Test (2014) 23:51–71
DOI 10.1007/s11749-013-0341-7

ORIGINAL PAPER

Assessing mean and median filters in multiple testing for large-scale imaging data

Chunming Zhang

Received: 15 January 2013 / Accepted: 31 July 2013 / Published online: 25 August 2013

Abstract A new multiple testing procedure, called the FDR_L procedure, was proposed by Zhang et al. (Ann Stat 39:613–642, 2011) for detecting the presence of spatial signals for large-scale 2D and 3D imaging data. In contrast to the conventional multiple testing procedure, the FDR_L procedure substitutes each p-value by a locally aggregated median filter of p-values. This paper examines the performance of another commonly used filter, mean filter, in the FDR_L procedure. It is demonstrated that when the p-values are independent and uniformly distributed under the true null hypotheses, (i) in view of estimating the resulting false discovery rate, the mean filter better alleviates the "*lack of identification phenomenon*" of the FDR_L procedure than the median filter; (ii) in view of signal detection, the median filter enjoys the "*edge-preserving property*" and lends support to its better performance in detecting sparse signals than the mean filter.

Keywords Brain fMRI · False discovery rate · Median · p-value · Sensitivity · Specificity

Mathematics Subject Classification (2010) 62H35 · 62G10 · 62P10 · 62E20

1 Introduction

The multiple testing procedure plays an important role in detecting the presence of spatial signals for large-scale imaging data. Typically, the spatial signals are sparse but clustered. See Zhang and Yu (2008) for an application of multiple testing to detecting regions of significant activation for brain fMRI data, and Efron (2010)'s monograph for more insightful discussions. Zhang et al. (2011) provided empirical

C. Zhang (✉)
University of Wisconsin-Madison, Madison, WI, USA
e-mail: cmzhang@stat.wisc.edu

evidence that for a range of commonly used control levels, the conventional false discovery rate (FDR) controlling procedure (Storey et al. 2004) can lack the ability to detect statistical significance, even if the p-values under the true null hypotheses are independent and uniformly distributed; more generally, ignoring the neighboring information of spatially structured data will tend to diminish the detection effectiveness of the FDR procedure. Zhang et al. (2011) first introduced a scalar quantity $\alpha_\infty^{\mathrm{FDR}}$ to characterize the extent to which the "*lack of identification phenomenon*" (LIP) of the FDR procedure occurs, then proposed a new multiple comparison procedure, called the FDR_L procedure, to accommodate the spatial information of neighboring p-values, via a locally aggregated filter of p-values. When the median filter is utilized, Zhang et al. (2011) investigated theoretical properties of the FDR_L procedure under weak dependence of p-values. It was shown that the FDR_L procedure alleviates the LIP of the FDR procedure, i.e. $\alpha_\infty^{\mathrm{FDR}_L} \le \alpha_\infty^{\mathrm{FDR}}$, thus substantially facilitating the selection of more stringent control levels.

It is natural to ask whether and to what extent the above conclusion for the median filter holds for other commonly used filters, for example, the mean filter, and if so, which one performs better. In the statistical estimation literature, it is well-known that the "sample mean" has an asymptotically smaller variance than the "sample median" in estimating a population mean (van-der Vaart 1998); conversely, the "sample median" is more robust against outlying observations than the "sample mean" (Brown 1983). Meanwhile, in the literature on image signal processing, it is widely believed that the median filter possesses the "edge-preserving property" better than the mean filter. Some new insight to this belief is offered in Arias-Castro and Donoho (2009). It remains unclear whether the median filter outperforms the mean filter in FDR controlling multiple testing procedures. As far as we know, this issue has not been clearly and carefully addressed in the existing literature on multiple testing procedures.

This paper aims to examine the performance of the mean filter in the FDR_L procedure, from viewpoints of both the estimation of FDR and the detection of signals. (It is anticipated that other filters can be studied similarly.) Theorems 1–2 confirm that similar to the median filter, the FDR_L procedure using the mean filter also reduces the extent of the LIP of the FDR procedure, when the p-values are independent and uniformly distributed under the true null hypotheses. There, technical manipulations are different from those used in Zhang et al. (2011) for the median filter. Furthermore, two recommendations are made for practical applications. (i) It is demonstrated that the mean filter better alleviates the "*lack of identification phenomenon*" of the FDR procedure than the median filter. Namely, $\alpha_\infty^{\mathrm{FDR}_L}$ associated with the mean filter is smaller than that associated with the median filter; (ii) the median filter enjoys the "*edge-preserving property*" and lends support to its better performance in detecting sparse signals than the mean filter. In cases with either more strongly correlated p-values or less sparse signals, our empirical results suggest that the distinction between the performance of the two types of filters diminishes.

Assessing and comparing the magnitude of LIP in part (i) are non-trivial. The distribution of the "sample mean" of i.i.d. standard uniform variables will be invoked in a suitable manner. The rest of the paper is organized as follows. Section 2 reviews the FDR and FDR_L procedures. Section 3.1 discusses the distribution of the mean filter applied to i.i.d. standard uniform random variables; Sect. 3.2 explores properties

Table 1 Outcomes from testing n null hypotheses

State \ Decision	Retain null	Reject null	Total
Null is true	U	V	n_0
Non-null is true	T	S	n_1
Total	W	R	n

of $\alpha_\infty^{\mathrm{FDR}_L}$ using the mean filter. Section 4 presents simulation comparisons of the FDR procedure and the mean and median filtered FDR_L procedures in 2D dependent data, and 1D data with serially clustered signals. Section 5 ends the paper with a brief discussion. The Appendix contains technical conditions and detailed derivations.

2 The FDR and FDR_L controlling procedures

We begin with a brief review of the FDR and FDR_L controlling procedures relevant to the study. For testing a family of null hypotheses, $\{H_{0,i}\}_{i=1}^n$, versus non-null hypotheses, $\{H_{1,i}\}_{i=1}^n$, associated with the p-values $\{p_i\}_{i=1}^n$, Table 1 summarizes the outcomes when applying some significance rule. The false discovery rate, defined as $\mathrm{FDR} = E(\frac{V}{R\vee 1})$, depicts the expected proportion of true null hypotheses rejected out of the total number of null hypotheses rejected (Benjamini and Hochberg 1995), where $R\vee 1 = \max\{R, 1\}$.

2.1 The FDR procedure

An empirical process definition of FDR, by means of $\mathrm{FDR}(t) = E\{\frac{V(t)}{R(t)\vee 1}\}$ for $t \in [0, 1]$, was introduced by Storey et al. (2004), where $V(t) = \sum_{i=1}^n \mathrm{I}(H_{0,i}$ is true, $p_i \le t)$ and $R(t) = \sum_{i=1}^n \mathrm{I}(p_i \le t)$, with $\mathrm{I}(\cdot)$ denoting an indicator function. Other quantities, $U(t)$, $T(t)$, $S(t)$ and $W(t)$, can be defined similarly. There, the point estimate of $\mathrm{FDR}(t)$ is given by

$$\widehat{\mathrm{FDR}}(t) = \frac{W(\lambda)t}{\{R(t)\vee 1\}(1-\lambda)}, \tag{2.1}$$

with the pointwise limit expressed as

$$\widehat{\mathrm{FDR}}^{\infty}(t) = \frac{[\pi_0\{1-G_0(\lambda)\} + \pi_1\{1-G_1(\lambda)\}]t}{\{\pi_0 G_0(t) + \pi_1 G_1(t)\}(1-\lambda)}, \tag{2.2}$$

where $\lambda \in (0, 1)$ is a tuning constant, $\pi_0 = \lim_{n\to\infty} n_0/n$, $\pi_1 = 1 - \pi_0$, and $G_0(t) = \lim_{n\to\infty} V(t)/n_0$ and $G_1(t) = \lim_{n\to\infty} S(t)/n_1$ are assumed to exist almost surely for each $t \in (0, 1]$. For a pre-chosen level α, a data-driven threshold for p-values is

$$t_\alpha(\widehat{\mathrm{FDR}}) = \sup\bigl\{0 \le t \le 1 : \widehat{\mathrm{FDR}}(t) \le \alpha\bigr\}.$$

A null hypothesis $H_{0,i}$ is rejected if $p_i \le t_\alpha(\widehat{\mathrm{FDR}})$.

The measure proposed in Zhang et al. (2011) of "*lack of identification phenomenon*" (LIP) for the above FDR controlling procedure is motivated by the observation that $t_\alpha(\widehat{\mathrm{FDR}})$ is a non-decreasing function of α. As α decreases below $\inf_{0<t\le 1}\widehat{\mathrm{FDR}}(t)$, the threshold $t_\alpha(\widehat{\mathrm{FDR}})$ will drop to zero and the FDR procedure can only reject those null hypotheses with p-values exactly equal to zero. This causes LIP. For the estimation method $\widehat{\mathrm{FDR}}(\cdot)$ in (2.1), Zhang et al. (2011) introduced a measure of LIP by

$$\alpha_\infty^{\mathrm{FDR}} = \inf_{0<t\le 1} \widehat{\mathrm{FDR}}^{\infty}(t),$$

where $\widehat{\mathrm{FDR}}^{\infty}(t)$ is defined in (2.2). Note that $\alpha_\infty^{\mathrm{FDR}} > 0$ implies the occurrence of the LIP, whereas $\alpha_\infty^{\mathrm{FDR}} = 0$ rules out the possibility of the LIP.

