The Annals of Statistics
2020, Vol. 48, No. 5, 2548–2574
https://doi.org/10.1214/19-AOS1897

ON SPIKE AND SLAB EMPIRICAL BAYES MULTIPLE TESTING

BY ISMAËL CASTILLO* AND ÉTIENNE ROQUAIN**

Laboratoire de Probabilités, Statistique et Modélisation, LPSM, Sorbonne Université, *ismael.castillo@upmc.fr; **etienne.roquain@upmc.fr

This paper explores a connection between empirical Bayes posterior distributions and false discovery rate (FDR) control. In the Gaussian sequence model this work shows that empirical Bayes-calibrated spike and slab posterior distributions allow a correct FDR control under sparsity. Doing so, it offers a frequentist theoretical validation of empirical Bayes methods in the context of multiple testing. Our theoretical results are illustrated with numerical experiments.

1. Introduction.

1.1. *Context.* In modern high-dimensional statistical models several aims are typically pursued, often at the same time: *testing* of hypotheses on the parameters of interest, *estimation* and *uncertainty* quantification, among others. Due to their flexibility, in particular in the choice of the prior, Bayesian posterior distributions are routinely used to provide solutions to a variety of such inference problems. However, although practitioners may often directly read off quantities, such as the posterior mean or credible sets once they have simulated posterior draws, the question of mathematical justification of the use of such quantities, in particular from a frequentist perspective, has recently attracted a lot of attention. While the seminal papers [27, 41] set the stage for the study of posterior estimation rates in general models, the case of estimation in high-dimensional models has been considered only recently from the point of view of estimation (see [19, 30, 49] among others), while results on frequentist coverage of credible sets are just starting to emerge; see, for example, [6, 48]. Some of the previous approaches rely on automatic data-driven calibration of the prior parameters, following the so-called *empirical Bayes* approach, notably [30], estimating the proportion of significant parameters, and [29], where the full distribution function of the unknowns is estimated.

Our interest here is on the issue of *multiple testing* of hypotheses. Typically, the problem is to identify the active variables among a large number of candidates. This task appears in a wide variety of applied fields as genomics, neuroimaging and astrophysics among others. Such data typically involve more than thousands of variables with only a small part of them being significant (sparsity).

In this context, a typical aim is to control the false discovery rate (FDR) (see (9) below), that is, to find a selection rule that ensures that the averaged proportion of errors among the selected variables is smaller than some prescribed level α. This multiple testing type I error rate, introduced in [7], became quickly popular with the development of high-throughput technologies because it is "scalable" with respect to the dimension—the more rejections are possible, the more false positives are allowed. A common way to achieve this goal is to compute the p-values (probability under the null that the test statistic is larger than the observed value) and to run the Benjamini–Hochberg (BH) procedure [7] which is often considered as

Received August 2018; revised June 2019.
MSC2020 subject classifications. 62C12, 62G10.
Key words and phrases. Frequentist properties of Bayesian procedures, false discovery rate, sparsity, multiple testing.

a benchmark procedure. In the last decades an extensive literature aimed at studying the BH method, by showing that it (or versions of it) controls the FDR in various frameworks; see [8, 9, 24, 38], among others.

In a fundamental work [2], Abramovich, Benjamini, Donoho and Johnstone proved that a certain hard thresholding rule deduced from the BH procedure—keeping only observations with significant p-values—satisfies remarkable risk properties; it is minimax adaptive simultaneously for a range of losses and sparsity classes over a broad range of sparsity parameters. In addition, similar results hold true for the misclassification risks; see [10, 35]. These results, in particular, suggest a link between FDR controlling procedures and adaptation to sparsity. Here, we shall follow a questioning that can be seen as "dual" to the former one. Starting from a commonly used Bayesian procedure that is known to optimally adapt to the sparsity in terms of risk over a broad range of sparsity classes (and even, under appropriate self-similarity-type conditions, to produce adaptive confidence sets), we ask whether a uniform FDR control can be guaranteed.

1.2. *Setting*. In this paper we consider the Gaussian sequence model. One observes, for $1 \leq i \leq n$,

$$X_i = \theta_{0,i} + \varepsilon_i \tag{1}$$

for an unknown n-dimensional vector $\theta_0 = (\theta_{0,i})_{1 \leq i \leq n} \in \mathbb{R}^n$ and ε_i i.i.d. $\mathcal{N}(0, 1)$. This model can be seen as a stylized version of a high-dimensional model. The problem is to test

$$H_{0,i} : \text{“}\theta_{0,i} = 0\text{”} \quad \text{against} \quad H_{1,i} : \text{“}\theta_{0,i} \neq 0\text{”},$$

simultaneously over $i \in \{1, \dots, n\}$. We also introduce the assumption that the vector θ_0 is s_n-sparse, that is, is supposed to belong to the set

$$\ell_0[s_n] = \{\theta \in \mathbb{R}^n : \#\{1 \leq i \leq n : \theta_i \neq 0\} \leq s_n\} \tag{2}$$

for some sequence $s_n \in \{0, 1, \dots, n\}$, typically much smaller than n, measuring the sparsity of the vector.

1.3. *Bayesian multiple testing methodology.* From the point of view of posterior distributions, one natural approach for testing is simply based on comparing posterior probabilities of the hypotheses under consideration. Yet, to do so, a choice of prior needs to be made, and for this reason it is important to carefully design a prior that is flexible enough to adapt to the unknown underlying structure (and, here, sparsity) of the model. This is one of the reasons behind the use of *empirical Bayes* approaches that aim at calibrating the prior in a fully automatic, data-driven, way. Empirical Bayes methods for multiple testing have been in particular advocated by Efron (see, e.g., [22] and references therein) in a series of works over the last 10–15 years, reporting excellent behaviour of such procedures—we describe two of them in more detail in the next paragraphs—in practice. Fully Bayes methods, that bring added flexibility by putting prior on sensible hyperparameters are another alternative. In the sequel, *Bayesian multiple testing procedures* will be referred to as BMT for brevity.

Several popular BMT procedures rely on two quantities that can be seen as possible Bayesian counterparts of standard p-values:

- the ℓ-value: the probability that the null is true conditionally on the fact that the test statistic is *equal* to the observed value; see, for example, [23];
- the q-value: the probability that the null is true conditionally on the fact that the test statistic is *larger* than the observed value, introduced in [43].

(Note that the ℓ-value is usually called "local FDR." Here, we used another terminology to avoid any confusion between the procedure and the FDR.) Obviously, these quantities are well defined only if the trueness/falseness of a null hypothesis is random which is obtained by introducing an appropriate prior distribution.

Once the prior is calibrated (in a data-driven way or not), the q-values (resp. ℓ-values) can be computed and combined to produce BMT procedures. For instance, existing strategies reject null hypotheses with:

- a ℓ-value smaller than a fixed cutoff $t = 0.2$ [21];
- a q-value smaller than the nominal level α [22];
- averaged ℓ-values smaller than the nominal level α [34, 45, 46].

For alternatives see, for example, [1, 39]. In particular, one popular fact is that the use of Bayesian quantities "automatically corrects for the multiplicity of the tests," see, for example, [43]. While using p-values requires to use a cutoff t that decreases with the dimension n, using ℓ-values/q-values can be used with a cutoff t close to the nominal level α and without any further correction. This is well known to be valid from a decision theoretic perspective for the Bayes FDR, that is, for the FDR integrated w.r.t. the prior distribution, as we recall in Proposition 1 below. When the hyperparameters are estimated from the data within the BMT, the Bayes FDR is still controlled, to some extent, as proved in [45, 46]. However, controlling the Bayes FDR does not give theoretical guarantees for the usual frequentist FDR, that is, for the FDR at the true value of the parameter, as the pointwise FDR may deviate from an integrated version thereof.

1.4. *Frequentist control of BMT.* In this paper our main aim is to study whether BMT procedures have valid frequentist multiple testing properties.

A first hint has already been given in [22, 43]. It turns out that the BH procedure can loosely be seen as a "plug-in version" of the procedure rejecting the q-values smaller than α (namely, the theoretical c.d.f. of the p-values is estimated by its empirical counterpart). Since the BH procedure controls the (frequentist) FDR, this might suggest a possible connection between BMT and successful frequentist multiple testing procedures.

In regard to the rapidly increasing literature on frequentist validity of Bayesian procedures from the *estimation* perspective, the multiple testing question for BMT procedures has been less studied so far from the theoretical, frequentist point of view. This is despite a number of very encouraging simulation performance results; see, for example, [15, 28, 32, 34]. A recent exception is the interesting preprint [37] that shows a frequentist FDR control for a BMT based on a continuous shrinkage prior; yet, this control holds under a certain signal-strength assumption only. One main question we ask in the present work is whether a fully *uniform* control (over sparse vectors) of the frequentist FDR is possible for some posterior-based BMT procedures. Also, while the constants in the risk bounds are not made explicit in [37], we would like to clarify whether the final FDR control is made at, or close to, the required level α. The FDR control results below will also be complemented by appropriate type II-error controls.

1.5. *Spike and slab prior distributions and sparse priors.* Let $w \in (0, 1)$ be a fixed hyperparameter. Let us define the prior distribution $\Pi = \Pi_{w,\gamma}$ on $\mathbb{R}^n$ as

$$\Pi_{w,\gamma} = \big((1 - w)\delta_0 + w\mathcal{G}\big)^{\otimes n}, \tag{3}$$

where $\mathcal{G}$ is a distribution with a symmetric density γ on $\mathbb{R}$. Such a prior is a tensor product of a mixture of a Dirac mass at 0 (spike), that reflects the sparsity assumption, and of an absolutely continuous distribution (slab), that models nonzero coefficients. This is arguably one of the most natural priors on sparse vectors and has been considered in many key contributions on Bayesian sparse estimation and model selection; see, for example, [25, 33].

Of course, an important question is that of the choice of w and γ. A popular choice of w is data driven and based on a marginal maximum likelihood empirical Bayes method (to be described in more details below). The idea is to make the procedure learn the intrinsic sparsity while also incorporating some automatic multiplicity correction, as discussed, for example, in [11, 40]. Following such an approach in a fundamental paper, Johnstone and Silverman [30] show that, provided γ has tails at least as heavy as Laplace, the posterior median of the empirical Bayes posterior is rate adaptive for a wide range of sparsity parameters and classes, is fast to compute and enjoys excellent behaviour in simulations (the corresponding R-package `EBayesThresh` [31] is widely used). Namely, if $\|\cdot\|$ denotes the Euclidian norm and $\hat{\theta} = \hat{\theta}(X)$ is the coordinate-wise median of the empirical Bayes posterior distribution, there exists $c_1 > 0$ such that

$$\sup_{\theta_0 \in \ell_0[s_n]} E_{\theta_0} \|\hat{\theta} - \theta_0\|^2 \leq c_1 s_n \log(n/s_n). \tag{4}$$

Thus, asymptotically (in the regime $s_n, n \to \infty$, $s_n/n \to 0$), it matches up to a constant the minimax risk for this problem ([20]). In the recent work [16], the convergence of the empirical Bayes full posterior distribution (not only aspects such as median or mean) is considered, and similar results can be obtained, under stronger conditions on the tails of γ (for instance, γ Cauchy works). More precisely, for $\Pi(\cdot \mid X) = \hat{\Pi}(\cdot \mid X)$ the empirical Bayes posterior, one can find a constant $C_1 > 0$ such that

$$\sup_{\theta_0 \in \ell_0[s_n]} E_{\theta_0} \int \|\theta - \theta_0\|^2 \, d\Pi(\theta \mid X) \leq C_1 s_n \log(n/s_n). \tag{5}$$

Further, under some conditions, one can show that certain credible sets from the posterior distributions are also adaptive confidence sets in the frequentist sense [18]. Alternatively, one can also follow a hierarchical approach and put a prior on w. The paper [19] obtains adaptive rates for such a fully Bayes procedure over a variety of sparsity classes and presents a polynomial time algorithm to compute certain aspects of the posterior.

