	0.132(0.030)
		$n = 200$	0(0)	0(0)	0(0)	0(0)	0.008(0.008)	0.030(0.015)	0.012(0.010)	0.066(0.022)	0.118(0.028)
		$n = 1000$	0(0)	0.004(0.006)	0.008(0.008)	0(0)	0.022(0.013)	0.044(0.018)	0.008(0.008)	0.040(0.017)	0.100(0.026)
M20		$n = 50$	0(0)	0(0)	0(0)	0.002(0.004)	0.004(0.006)	0.022(0.013)	0.010(0.009)	0.068(0.022)	0.112(0.028)
		$n = 200$	0.250(0.038)	0.420(0.043)	0.542(0.044)	0.958(0.018)	0.974(0.014)	0.982(0.012)	0.016(0.011)	0.060(0.021)	0.128(0.029)
		$n = 1000$	0.334(0.041)	0.466(0.044)	0.570(0.043)	0.976(0.013)	0.990(0.009)	0.998(0.004)	0.004(0.006)	0.038(0.017)	0.096(0.026)

Table 4: Percentages of rejections for testing  $H_0 : j = 2$ , with 500 simulations (1.96 times their estimated standard deviation in parenthesis) and  $B = 500$  bootstrap samples.

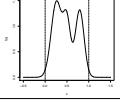
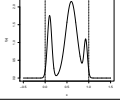
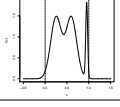
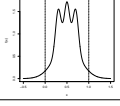
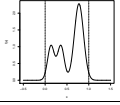
		$\alpha$	0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10
M21		$n = 50$ $n = 100$ $n = 200$	SI			FM			NP		
			0(0)	0.026(0.014)	0.058(0.020)	0.010(0.009)	0.052(0.019)	0.112(0.028)	0.014(0.010)	0.054(0.020)	0.112(0.028)
			0.002(0.004)	0.056(0.020)	0.154(0.032)	0.044(0.018)	0.148(0.031)	0.228(0.037)	0.020(0.012)	0.092(0.025)	0.168(0.033)
			0.024(0.013)	0.104(0.027)	0.208(0.036)	0.076(0.023)	0.202(0.035)	0.278(0.039)	0.050(0.019)	0.134(0.030)	0.194(0.035)
M22		$n = 50$ $n = 100$ $n = 200$	SI			FM			NP		
			0.002(0.004)	0.006(0.007)	0.022(0.013)	0.036(0.016)	0.142(0.031)	0.300(0.040)	0.080(0.024)	0.256(0.038)	0.402(0.043)
			0(0)	0.026(0.014)	0.180(0.034)	0.248(0.038)	0.610(0.043)	0.804(0.035)	0.266(0.039)	0.542(0.044)	0.706(0.040)
			0.052(0.019)	0.500(0.044)	0.806(0.035)	0.740(0.038)	0.960(0.017)	0.982(0.012)	0.890(0.027)	0.956(0.018)	0.980(0.012)
M23		$n = 50$ $n = 100$ $n = 200$	SI			FM			NP		
			0(0)	0.068(0.022)	0.206(0.035)	0.054(0.020)	0.196(0.035)	0.338(0.041)	0.050(0.019)	0.204(0.035)	0.336(0.041)
			0.082(0.024)	0.486(0.044)	0.684(0.041)	0.334(0.041)	0.658(0.042)	0.780(0.036)	0.326(0.041)	0.624(0.042)	0.746(0.038)
			0.570(0.043)	0.780(0.036)	0.840(0.032)	0.832(0.033)	0.906(0.026)	0.938(0.021)	0.722(0.039)	0.878(0.029)	0.934(0.022)
M24		$n = 50$ $n = 100$ $n = 200$	SI			FM			NP		
			0(0)	0.008(0.008)	0.038(0.017)	0.004(0.006)	0.030(0.015)	0.082(0.024)	0.008(0.008)	0.060(0.021)	0.134(0.030)
			0(0)	0.006(0.007)	0.026(0.014)	0.012(0.010)	0.052(0.019)	0.108(0.027)	0.034(0.016)	0.110(0.027)	0.176(0.033)
			0.006(0.007)	0.040(0.017)	0.108(0.027)	0.050(0.019)	0.160(0.032)	0.288(0.040)	0.096(0.026)	0.232(0.037)	0.334(0.041)
M25		$n = 50$ $n = 100$ $n = 200$	SI			FM			NP		
			0.002(0.004)	0.046(0.018)	0.138(0.030)	0.068(0.022)	0.200(0.035)	0.312(0.041)	0.018(0.012)	0.098(0.026)	0.148(0.031)
			0.004(0.006)	0.064(0.021)	0.176(0.033)	0.170(0.033)	0.344(0.042)	0.462(0.044)	0.098(0.026)	0.232(0.037)	0.320(0.041)
			0.008(0.008)	0.084(0.024)	0.198(0.035)	0.190(0.034)	0.404(0.043)	0.552(0.044)	0.106(0.027)	0.248(0.038)	0.356(0.042)

Table 5: Percentages of rejections for testing  $H_0 : j = 2$ , with 500 simulations (1.96 times their estimated standard deviation in parenthesis) and  $B = 500$  bootstrap samples.


		$\alpha$	0.01	0.05	0.10		$\alpha$	0.01	0.05	0.10
$n = 50$ $n = 200$ $n = 1000$	M9		NP (known support)			M16		NP (known support)		
			0(0)	0.014(0.010)	0.034(0.016)			0.002(0.004)	0.024(0.013)	0.076(0.023)
			0.012(0.010)	0.052(0.019)	0.090(0.025)			0.008(0.008)	0.058(0.020)	0.126(0.029)
			0.010(0.009)	0.040(0.017)	0.092(0.025)			0.016(0.011)	0.038(0.017)	0.080(0.024)

Table 6: Percentages of rejections for testing  $H_0 : j = 1$  (in model M9) and  $H_0 : j = 2$  (in model M16), with 500 simulations (1.96 times their estimated standard deviation in parenthesis) and  $B = 500$  bootstrap samples.

watermark, it is estimated that the symbol appears in between 4 and 10 stamps if the sheet was watermarked, and some authors suggested that this watermark appears only once of every 4 sheets. In order to get more information about this particular problem and to obtain some further references, see Izenman and Sommer (1988).

This particular example has been explored in several references in the literature for determining the number of groups. From a non-parametric point of view, some examples of its utilization for mode testing can be shown in Efron and Tibshirani (1994, Ch. 16), Izenman and Sommer (1988) or in Fisher and Marron (2001). Also it was analysed using non-parametric exploratory tools in Wilson (1983), Minnotte and Scott (1993) and in Chaudhuri and Marron (1999). Some parametric studies of the 1872 Hidalgo issue can be found in Basford et al. (1997) or in McLachlan and Peel (2000, Ch. 6).

Taking a subsample of 437 stamps on white wove, Wilson (1983) made a histogram and the conclusion was that only two kinds of paper were used, the *Papel Sellado* and the *La Croix-Freres*, and that there was not a third kind of paper. Izenman and Sommer (1988) revisited the example considering a more complete collection, with 485 stamps. A histogram with the same parameters as those used by Wilson (1983) (same starting point and bin width) is shown in Figure 1 (top-left panel), revealing the same features as those noticed in the original reference. Two groups are also shown by a kernel density estimator, shown in the same plot, considering a gaussian kernel and a rule of thumb bandwidth (see Wand and Jones, 1995, Ch. 3.2). However, both approximations (histogram and kernel density estimator) depend heavily on the bin width and bandwidth, respectively. Specifically, the use of an automatic rule for selecting the bandwidth value (focused on the global estimation of the entire density function) does not guarantee an appropriate recovery of the modes. In fact, using another automatic rule as the plug-in bandwidth (Figure 1, bottom-left panel), nine modes are observed. A histogram with a smaller bin width is also

included in this plot, exhibiting apparently more modes than the initial one.