2.2 The FDR_L procedure

For testing a set of 2D or 3D spatial signals $\{\mu(\upsilon) : \upsilon \in \mathcal{V}\}$, assume that the p-values are $\{p(\upsilon) : \upsilon \in \mathcal{V}\}$ corresponding to hypotheses, for e.g., $H_0(\upsilon) : \mu(\upsilon) = 0$ versus $H_1(\upsilon) : \mu(\upsilon) \ne 0$, and $\mathcal{V} = \mathcal{V}_0 \cup \mathcal{V}_1$, with $\mathcal{V}_0$ denoting the set of $\upsilon \in \mathcal{V}$ under $H_0(\upsilon)$ and $\mathcal{V}_1$ the set of $\upsilon \in \mathcal{V}$ under $H_1(\upsilon)$. Zhang et al. (2011) proposed the FDR_L procedure which uses a local aggregation of p-values at points adjacent to υ. The procedure consists of the steps below. Step 1: Choose a local neighborhood with size k. Step 2: At each υ, find the set N_υ of its neighborhood points, and the set $\{p(\upsilon') : \upsilon' \in N_\upsilon\}$ of the corresponding p-values. Step 3: At each υ, apply a transformation filter : $[0, 1]^k \mapsto [0, 1]$ to the set of p-values in Step 2, leading to a "locally aggregated" quantity,

$$p^*(\upsilon) = \mathrm{filter}\big(\big\{p\big(\upsilon'\big) : \upsilon' \in N_\upsilon\big\}\big). \tag{2.3}$$

Step 4: Determine a data-driven threshold for $\{p^*(\upsilon) : \upsilon \in \mathcal{V}\}$. The FDR based on p^*-values is defined by

$$\mathrm{FDR}_L(t) = E\left\{\frac{V^*(t)}{R^*(t) \vee 1}\right\},$$

where $V^*(t) = \sum_{\upsilon\in\mathcal{V}} \mathrm{I}\{H_0(\upsilon) \text{ is true, } p^*(\upsilon) \le t\}$ and $R^*(t) = \sum_{\upsilon\in\mathcal{V}} \mathrm{I}\{p^*(\upsilon) \le t\}$. Accordingly, the point estimate of $\mathrm{FDR}_L(t)$ is given by

$$\widehat{\mathrm{FDR}_L}(t) = \frac{W^*(\lambda)\widehat{\mathrm{G}}^*(t)}{\{R^*(t) \vee 1\}\{1 - \widehat{\mathrm{G}}^*(\lambda)\}}, \tag{2.4}$$

where $\widehat{\mathrm{G}}^*(\cdot)$ denotes an estimator of the sample distribution $\widetilde{G}^*(\cdot)$ of $\{p^*(\upsilon) : \upsilon \in \mathcal{V}_0\}$. Zhang et al. (2011) developed two estimators, $\widehat{G}^*(\cdot)$, called Method I, and $\widehat{G}_c^*(\cdot)$, called Method II (described in Sects. 3.2–3.3 of Zhang et al. (2011)). In brief, Method I uses the empirical distribution function of $\{p^*(\upsilon)\}$ and is useful for large-scale imaging data; Method II refines Method I via a mixture model approach and is useful for data of limited resolution. The pointwise limit of (2.4) is

$$\widehat{\mathrm{FDR}_L}^{\infty}(t) = \frac{[\pi_0\{1 - G_0^*(\lambda)\} + \pi_1\{1 - G_1^*(\lambda)\}]G^{*\infty}(t)}{\{\pi_0 G_0^*(t) + \pi_1 G_1^*(t)\}\{1 - G^{*\infty}(\lambda)\}}, \tag{2.5}$$

Table 2 Illustration of the FDR_L procedure

$\{p(\upsilon) : \upsilon \in \mathcal{V}\}$	
$\Downarrow$	$\longleftarrow$ via filter (median filter or mean filter) in (2.3)
$\{p^*(\upsilon) : \upsilon \in \mathcal{V}\}$	
$\Downarrow$	$\longleftarrow$ via estimation method (Method I or Method II)
$\widehat{\mathrm{G}}^*(\cdot)$	
$\Downarrow$	$\longleftarrow$ via (2.4)
$\widehat{\mathrm{FDR}_L}(\cdot)$	
$\Downarrow$	$\longleftarrow$ via (2.6)
$t_\alpha(\widehat{\mathrm{FDR}_L})$	
$\Downarrow$	
check whether $p^*(\upsilon) \le t_\alpha(\widehat{\mathrm{FDR}_L})$	

where $G_0^*(t) = \lim_{n\to\infty} V^*(t)/n_0$ and $G_1^*(t) = \lim_{n\to\infty} S^*(t)/n_1$ exist almost surely for each $t \in (0, 1]$, and $G^{*\infty}(t) = \lim_{n\to\infty} G^*(t)$, where $G^*(\cdot)$ denotes the cumulative distribution function of a "locally aggregated" p^*-value corresponding to the true null hypothesis. The data-driven threshold for p^*-values is

$$t_\alpha(\widehat{\mathrm{FDR}_L}) \equiv \sup\big\{0 \le t \le 1 : \widehat{\mathrm{FDR}_L}(t) \le \alpha\big\}. \tag{2.6}$$

A null hypothesis $H_0(\upsilon)$ is rejected if $p^*(\upsilon) \le t_\alpha(\widehat{\mathrm{FDR}_L})$. Table 2 illustrates the scheme of the FDR_L procedure. Zhang et al. (2011) showed that similar to the FDR procedure Storey et al. (2004), the "median filtered" FDR_L procedure asymptotically controls the FDR at level α.

Likewise, the measure of LIP for the estimation method $\widehat{\mathrm{FDR}_L}(\cdot)$ in (2.4) can be defined as

$$\alpha_\infty^{\mathrm{FDR}_L} = \inf_{0<t\le 1} \widehat{\mathrm{FDR}_L}^{\infty}(t), \tag{2.7}$$

where $\widehat{\mathrm{FDR}_L}^{\infty}(t)$ is given in (2.5).

3 Property of $\alpha_\infty^{\mathrm{FDR}_L}$ using the mean filter

3.1 Distribution of the mean filter

Our major theoretical results in Theorems 1–2 will concern the evaluation of $\alpha_\infty^{\mathrm{FDR}_L}$, which hinges on the distribution of the mean filter of p-values under the true null hypotheses. As seen from (2.5) and (A.2), such distribution corresponds to the distribution of $\overline{\mathrm{U}}_k = \sum_{i=1}^k \mathrm{U}_i/k$, where $\{\mathrm{U}_i\}_{i=1}^k \overset{\text{i.i.d.}}{\sim} \mathrm{Unif}(0, 1)$ and $k \ge 1$ is an integer, with $\mathrm{Unif}(0, 1)$ denoting the uniform distribution on the interval $(0, 1)$. Let $G_0^*(\cdot)$ denote the cumulative distribution function (C.D.F.) of $\sum_{i=1}^k \mathrm{U}_i/k$. This section will discuss explicit expressions for both its probability density function (p.d.f.), $dG_0^*(t)/dt$, and the C.D.F., $G_0^*(t)$.

For any integer $k \geq 1$, the p.d.f. of the sum, $\sum_{i=1}^{k} \mathrm{U}_i$, where $\{\mathrm{U}_i\}_{i=1}^{k} \overset{\text{i.i.d.}}{\sim}$ Unif(0, 1), can be obtained by the formula (Feller 1968; p. 175 of Sadooghi-Alvandi et al. 2009; p. 277 of Uspensky 1937),

$$\begin{cases} \frac{1}{(k-1)!} \sum_{j=0}^{k} (-1)^j \binom{k}{j} \{(x-j)_+\}^{k-1} \mathrm{I}(x \geq j), & \text{if } 0 < x < k, \\ 0, & \text{otherwise}, \end{cases} \tag{3.1}$$

where $x_+ = x\mathrm{I}(x \geq 0)$. According to (3.1), the mean, $\sum_{i=1}^{k} \mathrm{U}_i/k$, has the p.d.f.

$$\frac{dG_0^*(t)}{dt} = \begin{cases} \frac{k}{(k-1)!} \sum_{j=0}^{k} (-1)^j \binom{k}{j} \{(kt-j)_+\}^{k-1} \mathrm{I}(kt \geq j), & \text{if } 0 < t < 1, \\ 0, & \text{otherwise}. \end{cases} \tag{3.2}$$

In particular, for $0 < kt < 1$,

$$\frac{dG_0^*(t)}{dt} = \frac{k}{(k-1)!}(-1)^0 \binom{k}{0} \left\{(kt)_+\right\}^{k-1} = \frac{k}{(k-1)!} k^{k-1} t^{k-1}, \tag{3.3}$$

which is proportional to t^{k-1}. It follows from (3.2) that

$$\frac{d^2G_0^*(t)}{dt^2} = \begin{cases} \frac{k^2(k-1)}{(k-1)!} \sum_{j=0}^{k} (-1)^j \binom{k}{j} \{(kt-j)_+\}^{k-2} \mathrm{I}(kt \geq j), & \text{if } 0 < t < 1, \\ 0, & \text{otherwise}. \end{cases} \tag{3.4}$$

In addition, an integration of (3.2) yields

$$G_0^*(t) = \frac{1}{k!} \sum_{j=0}^{k} (-1)^j \binom{k}{j} \left\{(kt-j)_+\right\}^{k} \mathrm{I}(kt \geq j). \tag{3.5}$$

As a comparison, the distribution of the median filter of p-values under the true null hypotheses is that of Beta$((k+1)/2, (k+1)/2)$ when k is an odd integer. The behavior of its p.d.f. is proportional to $t^{(k-1)/2}(1-t)^{(k-1)/2} = (t-t^2)^{(k-1)/2}$, and controlled by $t^{(k-1)/2}$ for $t \approx 0$. This aspect distinguishes the tail behavior of the distribution of the median from that of the mean. Figure 1 plots the p.d.f. of the mean filter and the p.d.f. of the median filter for $k = 5$ and $k = 7$. In each case, the asymptotic normal density functions (van-der Vaart 1998) for the mean and median of $\{\mathrm{U}_1, \ldots, \mathrm{U}_k\}$, assuming $k \to \infty$, are overlaid. Apparently, the distribution of the mean filter converges faster to be normal.