Empirical Bayes approaches have also been successfully applied to a variety of different sparse priors such as empirically recentered Gaussian slabs, as in [5, 6] or the horseshoe [47, 48], both studied in terms of estimation and the possibility to construct adaptive confidence sets. In [29], an empirical Bayes approach based on the "empirical" c.d.f. of the θs is shown to allow for optimal adaptive estimation over various sparsity classes. For an overview on the rapidly growing literature on sparse priors, we refer to the discussion paper [48].

Yet, most of the previous results are concerned with estimation or confidence sets, although a few of them report empirical false discoveries, for example, [48], Figure 7, though without theoretical analysis.

1.6. *Aim and results of the paper.* Here, we wish to find, if this is at all possible, a posterior-based procedure using a prior Π (possibly an empirical Bayes one, that is, $\Pi = \hat{\Pi}$), that can perform *simultaneous inference* in that: (a) it behaves optimally up to constants in terms of the quadratic risk in the sense of (4) (or (5)); (b) its frequentist FDR at *any* sparse vector is bounded from above by (a constant times) a given nominal level. More precisely, given a nominal level $t \in (0, 1)$ and φ_t a multiple testing procedure deduced from Π (ℓ-values or q-values procedure, as listed in Section 1.3), we want to *validate* its use in terms of a uniform control of its false discovery rate $\mathrm{FDR}(\theta_0, \varphi_t)$ (see (9) below) over the whole parameter space. That is, we ask whether we can find $C_2 > 0$ independent of t such that, for n large enough,

$$\sup_{\theta_0 \in \ell_0[s_n]} \mathrm{FDR}(\theta_0, \varphi_t) \leq C_2 t. \tag{6}$$

Our main results are as follows for a sparsity $s_n = O(n^{\upsilon})$ with $\upsilon \in (0,1)$:

- Theorem 1 shows that (6) holds with C_2 arbitrary small for the BMT procedure rejecting the nulls whenever the corresponding ℓ-value is smaller than t.
- Theorem 2 shows that (6) holds for some $C_2 > 0$ for the BMT procedure rejecting the nulls whenever the corresponding q-value is smaller than t (with a slight modification if only few signals are detected).

These results hold for spike and slab priors, for γ being Laplace or Cauchy, or even for slightly more general heavy-tailed distributions. The hyperparameter $\hat{w}$ is chosen according to a certain empirical Bayes approach to be specified below (with minor modifications with respect to the choice of [30]). In addition, it is important to evaluate the amplitude of $C_2 > 0$ in (6). Our numerical experiments support the fact that, roughly, $C_2 = 1$. Furthermore, Theorem 3 shows that for some subset $\mathcal{L}_0[s_n] \subset \ell_0[s_n]$ (containing strong signals), we have for the q-value BMT, for any (sequence) $\theta_0 \in \mathcal{L}_0[s_n]$,

$$\lim_n \mathrm{FDR}(\theta_0, \varphi_t) = t, \tag{7}$$

so the FDR control is exactly achieved asymptotically in that case.

Finally, we provide a control of the type II error of the considered procedures by showing in Theorem 4 that if $\mathrm{FNR}(\theta_0, \varphi)$ denotes the average number of nondiscoveries of a procedure φ, for $\theta_0 \in \mathcal{L}_0[s_n]$ as above,

$$\lim_n \mathrm{FNR}(\theta_0, \varphi_t) = 0, \tag{8}$$

where φ_t can either be the ℓ-values or q-values procedure at level t.

It follows from these results (combined with previous results of [16, 30]) that the posterior distribution associated to a spike and slab prior, with γ Cauchy and a suitably empirical Bayes-calibrated w, is appropriate to perform several tasks: (6)–(7)–(8) (multiple testing), (5)–(4) (posterior concentration in L^2-distance). The posterior can also be used to build honest adaptive confidence sets ([18]). The present work, focusing on the multiple testing aspect, then completes the inference picture for spike and slab empirical Bayes posteriors, confirming their excellent behaviour in simulations.

1.7. *Organisation of the paper.* In Section 2 we introduce Bayesian multiple testing procedures associated to spike and slab posterior distributions as well as the considered empirical Bayes choice of w. In Section 3 our main results are stated, while Section 4 contains numerical experiments. Section 5 presents some related BMT procedures, and Section 6 gives a short discussion. Preliminaries for the proofs are given in Section 7, while the proof of Theorems 1 and 2 can be found in Section 8. The Supplementary Material [17] gathers a number of lemmas used in the proofs, as well as the proofs of Propositions 1–3 and Theorems 3, 4 and 5. The sections and equations of this supplement are referred to with an additional symbol "S-" in the numbering.

1.8. *Notation.* In this paper we use the following notation:

- for F, a c.d.f., we set $\overline{F} = 1 - F$
- $\phi(x) = (2\pi)^{-1/2} e^{-x^2/2}$ and $\Phi(x) = \int_{-\infty}^{x} \phi(u)\, du$;
- $u_n \asymp v_n$ means that there exist constants $c, C > 0$ such that $|v_n| c \le |u_n| \le C|v_n|$ for n large enough;
- $u_n \lesssim v_n$ means that there exist constants $C > 0$ such that $|u_n| \le C|v_n|$ for n large enough;
- $f(y) \asymp g(y)$, for $y \in A$ means that there exist constants $c, C > 0$ such that for all $y \in A$, $c|g(y)| \le |f(y)| \le C|g(y)|$;

- $f(y) \asymp g(y)$, as $y \to \infty$ means that there exist constants $c, C > 0$ such that $c|g(y)| \le |f(y)| \le C|g(y)|$ for y large enough;
- $u_n \sim v_n$ means $u_n - v_n = o(u_n)$.

Also, for $\tau \in \mathbb{R}^n$, the symbol E_τ (resp. P_τ) denotes the expectation (resp. probability) under $\theta_0 = \tau$ in the model (1). The support of $\theta_0 \in \mathbb{R}^n$ is denoted by $S_{\theta_0} = \{i : \theta_{0,i} \neq 0\}$ or sometimes S_0 for simplicity. The cardinality of the support S_{θ_0} is denoted by $\sigma_0 = |S_0|$.

1.9. *Relevance and novelty of the approach.* We now briefly emphasize connections with existing works, and discuss several merits of the proposed approach. First, studying theoretical properties of BMT procedures is motivated by the fact that they are routinely used in practice since Efron's seminal papers [22, 23]; in the context of genomic applications, we refer for instance to a recent series of works by Stephens and coauthors [26, 42] and references therein. Second, we note that just a few other procedures to date theoretically allow both estimation at minimax rate and uniform FDR control. Besides the BH procedure [2, 7], the SLOPE procedure [12, 44] also enjoys these two properties in a regression context. In addition, the Bayesian maximum a posteriori (MAP) rule [3] has a minimax estimation rate and shares connections with the BH rule [1] for some specific choice of the prior. Third, let us mention that Sun, Cai and coauthors have also investigated a generic ℓ-value-based approach (see Section 5.2 for more details) that allows to control the FDR in structured settings where the BH procedure can be suboptimal [13, 14, 46]. Nevertheless, the proposed FDR control is not uniform from the frequentist perspective and is restricted to a specific asymptotical setting. Interestingly, using the present spike and slab prior in these contexts seems promising to get uniform FDR control while improving upon the BH procedure. During the submission process of the manuscript, a first encouraging attempt has being made by the second author in the discussion part of the paper [14] (see page 218 therein).

To summarize, the present work aims at providing guarantees for a widely used class of ℓ-value/q-value-based BMT procedures, deploying a spike and slab prior with suitably heavy tails and empirical Bayes choice of the weight. Further, by doing so and combining with results from recent parallel investigations [16, 18], our work demonstrates that the corresponding posterior distribution produces simultaneously optimal estimation rates, confidence sets and uniform FDR control (as well as FNR control over appropriately large signals) thereby achieving a complete inference picture along the three canonical inferential goals of "estimation, testing (here, multiple) and confidence sets." We are not aware of any another method that produces, simultaneously, these (frequentist) inferences in the present setting.

2. Preliminaries.

2.1. *Procedure and FDR.* A multiple testing procedure is a measurable function of the form $\varphi(X) = (\varphi_i(X))_{1 \le i \le n} \in \{0, 1\}^n$, where each $\varphi_i(X) = 0$ (resp. $\varphi_i(X) = 1$) codes for accepting $H_{0,i}$ (resp. rejecting $H_{0,i}$). For any such procedure φ, we let

$$\mathrm{FDR}(\theta_0, \varphi) = E_{\theta_0}\left[\frac{\sum_{i=1}^n \mathbf{1}\{\theta_{0,i} = 0\}\varphi_i(X)}{1 \vee \sum_{i=1}^n \varphi_i(X)}\right]. \tag{9}$$

A procedure φ is said to control the FDR at level α if $\mathrm{FDR}(\theta_0, \varphi) \le \alpha$ for any θ_0 in $\mathbb{R}^n$. Note that under $\theta_0 = 0$, we have $\mathrm{FDR}(\theta_0, \varphi) = P_{\theta_0=0}(\exists i : \varphi_i(X) = 1)$ which means that an α-FDR controlling procedure provides in particular a (single) test of level α of the full null "$\theta_{0,i} = 0$ for all i." As already mentioned, in the framework of this paper our goal is a control of the FDR around the prespecified target level, as in (6) or (7) (where $t = \alpha$).

2.2. *Prior, posterior, ℓ-values and q-values.* Recall the definition of the prior distribution $\Pi = \Pi_{w,\gamma}$ from (3), and let

$$g(x) = \int \gamma(x-u)\phi(u)\,du. \tag{10}$$

The posterior distribution $\Pi[\cdot \mid X] = \Pi_{w,\gamma}[\cdot \mid X]$ of θ is explicitly given by

$$\theta|X \sim \bigotimes_{i=1}^{n} \ell_i(X)\delta_0 + \big(1-\ell_i(X)\big)\mathcal{G}_{X_i}, \tag{11}$$

where $\mathcal{G}_x$ is the distribution with density $\gamma_x(u) := \phi(x-u)\gamma(u)/g(x)$ and

$$\ell_i(X) = \ell(X_i; w, g); \tag{12}$$

$$\ell(x; w, g) = \Pi(\theta_1 = 0 \mid X_1 = x) = \frac{(1-w)\phi(x)}{(1-w)\phi(x) + wg(x)}. \tag{13}$$

The quantities $\ell_i(X)$, $1 \le i \le n$, given by (12), are called the *ℓ-values*. Note that, although we do not emphasize it in the notation for short, the ℓ-values also depend on w and g. The ℓ-value measures locally, for a given observation X_i, the probability that the latter comes from pure noise. This is why it is sometimes called "local-FDR"; see [23].