Given that the exploratory tools did not provide a formal way of determining if there are more than two groups, *Papel Sellado* and *La Croix-Freres*, Izenman and Sommer (1988) employed the multimodality test of Silverman (1981). Note that for this purpose, just FM, SI and the new proposal NP can be used, and the first two proposals present a poor calibration, as shown in the simulation study. Results from Izenman and Sommer (1988), applying SI with  $B = 100$ , are shown in Table 7 (note that with  $B = 500$ , different p-values are obtained). For  $\alpha = 0.05$ , the conclusions are the same, except in the crucial case of testing  $H_0 : j \leq 2$ , where for  $B = 500$ , there are no evidences to reject the null hypothesis. These differences may be caused by the approximations implemented by Izenman and Sommer (1988) to obtain the critical bandwidth. Both Efron and Tibshirani (1994, Ch. 16) (using  $B = 500$  bootstrap replicates) and Salgado-Ugarte et al. (1998) (employing  $B = 600$ ) obtained similar results to ours. Hence, the null hypothesis must not be rejected when the hypothesis is that the distribution has at most two modes, but it has to be rejected when  $H_0$  is that the distribution has at most six modes. This strange behaviour also happens in Izenman and Sommer (1988) analysis, when testing  $H_0 : j \leq 3$  and  $H_0 : j \leq 6$ .

Izenman and Sommer (1988) suggested non-rejecting the null hypothesis the first time that the p-value is higher than 0.4. The consideration of a *flexible* rule for rejecting the null hypothesis is justified by the fluctuations in the p-values of SI and, as Izenman and Sommer (1988) mentioned, by the “conservative” nature of this test. Under this premise, the result when applying SI would be that the null hypothesis is rejected until it is tested  $H_0 : j \leq 7$ . Hence, Izenman and Sommer (1988) conclude that the number of groups in the 1872 Hidalgo Issue is seven.

As shown in Section 3, SI does not present a good calibration and sometimes it can be

	$k$	1	2	3	4	5	6	7	8	9
SI	$B = 100$	0	0.04	0.06	0.01	0	0	0.44	0.31	0.82
	$B = 500$	0.006	0.348	0.092	0.016	0.006	0.002	0.494	0.308	0.616
FM		0	0.04	0	0	0	0	0.06	0.01	0.06
NP		0	0.022	0.004	0.506	0.574	0.566	0.376	0.886	0.808

Table 7: P-values obtained using different proposals for testing  $k$ -modality, with  $k$  between 1 and 9. Methods: SI, FM and NP. For SI method,  $B = 100$  (first row; Izenman and Sommer, 1988) and  $B = 500$ .

also anticonservative. It is not surprising that SI behaves differently when testing  $H_0 : j \leq k$  for  $k = 2, 3$ , with respect to the rest of cases until  $k = 7$ . Since NP has a good calibration behaviour, even with “small” sample sizes, this method is going to be used, first for testing the important case  $H_0 : j = 2$  vs.  $H_a : j > 2$  and then to figure out how many groups are there in the 1872 Hidalgo Issue.

The computation of the excess mass statistic requires a non-discrete sample and the original data (denoted as  $\mathcal{X}$ ) contained repeated values, the artificial sample  $\mathcal{Y} = \mathcal{X} + \mathcal{E}$  will be employed for testing the number of modes, where  $\mathcal{E}$  is a sample of size 485 from the  $U(-5 \cdot 10^{-4}, 5 \cdot 10^{-4})$  distribution. This modification of the data was also considered by Fisher and Marron (2001). The p-values obtained in their studio (using  $B = 200$  bootstrap replicates) are shown in Table 7: it is not clear which conclusion has to be made. They mentioned that their results are consistent with the previous studies, detecting 7 modes. But it should be noticed that, as shown in the simulation, FM does not present a good calibration behaviour.

Finally, the p-values obtained with NP are also shown in the Table 7, with  $B = 500$ . Similar results can be obtained employing the interval  $I = [0.04, 0.15]$  in NP with known

support, as Izenman and Sommer (1988) notice that the thickness of the stamps is always in this interval  $I$ . Employing a significance level  $\alpha = 0.05$  for testing  $H_0 : j = 2$ , it can be observed that the null hypothesis is rejected. It can be seen that the null hypothesis is rejected until  $k = 4$ , and then there is no evidences to reject  $H_0$  employing greater values of  $k$ . Then, applying our new procedure, the conclusion is that the number of groups in the 1872 Hidalgo Issue is four.

In order to compare the results obtained by Izenman and Sommer (1988) and the ones derived applying the new proposal, two kernel density estimators, with gaussian kernel and critical bandwidths  $h_4$  and  $h_7$  are depicted in Figure 1 (bottom-right panel). Izenman and Sommer (1988) conclude that seven modes were present, and argued that the stamps could be divided in, first, three groups (pelure paper with mode at 0.072 mm, related with the forged stamps; the medium paper in the point 0.080 mm; and the thick paper at 0.090 mm). Given the efforts made in the new issue in 1872 to avoid forged stamps, it seems quite reasonable to assume that the group associated with the pelure paper had disappeared in this new issue. In that case, the asymmetry in the first mode using  $h_4$  can be attributed to the modifications in the paper made by the manufacturers. Also, this first and asymmetric group, justifies the application of non-parametric techniques to determine the number of groups. It can be seen, in the Section 7 of Izenman and Sommer (1988) and in other references using mixtures of gaussian densities to model this data (see, for example, McLachlan and Peel, 2000, Ch. 6), that these parametric techniques have problems capturing this asymmetry, and they always determine that there are two modes in this first part of the density, one near the point 0.07 mm and another one near 0.08 mm. For the two modes near the points 0.10 and 0.11 mm, both corresponding to stamps produced in 1872. As Izenman and Sommer (1988) noticed, it seems that the stamps of 1872 were printed on two different paper types, one with the same characteristics as the

unwatermarked white wove paper used in the 1868 issue, and a second much thicker paper that disappeared completely by the end of 1872. Using this explanation, it seems quite reasonable to think that the two final modes using  $h_4$ , corresponds with the medium paper and the thick paper in this second block of stamps produced in 1872. Finally, for the two minor modes appearing near 0.12 and 0.13 mm, when  $h_7$  is used, Izenman and Sommer (1988) do not find an explanation and they mention that probably they could be artefacts of the estimation procedure. This seems to confirm the conclusions obtained with our new procedure. The reason of determining more groups than the four obtained with our proposal, seems to be quite similar to that of the model M20 in our simulation study. This possible explanation is that the spurious data in the right tail of the last mode are causing the rejection of  $H_0$ , when SI is used.

## 5 Discussion

Determining the number of modes in a distribution is a relevant practical problem in many applied sciences. The proposal presented in this paper provides a good performance for the testing problem (1), being in the case of a general number of modes  $k$  the only alternative with a reasonable behaviour. The totally nonparametric testing procedure can be extended to other contexts where a natural nonparametric estimator under the null hypothesis is available. For instance, the method can be adapted for dealing with periodic data, as it happens with the proposal by Fisher and Marron (2001).

In practical problems, where a large number of tests must be computed, obtaining a set of p-values is a crucial task. In this setting, such a computation should be accompanied by the application of FDR correction techniques. The proposal in this work, based on the use of critical bandwidth and excess mass ideas, and its combination with FDR, is

computationally feasible. With the aim of making this procedure accessible for the scientific community, and therefore, enabling its use in large size practical problems, an R package has been developed.



## A Models for simulation study

The specific formulas of those models considered in the simulation study carried out in Section 3 are given here with the notation  $\sum_{i=1}^l p_i \cdot \psi_i$ , where each  $\psi_i$  represents one of the component of the mixture and  $p_i$  are the weights of these different components, with  $i = 1, \dots, l$ , satisfying  $\sum_{i=1}^l p_i = 1$ . The unimodal density functions used as  $\psi_i$  are the following models, as defined in Johnson et al. (1995): Beta( $\theta_i, \phi_i$ ), Gamma( $\alpha_i, \beta_i$ ),  $N(\mu_i, \sigma_i^2)$  and Weibull( $\delta_i, c_i$ ). All the models were created in such a way that  $f(0) \approx f(1) \approx 0.1 \max_{x \in (0,1)} f(x)$ . The unimodal probability density functions are represented in Figure 3, the bimodal and trimodal models appear in Figure 4.