We would like to make a remark on the distribution of $G_0^*(\cdot)$. Using Corollary 2 of Ruiz (1996), an alternative expression for the p.d.f. of the sum, $\sum_{i=1}^{k} \mathrm{U}_i$, is obtained in the form,

$$\begin{cases} \frac{1}{2(k-1)!} \sum_{j=0}^{k} (-1)^j \binom{k}{j} (x-j)^{k-1} \mathrm{sign}(x-j), & \text{if } 0 < x < k, \\ 0, & \text{otherwise}, \end{cases} \tag{3.6}$$

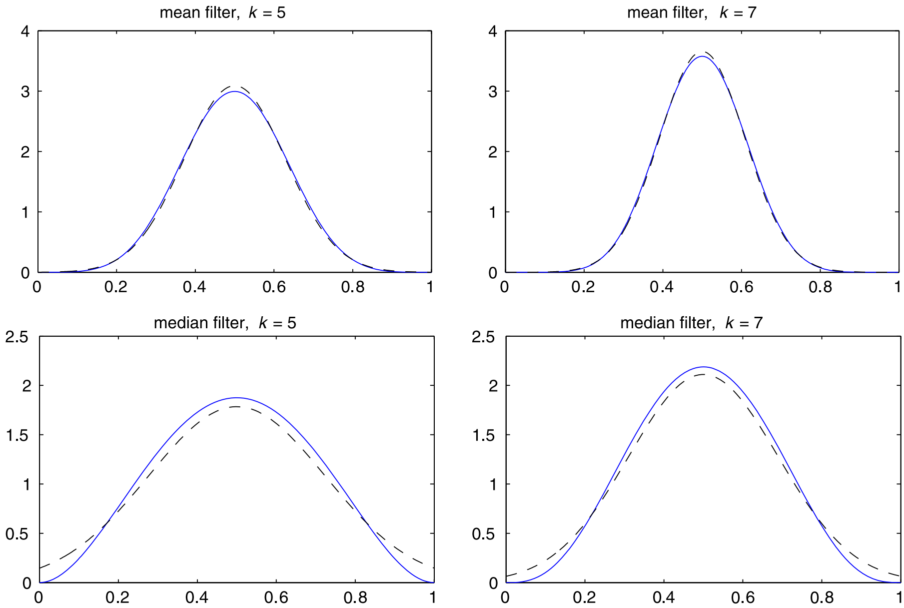

Fig. 1 Probability density functions for the mean filter (*at top panels*) and the median filter (*at bottom panels*). *Solid curve*: p.d.f. of the filter; *dashed curve*: asymptotic normal approximation

where $\mathrm{sign}(x)$ equals -1 if $x < 0$, 0 if $x = 0$, and $+1$ if $x > 0$. From (3.6), the p.d.f. of the mean, $\sum_{i=1}^{k} \mathrm{U}_i/k$, is given by

$$\frac{dG_0^*(t)}{dt} = \begin{cases} \frac{k}{2(k-1)!} \sum_{j=0}^{k} (-1)^j \binom{k}{j} (kt-j)^{k-1} \mathrm{sign}(kt-j), & \text{if } 0 < t < 1, \\ 0, & \text{otherwise.} \end{cases} \tag{3.7}$$

For the equivalent forms, (3.2) and (3.7), whichever is more convenient will be used in the proofs of Theorems 1–2.

3.2 Conditions for lack of identification phenomenon

For $\alpha_\infty^{\mathrm{FDR}_L}$ defined in (2.7) using the mean filter, results analogous to those using the median filter are established in Theorems 1–2. Theorem 1 presents conditions under which the LIP does or does not take place with the FDR and FDR_L procedures. Theorem 2 demonstrates that $\alpha_\infty^{\mathrm{FDR}} \geq \alpha_\infty^{\mathrm{FDR}_L}$ under mild conditions. Thus, similar to the median filter, the FDR_L procedure using the mean filter also reduces the extent of the LIP. Applications of Theorems 1–2 will be illustrated in a numerical example in Sect. 4.1.

Theorem 1 *Let $\{T(v) : v \in \mathcal{V} \subseteq \mathbb{Z}^d\}$ be the set of test statistics for testing the presence of the spatial signals $\{\mu(v) : v \in \mathcal{V} \subseteq \mathbb{Z}^d\}$. Consider the one-sided testing problem,*

$$H_0(v) : \mu(v) = 0 \quad \textit{versus} \quad H_1(v) : \mu(v) > 0.$$

For $j = 0$ and $j = 1$, respectively, assume that $T(\upsilon)$, corresponding to the true $H_j(\upsilon)$, are i.i.d. *random variables having a* C.D.F. F_j *with a* p.d.f. f_j. *Assume that the neighborhood size $k \geq 3$ used in the* FDR_L *procedure is an integer and that the proportion of boundary grid points within $\mathcal{V}_0$ shrinks to zero, as $n \to \infty$, i.e.,* $\lim_{n\to\infty} \#\mathcal{V}_0^{(0)}/n_0 = 1$, *where* $\mathcal{V}_0^{(0)} = \{\upsilon \in \mathcal{V} : \mu(\upsilon') = 0$ *for any* $\upsilon' \in N_\upsilon\}$. *Assume Condition* A *in the Appendix. Let* $x_0 = F_0^{-1}(1) = \inf\{t : F_0(t) = 1\}$.

I. *If* $\lim_{x\to x_0-} f_1(x)/f_0(x) = \infty$, *then* $\alpha_\infty^{\mathrm{FDR}} = 0$ *and* $\alpha_\infty^{\mathrm{FDR}_L} = 0$.
II. *If* $\limsup_{x\to x_0-} f_1(x)/f_0(x) < \infty$, *then* $\alpha_\infty^{\mathrm{FDR}} > 0$ *and* $\alpha_\infty^{\mathrm{FDR}_L} > 0$.

Theorem 2 *Assume conditions in Theorem 1. Suppose that $f_0(\cdot)$ is supported in an interval;* $f_1(x) \leq f_0(x)$ *for any* $x \leq F_0^{-1}(0.5)$; $1 - F_0(F_1^{-1}(0.5)) \leq \lambda \leq 0.5$. *Then* $\alpha_\infty^{\mathrm{FDR}} \geq \alpha_\infty^{\mathrm{FDR}_L}$.

As a comparison, it is interesting to observe that when the mean filter is used, Theorem 1 and Theorem 2 hold for any integer $k \geq 3$, whereas when the median filter is used, the corresponding Theorem 4.4 and Theorem 4.5 in Zhang et al. (2011) hold for any odd integer $k \geq 3$.

4 Simulation study

4.1 An illustrative example of $\alpha_\infty^{\mathrm{FDR}} > \alpha_\infty^{\mathrm{FDR}_L} > 0$

Consider a pixelated 2D image dataset consisting of $n = 50 \times 50$ pixels, illustrated in the left panel of Fig. 2, where the black rectangles represent the true significant regions $\mathcal{V}_1$ with $n_1 = 0.16 \times n$ pixels and the white background serves as the true non-significant regions $\mathcal{V}_0$ with $n_0 = n - n_1$ pixels. The data $Y(i, j)$ are simulated from the model,

$$Y(i, j) = \mu(i, j) + \epsilon(i, j), \quad i, j = 1, \ldots, 50,$$

where the signals are $\mu(i, j) = 0$ for $(i, j) \in \mathcal{V}_0$, and $\mu(i, j) = C$ for $(i, j) \in \mathcal{V}_1$ with a constant $C \in (0, \infty)$, and the error terms $\{\epsilon(i, j)\}$ are i.i.d. following the centered Exp(1) distribution. At each site (i, j), the observed data $Y(i, j)$ is the (shifted) survival time and used as the test statistic for testing $\mu(i, j) = 0$ versus $\mu(i, j) > 0$. Clearly, all test statistics are i.i.d., having the p.d.f.s $f_0(x) = \exp\{-(x + 1)\}\mathrm{I}(x + 1 > 0)$ under true null hypotheses, and $f_1(x) = \exp\{-(x + 1 - C)\}\mathrm{I}(x + 1 > C)$ under true non-null hypotheses. Following the notation in Theorem 1, it is easily seen that $x_0 = \infty$, and $\limsup_{x\to\infty} f_1(x)/f_0(x) = \exp(C) < \infty$. An appeal to Theorem 1 yields $\alpha_\infty^{\mathrm{FDR}} > 0$ and $\alpha_\infty^{\mathrm{FDR}_L} > 0$, and thus both the FDR and the FDR_L procedures will encounter the LIP. Moreover, if $C > \log(2)$ and $\exp(-C)/2 \leq \lambda \leq 0.5$, then sufficient conditions in Theorem 2 are satisfied and hence $\alpha_\infty^{\mathrm{FDR}} \geq \alpha_\infty^{\mathrm{FDR}_L}$.

