Annals of the Institute of Statistical Mathematics (2021) 73:311–336
https://doi.org/10.1007/s10463-020-00752-5

On the power of some sequential multiple testing procedures

Shiyun Chen[1] · **Ery Arias-Castro**[2]

Received: 18 January 2019 / Revised: 21 January 2020 / Published online: 2 April 2020

Abstract

We study an online multiple testing problem where the hypotheses arrive sequentially in a stream. The test statistics are independent and assumed to have the same distribution under their respective null hypotheses. We investigate two recently proposed procedures LORD and LOND, which are proved to control the FDR in an online manner. In some (static) model, we show that LORD is optimal in some asymptotic sense, in particular as powerful as the (static) Benjamini–Hochberg procedure to first asymptotic order. We also quantify the performance of LOND. Some numerical experiments complement our theory.

Keywords Online multiple testing · False discovery rate (FDR) control · Asymptotic optimality · False non-discovery rate (FNR) analysis

1 Introduction

Multiple testing is now a well-established area in statistics and arises in almost every scientific field (Dudoit and van der Laan 2007; Dickhaus 2014; Roquain 2011). The literature on multiple testing has focused almost exclusively on the offline setting, which assumes the entire batch of hypotheses and the corresponding P values are available.

In this paper, we consider, instead, an online multiple testing scenario where infinitely many hypotheses $\mathcal{H} = (\mathbb{H}_1, \mathbb{H}_2, \mathbb{H}_3, \ldots)$ arrive sequentially in a stream with corresponding P values, $P_1, P_2, P_3, \ldots$, and we are required to decide whether we accept or reject $\mathbb{H}_i$ only based on $P_1, \ldots, P_i$, without access to the total number

✉ Ery Arias-Castro
eariascastro@ucsd.edu

Shiyun Chen
shc176@ucsd.edu

[1] Amazon, Seattle, WA, USA

[2] University of California, San Diego, CA, USA

of hypotheses in the stream or future P values. Our purpose is to study the recent online multiple testing procedures introduced by Javanmard and Montanari (2018). They are known to control the FDR in an online manner, and we study their power properties in an asymptotic framework where the offline method of Benjamini and Hochberg (1995) plays the role of (oracle) benchmark.

1.1 The risk of a multiple testing procedure

Consider a setting where we want to test an ordered infinite sequence of null hypotheses, denoted $\mathcal{H} = (\mathbb{H}_1, \mathbb{H}_2, \mathbb{H}_3, \dots ,)$, where at each step i we have to decide whether to reject $\mathbb{H}_i$ having access to only previous decisions. The test that we use for $\mathbb{H}_i$ rejects for large positive values of a statistic X_i. Throughout, we assume that test statistics are all independent. Denote the collection of the first n hypotheses in the stream by $\mathcal{H}(n) = (\mathbb{H}_1, \dots, \mathbb{H}_n)$, and the vector of first n test statistics by $\mathbf{X}(n) = (X_1, \dots, X_n)$. Let $\boldsymbol{\Phi}_i$ denote the (true) survival function[1] of X_i and $\boldsymbol{\Phi}(n) = (\boldsymbol{\Phi}_1, \dots, \boldsymbol{\Phi}_n)$. We assume that the P value corresponding to X_i can be computed. The simplest such case is when $\mathbb{H}_i$ is a singleton, $\mathbb{H}_i = \{\boldsymbol{\Phi}_i^{\text{null}}\}$, and the corresponding null distribution $\boldsymbol{\Phi}_i^{\text{null}}$ is known. In that case, the ith P value is defined as $P_i = \boldsymbol{\Phi}_i^{\text{null}}(X_i)$, which is the probability of exceeding the observed value of the statistic under its null distribution.

Let $\mathcal{F}$ index all the false null hypotheses in the stream, and let $\mathcal{F}_n \subset [n] := \{1, \dots, n\}$ index the false null hypotheses in the first n hypotheses, meaning

$$\mathcal{F}_n = \{i \in [n] : \boldsymbol{\Phi}_i \notin \mathbb{H}_i\}. \tag{1}$$

A multiple testing procedure $\mathcal{R}$, for each $n \geq 1$, takes in $\mathbf{X}(n)$ and returns a subset of $\{1, \dots, n\}$ indicating the null hypotheses that the procedure rejects among the first n in the sequence. Given such a procedure $\mathcal{R}$, the false discovery rate is defined as the expected value of the false discovery proportion (Benjamini and Hochberg 1995)

$$\text{FDR}_n(\mathcal{R}) = \mathbb{E}_{\boldsymbol{\Phi}}\left[\frac{|\mathcal{R}(\mathbf{X}(n)) \setminus \mathcal{F}_n|}{|\mathcal{R}(\mathbf{X}(n))|}\right], \tag{2}$$

where we denoted the cardinality of a set $\mathcal{A} \subset [n]$ by $|\mathcal{A}|$ and use the convention that $0/0 = 0$. While the FDR of a multiple testing procedure is analogous to the level or size of a test procedure, the false non-discovery rate (FNR) plays the role of power and is defined as the expected value of the false non-discovery proportion[2]

[1] In this paper, the survival function of a random variable Y is defined as $y \mapsto \mathbb{P}(Y \geq y)$.

[2] We note that this definition is different from that of Genovese and Wasserman (2002). According to our definition, the FNR is the expected fraction of non-nulls that are not correctly rejected out of all non-nulls, while according to the definition of Genovese and Wasserman (2002), the FNR is the fraction of non-nulls that are not rejected out of all non-rejections. We find our definition to be more appropriate in the asymptotic setting that we consider.

$$\mathrm{FNR}_n(\mathcal{R}) = \mathbb{E}_{\boldsymbol{\Phi}}\left[\frac{|\mathcal{F}_n \backslash \mathcal{R}(\mathbf{X}(n))|}{|\mathcal{F}_n|}\right]. \tag{3}$$

In analogy with the risk of a test—which is defined as the sum of the probabilities of type I and type II error—we define the risk of a multiple testing procedure $\mathcal{R}$ as the sum of the false discovery rate and the false non-discovery rate

$$\mathrm{risk}_n(\mathcal{R}) = \mathrm{FDR}_n(\mathcal{R}) + \mathrm{FNR}_n(\mathcal{R}). \tag{4}$$

Remark 1 The procedure that never rejects and the one that always rejects both achieve a risk of 1, so that any method that has a risk exceeding 1 is useless.

1.2 More related work

1.2.1 Related work in online multiple testing and FDR control

The literature on multiple testing is by now vast. Only more recently, though, have multiple testing procedures been proposed for the online setting. The online multiple testing framework goes back to the work of Foster and Stine (2008), who propose a class of online multiple testing procedures referred to as *alpha-investing rules*, which provide in the online setting a uniform control of the marginal FDR (mFDR). Building on this early work, Aharoni and Rosset (2014) propose a broader class of online procedures called *generalized alpha-investing rules* (GAI), which are also shown to control the mFDR. Javanmard and Montanari (2015) propose two closely related procedures called LOND and LORD algorithms which control both FDR and mFDR in online testing. We refer to Sects. 4.1 and 4.2 for more details of rules and discuss their asymptotic risk in our context. In the published version of their work, Javanmard and Montanari (2018) study generalized alpha-investing rules and derive conditions under which they provide an online FDR control. They also propose other procedures for online control of the false discovery exceedance. In their paper, they establish some power lower bound for LORD under the Gaussian mixture model with a fixed fraction of contamination. We note that the lower bound becomes trivial (tending to zero) under the sparse Gaussian mixture model where the fraction of contamination is vanishing to zero, which is the regime we consider in the present paper. Ramdas et al. (2017) modify the GAI class (referred as to GAI++ in their paper) to improve the power of GAI rules (uniformly) while still controlling the FDR, and in particular propose LORD++, a variant of LORD. In follow-up work, Ramdas et al. (2018) demonstrate that using adaptivity à la (Storey 2002), a method called SAFFRON is able to improve on LORD++. (We note that these works appeared after ours was made available.)

In the present paper, we study the asymptotic power properties of the LORD and LOND methods, complementing the results of Javanmard and Montanari (2018) in the online multiple testing setup. To the best of our knowledge, this is the first time that the (asymptotic) power of online testing rules is established under the sparse Gaussian mixture model defined in (9).

This paper is a continuation of our previous work in the static[3] (offline) setting (Arias-Castro and Chen 2017), where an asymptotic oracle risk bound for multiple testing is obtained, and both the method of Benjamini and Hochberg (1995) and the distribution-free method of Foygel-Barber and Candès (2015) are proved to achieve that bound. Compared with the offline testing problem considered by Arias-Castro and Chen (2017), the online testing problem is more complicated in that an online procedure assigns a different threshold to each p value arriving sequentially, instead of using the same threshold as the BH method does.

Various other oracle bounds and corresponding optimality results for multiple testing procedures are available in the literature; see, for example, Genovese and Wasserman (2002), Sun and Cai (2007), Storey (2007), Bogdan et al. (2011), Neuvial and Roquain (2012), Meinshausen et al. (2011), Ji and Jin (2012), Jin and Ke (2016) and Butucea et al. (2018).

1.2.2 Related work in other sequential setting

There is a broader literature on sequential multiple testing problems. For example, Bartroff (2013) consider the setup of testing $K > 2$ null hypotheses $\mathbb{H}_1, \dots, \mathbb{H}_K$, where K is fixed. For each hypothesis $\mathbb{H}_k$, the data $X_1^{(k)}, \dots, X_n^{(k)}$ comes sequentially in a stream. So in their paper, "sequential" refers to the manner in which data is collected, rather than the manner in which the hypotheses are considered (which is the setting we consider). They propose variant of the BH procedure which can deal with this sequential sampling setup and prove its simultaneous FDR and FNR control (which are defined differently than we do here). See Bartroff (2014) and Bartroff and Song (2014) for general step-up and step-down procedures which control the simultaneous generalized type I and II error rates (including FDR and FWER) in this sequential sampling setting.