*Unimodal models:*

- M1:  $0.44 \cdot N(0.372, 0.03) + 0.44 \cdot N(0.67, 0.022) + 0.12 \cdot N(0.5, 0.2)$ .
- M2:  $0.9 \cdot N(0.5, 0.05) + 0.05 \cdot N(0.197, 0.01) + 0.05 \cdot N(0.803, 0.01)$ .
- M3:  $0.6 \cdot N(0.62, 0.04) + 0.2 \cdot N(0.218, 0.1) + 0.2 \cdot N(0.5, 0.00795)$ .
- M4:  $N(0.5, 0.05428)$ .
- M5:  $0.9 \cdot N(0.5, 0.0485) + 0.1 \cdot N(0.5, 0.47)$ .
- M6:  $0.6 \cdot N(0.5, 0.0502) + 0.2 \cdot N(0.3, 0.02) + 0.2 \cdot N(0.7, 0.02)$ .
- M7:  $0.5 \cdot \text{Beta}(10, 3) + 0.5 \cdot N(0.5, 0.137)$ .
- M8:  $0.6 \cdot N(0.4985, 0.0793) + 0.4 \cdot \text{Weibull}(3, 0.5)$ .
- M9:  $0.5 \cdot N(0.5, 0.3) + 0.45 \cdot N(0.5, 0.045) + 0.05 \cdot N(0.5, 0.000135)$ .
- M10:  $0.6 \cdot N(0.307, 0.0518) + 0.4 \cdot \text{Gamma}(4, 8)$ .

- M26:  $0.58 \cdot N(0.61, 0.035) + 0.2 \cdot N(0.232, 0.04) + 0.2 \cdot N(0.5, 0.00795) + 0.01 \cdot N(0.15, 0.0028) + 0.01 \cdot N(0.98, 0.0028)$ .

*Bimodal models:*

- M11:  $0.75 \cdot N(0.458, 0.0546) + 0.25 \cdot N(0.85, 0.0041)$ .
- M12:  $0.5 \cdot N(0.211, 0.012) + 0.3 \cdot N(0.75, 0.062) + 0.2 \cdot \text{Beta}(5, 2)$ .
- M13:  $0.95 \cdot N(0.3035, 0.02) + 0.05 \cdot N(0.96757, 0.0004)$ .
- M14:  $0.5 \cdot N(0.776, 0.0109) + 0.3 \cdot N(0.3, 0.04) + 0.1 \cdot N(0.25, 0.0025) + 0.1 \cdot N(0.35, 0.0025)$ .
- M15:  $0.3 \cdot N(0.13, 0.1) + 0.3 \cdot N(0.81, 0.1) + 0.2 \cdot \text{Gamma}(3, 9) + 0.2 \cdot \text{Beta}(7, 2)$ .
- M16:  $0.6 \cdot N(0.384, 0.01202) + 0.2 \cdot N(0.2, 0.05) + 0.2 \cdot N(0.9, 0.00272)$ .
- M17:  $0.5 \cdot N(0.3, 0.0197) + 0.5 \cdot N(0.7, 0.0197)$ .
- M18:  $0.5 \cdot N(0.18, 0.007) + 0.5 \cdot N(0.82, 0.007)$ .
- M19:  $0.5 \cdot N(0.06787, 0.001) + 0.5 \cdot N(0.93213, 0.001)$ .
- M20:  $0.48 \cdot N(0.06777, 0.001) + 0.48 \cdot N(0.93223, 0.001) + 0.02 \cdot \text{Beta}(1.1, 2.37558) + 0.02 \cdot \text{Beta}(2.37558, 1.1)$ .

*Trimodal models:*

- M21:  $0.45 \cdot N(0.26, 0.01476) + 0.33 \cdot N(0.79145, 0.01) + 0.22 \cdot N(0.5, 0.007)$ .
- M22:  $0.68 \cdot N(0.6, 0.0025) + 0.22 \cdot N(0.10245, 0.01588) + 0.1 \cdot N(0.93, 0.0015)$ .
- M23:  $0.45 \cdot N(0.25, 0.015) + 0.45 \cdot N(0.6, 0.015) + 0.1 \cdot N(0.95222, 0.00049)$ .

- M24:  $0.55 \cdot N(0.5, 0.08425) + 0.15 \cdot N(0.3, 0.004) + 0.15 \cdot N(0.5, 0.004) + 0.15 \cdot N(0.7, 0.004)$ .
- M25:  $0.6 \cdot N(0.7749, 0.011) + 0.2 \cdot N(0.1345, 0.006) + 0.2 \cdot N(0.36, 0.006)$ .

## B New proposal when the support is known

When the bounded support is known, an alternative approach for the new proposal can be used in order to get better results in practice. This new proposal consists in replacing the critical bandwidth of Silverman (1981) for the one of Hall and York (2001) in the definition of the calibration function  $g$ . If the number of modes in the entire support is equal to  $k$  when testing  $H_0 : j = k$  (with  $(k - 1)$  antimodes) in  $[a, b]$ , then no more changes are needed. If modes appear outside  $[a, b]$ , then the link function (9) can be used in order to preserve the required regularity conditions. Denoting as  $a < \hat{x}_1 < \dots < \hat{x}_{2k-1} < b$ , being  $\hat{x}_1$  and  $\hat{x}_{2k-1}$  modes, the points  $\hat{x}_0$  and  $\hat{x}_{2k}$ , needed to obtain the values in (11), will be redefined to remove the modes outside  $[a, b]$ . If there are modes lower than  $\hat{x}_1$ , then  $\hat{x}_0 = \min\{x : x \geq a \text{ and } \hat{f}'_{h_{HY,k}}(x) > 0\}$  and if there are modes greater than  $\hat{x}_{2k-1}$ , then  $\hat{x}_{2k} = \max\{x : x \leq b \text{ and } \hat{f}'_{h_{HY,k}}(x) < 0\}$ . Once this change is done, two extra values,  $\mathfrak{a} < \hat{x}_0$  and  $\mathfrak{b} > \hat{x}_{2k}$ , are needed in order to use the link function. The steps to obtain these two values will be defined later on and, from them, the  $g$  function defined in (8) can be modified in its tails as follows