Actual computations indicate that $\alpha_\infty^{\mathrm{FDR}_L}$ using the mean filter (with detailed derivations in the appendix) is smaller than that using the median filter. For example, set $\lambda = 0.1$; assume that the neighborhood in the FDR_L procedure is illustrated

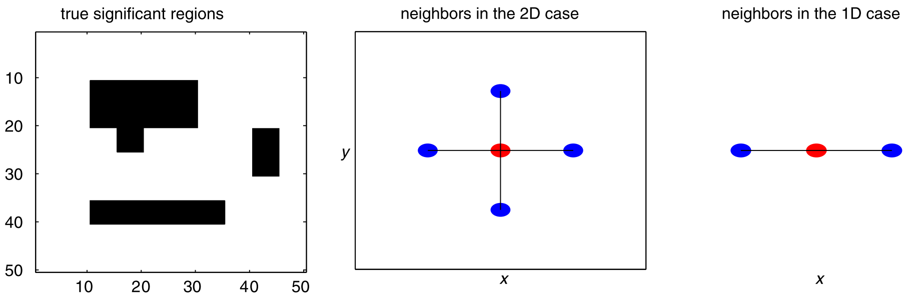

Fig. 2 *Left panel*: the true significant regions for the 2D simulated datasets. *Middle panel*: neighbors of a point at (x, y) used in the FDR_L procedure for 2D data. *Right panel*: neighbors of a point at x used in the FDR_L procedure for 1D data

Table 3 Comparing $\alpha_\infty^{\mathrm{FDR}}$, $\alpha_\infty^{\mathrm{FDR}_L}$ using the median filter and $\alpha_\infty^{\mathrm{FDR}_L}$ using the mean filter, where $k = 5$. Here $\lambda = 0.1$

C	$\alpha_\infty^{\mathrm{FDR}}$	$\alpha_\infty^{\mathrm{FDR}_L}$	
		Median filter	Mean filter
log(4)	0.6396	0.0858	0.005853973160543
log(8)	0.4130	0.0103	0.000160445958336
log(12)	0.3043	0.0030	0.000021098127390
log(16)	0.2471	0.0013	0.000005006765093
log(20)	0.2079	0.0007	0.000001640622308
log(24)	0.1795	0.0004	0.000000659329957
log(28)	0.1579	0.0002	0.000000305048585
log(32)	0.1409	0.0002	0.000000156462168
log(36)	0.1273	0.0001	0.000000086825394

in the middle panel of Fig. 2, that is, $k = 5$. Table 3 indicates that the mean filter combined with the FDR_L procedure better alleviates the LIP of the FDR procedure than the median filter. Whether or not the conclusion that $\alpha_\infty^{\mathrm{FDR}_L}$ using the median filter exceeds $\alpha_\infty^{\mathrm{FDR}_L}$ using the mean filter holds in general situations needs to be further investigated.

Figure 3 compares the regions detected as significant by the FDR and the "mean filtered" FDR_L procedures. As a comparison, a counterpart using the median filter is presented in Fig. 4. Comparing Figs. 3–4, it is clearly seen that for $\alpha \approx 0$, the mean filter better detects the significance than the median filter, in agreement with Table 3. On the other hand, the median filter better preserves the edges between regions declared to be insignificant and significant.

4.2 2D dependent data

For the sake of comparison, all simulation set-ups for dependent data are identical to those in Sect. 5 of Zhang et al. (2011), except that the mean filter replaces the median filter in the FDR_L procedure. It will be seen that in the presence of dependence, the distinction between the performance of the mean filter relative to median filter diminishes.

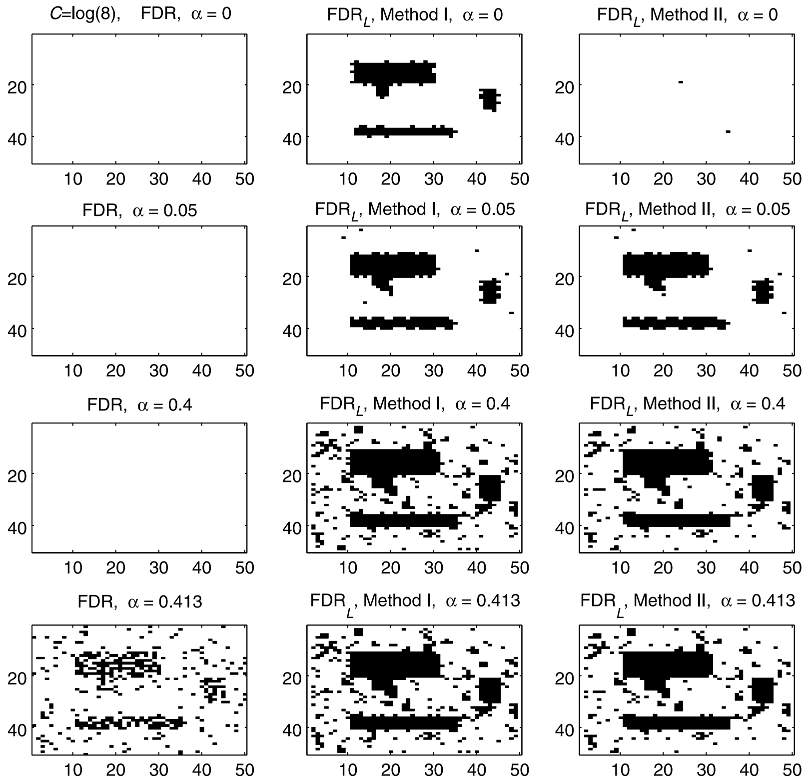

Fig. 3 Lack of identification phenomenon when α varies from 0 to $\alpha_{\infty}^{\mathrm{FDR}} = 0.4130$. The sites that are called statistically significant based on the realization are shown in *black*. *Left panels*: the FDR procedure. *Middle panels*: the FDR_L procedure using Method I. *Right panels*: the FDR_L procedure using Method II. Here $\lambda = 0.1$

4.2.1 Example 1: weakly correlated case

The data $Y(i, j)$ are generated according to the model,

$$Y(i, j) = \mu(i, j) + \epsilon(i, j), \quad i, j = 1, \ldots, 258, \tag{4.1}$$

where the signals are $\mu(i, j) = 0$ for $(i, j) \in \mathcal{V}_0$, $\mu(i, j) = 4$ in the larger black rectangle and $\mu(i, j) = 2$ in the smaller black rectangle. The errors $\{\epsilon(i, j)\}$ have zero-mean, unit-variance and are *spatially dependent*, by taking $\epsilon(i, j) = \{e(i-1, j) + e(i, j) + e(i+1, j) + e(i, j-1) + e(i, j+1)\}/\sqrt{5}$, where $\{e(i, j)\}_{i,j=0}^{259}$ are i.i.d. $N(0, 1)$. At each pixel (i, j), $Y(i, j)$ is used as the test statistic for testing $\mu(i, j) = 0$ against $\mu(i, j) > 0$.

To evaluate the performance of Method I and Method II in estimating $\widetilde{G}^*(t)$, Fig. 5 displays the plots of $\widehat{G}^*(t)$ versus $\widetilde{G}^*(t)$ and $\widehat{G}_c^*(t)$ versus $\widetilde{G}^*(t)$. The agreement with 45 degree lines indicates that both estimation methods developed for the median filter can be well applicable to the mean filter.

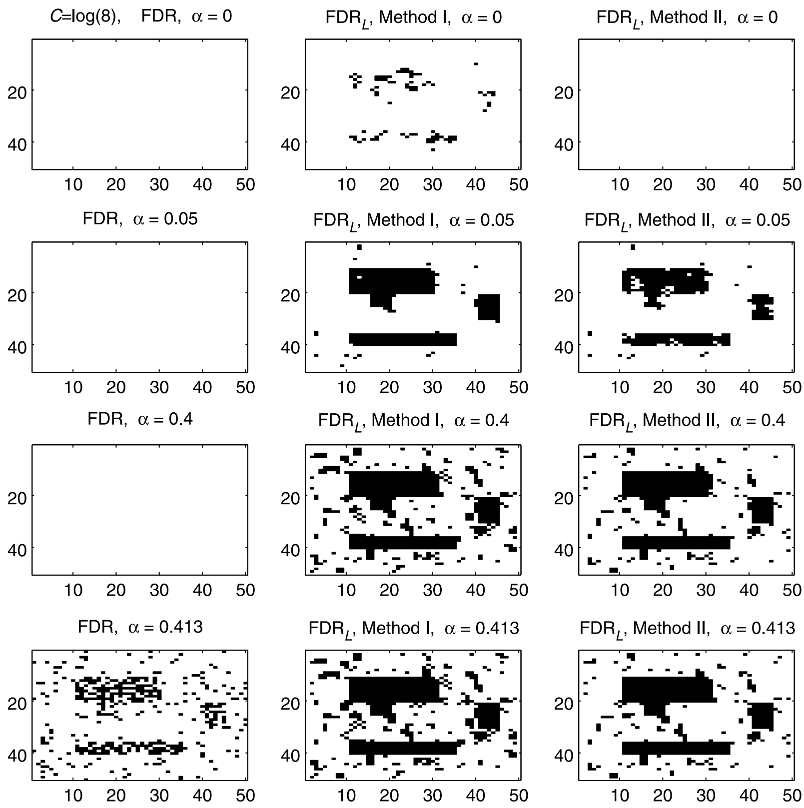

Fig. 4 The caption is similar to that for Fig. 3, except that the median filter is used