Another situation, referred to as *sequential selection*, is where the hypotheses are ordered based on side information about how important or promising each hypothesis is, and one is only permitted to reject an initial contiguous block, $\mathbb{H}_1, \dots, \mathbb{H}_k$, of hypotheses. A rejection rule in this setting amounts to a procedure for choosing the stopping point k. In this setting the total number of hypotheses is assumed known. In this context, G'Sell et al. (2016) develop two rules (FowardStop and StrongStop) to decide the first k hypotheses in the sequence to reject, which are shown to control the FDR. Foygel-Barber and Candès (2015) propose their own step-up procedure (SeqStep), which also guaranties FDR control under independence, while Li and Barber (2017) develop a broader class of ordered hypotheses testing procedures under such setting, called *accumulation tests*. Lei and Fithian (2016) propose an adaptive version of SeqStep. Fithian et al. (2015) consider sequential selection in the context of selecting variables in a linear regression modeling.

[3] By "static" we mean a setting where all the null hypotheses of interest are considered together. This is the more common setting considered in the multiple testing literature.

1.3 Content

The rest of the paper is organized as follows. In Sect. 3.1 we consider the normal location model and derive the performance of LORD under this model. Generalizing this model, in Sect. 3.2 we consider a nonparametric Asymptotic Generalized Gaussian model. We analyze the asymptotic performance of the LORD and LOND procedures of Javanmard and Montanari (2018) under this model in Sects. 4.1 and 4.2. We present some numerical experiments in Sect. 5. All proofs are gathered in Sect. 7.

2 Methods

We describe the LORD and LOND procedures of Javanmard and Montanari (2018), which are the methods we study in this paper. Recall that $\mathbb{H}_1, \mathbb{H}_2, \dots$ are tested sequentially and that P_i denotes the P value corresponding to the test of $\mathbb{H}_i$. These two procedures, and most others, work as follows: set a significance level α_i based on $P_1, \dots, P_{i-1}$ (except for α_1 which is set beforehand) and reject $\mathbb{H}_i$ if $P_i \le \alpha_i$. The LORD and LOND methods vary in how they set these thresholds, although they both start with a sequence of the form

$$\lambda_i \ge 0 \text{ such that } \sum_{i=1}^{\infty} \lambda_i = q, \tag{5}$$

where q denotes the desired FDR control level. In what follows, we stay close to the notation used in Javanmard and Montanari (2018).

2.1 The LORD method

Based on a chosen sequence (5), the LORD algorithm—which stands for (significance) Levels based On Recent Discovery—sets the sequential significance levels $(\alpha_i)_{i=1}^{\infty}$ as follows:

$$\alpha_i = \lambda_{i-t_i}, \quad t_i = \max\{l < i : \mathbb{H}_l \text{ is rejected}\}, \tag{6}$$

with $t_i := 0$ for all i before the time of first discovery.

In Javanmard and Montanari (2018) the LORD algorithm is shown to control FDR at a level less than or equal to q in an online fashion, specifically,

$$\sup_{n \ge 1} \mathrm{FDR}_n(\mathcal{R}) \le q, \tag{7}$$

if the P values are independent. More generally, Javanmard and Montanari (2018) study a class of monotone generalized alpha-investing procedures (which includes LORD as a special case) and prove that any rule in this class controls the cumulative

FDR at each stage provided the P values corresponding to true nulls are independent from the other P values.

2.2 The LOND method

Based on a chosen sequence (5), the LOND algorithm—which stands for (significance) Levels based On Number of Discovery—sets the sequential significance levels $(\alpha_i)_{i=1}^{\infty}$ as follows:

$$\alpha_i = \lambda_i(D(i-1)+1), \tag{8}$$

where $D(n)$ denotes the number of discoveries in $\mathcal{H}(n) = (\mathbb{H}_1, \dots, \mathbb{H}_n)$, with $D(0) := 0$.

In Javanmard and Montanari (2018) the LOND is shown to control FDR at level less than or equal to q everywhere in an online manner, the same as (7), if the P values are independent.

3 Models

In this paper, we study the FNR of each of the LORD and LOND methods of Javanmard and Montanari (2018) on the first n hypotheses as $n \to \infty$. As benchmark, we use the oracle that we considered previously (Arias-Castro and Chen 2017) for the static setting defined by these n hypothesis testing problems. For the reader not familiar with that paper, at least in the models that we consider, this turns out to be asymptotically equivalent to applying the Benjamini–Hochberg (BH) method to the first n hypotheses. Note that the latter accesses all the first n hypotheses at once and is thus not constrained to be sequential in nature.

The static setting we consider is that of a location mixture model. We assume that we know the null distribution function Φ, assumed to be continuous for simplicity. We then assume that the test statistics are independent with respective distribution $X_i \sim \Phi_i = \Phi(\cdot - \mu_i)$, where $\mu_i = 0$ under the null $\mathbb{H}_i$ and $\mu_i > 0$ otherwise. Both minimax and Bayesian considerations lead one to consider a prior on the μ_i's where a fraction ε of the μ_i's are randomly picked and set equal to some $\mu > 0$, while the others are set to 0. The prior is therefore defined based on ε and μ, which together control the signal strength. The P value corresponding $\mathbb{H}_i$ is $P_i := \bar{\Phi}(X_i)$, where $\bar{\Phi} := 1 - \Phi$ is the null survival function.

3.1 The normal model

As an emblematic example of the distributional models that we consider in this paper, let Φ denote the standard normal distribution. Assume as above that $X_i \sim \Phi$ under $\mathbb{H}_i$ and $X_i \sim \Phi(\cdot - \mu)$ otherwise. Thus, under each null hypothesis, the corresponding test statistic is standard normal, while that statistic is normal with mean μ and unit variance otherwise. This is the model we consider in Arias-Castro and Chen (2017) and the inspiration comes from a line of research on testing the global null $\bigcap_i \mathbb{H}_i$ in the static setting (Ingster 1997; Ingster and Suslina 2003; Donoho and Jin 2004). As in this line of work, we use the parameterization pioneered by Ingster (1997), namely

$$\varepsilon = n^{-\beta} \text{ and } \mu = \sqrt{2r \log n}. \tag{9}$$

In the static setting, we know from our previous work (Arias-Castro and Chen 2017) that any threshold-type procedure has risk tending to 1 as $n \to \infty$ when $r < \beta$ are fixed. We also know that the BH method with FDR control at $q \to 0$ slowly has risk tending to 0 when $r > \beta$ are fixed. In fact, these results are derived in the wider context of an asymptotically generalized Gaussian model, which we consider later. Thus, $r = \beta$ is the static selection boundary.

Remark 2 (Javanmard and Montanari 2018) compared the power of their procedures in terms of lower bounds on the total discovery rate under the same mixture model but with a fixed mixture weight ε. This is in keeping with the seminal work of Genovese and Wasserman (2002). In contrast, here we focus on a "sparser" setting where $\varepsilon = o(1)$, meaning that the fraction of false null hypotheses (i.e., true discoveries) is negligible compared to the total number of null hypotheses being tested.

3.2 Asymptotically generalized Gaussian model

Beyond the normal model, we follow (Arias-Castro and Chen 2017; Donoho and Jin 2004) and consider other location models where the base distribution has a polynomial right tail in log scale.

Definition 1 A survival function $\bar{\Phi} = 1 - \Phi$ is asymptotically generalized Gaussian (AGG) on the right with exponent $\gamma > 0$ if $\lim_{x\to\infty} x^{-\gamma} \log \bar{\Phi}(x) = -1/\gamma$.

The AGG class of distributions is nonparametric and quite general. It includes the parametric class of generalized Gaussian (GG) distributions with densities $\{\psi_\gamma, \gamma > 0\}$ given by $\log \psi_\gamma(x) \propto -|x|^\gamma/\gamma$, which comprises the normal distribution ($\gamma = 2$) and the double-exponential distribution ($\gamma = 1$). We assume that $\gamma \geq 1$ so that the null distribution has indeed a sub-exponential right tail.

Remark 3 We note that the scale (e.g., standard deviation) is fixed, but this is really without loss of generality as both the LORD and LOND methods are scale invariant. This is because the P values are scale invariant.

The model is the same as the one considered in Sect. 3.1 except that Φ is an AGG distribution with parameter $\gamma \geq 1$. As in our previous work (Arias-Castro and Chen 2017), we use the following parameterization

$$\varepsilon = n^{-\beta} \text{ and } \mu = (\gamma r \log n)^{1/\gamma}, \tag{10}$$

where $r \geq 0$ and $\beta \in (0, 1)$ are always assumed fixed.

4 Performance analysis

In this section, we analyze the performance of the LORD and LOND methods in the static setting described earlier. Recall that q denotes the desired FDR control level. Typically q is set to a small number, like $q = 0.10$. In this paper, we allow $q \to 0$ as $\varepsilon \to 0$, but slowly. Specifically, we always assume that

$$q = q(n) > 0 \text{ and } n^a q(n) \to \infty \text{ for all fixed } a > 0. \tag{11}$$

4.1 The performance of LORD

We first establish a performance bound for LORD. It happens that, despite required to control the FDR in an online fashion, LORD achieves the static selection boundary when desired FDR control is appropriately set.

Theorem 1 (Performance bound for LORD) *Consider a static AGG mixture model with exponent $\gamma \geq 1$ parameterized as in (10). Assume that we apply LORD with $(\lambda_i)_{i=1}^{\infty}$ defined as $\lambda_i \propto i^{-\nu}$ with $\sum_{i=1}^{\infty} \lambda_i = q$, where $\nu > 1$and q satisfies (11). If $r > \nu\beta$, the LORD procedure has $\text{FNR}_n \to 0$ as $n \to \infty$. In particular, if $q \to 0$, then it has risk tending to 0.*

Note that the latter part comes from the fact that the LORD procedure controls of the FDR at the desired level q as established in (Javanmard and Montanari 2018) in the more demanding online setting. In essence, therefore, LORD (with a proper choice of ν above) achieves the static oracle selection boundary $r = \beta$.