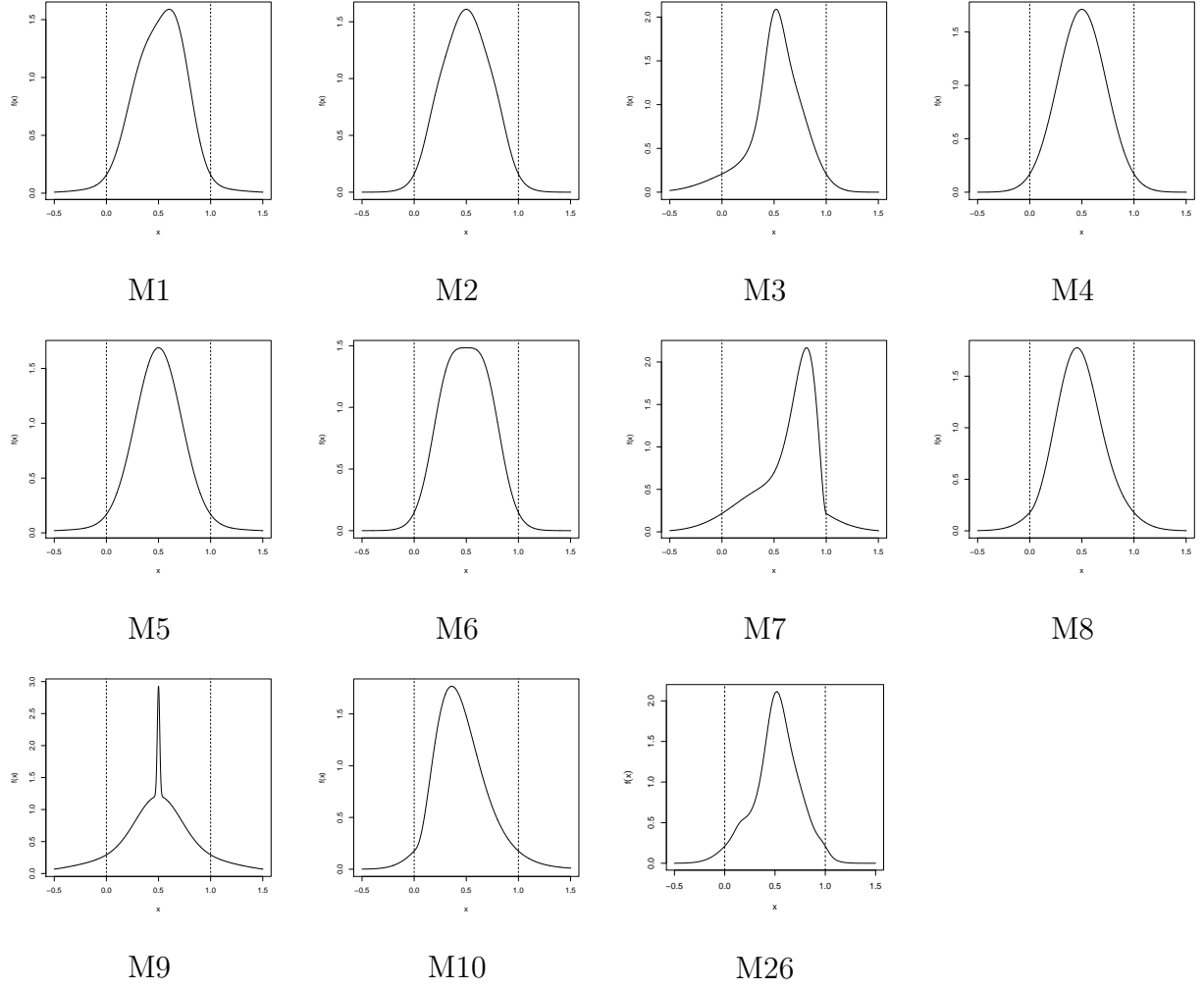


Figure 3: Unimodal density functions: M1–M10 and M26.

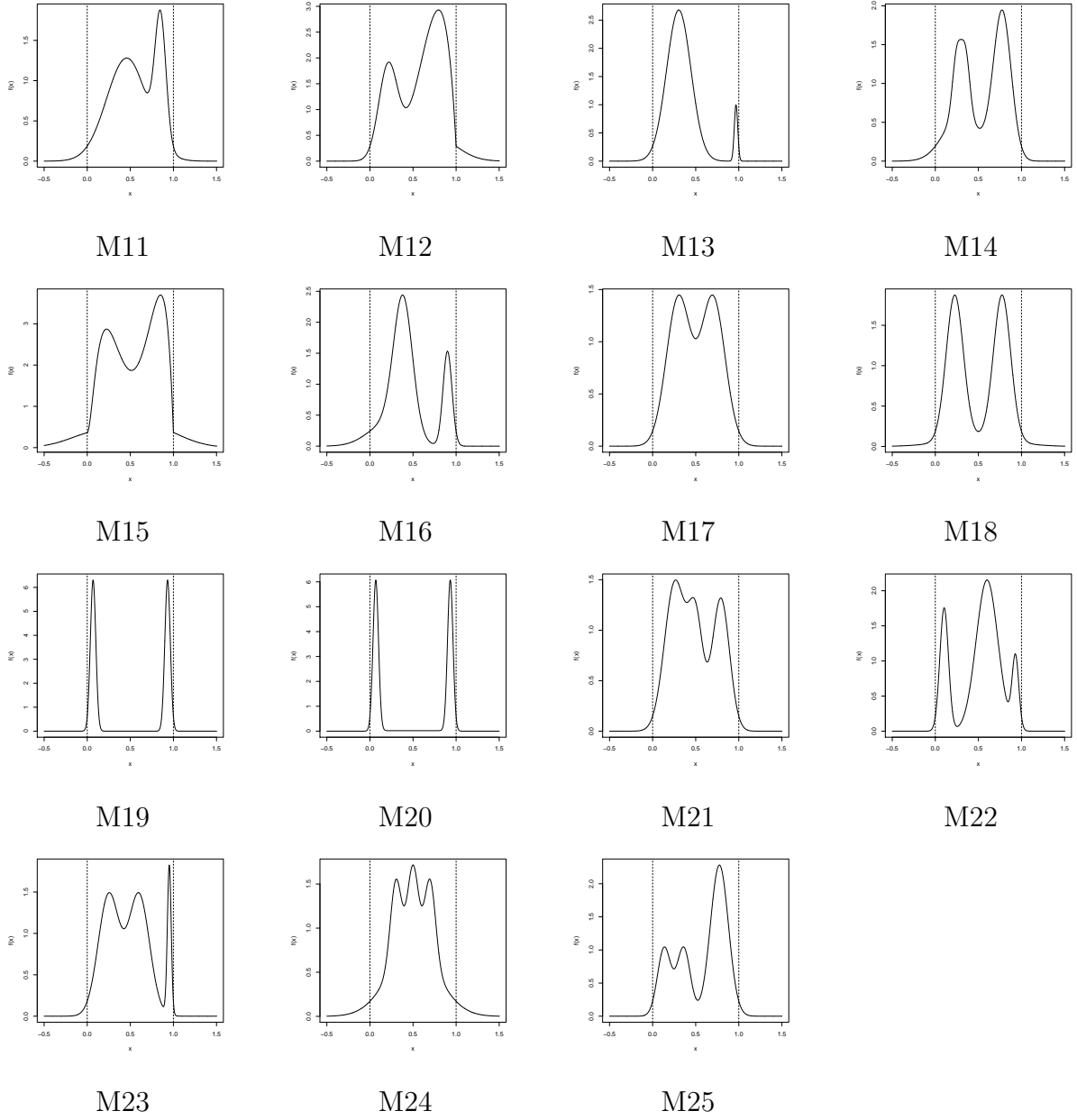


Figure 4: Density functions. M11–M20: bimodal models. M21–M25: trimodal models.

$$g(x; h_{\text{HY},k}, h_{\text{PI}}, \mathfrak{s}, \mathfrak{a}, \mathfrak{b}) = \begin{cases} 0 & \text{if } x \leq \mathfrak{a} \text{ and } \hat{f}_{h_{\text{HY},k}} \text{ has modes lower than } a, \\ l(x; \mathfrak{a}, \hat{x}_0, 0, \hat{f}_{h_{\text{HY},k}}(\hat{x}_0), 0, \hat{f}'_{h_{\text{HY},k}}(\hat{x}_0)) & \text{if } x \in (\mathfrak{a}, \hat{x}_0) \text{ and } \hat{f}_{h_{\text{HY},k}} \text{ has modes lower than } a, \\ J(x; \hat{x}_i, h_{\text{HY},k}, h_{\text{PI}}, \mathfrak{s}_i) & \text{if } x \in (\mathfrak{r}_i, \mathfrak{s}_i) \text{ for some } i \in \{1, \dots, (2k-1)\}, \\ L(x; \zeta_p, h_{\text{HY},k}) & \text{if } x \in (z_{(2p-1)}, z_{(2p)}) \text{ for some } p \in \{1, \dots, t\}, \\ & \text{and } \zeta_p \notin (\mathfrak{r}_i, \mathfrak{s}_i) \text{ for any } i \in \{1, \dots, (2k-1)\}, \\ l(x; \hat{x}_{2k}, \mathfrak{b}, \hat{f}_{h_{\text{HY},k}}(\hat{x}_{2k}), 0, \hat{f}'_{h_{\text{HY},k}}(\hat{x}_{2k}), 0) & \text{if } x \in (\hat{x}_{2k}, \mathfrak{b}) \text{ and } \hat{f}_{h_{\text{HY},k}} \text{ has modes greater than } b, \\ 0 & \text{if } x \geq \mathfrak{b} \text{ and } \hat{f}_{h_{\text{HY},k}} \text{ has modes greater than } b, \\ \hat{f}_{h_{\text{HY},k}}(x) & \text{otherwise,} \end{cases}$$

the functions  $J$  and  $L$  are defined as in Section 2.3, replacing the kernel density estimator  $\hat{f}_{h_k}$  by  $\hat{f}_{h_{\text{HY},k}}$  and changing the values of  $\hat{x}_0$  and  $\hat{x}_{2k}$  as it was pointed out. The neighborhood in which the  $J$  functions are defined is chosen by the same method as in the approach described in the main text. To guarantee that the calibration function is a density, it is also necessary to select correctly the values of  $\mathfrak{a}$  (if  $\hat{f}_{h_{\text{HY},k}}$  has modes lower than  $a$ ) and  $\mathfrak{b}$  (if it has modes greater than  $b$ ) to obtain an integral equal to one. An option is to employ  $\mathfrak{a}$  and  $\mathfrak{b}$  satisfying

$$\int_{-\infty}^{\hat{x}_0} g(x; h_{\text{HY},k}, h_{\text{PI}}, \mathfrak{s}, \mathfrak{a}, \mathfrak{b}) dx + \int_{\hat{x}_{2k}}^{\infty} g(x; h_{\text{HY},k}, h_{\text{PI}}, \mathfrak{s}, \mathfrak{a}, \mathfrak{b}) dx = \int_{-\infty}^{\hat{x}_0} \hat{f}_{h_k}(x) dx + \int_{\hat{x}_{2k}}^{\infty} \hat{f}_{h_k}(x) dx. \quad (14)$$