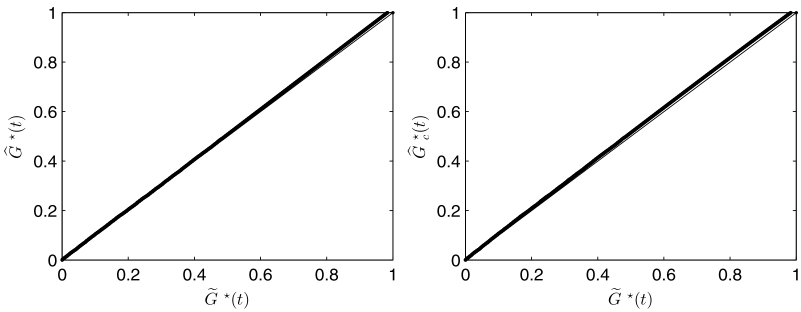

Fig. 5 *Left*: $\widehat{G}^*(t)$ versus $\widetilde{G}^*(t)$; *right*: $\widehat{G}_c^*(t)$ versus $\widetilde{G}^*(t)$; *straight line*: the 45 degree reference line. Here $\alpha = 0.01$ and $\lambda = 0.1$

To examine the overall performance of the estimated $\mathrm{FDR}(t)$ and $\mathrm{FDR}_L(t)$ for a same threshold $t \in [0, 1]$, we replicate the simulation 100 times. For notational convenience, denote by $\mathrm{FDP}(t) = V(t)/\{R(t) \vee 1\}$ and $\mathrm{FDP}_L(t) = V^*(t)/\{R^*(t) \vee 1\}$

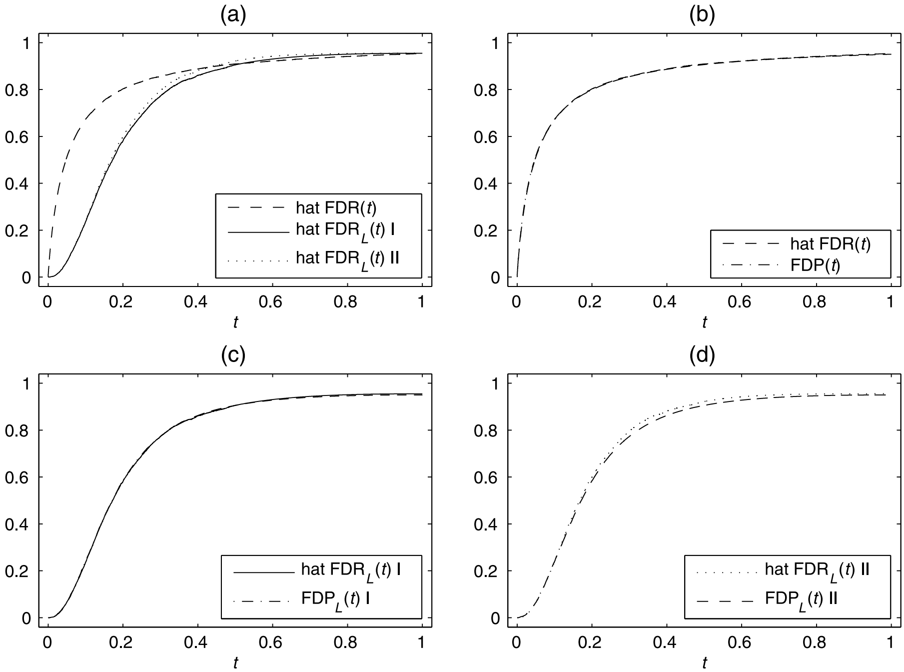

Fig. 6 Panel (**a**): compare the average values of $\widehat{\mathrm{FDR}}(t)$ and those of $\widehat{\mathrm{FDR}_L}(t)$ using Methods I and II. Panel (**b**): compare the average values of $\widehat{\mathrm{FDR}}(t)$ and those of $\mathrm{FDP}(t)$. Panel (**c**): compare the average values of $\widehat{\mathrm{FDR}_L}(t)$ using Method I and those of $\mathrm{FDP}_L(t)$. Panel (**d**): compare the average values of $\widehat{\mathrm{FDR}_L}(t)$ using Method II and those of $\mathrm{FDP}_L(t)$. Here $\lambda = 0.1$

the false discovery proportions of the FDR and FDR_L procedures, respectively. The average values (over 100 data) of $\widehat{\mathrm{FDR}}(t)$ and $\widehat{\mathrm{FDR}}_L(t)$ at each point t are plotted in Fig. 6(a). It is clearly observed that $\widehat{\mathrm{FDR}}_L(t)$ using both Methods I and II is below $\widehat{\mathrm{FDR}}(t)$, demonstrating that the FDR_L procedure produces the estimated false discovery rates lower than those of the FDR procedure. Meanwhile, Fig. 6 compares the average values of $\mathrm{FDP}(t)$ and those of $\widehat{\mathrm{FDR}}(t)$ in panel (b), and the average values of $\mathrm{FDP}_L(t)$ using Methods I and II and those of $\widehat{\mathrm{FDR}}_L(t)$ in panels (c) and (d), respectively. For each procedure, the two types of estimates are very close to each other, again lending support to the estimation procedure.

We randomly generate 100 sets of simulated data and perform the FDR and FDR_L procedures for each dataset, with the control levels α varying from 0 to 0.1. The choices $\lambda = 0.1$ and $\lambda = 0.4$ are considered. In either case, we observe from Fig. 7 that the average sensitivity (over the datasets) of the FDR_L procedure using Method I is consistently higher than that of the FDR procedure, whereas the average specificities of both procedures approach one and are nearly indistinguishable. In addition, the bottom panels indicate that the FDR procedure yields larger (average) false discovery proportions than the FDR_L procedure. It is apparent that the results in Fig. 7 are not very sensitive to the choice of λ.

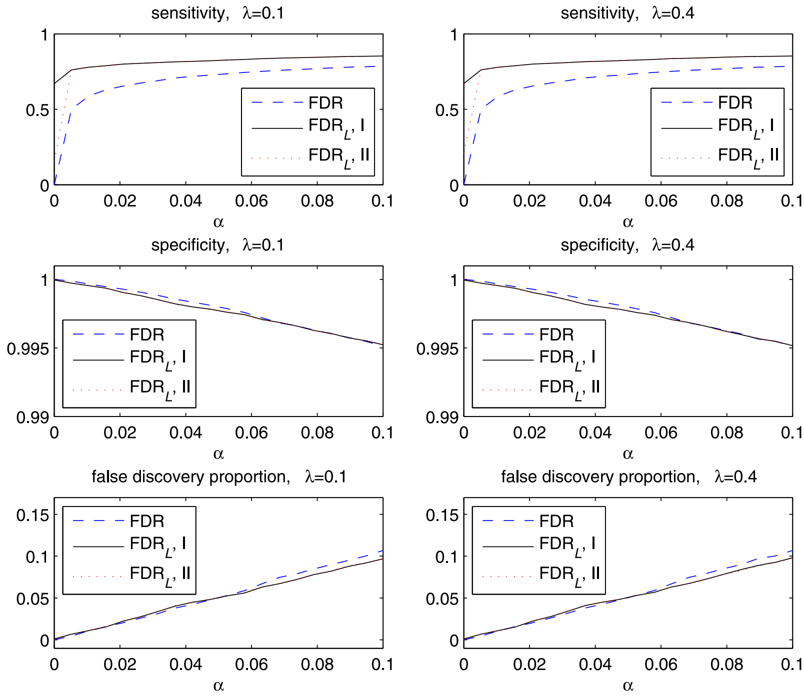

Fig. 7 Comparison of the average sensitivity (*top panels*), average specificity (*middle panels*) and average false discovery proportion (*bottom panels*). *Left panels*: $\lambda = 0.1$. *Right panels*: $\lambda = 0.4$

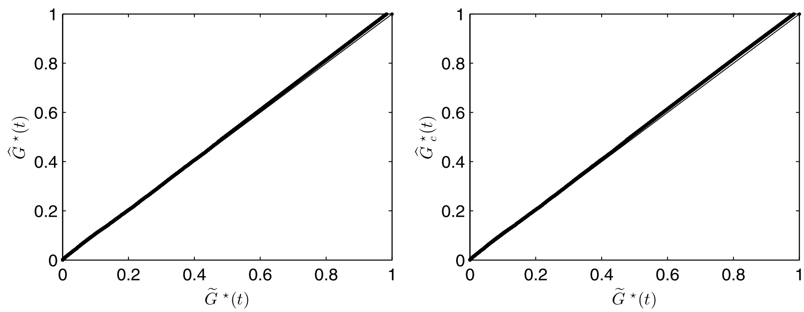

Fig. 8 The captions are similar to those for Fig. 5, except that Example 2 is studied

4.2.2 *Example* 2: *more strongly correlated case*

We generate one dataset according to the same model (4.1) as in Example 1, but with more strongly correlated errors, $\epsilon(i, j) = \sum_{i=0}^{6} \sum_{j=0}^{6} e(i, j)/7$, where $\{e(i, j)\}_{i,j=0}^{264}$ are i.i.d. $N(0, 1)$. As seen from Fig. 8, estimation Methods I and II for the "mean filtered" FDR$_L$ procedure with strongly correlated data perform as well as with weakly correlated data (given in Fig. 5).