Remark 4 Assume that, instead, we apply LORD with any decreasing sequence $(\lambda_i)_{i=1}^{\infty}$ satisfying $\sum_{i=1}^{\infty} \lambda_i = q$ and

$$i^{\nu} \lambda_i \to \infty, \text{ for any fixed } \nu > 1. \tag{12}$$

Then the conclusions of Theorem 1 remain valid. In particular, such a choice of sequence (e.g., $\lambda_i \propto (\log i)^2/i$) adapts to the (usually unknown) values of r and β. (We provide details in Sect. 7.)

4.2 The performance of LOND

We now turn to LOND and establish a performance bound under the same setting.

Theorem 2 *Consider a static AGG mixture model with exponent $\gamma \geq 1$ parameterized as in (10). Assume that we apply LOND with $(\lambda_i)_{i=1}^{\infty}$ defined as $\lambda_i \propto i^{-\nu}$ with $\sum_{i=1}^{\infty} \lambda_i = q$, where $\nu > 1$ and q satisfies (11). If $r > \beta + (\nu^{1/\gamma} - r^{1/\gamma})^{\gamma} + \nu - 1$, the*

LOND procedure has $\text{FNR}_n \to 0$ *as* $n \to \infty$. *In particular, if* $q \to 0$, *then it has risk tending to 0.*

In essence, LOND (with a proper choice of ν above) has risk tending to 0 when $r - (1 - r^{1/\gamma})^\gamma > \beta$. This is the best upper bound that we were able to establish for the LOND algorithm. We do not know if it is optimal or not. In particular, it's quite possible that LOND also achieves the static selection boundary.

Remark 5 The analog of Remark 4 applies here as well. (Technical details are omitted.)

5 Numerical experiments

In this section, we perform some simulations to study the numerical performance of LORD and LOND, and also to compare them with the (static) BH procedure. We consider the normal model and the double-exponential model. It is worth repeating that the BH procedure, which is a static procedure, requires knowledge of all P values to determine the significance level for testing the hypotheses. Hence, it does not address the scenario in online testing. In contrast, the sequential methods decide the significance level at each step based on previous outcomes and are required to control the FDR at each step.

In our experiments, for both LORD and LOND, we choose the sequence $(\lambda_i)_{i=1}^\infty$ as

$$\lambda_i = \frac{L}{i^\nu}, \quad \nu = 1.05, \tag{13}$$

with L set to ensure $\sum_{i=1}^\infty \lambda_i = q$, where (as before) q denotes the desired FDR level.

5.1 Fixed number of hypotheses

In this first set of experiments, the number of tests is chosen large at $n = 10^9$, to assess the accuracy of our asymptotic analysis, and then small at $n = 10^3$, to see how the various methods behave when the number of tests is more moderate. ("Appendix" describes additional simulation results with varying number of hypotheses and desired FDR level q.)

We draw m values of the test statistics from the alternative distribution $\Phi(\cdot - \mu)$, and the other $n - m$ from the null distribution Φ. All the models are parameterized as in (10). We choose a few values for the parameter β so as to exhibit different sparsity levels, while the parameter r takes values in a grid spanning [0, 1.5]. Each situation is repeated 500 times, and we report the average FDP and FNP for each procedure. The FDR control level is set at $q = 0.10$.

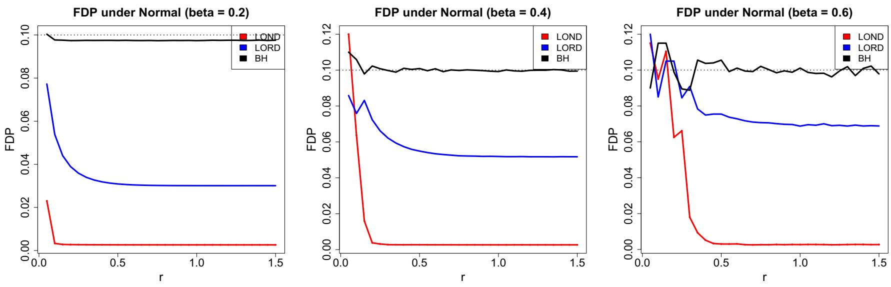

Fig. 1 Simulation results showing the FDP for the BH, LORD and LOND methods under the normal model in three distinct sparsity regimes. The black horizontal line delineates the desired FDR control level ($q = 0.1$). The number of the hypotheses is $n = 10^9$

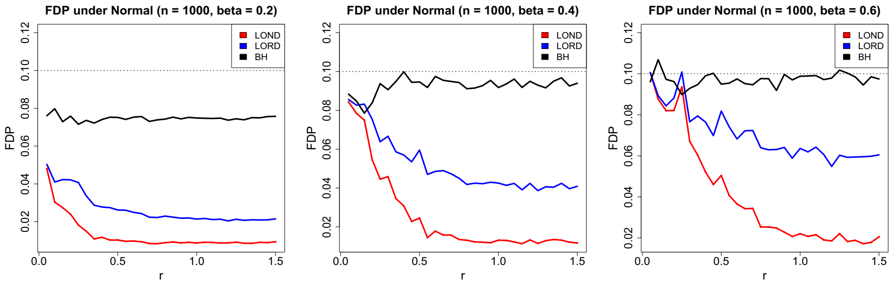

Fig. 2 Simulation results showing the FDP for the BH, LORD and LOND methods under the normal model in three distinct sparsity regimes. The black horizontal line delineates the desired FDR control level ($q = 0.1$). The number of the hypotheses is $n = 10^3$

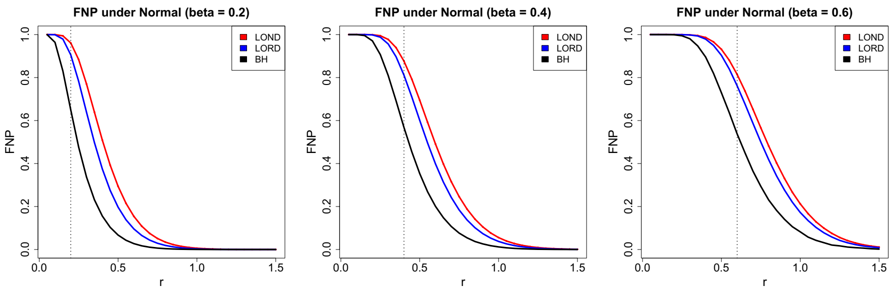

Fig. 3 Simulation results showing the FNP for the BH, LORD and LOND methods under the normal model in three distinct sparsity regimes. The black vertical line delineates the theoretical threshold ($r = \beta$). The number of the hypotheses is $n = 10^9$

5.1.1 Normal model

In this model, $\boldsymbol{\Phi}$ is the standard normal distribution. The simulation results are reported in Figs. 1 and 3 for $n = 10^9$ and in Figs. 2 and 4 for $n = 10^3$. In Figs. 1 and 2 we report the FDP. We see that LOND becomes more conservative than

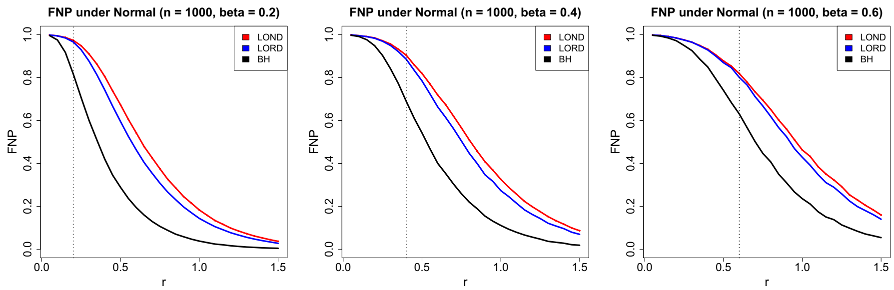

Fig. 4 Simulation results showing the FNP for the BH, LORD and LOND methods under the normal model in three distinct sparsity regimes. The black vertical line delineates the theoretical threshold ($r = \beta$). The number of the hypotheses is $n = 10^3$

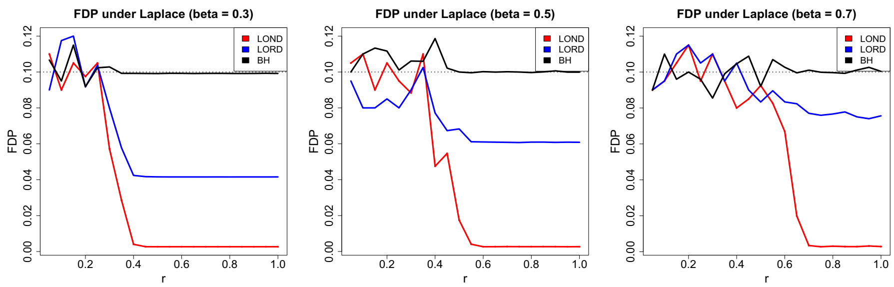

Fig. 5 Simulation results showing the FDP for the BH, LORD and LOND methods under the double-exponential model in three distinct sparsity regimes. The black horizontal line delineates the desired FDR control level ($q = 0.1$). The number of the hypotheses is $n = 10^9$

LORD as r increases under both scenarios. In Figs. 3 and 4 we report the FNP. We see that LOND is clearly less powerful than LORD in the regime $\beta = 0.2$, but performs comparably to LORD in the regime $\beta = 0.6$. This is in line with the theory that LOND can at least achieve the line $r = \beta + (1 - r^{1/\gamma})^{\gamma}$, which is getting closer to $r = \beta$ with increasing values of β. We notice that both LORD and LOND are clearly less powerful than BH in finite samples, even at $n = 10^9$, even though our theory says that LORD achieves the same selection boundary as BH in the large-sample limit. Also, due to the limitation in choice of ν (here $\nu = 1.05$), the selection boundary that LORD can achieve is $r = \nu\beta$ by theory, which explains why LORD lags behind BH—although we do see that as n increases from 10^3 to 10^9, the distance between LORD and BH is smaller, and also the transition from powerless (high FNP) to powerful (low FNP) is sharper (for all methods) around the selection boundary of $r = \beta$.