It can be the case where the equality (14) is not satisfy for any pair  $(\mathfrak{a}, \mathfrak{b})$ , being  $\mathfrak{a} \in (-\infty, \hat{x}_0)$  and  $\mathfrak{b} \in (\hat{x}_{2k}, \infty)$ . If this happens, the calibration function can be divided by the normalizing constant to correct the value of the integral. Another alternative can be to take other values of  $\hat{x}_0 < \hat{x}_1$  and  $\hat{x}_{2k} > \hat{x}_{2k-1}$ , such as  $\hat{f}_{h_{\text{HY},k}}(x) > 0$ , for all  $x \in [\hat{x}_0, \hat{x}_1)$ , and  $\hat{f}'_{h_{\text{HY},k}}(x) < 0$ , for all  $x \in (\hat{x}_{2k-1}, \hat{x}_{2k}]$ .

The approach considered in the simulation study (when the support is known) is, in the tails, try to find the value of  $\mathfrak{a}$  in the interval  $[\hat{x}_0 - b + a, \hat{x}_0)$  and the value of  $\mathfrak{b}$  in

$(\widehat{x}_{2k}, \widehat{x}_{2k} + b - a]$ . If for all the possible values of  $\mathbf{a}$  and  $\mathbf{b}$  the integral

$$q_2 = \int_{-\infty}^{\infty} g(x; h_{\text{HY},k}, h_{\text{PI}}, \boldsymbol{\varsigma}, \mathbf{a}, \mathbf{b}) dx,$$

is not equal to 1, then the solution is take  $\mathbf{a}$  and  $\mathbf{b}$  in such a way that  $q_2$  is as close as possible to one and, then, employ the quotient  $g(\cdot; h_{\text{HY},k}, h_{\text{PI}}, \boldsymbol{\varsigma}, \mathbf{a}, \mathbf{b})/q_2$  as the calibration function.

## C Testing $k$ -modality when the true density has less than $k$ modes

As it has already been mentioned, the methods proposed by Silverman (1981) and Fisher and Marron (2001) can be extended from unimodality to test a general null hypothesis as  $H_0 : j \leq k$ . Nevertheless, the proposal presented in this work just allows to test  $H_0 : j = k$  vs.  $H_0 : j > k$ . The reason why the  $k$ -modal test should not be used when the true underlying density has less than  $k$  modes is that the test statistic in the bootstrap resamples converge in distribution to a random variable, depending only on the values  $\widehat{d}_i$  with  $i = 1, \dots, (2k - 1)$  (see Cheng and Hall, 1998). When  $j < k$ , in the calibration function  $g$ , there exist  $(2k - 2j)$  turning points that they will not converge to any fixed value depending on the real density function. As the (asymptotic) distribution of the test statistic in the bootstrap resamples depends also on this  $(2k - 2j)$  values, one would expect that the sample distribution of the test statistic will not be correctly approximated with the bootstrap resamples.

Testing  $H_0 : j = k$  instead of  $H_0 : j \leq k$  is not in general an important limitation for practical purposes. As it is done in the stamp example in Section 4, the usual procedure is to perform a stepwise algorithm starting with one mode and, if the null hypothesis is

rejected, increasing the number of modes in the null hypothesis by one until there is no evidences for rejection. Despite this note of caution, it can be seen that, generally, testing  $H_0 : j = k$ , when  $j < k$ , reports also good calibration results.

In order to show the accuracy in practice when the bimodality test is employed in unimodal cases, Table 8 reports the percentages of rejections for significance levels  $\alpha = 0.01$ ,  $\alpha = 0.05$  and  $\alpha = 0.10$  for testing bimodality employing the proposals by Silverman (1981) (SI), Fisher and Marron (2001) (FM) and the new proposal (NP) presented in this paper. With this goal, samples of size  $n = 50$ ,  $n = 200$  and  $n = 1000$  were drawn from 10 unimodal distributions (models M1–M10). Again, for each choice of sampling distribution and sample size, 500 realizations were generated. Conditionally on each of those samples, for testing purposes, 500 resamples of size  $n$  were drawn from the population.

The conclusions from the results reported in Table 8 are quite similar to those given previously in Section 3 when  $H_0 : j = 1$  was tested. First, although SI still reports a percentage of rejections below the significance level, it is less conservative than for the initial results (testing  $H_0 : j = 1$ ). For all the available models and all sample sizes and significance levels, the percentage of rejections employing the bimodality test is greater or equal than the one obtained when the unimodality test was applied but lower than the significance level. In fact, taking into account the two largest sample sizes ( $n = 200$  and  $n = 1000$ ), the percentage of rejections is always strictly greater in the bimodal test with the exceptions of the models M3 ( $n = 50$  and  $\alpha = 0.05$ ), M6 ( $n = 200$  and  $\alpha = 0.10$ ) and M10 ( $n = 50$  and  $\alpha = 0.05$ ). Regarding FM, again a systematic behaviour cannot be concluded, being the results similar to those ones reported in Section 3 when unimodality was tested on the same models. Finally, the results for the new proposal seem to be again quite satisfactory, with a slightly conservative performance in some models, such as M2, M3 ( $n = 1000$  and  $\alpha = 0.10$ ), M4, M5 ( $n = 200$ ), M7, M9 and M10 ( $n = 50$ ). Observing these



results, it seems that, in practice, NP can be used for testing  $H_0 : j \leq k$ , but it should be kept in mind that a correct calibration is not guaranteed. An example of poor behaviour can be observed for (unimodal) model M26. Analysing the results reported in Table 9 for NP, it can be seen that, for  $n = 1000$ , when testing  $H_0 : j = 1$  the percentage of rejections is close to the significance level, whereas when testing  $H_0 : j = 2$ , the percentage of rejections is below  $\alpha$ , even employing the correction provided when the support is known.

## D Some theoretical results

The complementary results needed to prove that  $\hat{d}_i$  converge in probability to  $d_i$ , as  $n \rightarrow \infty$ , for  $i = 1, \dots, (2j - 1)$ , where  $j$  is the number of modes of  $f$ , are given in this part. These proves are based on the results of Romano (1988). In order to provide the different results, some further conditions must be considered:

- (C1)  $f$  is a bounded and derivable density function.
- (C2) The second derivative of  $f$  is continuous in a neighbourhood of the order points  $x_1, \dots, x_{2j-1}$ . For all these points,  $f$  satisfies  $f(x_i) \cdot f''(x_i) \neq 0$ ,  $f'(x_i) = 0$  and  $\text{Sign}\{f'(x_{i+})\} = -\text{Sign}\{f'(x_{i-})\}$ .  $f'(x) \neq 0$  if  $x \notin \{x_1, \dots, x_{2j-1}\}$ .
- (C3)  $f$  has a bounded support  $[a, b]$ .
- (C4)  $f'(a+) > 0$  and  $f'(b-) < 0$ .
- (C5)  $f$  is twice continuously differentiable on  $[a, b]$ .
- (C6) Denoting as  $a_n$  and  $b_n$  the fixed sequence of numbers such as  $(nb_n^2)/(\log(n)a_n) \rightarrow \infty$  and  $a_n \rightarrow 0$ . The bandwidth  $h_{(n)}$  satisfies  $b_n \leq h_{(n)} \leq a_n$  and  $f_{h_{(n)}}$  has  $j$  modes and  $(j - 1)$  antimodes.