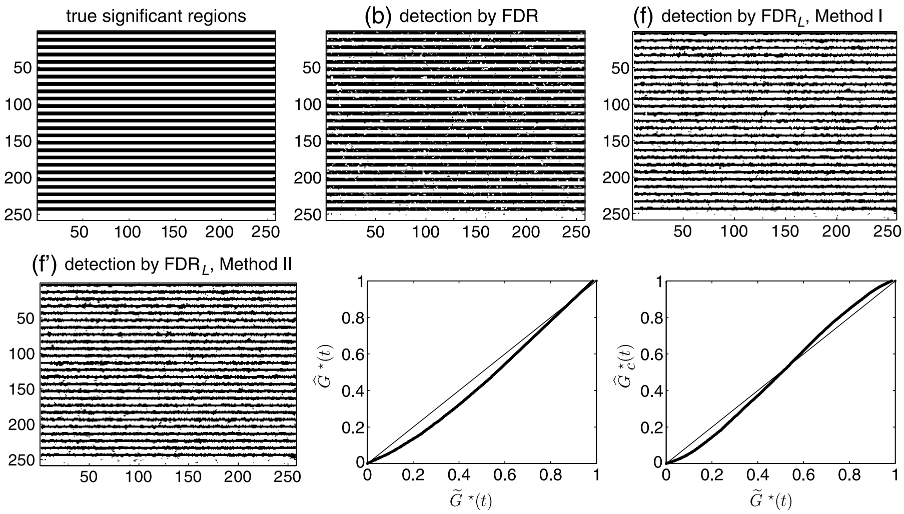

Fig. 9 Comparison of the FDR and FDR$_L$ procedure for Example 3. *Top left*: the true significant regions; *top middle*: the FDR procedure; *top right*: the FDR$_L$ procedure using Method I. *Bottom left*: the FDR$_L$ procedure using Method II; *bottom middle*: $\widehat{G}^*(t)$ versus $\widetilde{G}^*(t)$; *bottom right*: $\widehat{G}_c^*(t)$ versus $\widetilde{G}^*(t)$; *straight line*: the 45 degree reference line. Here $\alpha = 0.01$ and $\lambda = 0.1$

4.2.3 *Example 3: large proportion of boundary grid points*

The efficacy of the FDR$_L$ procedure is illustrated in Fig. 9 by a simulated data generated according to the same model (4.1) as in Example 1, but with a large proportion of boundary grid points, where $\mu(i, j) = 0$ for $(i, j) \in \mathcal{V}_0$ and $\mu(i, j) = 4$ for $(i, j) \in \mathcal{V}_1$. Similar plots using $\mu(i, j) = 2$ for $(i, j) \in \mathcal{V}_1$ are obtained in Fig. 10. It is observed that the FDR procedure slightly outperforms the FDR$_L$ procedure in the presence of strong signals, whereas the FDR$_L$ procedure better detects weaker signals. Note that detecting weaker signals, which arise more often in practical applications, is more challenging than detecting stronger signals. In this sense, there is no adverse effect of using the FDR$_L$ procedure to detect strong or weak signals.

4.3 1D serially clustered signals

Applications of the FDR$_L$ procedure are not limited to 2D and 3D data. In this example, we evaluate the performance of the FDR$_L$ procedure to 1D data with serially clustered signals. The data are independently generated according to $Y_i \sim N(\mu_i, 1)$, $i = 1, \ldots, 10000$, and Y_i is used for testing hypotheses $H_{0,i} : \mu_i = 0$ versus $H_{1,i} : \mu_i > 0$. The serial structure is designed as follows: indices i from true non-null hypotheses are partitioned into two groups, $\mathcal{I}_1 = \{1001, \ldots, 2000\}$ and $\mathcal{I}_2 = \{5001, \ldots, 6000\}$, where $\mu_i \equiv 1.5$ for $i \in \mathcal{I}_1$ and $\mu_i \equiv 2.0$ for $i \in \mathcal{I}_2$. The neighborhood in the FDR$_L$ procedure is illustrated in the right panel of Fig. 2, that is, $k = 3$. The calculated FDP and detection power, based on 500 sets of simulated data, are shown in Table 4. Both

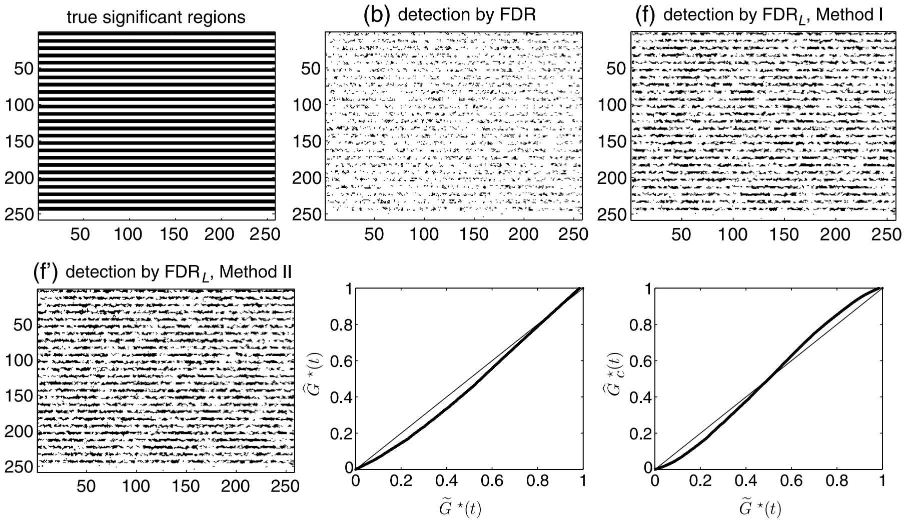

Fig. 10 The captions are similar to those for Fig. 9, except that the amount of signals is reduced from $\mu = 4$ to weaker signals $\mu = 2$

Table 4 Compare the calculated FDP and detection power, using the FDR$_L$ procedure with the mean and median filters, and the FDR procedure. Here $\lambda = 0.1$

α	FDR$_L$ procedure, Method I				FDR procedure	
	Mean filter		Median filter			
	FDP	Power	FDP	Power	FDP	Power
0.01	0.009	0.377	0.011	0.291	0.002	0.005
0.02	0.019	0.479	0.020	0.389	0.017	0.037
0.03	0.028	0.543	0.030	0.457	0.027	0.063
0.04	0.037	0.591	0.039	0.507	0.036	0.088
0.05	0.046	0.628	0.048	0.547	0.045	0.112
0.06	0.055	0.659	0.057	0.581	0.054	0.136
0.07	0.064	0.685	0.066	0.610	0.064	0.159
0.08	0.073	0.708	0.075	0.636	0.073	0.182
0.09	0.082	0.727	0.085	0.658	0.082	0.203
0.10	0.091	0.745	0.094	0.679	0.091	0.224
0.20	0.181	0.856	0.189	0.809	0.183	0.403
0.30	0.271	0.915	0.282	0.879	0.275	0.548

the FDR and the FDR$_L$ procedures maintain control of the false discovery rates. However, by capturing the structural information of the data, the FDR$_L$ procedure (Method I), using either the mean or median filter, is apparently more powerful than the FDR procedure. (Results using Method II of the FDR$_L$ procedure are similar and omitted.) Moreover, the mean filter increases detection power than the median filter.

5 Discussion

In the context of large-scale simultaneous inference, the hypotheses are often accompanied with a scientifically meaningful structure. Examples include spatial geometry and clustering/grouping by pattern similarity, which are, however, ignored by many conventional multiple testing procedures. The FDR_L procedure proposed by Zhang et al. (2011) makes an effort to incorporate spatial information associated with large-scale imaging data, by utilizing a locally median filter of p-values. This paper examines the performance of another commonly used filter, mean filter, in the FDR_L procedure. By making theoretical and empirical comparisons between the mean and median filters, we hope to provide valuable insight into understanding properties of the FDR_L procedure on the estimation of FDR and detection of sparse signals.

In summary, both the mean and the median filters in the FDR_L procedure reduce the extent of LIP of the FDR procedure when the p-values under the true nulls are independent and uniformly distributed. In practical applications, both filters have their own merits, while simultaneously achieving the goals of edge-preserving property, robustness property and detection power may not be feasible. If preserving the edges between significant and non-significant regions is a major concern, or if robustness to outlying points is cared, the median filter will be preferred. See comparisons in Figs. 3–4. If the detection power is essential, the mean filter will be a better choice. See comparisons in Tables 3–4.

A number of challenging issues remain to be explored. For the example in Table 3, $\alpha_\infty^{\mathrm{FDR}_L}$ using the mean filter does not exceed $\alpha_\infty^{\mathrm{FDR}_L}$ using the median filter. Does this relation hold in other cases? Moreover, large-scale multiple testing tasks often exhibit dependence, and leveraging the dependence among individual tests is an important but challenging problem in statistics. Whether and to which extent the FDR_L procedure deals with general types of dependence structure? We leave these issues for future research.

Acknowledgements The author thanks the Associate Editor and referee for insightful comments.
The research is supported by the US National Science Foundation grants DMS-1106586, DMS-1308872 and Wisconsin Alumni Research Foundation.

Appendix

Throughout the proof, since $\alpha_\infty^{\mathrm{FDR}}$ has been treated in the Appendix of Zhang et al. (2011), derivations will be confined to $\alpha_\infty^{\mathrm{FDR}_L}$ of the mean filter.

Condition A

A0 The neighborhood size k is an integer not depending on n.
A1 $\lim_{n\to\infty} n_0/n = \pi_0$ exists and $\pi_0 < 1$.