5.1.2 Double-exponential model

In this model, Φ is the double-exponential distribution with variance 1. The simulation results for $n = 10^9$ are reported in Figs. 5 (FDP) and 7 (FNP), and the results

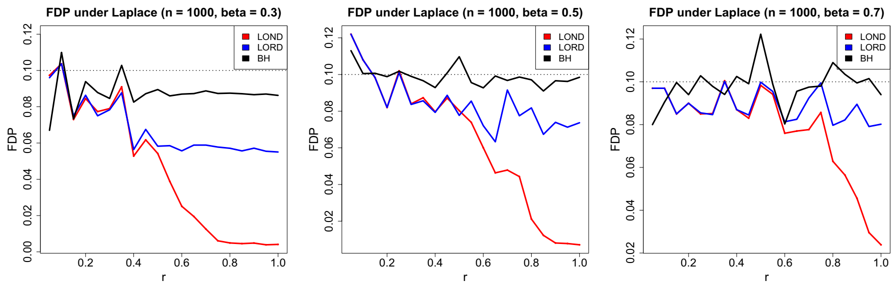

Fig. 6 Simulation results showing the FDP for the BH, LORD and LOND methods under the double-exponential model in three distinct sparsity regimes. The black horizontal line delineates the desired FDR control level ($q = 0.1$). The number of the hypotheses is $n = 10^3$

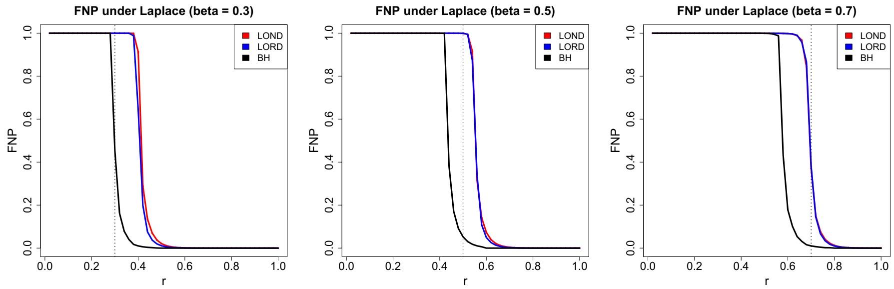

Fig. 7 Simulation results showing the FNP for the BH, LORD and LOND methods under the double-exponential model in three distinct sparsity regimes. The black vertical line delineates the theoretical threshold ($r = \beta$). The number of the hypotheses is $n = 10^9$

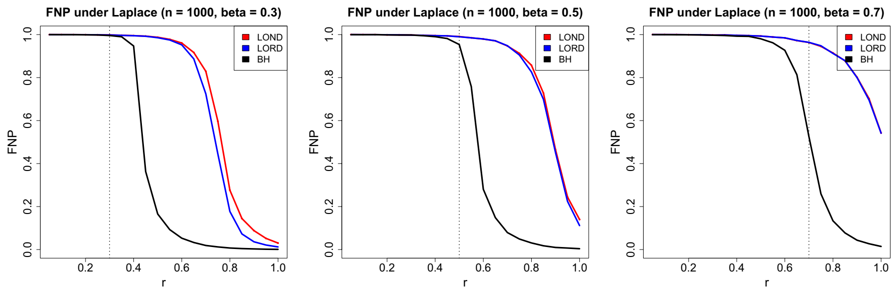

Fig. 8 Simulation results showing the FNP for the BH, LORD and LOND methods under the double-exponential model in three distinct sparsity regimes. The black vertical line delineates the theoretical threshold ($r = \beta$). The number of the hypotheses is $n = 10^3$

for $n = 10^3$ are reported in Figs. 6 (FDP) and 8 (FNP). Here, we observe that LOND becomes more conservative than LORD as r increases in terms of FDP. The LOND and LORD perform more comparably than in the normal setting in terms of FNP, especially when β is close to 1. This is again in line with our theoretical results. The transition from powerful to powerless appears sharper, in particular when n is large (Figs. 6, 7, 8).

6 Discussion

In this paper, we have considered the online multiple testing scenario where a possibly infinite sequence of hypotheses arrive in a stream, and decisions are made only based on previous results before the next hypothesis arrives. We investigate the asymptotic properties of the recent sequential multiple testing procedures, LORD and LOND, of Javanmard and Montanari (2015, 2018), which have been proved to control the FDR in an online manner. We have compared their performance with the (offline) BH method in the context of a sparse asymptotically generalized Gaussian mixture model, and have shown that LORD can achieve the same power as the BH method to first asymptotic order. We have also quantified the asymptotic performance of LOND in the same setting. Although we were not able to establish LOND as being as powerful as LORD (to first order), in our simulations their performances are comparable. (We note that Javanmard and Montanari (2018) introduce LORD but do not mention LOND.)

7 Proofs

We prove our results in this section. Let Φ denote the CDF of null distribution. Without loss of generality, we assume throughout that $\Phi(0) = 1/2$. Let $F(t)$ denote the CDF of the P values under alternatives so that

$$F(t) = \Phi(\mu - \Phi^{-1}(1-t)), \tag{14}$$

where Φ^{-1} is the inverse function of Φ. Let

$$G(t) = (1-\varepsilon)t + \varepsilon F(t), \tag{15}$$

which is the CDF of the P values from the mixture model. Let $\bar{F} = 1 - F$, which is the survival function of the P values under alternatives. Note that

$$\bar{F}(t) = 1 - F(t) = 1 - \Phi(\mu - \xi) = \bar{\Phi}(\mu - \xi), \tag{16}$$

where $\xi := \Phi^{-1}(1-t)$, or equivalently, $t = \bar{\Phi}(\xi)$. Because Φ is as in Definition 1, when $\xi \to \infty$, we have

$$t = \bar{\Phi}(\xi) = \exp\left\{-\frac{\xi^\gamma}{\gamma}(1+o(1))\right\} \to 0, \tag{17}$$

which also implies, when $t \to 0$, that

$$\xi = \Phi^{-1}(1-t) \sim (\gamma \log(1/t))^{1/\gamma}. \tag{18}$$

7.1 Discovery times (LORD)

We apply LORD to the static setting under consideration. Denote τ_l as the time of lth discovery (with $\tau_0 = 0$), and $\Delta_l = \tau_l - \tau_{l-1}$ as the time between the $(l-1)$th and lth discoveries. Assume a sequence satisfying (5) has been chosen. Given the update rule of (6), it can be seen that the inter-discovery times $\{\Delta_l : l \geq 1\}$ are IID.

To prove Theorem 1, we will use the following bound on the expected inter-discovery time.

Proposition 1 *Consider a static AGG mixture model with exponent* $\gamma \geq 1$ *parameterized as in* (10). *Assume that* $\beta \in (0, 1)$ *and* $r \geq 0$ *are both fixed. Assume that* $r > \beta$ *and let* $\nu > 1$ *be such that* $\nu < r/\beta$. *If we apply LORD with* $(\lambda_i)_{i=1}^{\infty}$ *defined as* $\lambda_i \propto i^{-\nu}$ *with* $\sum_{i=1}^{\infty} \lambda_i = q$,

$$\mathbb{E}(\Delta_l \wedge n) \leq 2n^{\beta} + C, \quad \text{for all } l > 0, \tag{19}$$

for some $C > 0$ that does not depend on n. The same holds if we apply LORD with $(\lambda_i)_{i=1}^{\infty}$ satisfying (12) and $\sum_{i=1}^{\infty} \lambda_i = q$.

We prove this result. Recall the definition of G in (15) and note that $G \geq \varepsilon F$. By the update rule of LORD algorithm, for all $m \geq 1$ we have

$$\begin{aligned}\mathbb{P}(\Delta_l > m) &= \prod_{i=\tau_{l-1}+1}^{\tau_{l-1}+m} (1 - G(\alpha_i)) = \prod_{i=\tau_{l-1}+1}^{\tau_{l-1}+m} (1 - G(\lambda_{i-\tau_{l-1}})) \\ &= \prod_{i=1}^{m} (1 - G(\lambda_i)) \leq \exp\left\{ -\sum_{i=1}^{m} G(\lambda_i) \right\} \leq \exp\left\{ -\varepsilon \sum_{i=1}^{m} F(\lambda_i) \right\}.\end{aligned}$$

Let t^* be the value such that $\Phi^{-1}(1-t^*) = \mu$, i.e., $t^* = \Phi(-\mu) = n^{-r+o(1)}$ by the fact that Φ satisfies Definition 1. Then, for $t \geq t^*$, we get

$$\Phi^{-1}(1-t) \leq \Phi^{-1}(1-t^*) = \mu,$$

and then

$$F(t) = \Phi(\mu - \Phi^{-1}(1-t)) \geq \Phi(\mu - \Phi^{-1}(1-t^*)) = \Phi(\mu - \mu) = \Phi(0) = 1/2,$$

so that if $\lambda_i = Li^{-\nu} \geq t^*$, i.e., $i \leq n_1 := \lfloor (L/t^*)^{1/\nu} \rfloor = n^{r/\nu + o(1)}$, we have $F(\lambda_i) \geq \Phi(0) = 1/2$.

Remark 6 If instead $(\lambda_i)_{i=1}^{\infty}$ satisfies (12) then $i^{\nu}\lambda_i \to \infty$ as $i \to \infty$, so that exists a constant $L > 0$ such that $\lambda_i \geq Li^{-\nu}$ for all i, and this is all that we need to proceed.