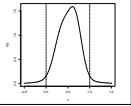
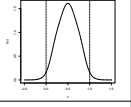
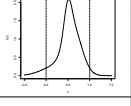
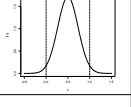
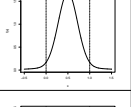
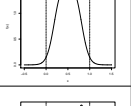
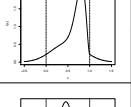
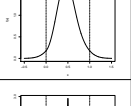
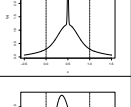
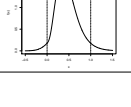
		$\alpha$	0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10
			SI			FM			NP		
M1		$n = 50$	0(0)	0.004(0.006)	0.042(0.018)	0.004(0.006)	0.040(0.017)	0.120(0.028)	0.014(0.010)	0.054(0.020)	0.120(0.028)
		$n = 200$	0(0)	0.010(0.009)	0.086(0.025)	0.008(0.008)	0.094(0.026)	0.180(0.034)	0.010(0.009)	0.042(0.018)	0.096(0.026)
		$n = 1000$	0(0)	0.012(0.010)	0.045(0.018)	0.010(0.009)	0.052(0.020)	0.138(0.030)	0.018(0.011)	0.052(0.020)	0.092(0.025)
M2		$n = 50$	0(0)	0.002(0.004)	0.018(0.012)	0.002(0.004)	0.020(0.012)	0.046(0.018)	0.006(0.007)	0.060(0.021)	0.126(0.029)
		$n = 200$	0(0)	0.004(0.006)	0.032(0.015)	0(0)	0.014(0.010)	0.060(0.021)	0.008(0.008)	0.032(0.015)	0.088(0.025)
		$n = 1000$	0(0)	0.010(0.009)	0.020(0.012)	0(0)	0.016(0.011)	0.038(0.017)	0.010(0.009)	0.042(0.018)	0.072(0.023)
M3		$n = 50$	0(0)	0(0)	0.020(0.012)	0.010(0.009)	0.054(0.020)	0.146(0.031)	0.004(0.006)	0.032(0.015)	0.064(0.021)
		$n = 200$	0(0)	0.012(0.010)	0.030(0.015)	0.008(0.008)	0.064(0.021)	0.150(0.031)	0.004(0.006)	0.028(0.014)	0.066(0.022)
		$n = 1000$	0(0)	0.002(0.004)	0.020(0.012)	0.002(0.004)	0.042(0.018)	0.110(0.027)	0.006(0.007)	0.034(0.016)	0.062(0.021)
M4		$n = 50$	0(0)	0.002(0.004)	0.016(0.011)	0.004(0.006)	0.020(0.012)	0.052(0.019)	0.010(0.009)	0.066(0.022)	0.112(0.028)
		$n = 200$	0(0)	0.002(0.004)	0.012(0.010)	0.002(0.004)	0.008(0.008)	0.026(0.014)	0.012(0.010)	0.028(0.014)	0.050(0.019)
		$n = 1000$	0(0)	0.012(0.010)	0.032(0.015)	0(0)	0.014(0.010)	0.048(0.019)	0.006(0.007)	0.046(0.018)	0.102(0.027)
M5		$n = 50$	0(0)	0.034(0.016)	0.100(0.026)	0.056(0.020)	0.168(0.033)	0.244(0.038)	0.008(0.008)	0.044(0.018)	0.094(0.026)
		$n = 200$	0.002(0.004)	0.032(0.015)	0.114(0.028)	0.130(0.029)	0.298(0.040)	0.414(0.043)	0.002(0.004)	0.026(0.014)	0.084(0.024)
		$n = 1000$	0.004(0.006)	0.008(0.008)	0.048(0.019)	0.072(0.023)	0.228(0.037)	0.364(0.042)	0.010(0.009)	0.046(0.018)	0.094(0.026)
M6		$n = 50$	0(0)	0.002(0.004)	0.014(0.010)	0(0)	0.012(0.010)	0.036(0.016)	0.010(0.009)	0.048(0.019)	0.110(0.027)
		$n = 200$	0(0)	0.018(0.012)	0.052(0.019)	0.010(0.009)	0.042(0.018)	0.072(0.023)	0.008(0.008)	0.058(0.020)	0.108(0.027)
		$n = 1000$	0.002(0.004)	0.024(0.013)	0.074(0.023)	0.002(0.004)	0.030(0.015)	0.082(0.024)	0.014(0.010)	0.052(0.019)	0.098(0.026)
M7		$n = 50$	0(0)	0.014(0.010)	0.052(0.019)	0.070(0.022)	0.186(0.034)	0.338(0.041)	0.002(0.004)	0.016(0.011)	0.042(0.018)
		$n = 200$	0(0)	0.010(0.009)	0.044(0.018)	0.064(0.021)	0.188(0.034)	0.320(0.041)	0.004(0.006)	0.034(0.016)	0.078(0.024)
		$n = 1000$	0(0)	0.002(0.004)	0.028(0.014)	0.038(0.017)	0.160(0.032)	0.262(0.039)	0.008(0.008)	0.042(0.018)	0.092(0.025)
M8		$n = 50$	0(0)	0.004(0.006)	0.024(0.013)	0.002(0.004)	0.024(0.013)	0.050(0.019)	0.008(0.008)	0.034(0.016)	0.084(0.024)
		$n = 200$	0(0)	0.008(0.008)	0.034(0.016)	0(0)	0.024(0.013)	0.064(0.021)	0.006(0.007)	0.040(0.017)	0.076(0.023)
		$n = 1000$	0(0)	0.008(0.008)	0.028(0.014)	0(0)	0.020(0.012)	0.058(0.020)	0.008(0.008)	0.044(0.018)	0.102(0.027)
M9		$n = 50$	0(0)	0.010(0.009)	0.038(0.017)	0.014(0.010)	0.064(0.021)	0.132(0.030)	0.004(0.006)	0.024(0.013)	0.066(0.022)
		$n = 200$	0(0)	0.010(0.009)	0.030(0.015)	0.036(0.016)	0.222(0.036)	0.376(0.042)	0.010(0.009)	0.044(0.018)	0.092(0.025)
		$n = 1000$	0(0)	0.006(0.007)	0.022(0.013)	0.032(0.015)	0.188(0.034)	0.334(0.041)	0.010(0.009)	0.048(0.019)	0.110(0.027)
M10		$n = 50$	0(0)	0(0)	0.014(0.010)	0(0)	0.026(0.014)	0.066(0.022)	0.002(0.004)	0.026(0.014)	0.072(0.023)
		$n = 200$	0(0)	0.014(0.010)	0.034(0.016)	0.004(0.006)	0.034(0.016)	0.082(0.024)	0.008(0.008)	0.042(0.018)	0.098(0.026)
		$n = 1000$	0(0)	0.008(0.008)	0.056(0.020)	0.006(0.007)	0.048(0.019)	0.096(0.026)	0.002(0.004)	0.042(0.018)	0.088(0.025)

Table 8: Percentages of rejections for testing  $H_0 : j \leq 2$  (or  $H_0 : j = 2$  in the NP case), with 500 simulations (1.96 times their estimated standard deviation in parenthesis) and  $B = 500$  bootstrap samples.

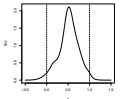
		$\alpha$	0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10
M26			NP (unknown support)			NP (unknown support)			NP (known support)		
		$n = 50$	0.008(0.008)	0.032(0.015)	0.066(0.022)	0.004(0.006)	0.028(0.014)	0.048(0.019)	0.002(0.004)	0.022(0.013)	0.076(0.023)
		$n = 200$	0(0)	0.022(0.013)	0.068(0.022)	0.006(0.007)	0.030(0.015)	0.074(0.023)	0.004(0.006)	0.034(0.016)	0.062(0.021)
		$n = 1000$	0.004(0.006)	0.040(0.017)	0.080(0.024)	0(0)	0.020(0.012)	0.054(0.020)	0(0)	0.026(0.014)	0.052(0.019)

Table 9: Percentages of rejections for testing  $H_0 : j = 1$  (first column) and  $H_0 : j = 2$  (second and third column), with 500 simulations (1.96 times their estimated standard deviation in parenthesis) and  $B = 500$  bootstrap samples.