Proof of Theorem 1 Note that for the original p-values,

$$G_0(t) = t, \qquad G_1(t) = 1 - F_1\big(F_0^{-1}(1-t)\big). \tag{A.1}$$

Also, for the "mean filtered" $p^*(\upsilon)$-values, for any $\upsilon \in \mathcal{V}_j$, where $j = 0, 1$, we obtain

$$\begin{aligned}
&\mathrm{P}\{p^*(\upsilon) \le t\} \\
&\quad = \mathrm{P}\{\mathrm{mean}\big(p_1(\upsilon), \ldots, p_k(\upsilon)\big) \le t\} \\
&\quad = \mathrm{P}\{\mathrm{mean}\big(G_j\big(p_1(\upsilon)\big), \ldots, G_j\big(p_k(\upsilon)\big)\big) \le G_j(t)\} \\
&\quad = \mathrm{P}\{\mathrm{mean}(\mathrm{U}_1, \ldots, \mathrm{U}_k) \le G_j(t)\} \\
&\quad = G_0^*\big(G_j(t)\big) = \begin{cases} G_0^*(t), & \text{if } \upsilon \in \mathcal{V}_0, \\ G_0^*(G_1(t)) = G_1^*(t), & \text{if } \upsilon \in \mathcal{V}_1. \end{cases}
\end{aligned} \tag{A.2}$$

Thus, for $p^*(\upsilon)$-values under true $H_0(\upsilon)$, $G_0^*(t)$ is the C.D.F. corresponding to the p.d.f. in (3.7); for $p^*(\upsilon)$-values under true $H_1(\upsilon)$, $G_1^*(t)$ is the C.D.F. corresponding to the p.d.f. in (3.7) with t replaced by $G_1(t)$.

Part I. So for the FDR_L procedure, by (3.2),

$$\begin{aligned}
\frac{dG_0^*(t)}{dt} &= \frac{k}{(k-1)!} \sum_{j=0}^{k} (-1)^j \binom{k}{j} \{(kt - j)_+\}^{k-1}\, \mathrm{I}(kt \ge j), \\
\frac{dG_1^*(t)}{dt} &= \frac{dG_0^*(G_1(t))}{dG_1(t)} \frac{dG_1(t)}{dt}, \\
&= \frac{k}{(k-1)!} \sum_{j=0}^{k} (-1)^j \binom{k}{j} \big[\{kG_1(t) - j\}_+\big]^{k-1}\, \mathrm{I}\{kG_1(t) \ge j\} \frac{dG_1(t)}{dt}.
\end{aligned}$$

Applying L'Hospital's rule and the fact $\lim_{t\to 0+} G_1(t) = 0$,

$$\lim_{t\to 0+} \frac{G_1(t)}{t} = \lim_{t\to 0+} \frac{dG_1(t)}{dt} = \lim_{t\to 0+} \frac{f_1(F_0^{-1}(1-t))}{f_0(F_0^{-1}(1-t))} = \lim_{x\to x_0-} \frac{f_1(x)}{f_0(x)} = \infty, \tag{A.3}$$

where $x = F_0^{-1}(1-t)$. Note that $\widehat{\mathrm{FDR}_L}^{\infty}(t)$ is a decreasing function of $G_1^*(t)/G_0^*(t)$. Using $\lim_{t\to 0+} G_1^*(t) = 0$, $\lim_{t\to 0+} G_0^*(t) = 0$, and (3.3),

$$\begin{aligned}
\lim_{t\to 0+} \frac{G_1^*(t)}{G_0^*(t)} &= \lim_{t\to 0+} \frac{\frac{dG_1^*(t)}{dt}}{\frac{dG_0^*(t)}{dt}} \\
&= \lim_{t\to 0+} \frac{\sum_{j=0}^{k} (-1)^j \binom{k}{j} [\{kG_1(t) - j\}_+]^{k-1}\, \mathrm{I}\{kG_1(t) \ge j\}}{\sum_{j=0}^{k} (-1)^j \binom{k}{j} \{(kt - j)_+\}^{k-1}\, \mathrm{I}(kt \ge j)} \frac{dG_1(t)}{dt} \\
&= \lim_{t\to 0+} \left\{\frac{G_1(t)}{t}\right\}^{k-1} \lim_{t\to 0+} \frac{dG_1(t)}{dt}
\end{aligned} \tag{A.4}$$

which together with (A.3) shows $\lim_{t\to 0+} G_1^*(t)/G_0^*(t) = \lim_{t\to 0+} \{G_1(t)/t\}^k = \infty$. Thus, $\sup_{0<t\le 1} G_1^*(t)/G_0^*(t) = \infty$, that is, $\alpha_\infty^{\mathrm{FDR}_L} = 0$ for the FDR_L procedure.

Part II. Following $\widehat{\mathrm{FDR}_L}^{\infty}(t)$, we conclude that $\alpha_\infty^{\mathrm{FDR}_L} \neq 0$ if

$$\sup_{0<t\leq 1} G_1^*(t)/G_0^*(t) < \infty. \tag{A.5}$$

We first verify (A.5) for the FDR_L procedure. Assume (A.5) fails, i.e., $\sup_{0<t\leq 1} G_1^*(t)/G_0^*(t) = \infty$. Note that for any $\delta > 0$, the function $G_1^*(t)/G_0^*(t)$, for $t \in [\delta, 1]$, is continuous and bounded away from ∞, thus, $\sup_{0<t\leq 1} G_1^*(t)/G_0^*(t) = \infty$ only if there exists a sequence $t_1 > t_2 > \cdots > 0$, such that $\lim_{m\to\infty} t_m = 0$ and $\lim_{m\to\infty} G_1^*(t_m)/G_0^*(t_m) = \infty$. For each m, recall that both $G_1^*(t)$ and $G_0^*(t)$ are continuous on $[0, t_m]$, and differentiable on $(0, t_m)$. Applying Cauchy's mean-value theorem, there exists $\xi_m \in (0, t_m)$ such that $G_1^*(t_m)/G_0^*(t_m) = \{G_1^*(t_m) - G_1^*(0)\}/\{G_0^*(t_m) - G_0^*(0)\} = \frac{dG_1^*(t)/dt}{dG_0^*(t)/dt}|_{t=\xi_m}$. Since $\lim_{m\to\infty} G_1^*(t_m)/G_0^*(t_m) = \infty$, it follows that $\limsup_{t\to 0+} \frac{dG_1^*(t)/dt}{dG_0^*(t)/dt} = \infty$, which combined with (A.4) implies

$$\limsup_{t\to 0+} \frac{dG_1(t)}{dt} = \infty. \tag{A.6}$$

On the other hand, the condition $\limsup_{x\to x_0-} f_1(x)/f_0(x) < \infty$ indicates that

$$\limsup_{t\to 0+} \frac{dG_1(t)}{dt} = \limsup_{t\to 0+} \frac{f_1(F_0^{-1}(1-t))}{f_0(F_0^{-1}(1-t))} = \limsup_{x\to x_0-} \frac{f_1(x)}{f_0(x)} < \infty, \tag{A.7}$$

where $x = F_0^{-1}(1-t)$. Clearly, (A.7) contradicts (A.6). The proof is completed. $\square$

Proof of Theorem 2 We first show Lemma 1.

Lemma 1 *Let $B(t)$ be the* C.D.F. *of the* p.d.f. *corresponding to* (3.7) *with $k \geq 3$. Then* I. *for $t \in (0, 0.5)$, $B(t)/t$ is a strictly increasing function and $B(t) < t$;* II. *for $t \in (0.5, 1)$, $B(t) > t$;* III. *for $t_1 \in (0, 0.5]$ and $t_2 \in [t_1, 1]$, $B(t_1)/t_1 \leq B(t_2)/t_2$.*

Proof Since (3.7) is symmetric with respect to 0.5, we deduce that $B(t) = 1 - B(1-t)$ and $B'(t) = B'(1-t)$, i.e. the p.d.f. $B'(t)$ is symmetric with respect to 0.5. It follows that

$$B''(t) = (-1)B''(1-t), \tag{A.8}$$

namely, $B''(t)$ is antisymmetric with respect to 0.5. More precisely, it is easy to see from (3.7) that

$$B''(t) = \frac{k^2(k-1)}{2(k-1)!}\sum_{j=0}^{k}(-1)^j\binom{k}{j}(kt-j)^{k-2}\mathrm{sign}(kt-j).$$

Hence from (3.4) and (A.8), the possible roots of $B''(t)$ are at $\{0, 0.5, 1\}$. For positive t close to 0, $B''(t)$ is a positive polynomial function of degree $k-2$.

To show part I, define $\mathrm{F}_1(t) = B(t)/t$. Then $\mathrm{F}_1'(t) = \{B'(t)t - B(t)\}/t^2$, where $d\{B'(t)t - B(t)\}/dt = B''(t)t$. For $t \in (0, 0.5)$, (A.8) and the above analysis indicate $B''(t) > 0$, i.e., $B'(t)t - B(t)$ is strictly increasing, implying $B'(t)t - B(t) > B'(0)0 - B(0) = 0$. Hence for $t \in (0, 0.5)$, $B(t)/t$ is strictly increasing and therefore $B(t)/t < B(0.5)/0.5 = 1$.

For part II, define $\mathrm{F}_2(t) = B(t) - t$. Then $\mathrm{F}_2''(t) = B''(t)$. By (A.8), $B''(t) < 0$ for $t \in (0.5, 1)$, thus $\mathrm{F}_2(t)$ is strictly concave, giving $\mathrm{F}_2(t) > \max\{\mathrm{F}_2(0.5), \mathrm{F}_2(1)\} = 0$.