Thus, for $m \leq n_1$,

$$\sum_{i=1}^{m} F(\lambda_i) \geq m/2,$$

and for $m > n_1$,

$$\sum_{i=1}^{m} F(\lambda_i) \geq \sum_{i=1}^{n_1} F(\lambda_i) \geq n_1/2.$$

Thus,

$$\mathbb{P}(\Delta_l > m) \leq \exp\{-\varepsilon(m \wedge n_1)/2\}.$$

Next, we bound $\mathbb{E}(\Delta_l \wedge n)$. Due to the fact that $\{\Delta_l \wedge n > m\} = \{\Delta_l > m\}$ for $1 \leq m \leq n-1$, and $\{\Delta_l \wedge n > m\} = \emptyset$ if $m \geq n$, we have

$$\begin{aligned}\mathbb{E}(\Delta_l \wedge n) &= \sum_{m=0}^{\infty} \mathbb{P}(\Delta_l \wedge n > m) \\ &= \sum_{m=1}^{n-1} \mathbb{P}(\Delta_l > m) + 1 \\ &\leq \sum_{m=1}^{n-1} \exp\{-\varepsilon(m \wedge n_1)/2\} + 1.\end{aligned}$$

We split the summation over $1 \leq m \leq n_1$ and $n_1 + 1 \leq m \leq n$ and derive the corresponding upper bound separately. For the first part,

$$\sum_{m=1}^{n_1} \exp\{-\varepsilon(m \wedge n_1)/2\} = \sum_{m=1}^{n_1} \exp\{-\varepsilon m/2\} \leq \frac{1}{\exp\{\varepsilon/2\} - 1} < \frac{2}{\varepsilon} = 2n^{\beta}.$$

For the second part,

$$\sum_{m=n_1+1}^{n-1} \exp\{-\varepsilon(m \wedge n_1)/2\} = \sum_{m=n_1+1}^{n-1} \exp\{-\varepsilon n_1/2\} \leq n \exp\{-\varepsilon n_1/2\} = o(1),$$

since $\varepsilon n_1 = n^{r/\nu-\beta+o(1)}$ and $\frac{r}{\nu} > \beta$. Combining the above two bounds, we obtain

$$\mathbb{E}(\Delta_l \wedge n) \leq 2n^{\beta} + o(1) + 1.$$

This establishes Proposition 1. □

7.2 Proof of Theorem 1

Note the number of false nulls is $m = |\mathcal{F}_n| = \varepsilon n \sim n^{1-\beta}$. The false non-discovery rate of LORD (denoted FNR_n) is as follows:

$$\begin{aligned}
\mathrm{FNR}_n &= \mathbb{E}\left(\frac{\sum_{i=1}^{n} \mathbb{I}\{i \notin \mathcal{H}_0(n) : P_i \geq \alpha_i\}}{m} \right) \\
&= \frac{\sum_{i=1}^{n} \mathbb{E}[\mathbb{E}(\mathbb{I}\{i \notin \mathcal{H}_0(n) : P_i \geq \alpha_i\} \mid \alpha_i)]}{m} \\
&= \frac{\sum_{i=1}^{n} \mathbb{E}[\mathbb{P}(i \notin \mathcal{H}_0(n), P_i \geq \alpha_i \mid \alpha_i)]}{m} \\
&= \frac{\sum_{i=1}^{n} \mathbb{E}[\varepsilon \bar{F}(\alpha_i)]}{\varepsilon n} \\
&= \frac{\sum_{i=1}^{n} \mathbb{E}[\bar{F}(\alpha_i)]}{n}.
\end{aligned}$$

So, it suffices to bound the RHS of the equation.

Let $D(n)$ be the number of discoveries in first n hypotheses $\mathcal{H}(n)$ by applying LORD with the sequence (λ_i). Let $\tilde{\Delta}_l = \Delta_l = \tau_l - \tau_{l-1}$, for $1 \leq l \leq D(n)$, and $\tilde{\Delta}_{D(n)+1} = n - \tau_{D(n)}$. Due to the fact that $0 \leq \tilde{\Delta}_l \leq (\Delta_l \wedge n)$, for $1 \leq l \leq D(n) + 1$, we have for any fixed $\delta > 0$,

$$\mathbb{P}(\tilde{\Delta}_l \geq \frac{\mathbb{E}(\Delta_l \wedge n)}{\delta}) \leq \frac{\mathbb{E}(\tilde{\Delta}_l)}{\mathbb{E}(\Delta_l \wedge n)} \cdot \delta \leq \delta, \quad \text{for } 1 \leq l \leq D(n) + 1,$$

by Markov Inequality. Note that $(\Delta_l \wedge n)$'s are IID. We define $M_n := \lceil \mathbb{E}(\Delta)/\delta \rceil$, where $\Delta \overset{d}{=} \Delta_l \wedge n$ for all $l > 0$.

For any $i \in \mathcal{H}(n)$, there exists only one $j = j(i) \in \{1, 2, \ldots, D(n) + 1\}$ such that $i \in (\tau_{j-1}, \tau_j \wedge n]$, and

$$\begin{aligned}
\mathbb{E}\left[\bar{F}(\alpha_i)\right] &= \mathbb{E}\left[\bar{F}(\alpha_i) \cdot \mathbb{I}\{\tilde{\Delta}_{j(i)} \geq M_n\}\right] + \mathbb{E}\left[\bar{F}(\alpha_i) \cdot \mathbb{I}\{\tilde{\Delta}_{j(i)} < M_n\}\right] \\
&\leq \delta + \mathbb{E}\left[\bar{F}(\alpha_i) \cdot \mathbb{I}\{\tilde{\Delta}_{j(i)} < M_n\}\right],
\end{aligned}$$

so that

$$\sum_{i=1}^{n} \mathbb{E}\left[\bar{F}(\alpha_i)\right] \leq n\delta + \mathbb{E}\left[\sum_{i=1}^{n} \bar{F}(\alpha_i) \cdot \mathbb{I}\{\tilde{\Delta}_{j(i)} < M_n\}\right]. \tag{20}$$

By Proposition 1, there is $C > 0$ not depending on n such that

$$\mathbb{E}(\Delta) \leq 2n^{\beta} + C, \quad \text{for all } l > 0. \tag{21}$$

And thus, there is some $L' > 0$ (constant in n) such that, for $1 \leq i \leq M_n$,

$$\lambda_i = Li^{-\nu} \geq L \cdot (M_n)^{-\nu} = L \cdot \lceil \mathbb{E}(\Delta)/\delta \rceil^{-\nu} \geq L \cdot \left\lceil (2n^{\beta} + C)/\delta \right\rceil^{-\nu} \geq L' n^{-\beta\nu}. \tag{22}$$

Remark 7 If instead $(\lambda_i)_{i=1}^{\infty}$ satisfies (12), then $i^{\nu}\lambda_i \to \infty$ as $i \to \infty$, so that exists a constant $L > 0$ such that $\lambda_i \geq Li^{-\nu}$ for all i, and this is all that we need to proceed.

Since $\bar{F}$ is a decreasing function, the second term in RHS of (20) can be bounded as

$$\begin{aligned}\mathbb{E}\left[\sum_{i=1}^{n}\bar{F}(\alpha_i)\cdot\mathbb{I}\{\tilde{\Delta}_{j(i)}<M_n\}\right]&=\mathbb{E}\left[\sum_{j=1}^{D(n)+1}\sum_{i=\tau_{j-1}+1}^{\tau_j\wedge n}\bar{F}(\alpha_i)\cdot\mathbb{I}\{\tilde{\Delta}_j<M_n\}\right]\\&=\mathbb{E}\left[\sum_{j=1}^{D(n)+1}\sum_{i=1}^{\tilde{\Delta}_j}\bar{F}(\lambda_i)\cdot\mathbb{I}\{\tilde{\Delta}_j<M_n\}\right]\\&\le\mathbb{E}\left[\sum_{j=1}^{D(n)+1}\sum_{i=1}^{\tilde{\Delta}_j}\bar{F}(L'n^{-\beta\nu})\cdot\mathbb{I}\{\tilde{\Delta}_j<M_n\}\right]\\&\le\mathbb{E}\left[\sum_{i=1}^{n}\bar{F}(L'n^{-\beta\nu})\right]\le n\cdot\bar{F}(L'n^{-\beta\nu}).\end{aligned}$$

Combining these bounds, we obtain

$$\mathrm{FNR}_n(\mathcal{R})=\frac{\sum_{i=1}^{n}\mathbb{E}[\bar{F}(\alpha_i)]}{n}\le\delta+\bar{F}(L'n^{-\beta\nu}).$$

Since $L'n^{-\beta\nu}\to 0$ as $n\to\infty$, by Eq. (18) we have

$$\xi_n:=\Phi^{-1}(1-L'n^{-\beta\nu})=(\gamma\beta\nu\log n)^{1/\gamma}(1+o(1)),$$

so that

$$\begin{aligned}\mu-\xi_n&=(\gamma r\log n)^{1/\gamma}-(\gamma\beta\nu\log n)^{1/\gamma}(1+o(1))\\&\sim(r^{1/\gamma}-(\beta\nu)^{1/\gamma})(\gamma\log n)^{1/\gamma}\to\infty,\quad\text{as } n\to\infty,\end{aligned}$$

since $r>\beta\nu$. Therefore, $\bar{F}(L'n^{-\beta\nu})=\bar{\Phi}(\mu-\xi_n)\to 0$ as $n\to\infty$. Hence,

$$\limsup_{n\to\infty}\mathrm{FNR}_n\le\delta.$$

This being true for any $\delta>0$, necessarily, $\mathrm{FNR}_n\to 0$ as $n\to\infty$. This establishes Theorem 1. □

7.3 Discovery times (LOND)

We apply LOND to the static setting under consideration. Denote τ_l as the time of lth discovery (with $\tau_0=0$), and $\Delta_l=\tau_l-\tau_{l-1}$ as the time between the $(l-1)$th and lth discoveries. Assume a sequence satisfying (5) has been chosen. Given the update rule of (8), it can be seen that the inter-discovery times $\{\Delta_l : l\ge 1\}$ are i.i.d.

To prove Theorem 2, we will use the following bound on the expected discovery times.

Proposition 2 *Consider a static AGG mixture model with exponent $\gamma \geq 1$ parameterized as in (10). Assume that $\beta \in (0,1)$ and $r \in [0,1]$ are both fixed. For any $\nu > 1$, if we apply LOND with $(\lambda_i)_{i=1}^{\infty}$ defined as $\lambda_i \propto i^{-\nu}$ with $\sum_{i=1}^{\infty} \lambda_i = q$,*

$$\mathbb{E}(\tau_l \wedge n) \leq l \cdot n^{\beta+(\nu^{1/\gamma}-r^{1/\gamma})^{\gamma}+b_n}, \quad \text{for all } l > 0, \tag{23}$$

where $b_n \to 0$ as $n \to \infty$.