If one has in mind a range of values, that is, those that exceed a given amplitude—, a different measure is given by the *q-values* defined by:

$$q_i(X) = q(X_i; w, g); \tag{14}$$

$$q(x; w, g) = \Pi\big(\theta_1 = 0 \mid |X_1| \ge |x|\big) = \frac{(1-w)\overline{\Phi}(|x|)}{(1-w)\overline{\Phi}(|x|) + w\overline{G}(|x|)}; \tag{15}$$

$$\overline{G}(s) = \int_s^{+\infty} g(x)\,dx. \tag{16}$$

The identity (15) relating the q-value to $\overline{\Phi}$, $\overline{G}$ is proved in Section S-3.

2.3. *Assumptions.* We follow throughout the paper assumptions similar to those of [30]. The prior γ is assumed to be unimodal, symmetric and so that

$$\big|\log\gamma(x) - \log\gamma(y)\big| \le \Lambda|x-y|, \quad x, y \in \mathbb{R}; \tag{17}$$

$$\gamma(y)^{-1}\int_y^{\infty} \gamma(u)\,du \asymp y^{\kappa-1}, \quad \text{as } y \to \infty, \kappa \in [1,2]; \tag{18}$$

$$y \in \mathbb{R} \to y^2\gamma(y) \text{ is bounded.} \tag{19}$$

Conditions (17), (18) and (19) above are, for instance, true when γ is Cauchy ($\kappa = 2$, $\Lambda = 1$) or Laplace ($\kappa = 1$, Λ is the scaling parameter). As we show in Remark 1 in the Supplementary Material, explicit expressions exist for g; see (10), in the Laplace case. In the Cauchy case the integral is not explicit, but, in practice, (to avoid approximating the integral) one can work with the quasi-Cauchy prior; see [31], that satisfies the above conditions and corresponds to:

$$\gamma(x) = (2\pi)^{-1/2}\big(1 - |x|\overline{\Phi}(x)/\phi(x)\big); \tag{20}$$

$$g(x) = (2\pi)^{-1/2}x^{-2}\big(1 - e^{-x^2/2}\big). \tag{21}$$

The condition (19) is mostly for simplicity to get unified proofs, but heavier tails could be considered as well by adapting estimates of [18].

2.4. *Bayesian multiple testing procedures* (*BMT*). We define the multiple procedures defined from the ℓ-values/q-values in the following way:

$$\varphi_i^{\ell\text{-val}}(t; w, g) = \mathbb{1}_{\{\ell_i(X) \le t\}}, \quad 1 \le i \le n; \tag{22}$$

$$\varphi_i^{q\text{-val}}(t; w, g) = \mathbb{1}_{\{q_i(X) \le t\}}, \quad 1 \le i \le n, \tag{23}$$

where $t \in (0, 1)$ is some threshold that possibly depends on X. As we will see in Section 7.2, these two procedures, denoted $\varphi^{\ell\text{-val}}(t)$, $\varphi^{q\text{-val}}(t)$ for brevity, simply correspond to (hard) thresholding procedures that select the $|X_i|$'s larger than some (*random*) threshold. The value of the threshold is driven by the posterior distribution in a very specific way. It depends on γ, t and on the whole data vector X through the empirical Bayes choice of the hyperparameter w that automatically "scales" the procedure according to the sparsity of the data.

2.5. *Controlling the Bayes FDR*. If the aim is to control the FDR at some level α, a first result indicates that choosing $t = \alpha$ in $\varphi^{\ell\text{-val}}(t)$ and $\varphi^{q\text{-val}}(t)$ may be appropriate because the corresponding procedures control the Bayes FDR, that is, the FDR where the parameter θ has been integrated with respect to the prior distribution (see, e.g., [39]). More formally, for any multiple testing procedure φ and hyperparameters w and γ, define

$$\mathrm{BFDR}(\varphi; w, \gamma) = \int_{\mathbb{R}^n} \mathrm{FDR}(\theta, \varphi)\, d\Pi_{w,\gamma}(\theta). \tag{24}$$

Then, the following result holds:

PROPOSITION 1. *Let $\alpha \in (0, 1)$ and $w \in (0, 1)$, and consider any density γ satisfying the assumptions of Section 2.3. Let $\varphi^\ell = \varphi^{\ell\text{-val}}(\alpha; w, g)$ as defined in (22) and $\varphi^q = \varphi^{q\text{-val}}(\alpha; w, g)$ as defined in (23). Then, we have*

$$\mathrm{BFDR}(\varphi^\ell; w, \gamma) \le \alpha P(\exists i : \ell_i(X) \le \alpha) \tag{25}$$

$$\le \alpha P(\exists i : q_i(X) \le \alpha) = \mathrm{BFDR}(\varphi^q; w, \gamma) \le \alpha. \tag{26}$$

This result can be certainly considered well known, as (25) (resp. (26)) is similar in essence to Theorem 4 of [46] (resp., Theorem 1 of [43]). It is essentially a consequence of Fubini's theorem; see Section S-2.1 for a proof. While Proposition 1 justifies the use of ℓ/q-values from the purely Bayesian perspective, it does not bring any information about $\mathrm{FDR}(\theta_0, \varphi^\ell)$ and $\mathrm{FDR}(\theta_0, \varphi^q)$ at an arbitrary sparse vector $\theta_0 \in \mathbb{R}^n$.

2.6. *Marginal maximum likelihood*. In order to choose the hyperparameter w, we explore now the choice made in [30], following the popular marginal maximum likelihood method. Let us introduce the auxiliary functions

$$\beta(x) = \frac{g}{\phi}(x) - 1; \qquad \beta(x, w) = \frac{\beta(x)}{1 + w\beta(x)}. \tag{27}$$

A useful property is that β is increasing on $[0, \infty)$ from $\beta(0) \in (-1, 0)$ to infinity; see Section 7.1. The marginal likelihood for w is, by definition, the marginal density of X, given w, in the Bayesian setting. Its logarithm is equal to

$$L(w) = \sum_{i=1}^n \log \phi(X_i) + \sum_{i=1}^n \log\bigl(1 + w\beta(X_i)\bigr),$$

which is a differentiable function on $[0, 1]$. The derivative $\mathcal{S}$ of L, the score function, can be written as

$$\mathcal{S}(w) = \sum_{i=1}^n \frac{\beta(X_i)}{1 + w\beta(X_i)} = \sum_{i=1}^n \beta(X_i, w). \tag{28}$$

The function $w \in [0,1] \to \mathcal{S}(w)$ is (a.s.) decreasing, and thus $w \in [0,1] \to L(w)$ is (a.s.) strictly concave. Hence, almost surely, the maximum of the function L on a compact interval exists, is unique and we can define the marginal maximum likelihood estimator $\hat{w}$ by

$$\hat{w} = \underset{w \in [\frac{1}{n}, 1]}{\operatorname{argmax}} L(w) \quad \text{(a.s.)}. \tag{29}$$

This choice of $\hat{w}$ is close to the one in [30]. The only difference is in the lower bound, here $1/n$, of the maximisation interval which differs from the choice in [30] by a slowly varying term. This difference is important for multiple testing in case of weak or zero signal (in contrast to the estimation task for which this different choice does not modify the results). Another slightly different choice of interval, still close to $[1/n, 1]$, will also be of interest below. In addition, if $\hat{w} \in (1/n, 1)$, it solves the equation $\mathcal{S}(w) = 0$ in w. However, note that, in general, the maximiser $\hat{w}$ can be at the boundary and thus may not be a zero of $\mathcal{S}$.

3. Main results.

Let us first describe the ℓ-value algorithm.

Algorithm `EBayesL`

`Input:` $X_1, \dots, X_n$, slab prior γ, target confidence t

`Output:` BMT procedure $\varphi^{\ell\text{-val}}$

1. Find the maximiser $\hat{w}$ given by (29).
2. Compute $\hat{\ell}_i(X) = \ell(X_i; \hat{w}, g)$ given by (13).
3. Return, for $1 \le i \le n$,

$$\varphi_i^{\ell\text{-val}} = \mathbf{1}\{\hat{\ell}_i(X) \le t\}. \tag{30}$$

THEOREM 1. *Consider the parameter space $\ell_0[s_n]$ given by (2) with sparsity $s_n \le n^{\upsilon}$ for some $\upsilon \in (0,1)$. Let γ be a unimodal symmetric slab density that satisfies (17)–(19) with κ as in (18). Then, the algorithm* `EBayesL` *produces as output the BMT $\varphi^{\ell\text{-val}}$ defined in (30) that satisfies the following: There exists a constant $C = C(\gamma, \upsilon)$ such that for any $t \le 3/4$, there exists an integer $N_0 = N_0(\gamma, \upsilon, t)$ such that, for any $n \ge N_0$,*

$$\sup_{\theta_0 \in \ell_0[s_n]} \mathrm{FDR}(\theta_0, \varphi^{\ell\text{-val}}) \le C \frac{\log\log n}{(\log n)^{\kappa/2}}. \tag{31}$$

Theorem 1 is proved in Section 8. The proof relies mainly on two different arguments: first, a careful analysis of the concentration of $\hat{w}$, which requires to distinguish between two regimes (weak/moderate or strong signal, basically); second, the study of the FDR of the ℓ-value procedure taken at some sparsity parameter w (not random but depending on n) in each of these two regimes. This requires to analyse the mathematical behavior of a number of functions of w, θ_0, uniformly over a wide range of possible sparsities, which is one main technical difficulty of our results. In particular, the concentration of $\hat{w}$ is obtained uniformly over all sparse vectors with polynomial sparsity, without any strong-signal or self-similarity-type assumption, as would typically be the case for obtaining adaptive confidence sets. Such assumptions would, of course, simplify the analysis significantly, but the point here is precisely that a uniform FDR control is possible for rate-adaptive procedures without any assumption on the true sparse signal. The uniform concentration of $\hat{w}$ is expressed implicitly and requires sharp estimates, contrary to rate results for which a concentration in a range of values is typically sufficient. In particular, some of our lemmas in the Supplementary Material [17] are refined versions of lemmas in [30].

As a corollary, (31) entails

$$\overline{\lim_n} \sup_{\theta_0 \in \ell_0[s_n]} \mathrm{FDR}(\theta_0, \varphi^{\ell\text{-val}}) = 0,$$

and this for any chosen threshold $t \in (0, 1)$ in $\varphi^{\ell\text{-val}}$. From a pure α-FDR controlling point of view, while making a vanishing small proportion of errors is obviously desirable, it implies that $\varphi^{\ell\text{-val}}$ is, as far as the FDR is concerned, somewhat conservative, in the sense that it does not spend all the allowed type I errors (0 instead of α) and thus will make too few (true) discoveries at the end. It turns out that in the present setting ℓ-values are not quite on the "exact" scale for FDR control. An alternative is to consider the q-value scale, as we now describe.

Algorithm `EBayesq`

`Input:` $X_1, \dots, X_n$, slab prior γ, target confidence t

`Output:` BMT procedure $\varphi^{q\text{-val}}$

1. Find the maximiser $\hat{w}$ given by (29).
2. Compute $\hat{q}_i(X) = q(X_i; \hat{w}, g)$.
3. Return, for $1 \le i \le n$,

$$\varphi_i^{q\text{-val}} = \mathbf{1}\{\hat{q}_i(X) \le t\}. \tag{32}$$

We also consider the following variant of the procedure `EBayesq`, which is mostly the same except that it does not allow for too small estimated weight $\hat{w}$. Set, for L_n tending slowly to infinity,

$$\omega_n = \frac{L_n}{n\overline{G}(\sqrt{2.1 \log n})}. \tag{33}$$

For instance, for γ Cauchy or quasi-Cauchy, we have $\omega_n \asymp (L_n/n)\sqrt{\log n}$, while for γ Laplace(1), we have $\omega_n \asymp (L_n/n) \exp\{C\sqrt{\log n}\}$.