Last, we show part III. For $t_2 \in [t_1, 0.5]$, part I indicates that $B(t_1)/t_1 \le B(t_2)/t_2$; for $t_2 \in [0.5, 1]$, part II indicates that $B(t_2)/t_2 \ge 1$ which, combined with $B(t_1)/t_1 \le 1$ from part I, yields $B(t_1)/t_1 \le B(t_2)/t_2$. □

We now prove Theorem 2. It suffices to show that

$$\{1 - G_1(\lambda)\}/(1 - \lambda) \ge \{1 - G_1^*(\lambda)\}/\{1 - G_0^*(\lambda)\}, \tag{A.9}$$

$$\sup_{0<t\le 1} G_1(t)/t \le \sup_{0<t\le 1} G_1^*(t)/G_0^*(t). \tag{A.10}$$

To verify (A.9), it suffices to show that $G_1(\lambda) \le G_1^*(\lambda)$ and $\lambda \ge G_0^*(\lambda)$. Following (A.2), for $0 \le t \le 1$,

$$G_1^*(t) = G_0^*\big(G_1(t)\big). \tag{A.11}$$

Applying (A.11), (A.1), $1 - F_0(F_1^{-1}(0.5)) \le \lambda$ and part II of Lemma 1 yields $G_1(\lambda) \le G_1^*(\lambda)$; applying $\lambda \le 0.5$ and part I of Lemma 1 implies $\lambda \ge G_0^*(\lambda)$. This shows (A.9).

To verify (A.10), let $M = \sup_{0<t\le 1} G_1(t)/t$. Since $G_1(1)/1 = 1$, we have $M \ge 1$ which will be discussed in two cases. Case 1: if $M = 1$, then

$$\sup_{0<t\le 1} \frac{G_1^*(t)}{G_0^*(t)} \ge \frac{G_1^*(1)}{G_0^*(1)} = 1 = \sup_{0<t\le 1} \frac{G_1(t)}{t}. \tag{A.12}$$

Case 2: if $M > 1$, then there exist $t_0 \in [0, 1]$ and $t_n \in (0, 1)$ such that $\lim_{n\to\infty} t_n = t_0$, and

$$\lim_{n\to\infty} \{G_1(t_n)/t_n\} = \sup_{0<t\le 1} \{G_1(t)/t\} = M > 1. \tag{A.13}$$

Thus, there exists N_1 such that for all $n > N_1$,

$$G_1(t_n) > t_n. \tag{A.14}$$

Cases of $t_0 = 1$, $t_0 = 0$ and $t_0 \in (0, 1)$ will be discussed separately. First, if $t_0 = 1$, then $M = \lim_{n\to\infty}\{G_1(t_n)/t_n\} = \lim_{n\to\infty} G_1(t_n) \le 1$, which contradicts (A.13). Thus $t_0 < 1$. Second, if $t_0 = 0$, then there exists N_2 such that $t_n < 0.5$ for all $n > N_2$. Thus for all $n > N \equiv \max\{N_1, N_2\}$, applying (A.11), (A.14) and part III of Lemma 1, we have

$$G_1^*(t_n)/G_1(t_n) = G_0^*\big(G_1(t_n)\big)/G_1(t_n) \ge G_0^*(t_n)/t_n.$$

This together with (A.13) shows

$$\sup_{0<t\le 1}\frac{G_1^*(t)}{G_0^*(t)} \ge \limsup_{n\to\infty}\frac{G_1^*(t_n)}{G_0^*(t_n)} \ge \lim_{n\to\infty}\frac{G_1(t_n)}{t_n} = M = \sup_{0<t\le 1}\frac{G_1(t)}{t}. \tag{A.15}$$

Third, for $t_0 \in (0, 1)$, since both F_0 and F_1 are differentiable and f_0 is supported in a single interval, $G_1(t)/t = \{1 - F_1(F_0^{-1}(1-t))\}/t$ is differentiable in $(0, 1)$. Thus,

$$\sup_{0<t\le 1} G_1(t)/t = G_1(t_0)/t_0 = M, \tag{A.16}$$

and $d\{G_1(t)/t\}/dt|_{t=t_0} = 0$. Notice

$$\left.\frac{d\{G_1(t)/t\}}{dt}\right|_{t=t_0} = \frac{\frac{dG_1(t)}{dt}|_{t=t_0} - G_1(t_0)/t_0}{t_0} = \frac{\frac{dG_1(t)}{dt}|_{t=t_0} - M}{t_0} = 0. \tag{A.17}$$

If $t_0 > 0.5$, then $F_0^{-1}(1-t_0) \le F_0^{-1}(0.5)$. By (A.3) and the assumption on f_0 and f_1, $dG_1(t)/dt|_{t=t_0} = f_1(F_0^{-1}(1-t_0))/f_0(F_0^{-1}(1-t_0)) \le 1$, which contradicts (A.17). Thus, $0 < t_0 \le 0.5$. This together with (A.11), (A.16), and part III of Lemma 1 gives

$$G_1^*(t_0)/G_1(t_0) = G_0^*\big(G_1(t_0)\big)/G_1(t_0) \ge G_0^*(t_0)/t_0.$$

This together with (A.16) shows,

$$\sup_{0<t\le 1}\frac{G_1^*(t)}{G_0^*(t)} \ge \frac{G_1^*(t_0)}{G_0^*(t_0)} \ge \frac{G_1(t_0)}{t_0} = M = \sup_{0<t\le 1}\frac{G_1(t)}{t}. \tag{A.18}$$

Combining (A.12), (A.15) and (A.18) completes the proof.

Calculation of $\alpha_\infty^{\mathrm{FDR}_L}$ of the mean filter in Table 3 of Sect. 4.1 From (A.1) and the conditions given in Sect. 4.1,

$$G_0(t) = t \quad \text{for } t \in [0, 1], \quad \text{and} \quad G_1(t) = \begin{cases} te^C, & \text{if } t \in [0, e^{-C}], \\ 1, & \text{if } t \in (e^{-C}, 1]. \end{cases} \tag{A.19}$$

Now we compute $\alpha_\infty^{\mathrm{FDR}_L}$ of the mean filter. Recall that the distribution $G_0^*(t)$ is that of $\overline{\mathrm{U}}_k = \sum_{i=1}^k \mathrm{U}_i/k$, where $\{\mathrm{U}_i\}_{i=1}^k \overset{\text{i.i.d.}}{\sim} \mathrm{Unif}(0, 1)$, and the distribution $G^*(t)$ is also that of $\overline{\mathrm{U}}_k$. Similarly, by (A.2), the distribution $G_1^*(t)$ is that of $\overline{\mathrm{U}}_k/e^C$.

Hence, $\widehat{\mathrm{FDR}_L}^{\infty}(t) = [\pi_0 + \pi_1\{1 - G_1^*(\lambda)\}/\{1 - G_0^*(\lambda)\}]/\{\pi_0 + \pi_1 G_1^*(t)/G_0^*(t)\}$, yielding

$$\alpha_\infty^{\mathrm{FDR}_L} = \frac{\pi_0 + \pi_1\{1 - G_1^*(\lambda)\}/\{1 - G_0^*(\lambda)\}}{\pi_0 + \pi_1 \sup_{0<t\le 1} G_1^*(t)/G_0^*(t)}.$$

Note that

$$\frac{G_1^*(t)}{G_0^*(t)} = \frac{\mathrm{P}(\overline{\mathrm{U}}_k \le te^C)}{\mathrm{P}(\overline{\mathrm{U}}_k \le t)} = \begin{cases} \frac{\mathrm{P}(\overline{\mathrm{U}}_k \le te^C)}{\mathrm{P}(\overline{\mathrm{U}}_k \le t)}, & \text{if } 0 \le t \le e^{-C}, \\ \frac{1}{\mathrm{P}(\overline{\mathrm{U}}_k \le t)}, & \text{if } e^{-C} < t \le 1. \end{cases}$$

Note that $1/\mathrm{P}(\overline{\mathrm{U}}_k \le t)$ is decreasing in t. By a graphical approach for $k = 5$, $\mathrm{P}(\overline{\mathrm{U}}_k \le te^C)/\mathrm{P}(\overline{\mathrm{U}}_k \le t)$ is also decreasing in t. Thus, $\sup_{t\in(0,1]} G_1^*(t)/G_0^*(t) = \lim_{t\to 0+} G_1^*(t)/G_0^*(t) = \lim_{t\to 0+}\{G_1(t)/t\}^k = e^{Ck}$ from (A.4) and (A.19). So $\alpha_\infty^{\mathrm{FDR}_L} = [\pi_0 + \pi_1\{1 - G_1^*(\lambda)\}/\{1 - G_0^*(\lambda)\}]/(\pi_0 + \pi_1 e^{Ck})$, where $G_0^*(\cdot)$ can be calculated from (3.5). The completes the derivation.

References

Arias-Castro E, Donoho DL (2009) Does median filtering truly preserve edges better than linear filtering? Ann Stat 37:1172–1206

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc B 57:289–300

Brown BM (1983) Statistical uses of the spatial median. J R Stat Soc B 45:25–30

Efron B (2010) Large-scale inference. Empirical Bayes methods for estimation, testing, and prediction. Cambridge University Press, Cambridge

Feller W (1968) An introduction to probability theory and its applications, Vol. I, 3rd edn. Wiley, New York

Ruiz S (1996) An algebraic identity leading to Wilson's theorem. Math Gaz 80:579–582

Sadooghi-Alvandi S, Nematollahi A, Habibi R (2009) On the distribution of the sum of independent uniform random variables. Stat Pap 50:171–175

Storey JD, Taylor JE, Siegmund D (2004) Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach. J R Stat Soc B 66:187–205

Uspensky JV (1937) Introduction to mathematical probability. McGraw-Hill, New York

van-der Vaart AW (1998) Asymptotic statistics. Cambridge University Press, Cambridge

Zhang CM, Yu T (2008) Semiparametric detection of significant activation for brain fMRI. Ann Stat 36:1693–1725

Zhang CM, Fan J, Yu T (2011) Multiple testing via FDR_L for large-scale imaging data. Ann Stat 39:613–642