We now prove this result. By the update rule of LOND algorithm, for all $l \geq 0$, and all $m \geq \tau_l + 1$, we have

$$\mathbb{P}(\tau_{l+1} > m \mid \tau_l) = \prod_{i=\tau_l+1}^{m} (1 - G((l+1)\lambda_i)) \leq \exp\{-\sum_{i=\tau_l+1}^{m} G((l+1)\lambda_i)\}.$$

Note τ_l is the time of lth discovery (with $\tau_0 = 0$) by LOND. Let $\tilde{\tau}_l = \tau_l \wedge n$. If $\tilde{\tau}_l = n$, we have $\mathbb{E}(\tilde{\tau}_{l+1} \mid \tilde{\tau}_l) = n = \tilde{\tau}_l$. Otherwise, if $\tilde{\tau}_l = \tau_l < n$,

$$\mathbb{E}(\tilde{\tau}_{l+1} \mid \tilde{\tau}_l) = \tau_l + 1 + \sum_{m=\tau_l+1}^{\infty} \mathbb{P}(\tau_{l+1} \wedge n > m \mid \tau_l) \tag{24}$$

$$= \tau_l + 1 + \sum_{m=\tau_l+1}^{n-1} \mathbb{P}(\tau_{l+1} > m \mid \tau_l) \tag{25}$$

$$\leq \tau_l + 1 + \sum_{m=\tau_l+1}^{n} \exp\{-\sum_{i=\tau_l+1}^{m} G((l+1)\lambda_i)\} \tag{26}$$

$$= \tilde{\tau}_l + 1 + \sum_{m=\tilde{\tau}_l+1}^{n} \exp\{-\sum_{i=\tilde{\tau}_l+1}^{m} G((l+1)\lambda_i)\}. \tag{27}$$

Next, we bound $\mathbb{E}(\tilde{\tau}_{l+1} \mid \tilde{\tau}_l)$. Let t^* be the value such that $\Phi^{-1}(1-t^*) = \mu$, i.e., $t^* = \Phi(-\mu) = n^{-r+o(1)}$ by the fact that Φ satisfies Definition 1. Then, for $t \geq t^*$, we get

$$\Phi^{-1}(1-t) \leq \Phi^{-1}(1-t^*) = \mu,$$

and,

$$F(t) = \Phi(\mu - \Phi^{-1}(1-t)) \geq \Phi(\mu - \Phi^{-1}(1-t^*)) = \Phi(\mu-\mu) = \Phi(0) = 1/2,$$

so that if $(l+1)\lambda_i = (l+1)Li^{-\nu} \geq t^*$, i.e., $i \leq n_1 := \lfloor ((l+1)L/t^*)^{1/\nu} \rfloor = n^{r/\nu+o(1)}$, we have $F((l+1)\lambda_i) \geq \Phi(0) = 1/2$.

We consider the following cases.

Case 1 $\tilde{\tau}_l < n_1 < n$. In this case, for $\tilde{\tau}_l + 1 \leq m \leq n_1$,

$$\sum_{i=\tilde{\tau}_l+1}^{m} G((l+1)\lambda_i) \geq \sum_{i=\tilde{\tau}_l+1}^{m} \varepsilon F((l+1)\lambda_i) \geq \varepsilon \cdot (m-\tilde{\tau}_l)/2,$$

and for $m > n_1$,

$$\begin{aligned}\sum_{i=\tilde{\tau}_l+1}^{m} G((l+1)\lambda_i) &\geq \sum_{i=\tilde{\tau}_l+1}^{m} \varepsilon F((l+1)\lambda_i) \\ &\geq \sum_{i=\tilde{\tau}_l+1}^{m} \varepsilon F((l+1)\lambda_m) = (m-\tilde{\tau}_l)\varepsilon F((l+1)\lambda_m),\end{aligned}$$

since $F(x)$ is non-decreasing.

We split the summation in (27) over $\tau_l + 1 \leq m \leq n_1$ and $n_1 + 1 \leq m \leq n$ and derive the corresponding upper bound separately. For the first part,

$$\begin{aligned}\sum_{m=\tilde{\tau}_l+1}^{n_1} \exp\{-\sum_{i=\tilde{\tau}_l+1}^{m} G((l+1)\lambda_i)\} &\leq \sum_{m=\tilde{\tau}_l+1}^{n_1} \exp\{-\varepsilon(m-\tilde{\tau}_l)/2\} = \sum_{m=1}^{n_1-\tilde{\tau}_l} \exp\{-\varepsilon m/2\} \\ &\leq \frac{1}{\exp\{\varepsilon/2\}-1} < \frac{2}{\varepsilon} = 2n^{\beta}.\end{aligned}$$

For the second part,

$$\begin{aligned}\sum_{m=n_1+1}^{n} \exp\left\{-\sum_{i=\tilde{\tau}_l+1}^{m} G((l+1)\lambda_i)\right\} &\leq \sum_{m=n_1+1}^{n} \exp\{-(m-\tilde{\tau}_l)\varepsilon F((l+1)\lambda_m)\} \\ &\leq \sum_{m=n_1+1}^{n} \exp\{-(m-n_1)\varepsilon F((l+1)\lambda_n)\} \\ &\leq \sum_{m=1}^{n-n_1} \exp\{-m\varepsilon F((l+1)\lambda_n)\} \\ &\leq \frac{1}{\exp\{\varepsilon F((l+1)\lambda_n)\}-1} \\ &< \frac{1}{\varepsilon F((l+1)\lambda_n)} \leq \frac{1}{\varepsilon F(\lambda_n)}.\end{aligned}$$

Case 2 $n_1 \leq \tilde{\tau}_l < n$.

For this case, we don't need to split the summation, since

$$\sum_{m=\tilde{\tau}_l+1}^{n} \exp\{-\sum_{i=\tilde{\tau}_l+1}^{m} G((l+1)\lambda_i)\} \leq \sum_{m=\tilde{\tau}_l+1}^{n} \exp\{-(m-\tilde{\tau}_l)\varepsilon F((l+1)\lambda_m)\}$$
$$\leq \sum_{m=1}^{n-\tilde{\tau}_l} \exp\{-m\varepsilon F((l+1)\lambda_n)\}$$
$$< \frac{1}{\varepsilon F((l+1)\lambda_n)} \leq \frac{1}{\varepsilon F(\lambda_n)}.$$

Case 3 $n_1 \geq n$. Since $\tilde{\tau}_l < n \leq n_1$, we have that

$$\sum_{m=\tilde{\tau}_l+1}^{n} \exp\{-\sum_{i=\tilde{\tau}_l+1}^{m} G((l+1)\lambda_i)\} \leq \sum_{m=\tilde{\tau}_l+1}^{n} \exp\{-(m-\tilde{\tau}_l)\varepsilon/2\}$$
$$\leq \sum_{m=1}^{n-\tilde{\tau}_l} \exp\{-m\varepsilon/2\} < \frac{2}{\varepsilon} = 2n^{\beta}.$$

Combining all the cases, we obtain

$$\mathbb{E}(\tilde{\tau}_{l+1} \mid \tilde{\tau}_l) \leq \tilde{\tau}_l + 1 + \frac{2}{\varepsilon} + \frac{1}{\varepsilon F(\lambda_n)},$$

where $F(\lambda_n) = \bar{\Phi}(\xi_n - \mu)$, and $\xi_n := \Phi^{-1}(1-\lambda_n)$. Since $\lambda_n = Ln^{-\nu} \to 0$ as $n \to 0$, by Eq. (18), we have $\xi_n \sim (\gamma\nu \log n)^{1/\gamma}$, so that

$$\xi_n - \mu = (\gamma\nu \log n)^{1/\gamma}(1+o(1)) - (\gamma r \log n)^{1/\gamma}$$
$$\sim (\nu^{1/\gamma} - r^{1/\gamma})(\gamma \log n)^{1/\gamma} \to \infty, \quad \text{as } n \to \infty,$$

by the fact that $\nu > 1 \geq r$. By Definition 1,

$$F(\lambda_n) = \bar{\Phi}(\xi_n - \mu) = \exp\left\{-\frac{(\xi_n-\mu)^\gamma}{\gamma}(1+o(1))\right\} = n^{-(\nu^{1/\gamma}-r^{1/\gamma})^\gamma + o(1)}.$$

Thus, when n is large enough,

$$\mathbb{E}(\tilde{\tau}_{l+1} \mid \tilde{\tau}_l) \leq \tilde{\tau}_l + 1 + \frac{2}{\varepsilon} + \frac{1}{\varepsilon F(\lambda_n)} \leq \tilde{\tau}_l + n^{\beta + (\nu^{1/\gamma}-r^{1/\gamma})^\gamma + o(1)}, \quad \text{for all } l > 0,$$

where the $o(1)$ is uniform in l, and this further implies that

$$\mathbb{E}(\tilde{\tau}_{l+1}) = \mathbb{E}[\mathbb{E}(\tilde{\tau}_{l+1} \mid \tilde{\tau}_l)] \leq \mathbb{E}(\tilde{\tau}_l) + n^{\beta + (\nu^{1/\gamma}-r^{1/\gamma})^\gamma + o(1)}, \quad \text{for all } l > 0,$$

so that

$$\mathbb{E}(\tau_l \wedge n) \leq l \cdot n^{\beta + (\nu^{1/\gamma}-r^{1/\gamma})^\gamma + o(1)}, \quad \text{for all } l > 0.$$

□

7.4 Proof of Theorem 2

It suffices to consider the case where $r \in [0,1]$ since the observations from $\mathbb{H}_0$ almost never get substantially larger than $(\gamma \log n)^{1/\gamma}$. For $r \in [0,1]$, if $r - (1 - r^{1/\gamma})^\gamma > \beta$, we can choose $\nu > 1$ close to 1 and $\eta > 0$ close to 0 such that $r > \rho := \beta + (\nu^{1/\gamma} - r^{1/\gamma})^\gamma + \nu - 1 + \eta$. By Proposition 2, when n is large enough,

$$\mathbb{E}(\tau_l \wedge n) \le l \cdot n^{\beta + (\nu^{1/\gamma} - r^{1/\gamma})^\gamma + \eta}, \quad \text{for all } l > 0.$$

Fix $\delta > 0$ and let $n_2 := \lceil n^{\beta + (\nu^{1/\gamma} - r^{1/\gamma})^\gamma + \eta} / \delta \rceil$. Note $n_2 = o(n)$, since $1 \ge r > \beta + (\nu^{1/\gamma} - r^{1/\gamma})^\gamma + \eta$.