Algorithm `EBayesq.0`

`Input:` $X_1, \dots, X_n$, slab prior γ, target confidence t, sequence L_n

`Output:` BMT procedure $\varphi^{q\text{-val.0}}$

1.–2. Same as for `EBayesq`, returning $\hat{q}_i(X)$.

3. Return, for $1 \le i \le n$ and ω_n, as in (33),

$$\varphi_i^{q\text{-val.0}} = \mathbf{1}\{\hat{q}_i(X) \le t\}\mathbf{1}\{\hat{w} > \omega_n\}. \tag{34}$$

THEOREM 2. *Consider the same setting as Theorem 1. Then, the algorithm* `EBayesq` *produces the BMT procedure $\varphi^{q\text{-val}}$ in (32) that satisfies the following: there exists a constant $C = C(\gamma, \upsilon)$ such that for any $t \le 3/4$, there exists an integer $N_0 = N_0(\gamma, \upsilon, t)$ such that, for any $n \ge N_0$,*

$$\sup_{\theta_0 \in \ell_0[s_n]} \mathrm{FDR}(\theta_0, \varphi^{q\text{-val}}) \le Ct \log(1/t). \tag{35}$$

In addition, the algorithm `EBayesq.0` *produces the BMT procedure $\varphi^{q\text{-val.0}}$ in (34) that satisfies, for ω_n as in (33) with $L_n \to \infty$, $L_n \le \log n$, $t \le 3/4$ and C, N_0 as before (but with possibly different numerical values) for any $n \ge N_0$,*

$$\sup_{\theta_0 \in \ell_0[s_n]} \mathrm{FDR}(\theta_0, \varphi^{q\text{-val.0}}) \le Ct. \tag{36}$$

The proof of Theorem 2 is technically close to that of Theorem 1 and is given in Section 8; see also Section 8.2 for an informal heuristic that serves as guidelines for the proof. The statements of Theorem 2 are, however, of different nature because the q-value threshold t appears explicitly in the bounds (35)–(36) that do not vanish as n tends to infinity.

The two bounds (35) and (36) differ from a $\log(1/t)$ term which may become significant for small t. This term appears in the case where the signal is weak (only few rejected nulls) and for which the calibration $\hat{w}$ is slightly too large. This may not be the case using a different type of sparsity-adaptation or a different estimate $\hat{w}$. Indeed, this phenomenon disappears when using `EBayesq.0`, since $\hat{w}$ is then set to 0 when it is not large enough, in which case the FDR control is shown to be guaranteed, and we retrieve a dependence in terms of a constant times the target level t.

A consequence of Theorem 2 is that an α-FDR control can be achieved with `EBayesq`/`EBayesq.0` procedures by taking $t = t(\alpha)$ sufficiently small (although not tending to zero). Again, it is important to know how small the constant $C > 0$ can be taken in (35) and (36). When the signal is strong enough, the following result shows that $C = 1$ and the $\log(1/t)$ factor can be removed in (35).

Let us first introduce a set $\mathcal{L}_0[s_n]$ of "large" signals for arbitrary $a > 1$,

$$\mathcal{L}_0[s_n] = \{\theta \in \ell_0[s_n] : |\theta_i| \geq a\sqrt{2\log(n/s_n)} \text{ for } i \in S_\theta, |S_\theta| = s_n\}. \tag{37}$$

THEOREM 3. *Consider $\mathcal{L}_0[s_n] = \mathcal{L}_0[s_n; a]$ defined by (37) with an arbitrary $a > 1$, for $s_n \to \infty$ and $s_n \leq n^\upsilon$ for some $\upsilon \in (0, 1)$. Assume that γ is a unimodal symmetric slab density that satisfies (17)–(19) with κ as in (18). Then, for any prespecified level $t \in (0, 1)$,* `EBayesq` *produces the BMT procedure $\varphi^{q\text{-val}}$ in (32) such that*

$$\lim_n \sup_{\theta_0 \in \mathcal{L}_0[s_n]} \mathrm{FDR}(\theta_0, \varphi^{q\text{-val}}) = \lim_n \inf_{\theta_0 \in \mathcal{L}_0[s_n]} \mathrm{FDR}(\theta_0, \varphi^{q\text{-val}}) = t. \tag{38}$$

In addition, `EBayesq.0` *with $L_n \to \infty$, satisfies the same property whenever $s_n/n \geq 2\omega_n$, for ω_n as in (33) which is, in particular, the case if s_n grows faster than a given power of n and $L_n \leq \log n$.*

Theorem 3, although focused on a specific regime, shows that empirical Bayes procedures are able to produce an asymptotically exact FDR control. Again, this may look surprising at first, as the prior slab density γ is not particularly linked to the true value of the parameter $\theta_0 \in \mathcal{L}_0[s_n]$ in (38). This puts forward a strong adaptive property of the spike and slab prior for multiple testing.

We conclude this section by giving results on the type II risk of the introduced multiple testing procedures. This is done by controlling the average number of false negatives (also called false nondiscoveries) among the nonzero coordinates which is called below False Negative Rate (FNR). For a given multiple testing procedure φ, following [4], we let

$$\mathrm{FNR}(\theta_0, \varphi) = E_{\theta_0}\left[\frac{\sum_{i=1}^n \mathbf{1}\{\theta_{0,i} \neq 0\}(1 - \varphi_i(X))}{1 \vee \sum_{i=1}^n \mathbf{1}\{\theta_{0,i} \neq 0\}}\right]. \tag{39}$$

Clearly, in the present setting, controlling this quantity is only possible under signal strength assumptions. Below, we provide such a control over the class $\mathcal{L}_0[s_n]$ defined in (37) above, and for the procedures $\varphi^{\ell\text{-val}}$ and $\varphi^{q\text{-val}}$ (results for $\varphi^{q\text{-val.0}}$ are the same as for $\varphi^{q\text{-val}}$ under the conditions of Theorem 3 and are omitted).

THEOREM 4. *Let $t \in (0, 1)$ be any prespecified level. Consider the setting and notation of Theorem 3, and recall the ℓ-values procedure from Theorem 1. The BMT procedures $\varphi^{\ell\text{-val}}$ and $\varphi^{q\text{-val}}$ verify*

$$\lim_n \sup_{\theta_0 \in \mathcal{L}_0[s_n]} \mathrm{FNR}(\theta_0, \varphi^{\ell\text{-val}}) = \lim_n \sup_{\theta_0 \in \mathcal{L}_0[s_n]} \mathrm{FNR}(\theta_0, \varphi^{q\text{-val}}) = 0. \tag{40}$$

COROLLARY 1. *In the setting of Theorem 4, for any prespecified level* $t \in (0,1)$, *the multiple testing procedures* $\varphi^{\ell\text{-val}}$ *and* $\varphi^{q\text{-val}}$ *satisfy*

$$\lim_n \Big[\sup_{\theta_0 \in \mathcal{L}_0[s_n]} \mathrm{FDR}(\theta_0, \varphi^{\ell\text{-val}}) + \sup_{\theta_0 \in \mathcal{L}_0[s_n]} \mathrm{FNR}(\theta_0, \varphi^{\ell\text{-val}}) \Big] = 0, \tag{41}$$

$$\lim_n \sup_{\theta_0 \in \mathcal{L}_0[s_n]} \mathrm{FDR}(\theta_0, \varphi^{q\text{-val}}) = t, \qquad \lim_n \sup_{\theta_0 \in \mathcal{L}_0[s_n]} \mathrm{FNR}(\theta_0, \varphi^{q\text{-val}}) = 0. \tag{42}$$

Let us consider, similarly to [4], the (multiple testing) classification risk $\mathfrak{R}(\theta_0, \varphi) = \mathrm{FDR}(\theta_0, \varphi) + \mathrm{FNR}(\theta_0, \varphi)$ for any $\theta_0 \in \mathbb{R}^n$ and procedure φ. It follows from Corollary 1 that for any $a > 1$ and $t < 1$,

$$\lim_n \sup_{\theta_0 \in \mathcal{L}_0[s_n; a]} \{\mathfrak{R}(\theta_0, \varphi^{\ell\text{-val}})\} = 0, \qquad \lim_n \sup_{\theta_0 \in \mathcal{L}_0[s_n; a]} \{\mathfrak{R}(\theta_0, \varphi^{q\text{-val}})\} = t,$$

so the procedure $\varphi^{\ell\text{-val}}$ is consistent for this risk on this range of signals, while $\varphi^{q\text{-val}}$ controls it at level $t < 1$.

We can legitimately ask if this property is optimal in some sense. We establish below that the classification task is impossible (i.e., the risk is at least 1) below the boundary $\sqrt{2\log(n/s_n)}$, at least over a fairly large class of procedures.

Define the class $\mathcal{C}$ of two-sided, thresholding-based multiple testing procedures φ of the form

$$\varphi_i(X) = \mathbf{1}\{X_i \geq \tau_1(X) \text{ or } -X_i \geq \tau_2(X)\}, \quad 1 \leq i \leq n$$

for some measurable $\tau_1(X), \tau_2(X) \geq 0$. The following result adapts a result of [4] to the two-sided context:

PROPOSITION 2. *Consider* $\mathcal{L}_0[s_n] = \mathcal{L}_0[s_n; a]$, *defined by* (37) *with an arbitrary* $a < 1$ *for* $s_n \to \infty$, *and* $s_n \leq n^{\upsilon}$ *for some* $\upsilon \in (0,1)$. *Consider the class of two-sided, thresholding-based multiple testing procedures* $\mathcal{C}$ *defined above. Then, for* $\mathfrak{R}$ *is the FDR+FNR classification risk defined above*,

$$\underline{\lim}_n \inf_{\varphi \in \mathcal{C}} \sup_{\theta_0 \in \mathcal{L}_0[s_n; a]} \mathfrak{R}(\theta_0, \varphi) \geq 1.$$

The proof of Proposition 2 is given in Section S-6. Let us underline that, therein, much sharper results are provided which allow to derive explicit convergence rates for the classification impossibility for a signal strength just below $\sqrt{2\log(n/s_n)}$.

Finally, we have established that the procedures $\varphi^{\ell\text{-val}}$ and $\varphi^{q\text{-val}}$ both achieve asymptotically the optimal classification boundary $\sqrt{2\log(n/s_n)}$. They asymptotically control the risk on $\mathcal{L}[s_n; a]$ for arbitrary $a > 1$ (at levels 0 and t respectively), while any such control is impossible if $a < 1$.

REMARK 1. Our results can be extended to the case where g is not of the form (10) (i.e., not necessarily of the form of a convolution with the standard Gaussian) but satisfies some weaker properties; see Section 7.1. This extended setting corresponds to a "quasi-Bayesian" approach where the ℓ-values (resp., q-values) are directly given by the formulas (12) (resp., (14)) without specifying a slab prior γ.