(C7) Let  $S_n = S_n(X_1, \dots, X_n)$  be any statistic such that for some real values  $s_1$  and  $s_2$ ,

$$0 < s_1 \leq \liminf_{n \rightarrow \infty} S_n \leq \limsup_{n \rightarrow \infty} S_n \leq s_2 < \infty \quad a.s.$$

(C8) Let  $\eta_{(n;l)}$  be a fixed sequence of numbers with  $\eta_{(n;l)} \rightarrow 0$  and  $n\eta_{(n;l)}^{2l+1}/\log(n) \rightarrow \infty$ .

Under assumptions (C1), (C2) and (C6), denoting as  $\hat{x}_i$  to the nearest mode or antimode of  $\hat{f}_{h(n)}$  to  $x_i$ , for  $i \in \{1, \dots, 2j - 1\}$ , then

$$(R1) \quad \hat{x}_i \rightarrow x_i \quad a.s.$$

$$(R2) \quad \hat{f}_{h(n)}(\hat{x}_i) \rightarrow f(\hat{x}_i) \quad a.s$$

Also, if (R1) is satisfied, employing the Proposition 2.1. of Romano (1988), it can be seen that under the assumptions (C2), (C7) and (C8), taking the bandwidth  $h_{(n;l)} = \eta_{(n;l)} \cdot S_n$ . Given the order of the derivative  $l$ , then, for all  $i \in \{1, \dots, 2j - 1\}$ ,

$$(R3) \quad \hat{f}_{h(n;l)}^{(l)}(\hat{x}_i) \rightarrow f^{(l)}(x_i) \quad a.s..$$

Note that under (C7) and (C8), taking  $a_n = \eta_{(n;0)} \cdot s_2$  and  $b_n = \eta_{(n;0)} \cdot s_1$ , the first part of condition (C6) is satisfied, so employing the bandwidth  $h_{(n;0)} = \eta_{(n;0)} \cdot S_n$ , if it has  $j$  modes and  $(j - 1)$  antimodes, (C6) can be replaced by (C7) and (C8).

Although in this document the results are given under more general conditions, for the important case of the critical bandwidth  $h_j$ , employing the results of Mammen et al. (1992), it can be seen that under (C1)–(C5),  $h_j$  always satisfy (C6). Based on this fact, combined with previous results, it can be concluded that, under (C1)–(C5),  $\hat{d}_i$  converge in probability to  $d_i$ , as  $n \rightarrow \infty$ , for  $i = 1, \dots, (2j - 1)$ . Note that for having this last result no more conditions than those ones given as *Assumptions* in Mammen et al. (1992) are needed.

To prove the results given in (R1) and (R2), in this document, the focus will be when  $x_i$  is a mode, that is,  $i = 1, 3, \dots, 2j - 1$ . The proof for the antimodes is analogue to the given one. To begin the proof, first, it will be introduced the function

$$f_h(t) = \frac{1}{h} \int_{-\infty}^{\infty} K\left(\frac{t-y}{h}\right) f(y) dy.$$

Taking  $x_0 = 2a - x_1$  and  $x_{2j} = 2b - x_{2j-1}$  (note that condition (C3) is not needed in this case, if the support is unbounded these values can be replace conveniently with other values  $x_0 < x_1$  and  $x_{2j} > x_{2j-1}$ ), let  $\delta_i$  be the values such as  $\delta_i \leq \min\{(x_i - x_{i-1})/2, (x_{i+1} - x_i)/2\}$  and also the following conditions are satisfied (see Romano, 1988, Proposition 5.1. and Corollary 5.1.)

$$\lim_{n \rightarrow \infty} \sup_{\{h: b_n \leq h \leq a_n\}} \sup_{\{t: |t-x_i| < \delta_i\}} |f_h(t) - f(t)| \leq \lim_{n \rightarrow \infty} \sup_{\{h: 0 < h \leq a_n\}} \sup_{\{t: |t-x_i| < \delta_i\}} |f_h(t) - f(t)| = 0. \quad (15)$$

$$\sup_{\{h: b_n \leq h \leq a_n\}} \sup_{\{t: |t-x_i| < \delta_i\}} |\hat{f}_h(t) - f_h(t)| \leq \sup_{\{h: b_n \leq h \leq a_n\}} \sup_t |\hat{f}_h(t) - f_h(t)| \rightarrow 0 \quad c.s. \quad (16)$$

Combining (15) and (16), it is obtained that

$$\mathbb{P} \left( \lim_{n \rightarrow \infty} \sup_{\{h: b_n \leq h \leq a_n\}} \sup_{\{t: |t-x_i| < \delta_i\}} |\hat{f}_h(t) - f(t)| = 0 \right) = 1 \quad (17)$$

Using that by definition, fixed  $i \in \{1, 3, \dots, 2j-1\}$ , for any  $\delta \in (0, \delta_i)$ ,

$$\sup_{t: \delta < |t-x_i| < \delta_i} f(t) < f(x_i),$$

the continuity of  $f$ , its Taylor expansion and  $\int_{-\infty}^{\infty} K(z)dz = 1$ , it can be seen that

$$\lim_{n \rightarrow \infty} \sup_{\{h: b_n \leq h \leq a_n\}} \sup_{\{t: \delta \leq |t-x_i| < \delta_i\}} f_h(t) \leq \lim_{n \rightarrow \infty} \sup_{\{h: 0 < h \leq a_n\}} \sup_{\{t: \delta \leq |t-x_i| < \delta_i\}} \int_{-\infty}^{\infty} K(z) f(t-zh) dz < f(x_i). \quad (18)$$

Combining (16) and (18) yields

$$\mathbb{P} \left( \lim_{n \rightarrow \infty} \sup_{\{h: b_n \leq h \leq a_n\}} \sup_{\{t: \delta \leq |t-x_i| < \delta_i\}} \hat{f}_h(t) < f(x_i) \right) = 1 \quad (19)$$

Denoting as  $\hat{l}_{i,h}$  the point where the supreme of  $\hat{f}_h$  is obtained, for a given value of  $h$  satisfying  $b_n \leq h \leq a_n$  in the interval  $(x_i - \delta, x_i + \delta)$  and then combining (17) and (19), it yields

$$\mathbb{P} \left( \lim_{n \rightarrow \infty} \sup_{\{h: b_n \leq h \leq a_n\}} |\hat{l}_{i,h} - x_i| \geq \delta \right) = 0$$

Since this can be done for any  $\delta \in (0, \delta_i)$ , this proves

$$\sup_{\{h: b_n \leq h \leq a_n\}} |\hat{l}_{i,h} - x_i| \rightarrow 0 \quad a.s.,$$

now, employing (17) and (19) with the fact that  $\hat{f}_h$  is continuous, for each  $i \in \{1, 3, \dots, (2j-1)\}$ , yields that  $\hat{l}_{i,h}$  is a mode of  $\hat{f}_h$  and it is satisfied

$$\sup_{\{h: b_n \leq h \leq a_n\}} \sup_{\{t: |t-x_i| < \delta_i\}} \hat{f}_h(t) \rightarrow f(t) \quad a.s., \quad (20)$$

Finally, using that  $\hat{f}_{h(n)}$  has  $j$  modes and  $\hat{x}_i$  was defined as the closest mode of  $\hat{f}_{h(n)}$  to  $x_i$  necessarily

$$|\hat{x}_i - x_i| = |\hat{l}_{i,h(n)} - x_i| \leq \sup_{\{h: b_n < h \leq a_n\}} |\hat{l}_{i,h} - x_i| \rightarrow 0 \quad c.s. \quad (21)$$

Results (R1) and (R2) are derived from (20) and (21).  $\square$

## E Numerical approximations

Details of the numerical approaches used in this paper can be found below. All the functions were implemented in the statistical software R Core Team (2016).