For $n_2 \le i \le n$, let $\zeta_i := i\,\delta\, n^{-\beta - (\nu^{1/\gamma} - r^{1/\gamma})^\gamma - \eta}$, we get

$$\begin{aligned}
\mathbb{P}(D(i) < \zeta_i) &= \mathbb{P}(\tau_{\lceil \zeta_i \rceil} > i) \le \mathbb{P}(\tau_{\lceil \zeta_i \rceil} \ge i) = \mathbb{P}(\tau_{\lceil \zeta_i \rceil} \wedge n \ge i) \\
&\le \frac{\mathbb{E}(\tau_{\lceil \zeta_i \rceil} \wedge n)}{i} \le \frac{\lceil \zeta_i \rceil \cdot n^{\beta + (\nu^{1/\gamma} - r^{1/\gamma})^\gamma + \eta}}{i} \\
&< \frac{(\zeta_i + 1) \cdot n^{\beta + (\nu^{1/\gamma} - r^{1/\gamma})^\gamma + \eta}}{i} \\
&= \delta + \frac{n^{\beta + (\nu^{1/\gamma} - r^{1/\gamma})^\gamma + \eta}}{i} < 2\delta.
\end{aligned}$$

By Rule (8) defining the LOND algorithm,

$$\mathbb{E}[\bar{F}(\alpha_i)] = \mathbb{E}[\bar{F}(\lambda_i(D(i-1)+1))] \le \mathbb{E}[\bar{F}(\lambda_i D(i))],$$

due to the fact that $D(i-1) + 1 \ge D(i)$ and that $\bar{F}(x)$ is a non-increasing function, so that LOND's false non-discovery rate (denoted FNR_n) is bounded as follows

$$\text{FNR}_n = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[\bar{F}(\alpha_i)] \le \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[\bar{F}(\lambda_i D(i))].$$

For $1 \le i \le n_2$,

$$\frac{1}{n} \sum_{i=1}^{n_2} \mathbb{E}[\bar{F}(\lambda_i D(i))] \le \frac{n_2}{n}.$$

And for $n_2 + 1 \le i \le n$,

$$\begin{aligned}
\mathbb{E}[\bar{F}(\lambda_i D(i))] &= \mathbb{E}[\bar{F}(\lambda_i D(i)) \cdot \mathbb{I}\{D(i) < \zeta_i\}] + \mathbb{E}[\bar{F}(\lambda_i D(i)) \cdot \mathbb{I}\{D(i) \ge \zeta_i\}] \\
&\le 2\delta + \bar{F}(\lambda_i \zeta_i),
\end{aligned}$$

and since $\nu > 1$, we have

$$\lambda_i \zeta_i = L\delta \cdot i^{1-\nu} \cdot n^{-\beta - (\nu^{1/\gamma} - r^{1/\gamma})^\gamma - \eta} \ge L\delta \cdot n^{1-\nu} \cdot n^{-\beta - (\nu^{1/\gamma} - r^{1/\gamma})^\gamma - \eta} = \lambda_n \zeta_n,$$

which implies that,

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[\bar{F}(\lambda_i D(i))] \le \frac{n_2}{n} + \frac{n-n_2}{n}(2\delta + \bar{F}(\lambda_n\zeta_n)) \le 2\delta + \bar{F}(\lambda_n\zeta_n) + o(1).$$

Since $\lambda_n\zeta_n = L\delta n^{-\rho} \to 0$ as $n \to \infty$, by Eq. (18)

$$\xi_n := \Phi^{-1}(1-\lambda_n\zeta_n) = (\gamma\rho\log n)^{1/\gamma}(1+o(1)),$$

then

$$\mu - \xi_n = (\gamma r \log n)^{1/\gamma} - (\gamma\rho\log n)^{1/\gamma}(1+o(1)) \\ \sim (r^{1/\gamma} - \rho^{1/\gamma})(\gamma\log n)^{1/\gamma} \to \infty, \quad \text{as } n \to \infty,$$

since $r > \rho$. Therefore, $\bar{F}(\lambda_n\zeta_n) = \bar{\Phi}(\mu - \xi_n) \to 0$ as $n \to \infty$. Hence,

$$\limsup_{n\to\infty} \text{FNR}_n \le 2\delta.$$

This being true for any $\delta > 0$, necessarily, $\text{FNR}_n \to 0$ as $n \to \infty$. □

Acknowledgements We would like to thank Jiaqi Gu from Department of Computer Science, University of California, Los Angeles, for his help with the numerical experiments in Sect. 5. This work was partially supported by grants from the US National Science Foundation (DMS 1223137).

Appendix: Simulations with varying number of hypotheses

In this second set of experiments, we examine the performance of the same methods as the number of hypotheses, n, varies.

FNR of LORD with a fixed level

In this subsection, we present numerical experiments meant to illustrate the theoretical results we derived about asymptotic FNR of LORD. We fix $q = 0.1$ and choose a few values for the parameter β so as to exhibit different sparsity levels, while the parameter r takes values in a grid of spanning [0, 1.5]. We plot the average FNP of

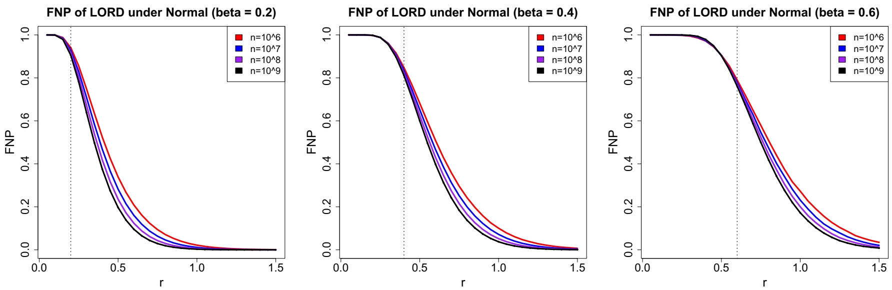

Fig. 9 Simulation results showing the FNP for LORD under the normal model in three distinct sparsity regimes with different test size. The black vertical line delineates the theoretical threshold ($r = \beta$)

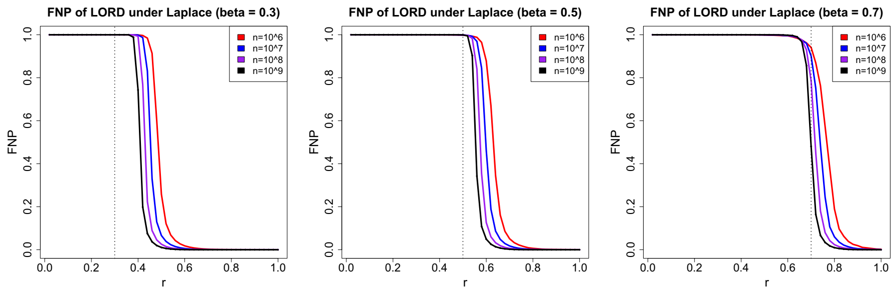

Fig. 10 Simulation results showing the FNP for LORD under the double-exponential model in three distinct sparsity regimes with different test size. The black vertical line delineates the theoretical threshold ($r = \beta$)

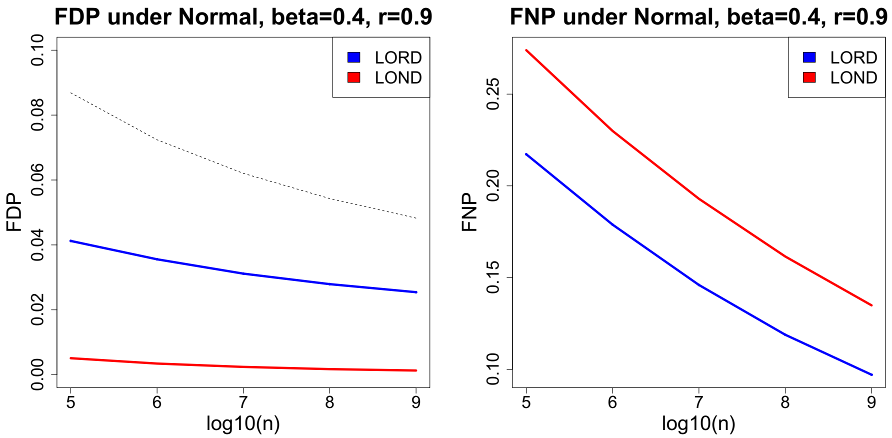

Fig. 11 FDP and FNP for the LORD and LOND methods under the normal model with $(\beta, r) = (0.4, 0.9)$ and varying test size n. The black line delineates the desired FDR control level ($q = q_n$)

LORD procedure with different $n \in \{10^6, 10^7, 10^8, 10^9\}$. The simulation results are reported in Figs. 9 and 10. Each situation is repeated 200 times. We observe that in the normal model when $r > \beta$, the FNP decreases as n is getting larger. In the double-exponential model, as n increases, the FNP transition lines are getting closer the theoretical thresholds $r = \beta$, especially when $\beta = 0.7$.

Varying level

Here, we explore the effect of letting the desired FDR control level q tend to 0 as n increases in accordance with (11). Specifically, we set it as $q = q_n = 1/\log n$. We choose n on a log scale, specifically, $n \in \{10^6, 10^7, 10^8, 10^9\}$. Each time, we fix a value of (β, r) such that $r > \beta$.