4. Numerical experiments. In this section our theoretical findings are illustrated via numerical experiments. A motivation here is also to evaluate how the parameters s_n, $\theta_0 \in \ell_0[s_n]$, and the hyperparameter γ (or g) affect the FDR control, in particular, the value of the constant in the bound of Theorem 2.

For this we consider $n = 10^4$, $s_n \in \{10, 10^2, 10^3\}$ and the following two possible scenarios for $\theta_0 \in \ell_0[s_n]$:

- constant alternatives: $\theta_{0,i} = \mu$ if $1 \le i \le s_n$ and 0 otherwise; or
- randomized alternatives: $\theta_{0,i}$ i.i.d. uniformly distributed on $(0, 2\mu)$ if $1 \le i \le s_n$ and 0 otherwise.

The parameter range for μ is taken equal to $\{0.01, 0.5, 1, 2, \dots, 10\}$. The marginal likelihood estimator $\hat{w}$ given by (29) is computed by using a modification of the function `wfromx` of the package `EbayesThresh` [31] that accommodates the lower bound $1/n$ in our definition (instead of $w_n = \zeta^{-1}(\sqrt{2\log n})$, see (55), in the original version). The parameter γ is either given by the quasi-Cauchy prior (20)–(21) or by the Laplace prior of scaling parameter $a = 1/2$ (see Remark 1 in the Supplementary Material for more details). For any of the above parameter combinations, the FDR of the procedures `EBayesL`, `EBayesq` (defined in Section 3) is evaluated empirically via 2000 replications.

Figure 1 displays the FDR of the procedures `EBayesL` (ℓ-values) and `EBayesq` (q-values). Concerning `EBayesL`, in all situations the FDR is small while not exactly equal to the value 0 which seems to indicate that the bound found in Theorem 1 is not too conservative. Moreover, the quasi-Cauchy version seems more conservative than the Laplace version, which corroborates our theoretical findings (in our bound (31); we have the factor $(\log n)^{-1}$ for quasi-Cauchy and $(\log n)^{-1/2}$ for Laplace). As for `EBayesq`, when the signal is large, the FDR curves are markedly close to the threshold value t when s_n/n is small, which is in line with Theorem 3. However, for a weak sparsity $s_n/n = 0.1$, the FDR values are slightly inflated (above the threshold t) which seems to indicate that the asymptotical regime is not yet reached for this value. Looking now at the whole range of signal strengths, one notices the presence of a "bump" in the regime of intermediate values of μ, especially for the Laplace prior. However, this bump seems to disappear when s_n/n decreases. We do not known presently whether this bump is vanishing with n or if this corresponds to a necessary additional constant $C = C(\gamma, \upsilon) > 1$ (or $\log(1/t)$) in the achieved FDR level, but we suspect that this is related to the fact that the intermediate regime was the most challenging part of our proofs. Overall, the Cauchy slab prior seems to have a particularly suitable behavior. This was not totally surprising for us, as it already showed more stability than the Laplace prior in the context of estimation with the full empirical Bayes posterior distribution, as seen in [16].

Finally, we provide additional experiments in the Supplementary Material; see Section S-10. The findings can be summarized as follows:

- The curves behave qualitatively similarly for randomized alternatives (second scenario).
- The procedure `EBayesq`.0 (with $L_n = \log\log n$) has a global behavior similar to `EBayesq` with more conservativeness for weak signal (as expected).
- It is possible to uniformly improve `EBayesq`.0 by considering the following modification (named `EBayesq.hybrid` below): if $w \le \omega_n$, instead of rejecting no null, `EBayesq.hybrid` performs a standard Bonferroni correction, that is, rejects the $H_{0,i}$'s such that $p_i(X) \le t/n$. Note that a careful inspection of the proof of Theorem 2 (`EBayesq`.0 part) shows that the bound (36) is still valid for `EBayesq.hybrid`.

5. Further procedures. Two other popular Bayesian multiple testing procedures are now briefly discussed as well as their links to both ℓ- and q-value procedures.

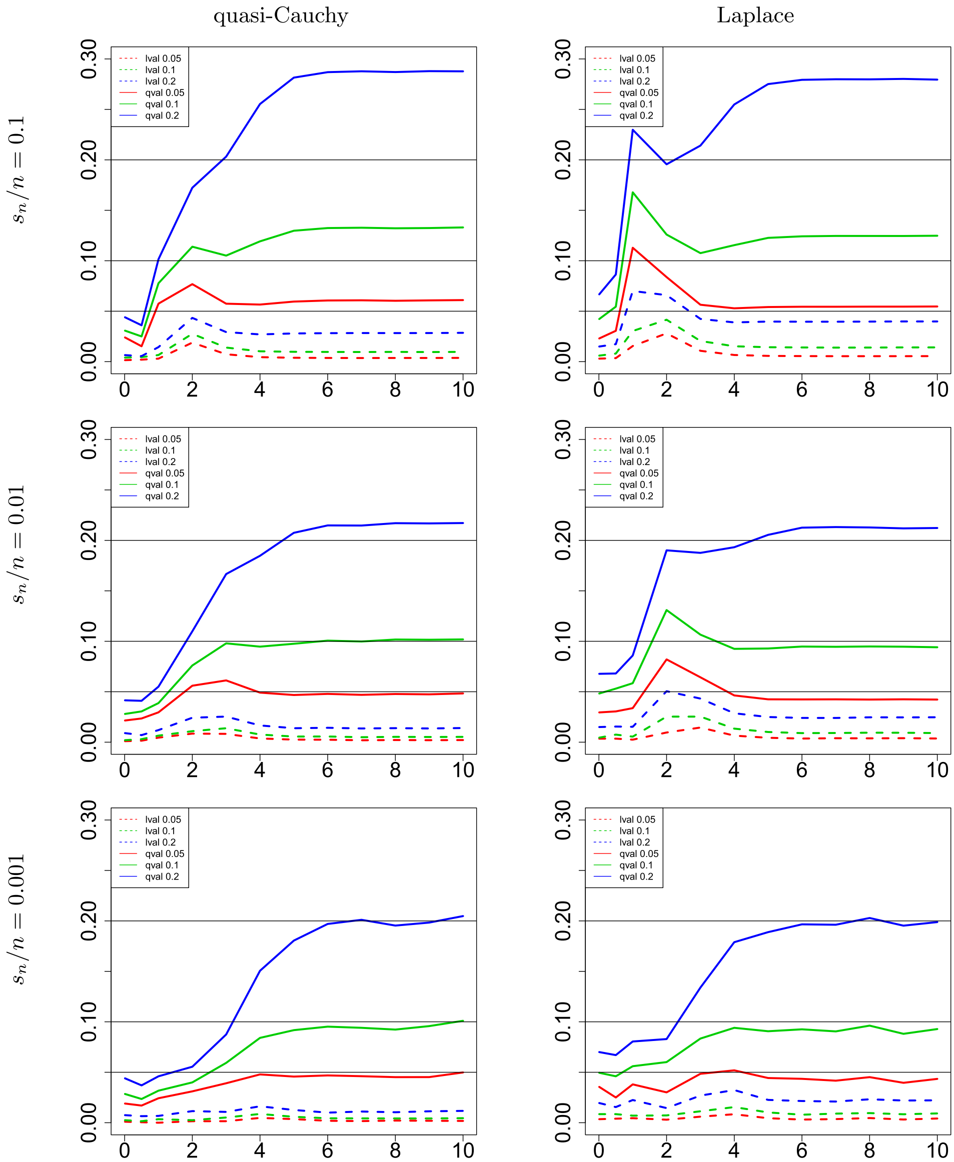

FIG. 1. *FDR of* EBayesL *and* EBayesq *procedures with threshold* $t \in \{0.05, 0.1, 0.2\}$; $n = 10{,}000$; 2000 *replications*; *alternative all equal to* μ (*on the* X-*axis*).

5.1. *MCI procedures*. Given a posterior distribution, one may test the presence of signal on a coordinate by looking at whether 0 belongs to a certain interval on this coordinate with high posterior probability. We refer to any such procedure based on marginal *credible* intervals as MCI procedure for short. Let $\mathcal{I}_i(X) = \mathcal{I}_i(t, X)$ be an interval with credibility, at least $1 - t$ for coordinate i for the empirical Bayes posterior, then, by definition, $\Pi[\mathcal{I}_i(X) \mid X] \geq 1 - t$. Hence, $0 \notin \mathcal{I}_i(X)$ implies, for $\hat{\ell}_i(X)$ as in (30),

$$\hat{\ell}_i(X) = \Pi[\theta_i = 0 \mid X] \leq \Pi\big[\theta_i \notin \mathcal{I}_i(X) \mid X\big] \leq 1 - (1 - t) = t.$$

One deduces that any MCI procedure at level $1 - t$ is more conservative than the ℓ-value procedure at level $t > 0$. For a natural quantile-based MCI procedure and spike-and-slab priors, it can be shown that the converse is also true up to taking a slightly lower level, say $t - \epsilon$, any $\epsilon > 0$, for the ℓ-value procedure; see Section S-7. This property means that in the present setting this quantile-based MCI procedure is essentially equivalent to the ℓ-value procedure which leads to Theorem 5 below and proved in Section S-7.

Let $z_i^t(X)$ denote the quantile at level $t \in (0,1)$ of the marginal empirical Bayes posterior distribution of the ith coordinate,

$$z_i^t(X) = \inf\{z \in \mathbb{R} : \Pi_{\hat{w},\gamma}[\theta_i \le z \mid X] \ge t\},$$

and define a procedure φ^m at level t as follows. For $i = 1, \dots, n$,

$$\varphi_i^m = \mathbf{1}\{0 \notin [z_i^t(X), z_i^{1-t}(X)]\}, \quad t \in (0, 1/2), \tag{43}$$
$$= \mathbf{1}\{0 < z_i^t(X)\} + \mathbf{1}\{z_i^{1-t}(X) < 0\} = \mathbf{1}\{0 \notin \mathcal{I}_i(X)\},$$

where $\mathcal{I}_i(X) = (z_i^t(X), +\infty)$ if $X_i \ge 0$, and $\mathcal{I}_i(X) = (-\infty, z_i^{1-t}(X))$ if $X_i < 0$. Note that such an interval $\mathcal{I}_i(X)$ is an MCI at level $1-t$, as its credibility is indeed $1-t$ in both cases.

THEOREM 5. *For $t < 1/2$, under the assumptions of Theorem 1, the conclusion of Theorem 1 holds for the MCI procedure φ^m at level t.*

In particular, the FDR of the φ^m procedure goes to 0 uniformly over sparse vectors. Control of FDR+FNR can be obtained as well, in a similar way as for the ℓ-value procedure in Section 3. The procedure φ^m can be shown to be very close to the ℓ-values procedure at level t; see Section S-7 for a justification and the proof of Theorem 5.

5.2. *Averaging ℓ-values.* Another type of procedures, advocated by Sun and Cai in a series of works (e.g., [45, 46]), are those based on averaged ℓ-values. In the Bayesian spike and slab context, it gives rise to the procedure, denoted here by `SC` (at a target confidence t), that rejects the $\hat{k}$ smallest ℓ-values, where $\hat{k}$ is the maximum of the k such that $k^{-1}\sum_{k'=1}^{k} \hat{\ell}_{(k')}(X) \le t$, where $\hat{\ell}_{(1)}(X) \le \dots \le \hat{\ell}_{(n)}(X)$ are the ordered elements of $\{\hat{\ell}_i(X), 1 \le i \le n\}$, the latter being the empirical Bayes ℓ-values used in `EBayesL` (Section 3). We provide insight into the behavior of `SC` in Section S-8, both theoretically and numerically. In a nutshell, we observe a qualitative behavior similar to `EBayesq` with an FDR tending to t under strong signal strength. Nevertheless, the convergence rate to the target level t seems slow (decreasing at a logarithmic order in n/s_n) because of a specific remainder term; see Lemma S-39.