### Practical computation of the critical bandwidth

To obtain both the critical bandwidth of Silverman (1981) and that one of Hall and York (2001), a binary search procedure was used. In each step of the algorithm, denoting  $h_a$  as the bandwidth in such a way than  $\hat{f}_{h_a}$  has at most  $k$  modes and  $h_b$  the bandwidth for which the kernel density estimation has more than  $k$  modes (both in the interval  $I$  if the critical bandwidth of Hall and York (2001) is being calculated). Then the dichotomy algorithm is stopped when  $(h_b - h_a) < (h_{a_0}/2^{10})$ , where  $h_{a_0}$  is the initial value of  $h_a$ . The last calculated value of  $h_a$  is the one employed as the critical bandwidth.

### Practical computation of the excess mass

To obtain the excess mass defined by Müller and Sawitzki (1991) when the null hypothesis of unimodality is being tested, the following result is used: the value of the excess mass statistic is exactly twice the value of the *dip* statistic introduced by Hartigan and Hartigan (1985). The value of the *dip* was obtained using the `dipTest` package implemented by Maechler (2015).

For the general case, the following algorithm will be employed to obtain the excess mass statistic when the null hypothesis  $H_0 : j = k$ , with  $k > 1$ , is tested. First, assume that  $d_k(p)$  is the minimum distance of the union of  $k$  intervals containing  $p$  data points. To get the possible values of  $\lambda$  corresponding to  $E_{n,k}(\mathbb{P}_n, \lambda)$ , and candidates to minimize the difference

$$D_{n,k+1}(\lambda) = \{E_{n,k+1}(\mathbb{P}_n, \lambda) - E_{n,k}(\mathbb{P}_n, \lambda)\},$$

the first step begins in  $q_k(1) = n$  (where the number in parenthesis is the current iteration). Then, it search from  $q_k(1) - 1$  until  $(k + 1)$  the integer  $q_k(2)$  minimizing the following expression

$$\lambda_k(1) = \min_{q_k(2)} \frac{q_k(1) - q_k(2)}{n(d_k(q_k(1)) - d_k(q_k(2)))},$$

the value of  $\lambda_k(1)$  is one of the possible values of  $\lambda$  minimizing  $D_{n,k+1}(\lambda)$ . Then, if  $q_k(2) = (k + 1)$  is the value minimizing the previous expression, the algorithm will be stopped, otherwise it is continued until  $q_k(t_k) = (k + 1)$  (where  $t_k$  is the realised number of iterations). After obtaining the vector  $\boldsymbol{\lambda}_k$  of possible values of  $\lambda$  minimizing  $D_{n,k+1}(\lambda)$ , the algorithm is repeated for  $(k + 1)$  to get the vector  $\boldsymbol{\lambda}_{k+1}$ . Then the excess mass statistic is easy to obtain using that

$$\Delta_{n,k+1} = \min_{\boldsymbol{\lambda}_k \cup \boldsymbol{\lambda}_{k+1}} D_{n,k+1}(\lambda).$$

To get the values of  $d_k(p)$ , one can obtain the exact result of the excess mass employing the algorithm provided by Müller and Sawitzki (1991). A similar algorithm was implemented and employed to get the exact results in Section 4. The problem of employing this algorithm is the high computing cost for an extensive study. For this reason an approximation was employed in Section 3 to get the excess mass statistic,  $\Delta_{n,3}$ . The new algorithm

consist in, first, calculate  $d_1(p)$  to obtain the exact values of  $\lambda_1$  and secondly create a grid of  $l$  possible values between each  $\lambda_1(j)$  and  $\lambda_1(j+1)$ , with  $j$  and entire value between 1 and  $(t_k - 1)$ . Finally, to get the test statistic, that value of  $\lambda$  belonging the entire grid and minimising  $D_{n,3}(\lambda)$  is chosen. The employed size  $l$  was:  $l = 100$  when  $n = 50$ ,  $l = 40$  for  $n = 100$ ,  $l = 20$  if  $n = 200$  and  $l = 5$  when  $n = 1000$ . This selection of points represents a balance between the accuracy and the computation time, as, in general, if  $n$  is large then the length of the vector  $t_k$  is also large.

## References

- Basford, K. E., G. J. McLachlan, and M. G. York (1997). Modelling the distribution of stamp paper thickness via finite normal mixtures: The 1872 hidalgo stamp issue of mexico revisited. *Journal of Applied Statistics* 24, 169–180.
- Chaudhuri, P. and J. S. Marron (1999). Sizer for exploration of structures in curves. *Journal of the American Statistical Association* 94, 807–823.
- Cheng, M. Y. and P. Hall (1998). Calibrating the excess mass and dip tests of modality. *Journal of the Royal Statistical Society. Series B* 60, 579–589.
- Efron, B. and R. J. Tibshirani (1994). *An Introduction to the Bootstrap*. United States of America: Chapman and Hall.
- Fisher, N. I. and J. S. Marron (2001). Mode testing via the excess mass estimate. *Biometrika* 88, 419–517.
- Good, I. J. and R. A. Gaskins (1980). Density estimation and bump-hunting by the



- penalized likelihood method exemplified by scattering and meteorite data. *Journal of the American Statistical Association* 75, 42–56.
- Hall, P. and M. York (2001). On the calibration of silverman’s test for multimodality. *Statistica Sinica* 11, 515–536.
- Hartigan, J. A. and P. M. Hartigan (1985). The dip test of unimodality. *Journal of the American Statistical Association* 86, 738–746.
- Izenman, A. J. and C. J. Sommer (1988). Philatelic mixtures and multimodal densities. *Journal of the American Statistical Association* 83, 941–953.
- Johnson, N. L., S. Kotz, and N. Balakrishnan (1995). *Continuous Univariate Distributions*, Volume 1 and 2. New York: Wiley Series in Probability and Statistics.
- Maechler, M. (2015). *dipTest: Hartigan’s Dip Test Statistic for Unimodality - Corrected*. R package version 0.75-7.
- Mammen, E., J. S. Marron, and N. I. Fisher (1992). Some asymptotics for multimodality tests based on kernel density estimates. *Probability Theory and Related Fields* 91, 115–132.
- Marron, J. S. and H. P. Schmitz (1992). Simultaneous density estimation of several income distributions. *Econometric Theory* 8, 476–488.
- McLachlan, G. and D. Peel (2000). *Finite Mixture Models*. United States of America: John Wiley & Sons.
- Minnotte, M. C., D. J. Marchette, and E. J. Wegman (1998). The bumpy road to the mode forest. *Journal of Computational and Graphical Statistics* 7, 239–251.

- Minnotte, M. C. and D. W. Scott (1993). The mode tree: A tool for visualization of nonparametric density features. *Journal of Computational and Graphical Statistics* 2, 51–68.
- Mitchell, J. F., K. A. Sundberg, and J. H. Reynolds (2007). Differential attention-dependent response modulation across cell classes in macaque visual area v4. *Neuron* 55, 131–141.
- Müller, D. W. and G. Sawitzki (1991). Excess mass estimates and tests for multimodality. *The Annals of Statistics* 13, 70–84.
- Olden, J. D., Z. S. Hogan, and M. Zanden (2007). Small fish, big fish, red fish, blue fish: size-biased extinction risk of the world’s freshwater and marine fishes. *Global Ecology and Biogeography* 16, 694–701.
- R Core Team (2016). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Roeder, K. (1990). Density estimation with confidence sets exemplified by superclusters and voids in the galaxies. *Journal of the American Statistical Association* 85, 617–624.
- Romano, J. P. (1988). On weak convergence and optimality of kernel density estimates of the mode. *The Annals of Statistics* 16, 629–647.
- Salgado-Ugarte, I. H., M. Shimizu, and T. Taniuchi (1998). Nonparametric assessment of multimodality for univariate data. *Stata Technical Bulletin* 7, 27–35.
- Silverman, B. W. (1981). Using kernel density estimates to investigate multimodality. *Journal of the Royal Statistical Society. Series B* 43, 97–99.

Wand, M. P. and M. C. Jones (1995). *Kernel Smoothing*. Great Britain: Chapman and Hall.

Wilson, I. G. (1983). Add a new dimension to your philately. *The American Philatelist* 97, 342–349.