In the first setting, we set $(\beta, r) = (0.4, 0.9)$ for normal model and $(\beta, r) = (0.4, 0.7)$ for double-exponential model. The simulation results are reported in Figs. 11

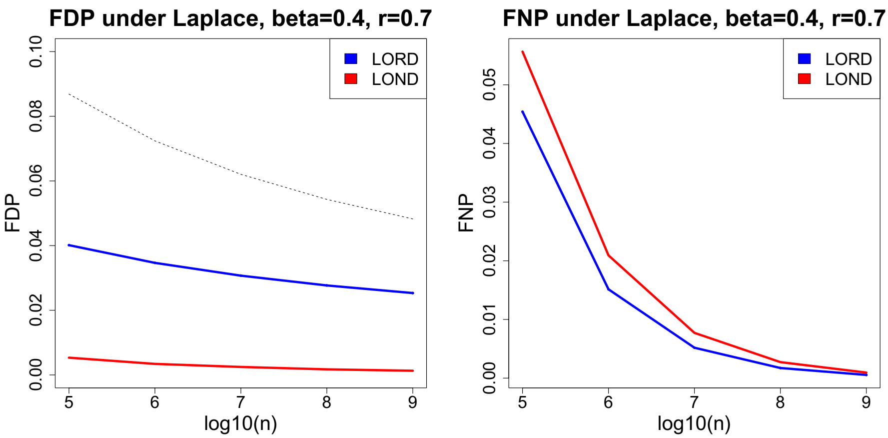

Fig. 12 FDP and FNP for the LORD and LOND methods under the double-exponential model with $(\beta, r) = (0.4, 0.7)$ and varying test size n. The black line delineates the desired FDR control level $(q = q_n)$

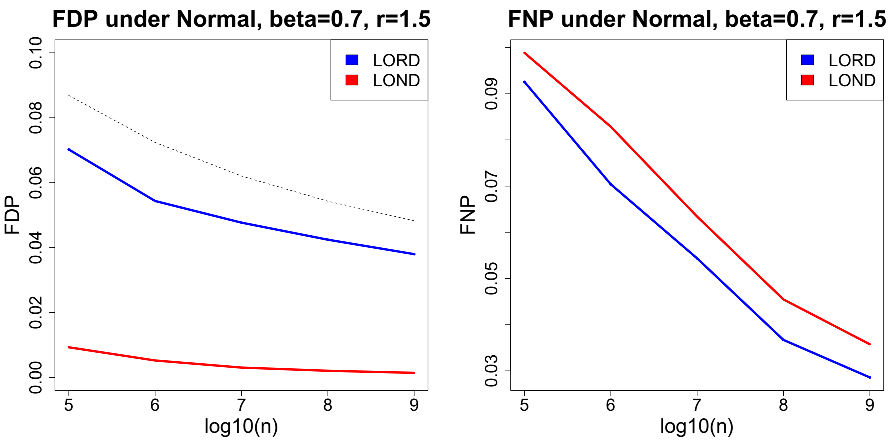

Fig. 13 FDP and FNP for the LORD and LOND methods under the normal model with $(\beta, r) = (0.7, 1.5)$ and varying test size n. The black line delineates the desired FDR control level $(q = q_n)$

and 12. We see that, in both models, the risks of the two procedures decrease to zero as the test size gets larger. LORD clearly dominates LOND (in terms of FNP). Both methods have FDP much lower than the level q_n, and in particular, LOND is very conservative.

In the second setting, we set $(\beta, r) = (0.7, 1.5)$ for normal model and $(\beta, r) = (0.7, 0.9)$ for double-exponential model. The simulation results are reported in Figs. 13 and 14. In this sparser regime, we can see that although LORD still dominates, the difference in FNP between two methods is much smaller than that in dense regime, especially in the double-exponential model. Both methods have FDP lower than the level q_n, and in particular, LOND is very conservative.

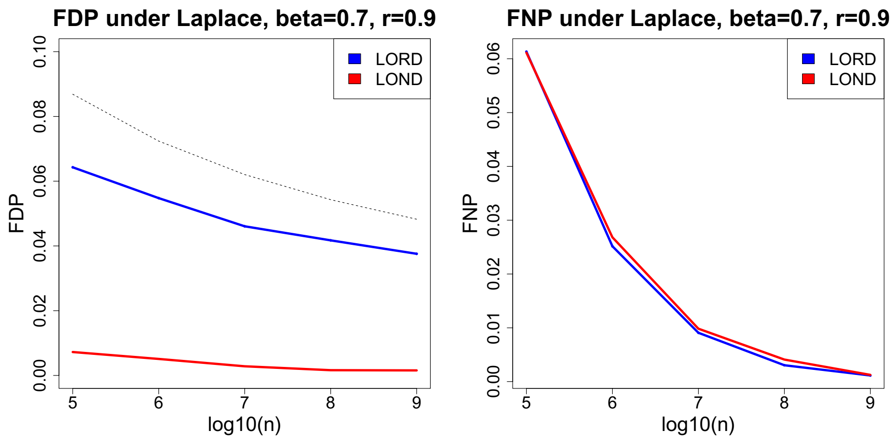

Fig. 14 FDP and FNP for the LORD and LOND methods under the double-exponential model with $(\beta, r) = (0.7, 0.9)$ and varying test size n

References

Aharoni, E., Rosset, S. (2014). Generalized α-investing: Definitions, optimality results and application to public databases. *Journal of the Royal Statistical Society*: *Series B (Statistical Methodology)*, *76*(4), 771–794.

Arias-Castro, E., Chen, S. (2017). Distribution-free multiple testing. *Electronic Journal of Statistics*, *11*(1), 1983–2001.

Bartroff, J. (2014). Multiple hypothesis tests controlling generalized error rates for sequential data. arXiv preprint arXiv:14065933

Bartroff, J., Song, J. (2013). Sequential tests of multiple hypotheses controlling false discovery and nondiscovery rates. arXiv preprint arXiv:13113350

Bartroff, J., Song, J. (2014). Sequential tests of multiple hypotheses controlling type I and II familywise error rates. *Journal of Statistical Planning and Inference*, *153*, 100–114.

Benjamini, Y., Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B (Methodological)*, *57*(1), 289–300.

Bogdan, M., Chakrabarti, A., Frommlet, F., Ghosh, J. K. (2011). Asymptotic Bayes-optimality under sparsity of some multiple testing procedures. *The Annals of Statistics*, *39*, 1551–1579.

Butucea, C., Ndaoud, M., Stepanova, N. A., Tsybakov, A. B., et al. (2018). Variable selection with Hamming loss. *The Annals of Statistics*, *46*(5), 1837–1875.

Dickhaus, T. (2014). *Simultaneous statistical inference*. Germany: Springer.

Donoho, D., Jin, J. (2004). Higher criticism for detecting sparse heterogeneous mixtures. *The Annals of Statistics*, *32*(3), 962–994.

Dudoit, S., van der Laan, M. J. (2007). *Multiple testing procedures with applications to genomics*. New York: Springer.

Fithian, W., Taylor, J., Tibshirani, R., Tibshirani, R. (2015). Selective sequential model selection. arXiv preprint arXiv:151202565

Foster, D. P., Stine, R. A. (2008). α-investing: A procedure for sequential control of expected false discoveries. *Journal of the Royal Statistical Society*: *Series B (Statistical Methodology)*, *70*(2), 429–444.

Foygel-Barber, R., Candès, E. J. (2015). Controlling the false discovery rate via knockoffs. *The Annals of Statistics*, *43*(5), 2055–2085.

Genovese, C., Wasserman, L. (2002). Operating characteristics and extensions of the false discovery rate procedure. *Journal of the Royal Statistical Society*: *Series B (Statistical Methodology)*, *64*(3), 499–517.

G'Sell, M. G., Wager, S., Chouldechova, A., Tibshirani, R. (2016). Sequential selection procedures and false discovery rate control. *Journal of the Royal Statistical Society*: *Series B (Statistical Methodology)*, *78*(2), 423–444.

Ingster, Y. I. (1997). Some problems of hypothesis testing leading to infinitely divisible distributions. *Mathematical Methods of Statistics*, *6*(1), 47–69.

Ingster, Y. I., Suslina, I. A. (2003). *Nonparametric goodness-of-fit testing under Gaussian models. Lecture notes in statistics*, Vol. 169. New York: Springer.

Javanmard, A., Montanari, A. (2015). On online control of false discovery rate. arXiv preprint arXiv:150206197

Javanmard, A., Montanari, A. (2018). Online rules for control of false discovery rate and false discovery exceedance. *The Annals of Statistics*, *46*(2), 526–554.

Ji, P., Jin, J., et al. (2012). UPS delivers optimal phase diagram in high-dimensional variable selection. *The Annals of Statistics*, *40*(1), 73–103.

Jin, J., Ke, Z. T. (2016). Rare and weak effects in large-scale inference: Methods and phase diagrams. *Statistica Sinica*, *26*, 1–34.

Lei, L., Fithian, W. (2016). Power of ordered hypothesis testing. In *International conference on machine learning* (pp. 2924–2932).

Li, A., Barber, R. F. (2017). Accumulation tests for fdr control in ordered hypothesis testing. *Journal of the American Statistical Association*, *112*(518), 837–849.

Meinshausen, N., Maathuis, M. H., Bühlmann, P., et al. (2011). Asymptotic optimality of the westfall-young permutation procedure for multiple testing under dependence. *The Annals of Statistics*, *39*(6), 3369–3391.

Neuvial, P., Roquain, E. (2012). On false discovery rate thresholding for classification under sparsity. *The Annals of Statistics*, *40*(5), 2572–2600.

Ramdas, A., Yang, F., Wainwright, M. J., Jordan, M. I. (2017). Online control of the false discovery rate with decaying memory. In *Advances in neural information processing systems* (pp. 5650–5659).

Ramdas, A., Zrnic, T., Wainwright, M., Jordan, M. (2018). SAFFRON: An adaptive algorithm for online control of the false discovery rate. In *International conference on machine learning* (pp. 4283–4291).

Roquain, E. (2011). Type I error rate control in multiple testing: A survey with proofs. *Journal de la Société Française de Statistique*, *152*(2), 3–38.

Storey, J. D. (2002). A direct approach to false discovery rates. *Journal of the Royal Statistical Society*: *Series B (Statistical Methodology)*, *64*(3), 479–498.

Storey, J. D. (2007). The optimal discovery procedure: A new approach to simultaneous significance testing. *Journal of the Royal Statistical Society*: *Series B (Statistical Methodology)*, *69*(3), 347–368.

Sun, W., Cai, T. T. (2007). Oracle and adaptive compound decision rules for false discovery rate control. *Journal of the American Statistical Association*, *102*(479), 901–912.