6. Discussion. Our results show that spike and slab priors produce posterior distributions with particularly suitable multiple testing properties. One main challenge in deriving the results was to build bounds that are uniform over sparse vectors. We demonstrate that such a uniform control is possible up to a constant term away from the target control level. This constant is very close to 1 in simulations, and can even be shown to be 1 asymptotically for some subclass of sparse vectors.

The results of the paper are meant as a theoretical validation of the common practical use of posterior-based quantities for (frequentist) FDR control. While the main purpose here was validation, it is remarkable that a uniform control of the FDR very close to the target level can be obtained for the spike and slab BMT procedure in the present unstructured sparse high-dimensional model.

While many studies focused on controlling the Bayes FDR with Bayesian multiple testing procedures, this work paves the way for a frequentist FDR analysis of such procedures in different settings. In our study, perhaps the most surprising fact is how well marginal maximum likelihood estimation combines with FDR control under sparsity. As shown in our proof (and summarized in our heuristic), the score function is linked to a peculiar equation that makes perfectly the link between the numerator and the denominator in the FDR of the q-value-based multiple testing procedure. This phenomenon has not been noticed before to the best of our knowledge. We suspect that this link is only part of a more general picture in which the concentration of the score process in general sparse high-dimensional models plays a central role. While this exceeds the scope of this paper, generalizing our results to such settings is a very interesting direction for future work.

7. Preliminaries for the proofs.

7.1. *Working with general g.* As noted in Remark 1, the results of Theorems 1, 2 and 3 are also true under slightly more general assumptions that do not impose that g is coming from a γ by a convolution product. Namely, let us assume that

$$\begin{array}{c} g \text{ is a positive, symmetric, differentiable density} \\ \text{that decreases on a vicinity of } +\infty \end{array} \tag{44}$$

(g decreasing on a vicinity of $+\infty$ means that $x \to g(x)$ is decreasing for $x > M$, for a suitably large constant $M = M(g)$). Assume moreover that

$$\left|(\log g)'(y)\right| \le \Lambda \quad \text{for all } y \in \mathbb{R}, \Lambda > 0; \tag{45}$$

$$\overline{G}(y) \asymp g(y) y^{\kappa-1}, \quad \text{as } y \to \infty, \text{ for some } \kappa \in [1, 2]; \tag{46}$$

$$y \in \mathbb{R} \to (1 + y^2) g(y) \text{ is bounded}; \tag{47}$$

$$g/\phi \text{ is increasing on } [0, \infty) \text{ from } (g/\phi)(0) < 1 \text{ to } \infty; \tag{48}$$

By Lemma S-9, it is worth to note that (48) implies

$$\overline{G}/\overline{\Phi} \text{ is increasing on } [0, \infty) \text{ from } 1 \text{ to } \infty. \tag{49}$$

In the case where g is of the form of a convolution with γ, as in (10), conditions (45), (46) and (47) are easy consequences of the fact $g(y) \asymp \gamma(y)$ as $y \to \infty$; condition (48) follows from the fact that for all fixed $u > 0$, the function $x \in [0, \infty) \to (\phi(x+u) + \phi(x-u))/\phi(x)$ is increasing; see Lemma 1 of [30] for a detailed derivation.

A consequence of (45) is that g and $\overline{G}$ have at least Laplace tails

$$g(y) \ge g(0) e^{-\Lambda y}, \quad y \ge 0; \tag{50}$$

$$\overline{G}(y) \ge g(0) \Lambda^{-1} e^{-\Lambda y}, \quad y \ge 0. \tag{51}$$

7.2. *BMT as thresholding-based procedures.* Recall the definitions (22) and (23). Let, for any w and t in $[0, 1)$,

$$r(w, t) = \frac{wt}{(1-w)(1-t)}. \tag{52}$$

The following quantity plays the role of threshold for ℓ-values,

$$\xi = (\phi/g)^{-1} : \left(0, (\phi/g)(0)\right] \to [0, \infty), \tag{53}$$

that is, ξ is the decreasing continuous inverse of ϕ/g (that exists thanks to (48)). Simple algebra shows that, for $w, t \in [0, 1)$ with $r(w, t) \le \phi(0)/g(0)$,

$$\ell_i(X) \le t \quad \Leftrightarrow \quad |X_i| \ge \xi\big(r(w, t)\big). \tag{54}$$

When u becomes small, the order magnitude of $\xi(u)$ is given in Lemma S-12. $\xi(u)$ slightly exceeds $(-2 \log u)^{1/2}$, but not by much, which comes from the fact that g has heavy tails.

Another quantity close to ξ we shall use in the sequel is the threshold ζ introduced in [30] and defined as, for any $w \in (0, 1]$,

$$\zeta(w) = \beta^{-1}(w^{-1}). \tag{55}$$

Combining the definitions leads (see (S-8) for details) to $\zeta(w) = \xi(w/(1+w))$ and $\xi(w) \le \zeta(w)$. Similarly, let us introduce a threshold for q-values as

$$\chi = (\overline{\Phi}/\overline{G})^{-1} : (0, 1] \to [0, \infty), \tag{56}$$

which is the decreasing continuous inverse of $\overline{\Phi}/\overline{G}$ (that exists thanks to (49)). For all $w \in [0, 1)$ and $t \in [0, 1)$ with $r(w, t) \le 1$,

$$q_i(X) \le t \quad \Leftrightarrow \quad |X_i| \ge \chi\big(r(w, t)\big). \tag{57}$$

Lemma S-13 shows that, for small u, the order of magnitude of $\chi(u)$ is slightly more than $\overline{\Phi}^{-1}(u)$ but not by much which comes from the fact that $\overline{G}$ has heavy tails. Also, Lemma S-10 together with (54)–(57) imply

$$\chi(u) \le \xi(u) \quad \text{for } u \le 1. \tag{58}$$

7.3. *Single type I error rates.* The single type I error rates of our procedures are evaluated by the following result (proved in Section S-2.2):

PROPOSITION 3. *Consider any function g satisfying the assumptions of Section 7.1. Consider $r(\cdot, \cdot)$ as in (52), ξ as in (53) and χ as in (56). Then, the following bounds hold. For all t, w such that $r(w, t) \le (\phi/g)(0)$,*

$$P_{\theta_0=0}\big(\ell_i(X) \le t\big) \le 2r(w, t)\frac{g(\xi(r(w, t)))}{\xi(r(w, t))}. \tag{59}$$

Also, for all t, w such that $r(w, t) \le (\phi/g)(1)$,

$$P_{\theta_0=0}\big(\ell_i(X) \le t\big) \ge r(w, t)\frac{g(\xi(r(w, t)))}{\xi(r(w, t))}. \tag{60}$$

For q-values we have, for all t, w such that $r(w, t) \le 1$,

$$P_{\theta_0=0}\big(q_i(X) \le t\big) = r(w, t)2\overline{G}\big(\chi\big(r(w, t)\big)\big). \tag{61}$$

As a result, for a fixed w, we see that heavier tails of g result in larger type I error rate. This is well expected, as the heavier the tails of g, the more mass the prior puts on large values.

8. Proof of the main results.

8.1. *Notation.* The following moments are useful when studying the score function $\mathcal{S}$. Let us set

$$\tilde{m}(w) = -E_0\beta(X, w) = \int_{-\infty}^{\infty} \beta(t, w)\phi(t)\, dt \tag{62}$$

and further denote

$$m_1(\tau, w) = E_\tau\big[\beta(X, w)\big] = \int_{-\infty}^{\infty} \beta(t, w)\phi(t - \tau)\, dt. \tag{63}$$

$$m_2(\tau, w) = E_\tau\big[\beta(X, w)^2\big] = \int_{-\infty}^{\infty} \big(\beta(t, w)\big)^2\phi(t - \tau)\, dt. \tag{64}$$

These expectations are well defined and studied in detail in Section S-5, refining previous results established in [30].

In order to study the FDR of a procedure φ, we introduce the notation

$$V(\varphi) = \sum_{i:\theta_{0,i}=0} \varphi_i, \qquad S(\varphi) = \sum_{i:\theta_{0,i}\neq 0} \varphi_i, \tag{65}$$

counting for φ the number of false and true discoveries, respectively.

8.2. *Heuristic*. Why should the marginal empirical Bayes choice of w lead to a correct control of the FDR? Here is an informal argument that will give a direction for our proofs. We consider the case of $\varphi^{q\text{-val}}$ here as it is expected to reject more nulls than $\varphi^{\ell\text{-val}}$ and thus to have a larger FDR.

First, let us note that, when there is enough signal, one can expect $\hat{w}$ to be approximately equal to the solution $w^\star$ of the score equation in expectation $E_{\theta_0}(\mathcal{S}(w^\star)) = 0$, that is, by using (28),

$$\sum_{i:\theta_{0,i}\neq 0} m_1(\theta_{0,i}, w^\star) = (n - s_n)\tilde{m}(w^\star),$$

where $\tilde{m}$ and m_1 are defined by (62) and (63), respectively, if there θ_0 has exactly s_n nonzero coordinates. As seen in Section S-5, up to log terms,

$$\sum_{i:\theta_{0,i}\neq 0} m_1(\theta_{0,i}, w^\star) \approx \sum_{i:\theta_{0,i}\neq 0} \frac{\overline{\Phi}(\zeta(w^\star) - \theta_{0,i}) + \overline{\Phi}(\zeta(w^\star) + \theta_{0,i})}{w^\star};$$
$$\tilde{m}(w^\star) \approx 2\overline{G}(\zeta(w^\star)).$$

Now, consider the FDR and assume that all quantities are well concentrated (in particular, take the expectation both in the numerator and denominator in (9)). Then, by using (61) we have, denoting $\varphi^{q\text{-val}}(\alpha; \hat{w}, g)$ the q-value procedure at level α with parameters $\hat{w}, g$,

$$\begin{aligned}
&\mathrm{FDR}(\theta_0, \varphi^{q\text{-val}}(\alpha; \hat{w}, g)) \\
&\quad\approx \mathrm{FDR}(\theta_0, \varphi^{q\text{-val}}(\alpha; w^\star, g)) \\
&\quad\approx \frac{\sum_{i:\theta_{0,i}=0} P_{\theta_{0,i}}(q_i^\star(X) \le \alpha)}{\sum_{i:\theta_{0,i}=0} P_{\theta_{0,i}}(q_i^\star(X) \le \alpha) + \sum_{i:\theta_{0,i}\neq 0} P_{\theta_{0,i}}(q_i^\star(X) \le \alpha)} \\
&\quad\approx \frac{(n - s_n) r(w^\star, \alpha) 2\overline{G}(\zeta(w^\star))}{(n - s_n) r(w^\star, \alpha) 2\overline{G}(\zeta(w^\star)) + \sum_{i:\theta_{0,i}\neq 0} P_{\theta_{0,i}}(q_i^\star(X) \le \alpha)},
\end{aligned}$$

where we denoted $q_i^\star(X) = q(X_i; w^\star, g)$ and we used that $\chi(r(w^\star, t))$ is close to $\zeta(w^\star)$, as seen in Section S-4. Now, by using the definition of $q_i^\star(X)$,

$$\begin{aligned}
\sum_{i:\theta_{0,i}\neq 0} P_{\theta_{0,i}}(q_i^\star(X) \le \alpha) &= \sum_{i:\theta_{0,i}\neq 0} \overline{\Phi}(\chi(r(w^\star, \alpha)) - \theta_{0,i}) + \overline{\Phi}(\chi(r(w^\star, \alpha)) + \theta_{0,i}) \\
&\approx \sum_{i:\theta_{0,i}\neq 0} \overline{\Phi}(\zeta(w^\star) - \theta_{0,i}) + \overline{\Phi}(\zeta(w^\star) + \theta_{0,i}),
\end{aligned}$$

where we used again $\chi(r(w^\star, t)) \approx \zeta(w^\star)$. Now, using the above properties of $w^\star$, the latter is

$$\approx w^\star \sum_{i:\theta_{0,i}\neq 0} m_1(\theta_{0,i}, w^\star) = (n - s_n) w^\star \tilde{m}(w^\star) \approx (n - s_n) w^\star 2\overline{G}(\zeta(w^\star)).$$

Putting the previous estimates together yields

$$\begin{aligned}
&\mathrm{FDR}\big(\theta_0, \varphi^{q\text{-val}}(\alpha; \hat{w}, g)\big) \\
&\approx \frac{(n-s_n)r(w^\star,\alpha)2\overline{G}(\zeta(w^\star))}{(n-s_n)r(w^\star,\alpha)2\overline{G}(\zeta(w^\star)) + (n-s_n)w^\star 2\overline{G}(\zeta(w^\star))} \\
&= \frac{r(w^\star,\alpha)}{r(w^\star,\alpha)+w^\star} = \frac{\frac{w^\star}{1-w^\star}\frac{\alpha}{1-\alpha}}{\frac{w^\star}{1-w^\star}\frac{\alpha}{1-\alpha}+w^\star} \approx \frac{\frac{\alpha}{1-\alpha}}{\frac{\alpha}{1-\alpha}+1} = \alpha.
\end{aligned}$$

We will see that this heuristic holds, up to some constant terms that may come in as constant multipliers of the target level α.

We note that one main challenge in the proof below is to show that the above estimates hold true for *any* sparse signal, in particular for "intermediate" signals θ_0 that are neither close to 0 nor large enough (e.g., do not belong to $\mathcal{L}_0[s_n]$ as in (37)). Among others, we prove in Lemma S-5 that $\hat{w} \in [w_2, w_1]$ with $w_2 \asymp w_1$, thereby obtaining a sharp concentration of the marginal maximum likelihood estimate (uniformly over sparse vectors) that was not observed before in high-dimensional settings, to the best of our knowledge. To derive some of the approximations $\approx$ above, we also sharpen several of the estimates for the moments $m_1, \tilde{m}$ obtained in [30]; see; for example, Lemmas S-24 and S-27 for sharp upper and lower bounds on m_1.

8.3. *Proof of Theorems* 1 *and* 2. We prove results for ℓ- and q-values together. The proof for `EBayesq`.0 is given at the end of this section. First, let w_0 be the solution of the equation,

$$nw_0\tilde{m}(w_0) = M \tag{66}$$

for M to be chosen below in the range $[1, \log n]$ (more precisely, equal to either $C\log(1/t)$ or $Ct^{-1}\log\log n$ for a constant C independent of t and large enough, both bounds belong to the previous interval for n large enough). For any $M \in [1, \log n]$, this equation has always a unique solution, as $\tilde{m}$ is continuous increasing (see Lemma S-21) so the map $w \to w\tilde{m}(w)$ increases from 0 at $w = 0$ to a constant at $w = 1$ and, in particular, has a continuous inverse. This implies that w_0 goes to 0 with n, which we use freely in the sequel. Also, we note that w_0 is larger than $1/n$ for C in the choice of M large enough. Indeed, $w_0 \ge \tilde{m}(1)^{-1}M/n$ by monotonicity of $\tilde{m}$. But $\tilde{m}(1)$ is at most a constant, so, provided M is large enough, $w_0 \ge 1/n$. Thus, w_0 is always inside the interval $[n^{-1}, 1]$ over which the maximiser $\hat{w}$ is defined.

Let $\nu \in (0,1)$ be a fixed constant and $\theta_0 \in \ell_0[s_n]$. Recall that S_0 denotes the support of θ_0 and that $\sigma_0 = |S_0|$ denotes the exact number of nonzero coefficients of θ_0, so that $0 \le \sigma_0 \le s_n$. The next equation, depending on the configuration θ_0 and on the just defined w_0, plays a key role in the proof:

$$\sum_{i\in S_0} m_1(\theta_{0,i}, w) = (1-\nu)(n-\sigma_0)\tilde{m}(w), \quad w \in [w_0, 1). \tag{67}$$

This equation may or may not have a solution, depending on the true θ_0 and the values of n and ν. We will now assume $n \ge N_0$ for some universal constant N_0 to be determined below.

8.3.1. *Case* 1: (67) *has no solution*. For a given value of n, let us consider the case where (67) has no solution in $w \in [w_0, 1)$.

First, the maps $w \in [0,1] \to \tilde{m}(w)$ and $w \in [0,1] \to m_1(\mu, w)$ ($\mu \in \mathbb{R}$) are continuous (see Lemmas S-21 and S-23) and, for any $\mu \in \mathbb{R}$,

$$|m_1(\mu,1)| \le \int \left|\frac{\beta(x)}{1+\beta(x)}\right| \phi(x-\mu)\,dx \le \max_{x\in\mathbb{R}} \left|\frac{\beta(x)}{1+\beta(x)}\right|,$$

so that $\sum_{i\in S_0} m_1(\theta_{0,i},1) \le C\sigma_0 < (1-\nu)(n-\sigma_0)\tilde{m}(1)$ for $n \ge N_0$, where we use $\sigma_0 \le s_n \le n^{\upsilon}$ and $\tilde{m}(1) > 0$ and $N_0 = N_0(g,\upsilon)$. This means

$$\sum_{i\in S_0} m_1(\theta_{0,i},w) < (1-\nu)(n-\sigma_0)\tilde{m}(w) \quad \text{for } w \in [w_0,1), \tag{68}$$

as otherwise by the intermediate value theorem (e.g., Theorem 4.23 in [36]) the graphs of the functions on the two sides of the previous inequality would have to cross on $[w_0,1)$ and (67) would have a solution. Lemma S-3 shows that, under (68), we have

$$P_{\theta_0}(\hat{w} > w_0) \le e^{-C_0\nu^2 M} \tag{69}$$

for some constant $C_0 = C_0(g,\upsilon)$. Now, consider φ being either $\varphi^{\ell\text{-val}}$ or $\varphi^{q\text{-val}}$, and denote by $\varphi(t;\hat{w},g)$ such a procedure with cut-off t and parameters $\hat{w}, g$, as defined in (30)–(32). Let us upper bound the FDR by the so-called family-wise error rate by distinguishing the two cases $\hat{w} \le w_0$ and $\hat{w} > w_0$:

$$\begin{aligned}\mathrm{FDR}\big(\theta_0,\varphi(t;\hat{w},g)\big) &\le P_{\theta_0}\big(\exists i : \theta_{0,i}=0, \varphi_i(t;\hat{w},g)=1\big)\\ &\le P_{\theta_0}\big(\exists i : \theta_{0,i}=0, \varphi_i(t;w_0,g)=1\big) + P_{\theta_0}(\hat{w} > w_0)\\ &\le (n-\sigma_0)P_{\theta_{0,i}=0}\big(\varphi_i(t;w_0,g)=1\big) + e^{-C_0\nu^2 M},\end{aligned} \tag{70}$$

where we use that $w \to \varphi_i(t;w,g)$ is nondecreasing; see Lemma S-7, together with a union bound.

ℓ-value part. Let $\xi_0 = \xi(r(w_0,t))$ and $\zeta_0 = \zeta(w_0)$; then, (59) leads to (provided $r(w_0,t) \le (\phi/g)(0)$, which holds for, for example, $t \le 3/4$ and $w_0 \le 1/4$)

$$\mathrm{FDR}\big(\theta_0,\varphi^{\ell\text{-val}}(t;\hat{w},g)\big) \le 2\frac{nw_0}{1-w_0}\frac{t}{1-t}\frac{g(\xi_0)}{\xi_0} + e^{-C_0\nu^2 M}.$$

Combining the definition of w_0 and Lemma S-23 and taking n large enough so that w_0 is appropriately small, with $t \le 3/4$,

$$\mathrm{FDR}\big(\theta_0,\varphi^{\ell\text{-val}}(t;\hat{w},g)\big) \le \frac{5M}{\xi_0}\frac{g(\xi_0)}{\overline{G}(\zeta_0)}t + e^{-C_0\nu^2 M}.$$

Noting that $|\xi_0-\zeta_0| \lesssim 1$, $g(\xi_0) \le Dg(\zeta_0)$ and $\overline{G}(\zeta_0) \asymp \zeta_0^{\kappa-1}g(\zeta_0)$ by Lemma S-16 and S-23, one obtains

$$\mathrm{FDR}\big(\theta_0,\varphi^{\ell\text{-val}}(t;\hat{w},g)\big) \le \frac{C(g)M}{\zeta_0^{\kappa}}t + e^{-C_0\nu^2 M}. \tag{71}$$

q-value part. For the q-value case we come back to (70) and use (61) instead of (59) to get, setting $\chi_0 = \chi(r(w_0,t))$,

$$\mathrm{FDR}\big(\theta_0,\varphi^{q\text{-val}}(t;\hat{w},g)\big) \le 2\frac{nw_0}{1-w_0}\frac{t}{1-t}\overline{G}(\chi_0) + e^{-C_0\nu^2 M}.$$

As a result, by (66) and Lemma S-23, one gets for n large enough, $t \le 3/4$,

$$\mathrm{FDR}\big(\theta_0,\varphi^{q\text{-val}}(t;\hat{w},g)\big) \le 5Mt\frac{\overline{G}(\chi_0)}{\overline{G}(\zeta_0)} + e^{-C_0\nu^2 M}.$$

Now, by the last assertion of Lemma S-16, the ratio in the last display is bounded by 2 (say), provided n is large enough, which gives

$$\mathrm{FDR}\big(\theta_0,\varphi^{q\text{-val}}(t;\hat{w},g)\big) \le 10Mt + e^{-C_0\nu^2 M}. \tag{72}$$

8.3.2. *Case* 2: (67) *has a solution.* In this case we denote the solution by $w_1 \in [w_0, 1)$ so that one can write

$$\sum_{i \in S_0} m_1(\theta_{0,i}, w_1) = (1-\nu)(n-\sigma_0)\tilde{m}(w_1). \tag{73}$$

Now, consider the slightly different equation in w

$$\sum_{i \in S_0} m_1(\theta_{0,i}, w) = (1+\nu)(n-\sigma_0)\tilde{m}(w), \quad w \in [0, 1). \tag{74}$$

Equation (74) always has a (unique) solution $w_2 \in [0, w_1)$. To see this, first note that the case $\theta_0 = 0$ is excluded from (73), as $m_1(0, w) = -\tilde{m}(w) < 0$ if $w \neq 0$. By Lemma S-21, $w \to m_1(\mu, w)$ and $w \to \tilde{m}(w)$ are continuous and, respectively, decreasing and increasing (both strictly), and $\tilde{m}(0) = 0$, while it can be seen that $m_1(\mu, 0) > 0$ if $\mu \neq 0$; see Lemma S-21. On the other hand, the value at $w = 1$ of the left-hand side of (74) is at most $\sigma_0 C/w \lesssim \sigma_0$, and so is of smaller order than $(1+\nu)(n-\sigma_0)\tilde{m}(1) \asymp n$.

The purpose of w_1, w_2 is to provide (implicit) deterministic upper and lower bounds for the random $\hat{w}$; this is the content of Lemma S-4. Additionally, the key Lemma S-5 shows that, in case where the solution w_1 of (73) exists, we have $w_1 \asymp w_2$; that is, the bounds are of the same order.

q-value part. Recall the notation (65). We focus on the case of q-values first. We come back to the case of ℓ-values at the end, its proof being similar. For simplicity, we write $V_q(w) = V(\varphi^{q\text{-val}}(t; w, g))$ and $S_q(w) = S(\varphi^{q\text{-val}}(t; w, g))$. By definition of the FDR,

$$\begin{aligned}
\mathrm{FDR}\big(\theta_0, \varphi^{q\text{-val}}(t; \hat{w}, g)\big) &= E_{\theta_0}\left[\frac{V_q(\hat{w})}{(V_q(\hat{w}) + S_q(\hat{w})) \vee 1}\right] \\
&\le E_{\theta_0}\left[\frac{V_q(\hat{w})}{(V_q(\hat{w}) + S_q(\hat{w})) \vee 1}\mathbf{1}\{w_2 \le \hat{w} \le w_1\}\right] \\
&\quad + P_{\theta_0}\big[\hat{w} \notin [w_2, w_1]\big].
\end{aligned}$$

The last expectation in the previous display is now bounded by, using first the monotonicity of the maps $w \to V_q(w)$, $w \to S_q(w)$, $x \to x/(1+x)$ and $x \to 1/(1+x)$, then bounding the indicator variable by 1 and finally combining with Lemma S-44 applied to the *independent* variables $T = V_q(w_1)$ and $U = S_q(w_2)$,

$$\begin{aligned}
&E_{\theta_0}\left[\frac{V_q(\hat{w})}{(V_q(\hat{w}) + S_q(\hat{w})) \vee 1}\mathbf{1}\{w_2 \le \hat{w} \le w_1\}\right] \\
&\quad \le E_{\theta_0}\left[\frac{V_q(w_1)}{(V_q(w_1) + S_q(w_2)) \vee 1}\right] \\
&\quad \le \exp\{-E_{\theta_0} S_q(w_2)\} + 12\frac{E_{\theta_0} V_q(w_1)}{E_{\theta_0} S_q(w_2)}.
\end{aligned}$$

Next, by using the definition of V_q, one writes

$$E_{\theta_0} V_q(w_1) = \sum_{i:\theta_{0,i}=0} 2\overline{\Phi}\big(\chi\big(r(w_1, t)\big)\big) = 2(n-\sigma_0)\overline{\Phi}\big(\chi\big(r(w_1, t)\big)\big).$$

Using the definition of χ, we have $\overline{\Phi}(\chi(u)) = \overline{G}(\chi(u))u$ for $u \in (0, 1)$, so

$$\overline{\Phi}\big(\chi\big(r(w_1, t)\big)\big) = r(w_1, t)\overline{G}\big(\chi\big(r(w_1, t)\big)\big).$$

Then, (S-21) in Lemma S-16 implies, for small enough w_1,

$$\overline{G}(\chi(r(w_1,t))) \leq 2\overline{G}(\zeta(w_1)).$$

Combining (S-3) in Lemma S-5, that is, $w_1/C \leq w_2 \leq w_1$, for a constant $C = C(\nu, \upsilon, g) > 0$ and Lemma S-18, we have (with, say, $\epsilon = 1/2$),

$$(1/2)\overline{G}(\zeta(w_1)) \leq \overline{G}(\zeta(w_1/C)) \leq \overline{G}(\zeta(w_2)).$$

Next, using Lemma S-23, one obtains $\overline{G}(\chi(r(w_1,t))) \leq 3\tilde{m}(w_2)$ so that

$$\begin{aligned} E_{\theta_0} V_q(w_1) &\leq 3(n-\sigma_0)\frac{w_1}{1-w_1}\tilde{m}(w_2)\frac{t}{1-t} \\ &\leq 3C(n-\sigma_0)\frac{w_2}{1-Cw_2}\tilde{m}(w_2)\frac{t}{1-t} \\ &\leq C^*(n-\sigma_0)w_2\tilde{m}(w_2)t, \end{aligned}$$

because $t \leq 3/4$ for some constant $C^* = C^*(\nu, \upsilon, g) > 0$. On the other hand, by definition of S_q, one can write

$$E_{\theta_0} S_q(w_2) = \sum_{i:\theta_{0,i}\neq 0} \overline{\Phi}(\chi(r(w_2,t)) - \theta_{0,i}) + \overline{\Phi}(\chi(r(w_2,t)) + \theta_{0,i}).$$

Let us introduce the set of indices, for $K_1 = 2/(1-\upsilon)$,

$$\mathcal{C}_0(w, K_1) = \left\{1 \leq i \leq n : |\theta_{0,i}| \geq \frac{\zeta(w)}{K_1}\right\}. \tag{75}$$

Moreover, $\chi(r(w_2,t)) \leq \zeta(w_2)$ by Lemma S-15. Hence,

$$\begin{aligned} E_{\theta_0} S_q(w_2) &\geq \sum_{i\in\mathcal{C}_0(w_2,K_1)} \overline{\Phi}(\zeta(w_2) - \theta_{0,i}) + \overline{\Phi}(\zeta(w_2) + \theta_{0,i}) \\ &\geq \sum_{i\in\mathcal{C}_0(w_2,K_1)} \overline{\Phi}(\zeta(w_2) - |\theta_{0,i}|). \end{aligned} \tag{76}$$

First, we apply Corollary S-25 with $K = K_1$, $w = w_2$ to bound each term in the sum in terms of m_1, noting that $|\theta_{0,i}| \geq \zeta(w_2)/K_1$ by definition of the set $\mathcal{C}_0(w_2, K_1)$. Next, one uses Lemma S-30, restricting the suprema to $w = w_2$ (which is in the prescribed interval by Lemmas S-1, S-2 and S-5) and $K = K_1$, to get for n large enough and constants $C = C(\upsilon, g) > 0$, $C' = C'(\upsilon, g) > 0$, $D = D(\upsilon, g) \in (0, 1)$,

$$\begin{aligned} \sum_{i\in\mathcal{C}_0(w_2,K_1)} \overline{\Phi}(\zeta(w_2) - |\theta_{0,i}|) &\geq Cw_2 \sum_{i\in\mathcal{C}_0(w_2,K_1)} m_1(\theta_{0,i}, w_2) \\ &\geq Cw_2\left\{\sum_{i\in S_0} m_1(\theta_{0,i}, w_2) - C'n^{1-D}\tilde{m}(w_2)\right\} \\ &= Cw_2\{(1+\nu)(n-\sigma_0)\tilde{m}(w_2) - C'n^{1-D}\tilde{m}(w_2)\}, \end{aligned}$$

where the last equality comes from (74). As a consequence, for n large enough, for a positive constant $C_* = C_*(\upsilon, g) > 0$, we have

$$E_{\theta_0} S_q(w_2) \geq C_*(n-\sigma_0)w_2\tilde{m}(w_2).$$

Combining the previous bounds leads to

$$E_{\theta_0}\left[\frac{V_q(\hat{w})}{V_q(\hat{w}) + S_q(\hat{w}) \vee 1}\mathbf{1}\{w_2 \leq \hat{w} \leq w_1\}\right] \leq e^{-C_*(n-\sigma_0)w_2\tilde{m}(w_2)} + 12\frac{C^*}{C_*}t.$$

As $w \to w\tilde{m}(w)$ is increasing, and $w_1/C \le w_2$ by Lemma S-5, we have $w_2\tilde{m}(w_2) \ge (w_1/C)\tilde{m}(w_1/C)$. Recall that $w_1 \ge w_0$ by definition, so Lemma S-23 together with (S-24) of Lemma S-18 imply

$$\tilde{m}(w_1/C) \ge (1/2)\tilde{m}(w_1) \ge (1/2)\tilde{m}(w_0).$$

Combining the obtained inequalities leads to

$$(n-\sigma_0)w_2\tilde{m}(w_2) \ge C'(n-\sigma_0)w_0\tilde{m}(w_0) \ge C'M, \tag{77}$$

where the last inequality follows from the definition of w_0. Now, turning to a bound on the FDR, Lemma S-4 and the above inequality imply, with $\nu = 1/2$,

$$P_{\theta_0}\big[\hat{w} \notin [w_1, w_2]\big] \le 2e^{-C_1\nu^2 n w_2 \tilde{m}(w_2)} \le 2e^{-CM} \tag{78}$$

for some $C = C(\upsilon, g) > 0$. Conclude that in the considered case, for some constants $c_1 = c_1(\upsilon, g), c_2 = c_2(\upsilon, g) > 0$,

$$\mathrm{FDR}\big(\theta_0, \varphi^{q\text{-val}}(t; \hat{w}, g)\big) \le c_2 t + 3e^{-c_1 M}. \tag{79}$$

ℓ-value part. In the case of ℓ-values, one can follow a similar argument. We write $V_\ell(w) = V(\varphi^{\ell\text{-val}}(t; w, g))$ and $S_\ell(w) = S(\varphi^{\ell\text{-val}}(t; w, g))$. Again, the maps $w \to V_\ell(w)$ and $w \to S_\ell(w)$ are monotone. So, as above for q-values,

$$\begin{aligned}\mathrm{FDR}\big(\theta_0, \varphi^{\ell\text{-val}}(t; \hat{w}, g)\big) \le \exp\{-E_{\theta_0} S_\ell(w_2)\} + 12\frac{E_{\theta_0} V_\ell(w_1)}{E_{\theta_0} S_\ell(w_2)} \\ + P_{\theta_0}\big[\hat{w} \notin [w_2, w_1]\big].\end{aligned}$$

By definition of V_ℓ and ξ, one can write

$$E_{\theta_0} V_\ell(w_1) = 2(n-\sigma_0)\overline{\Phi}\big(\xi\big(r(w_1, t)\big)\big).$$

The bound $\overline{\Phi}(u) \le \phi(u)/u$ for $u > 0$ (see Lemma S-40), combined with the definition of ξ and that $|\xi(r(w_1, t)) - \zeta(w_1)| \lesssim 1$ by Lemma S-16 leads to

$$E_{\theta_0} V_\ell(w_1) \le 3(n-\sigma_0)\zeta(w_1)^{-1} r(w_1, t) g\big(\xi\big(r(w_1, t)\big)\big).$$

Lemma S-16 then implies $g(\xi(r(w_1, t))) \le 2g(\zeta(w_1))$ (say), for n large enough. Using $w_1/C \le w_2 \le w_1$ and (S-24) in Lemma S-18, we have

$$(1/2)g\big(\zeta(w_1)\big) \le g\big(\zeta(w_1/C)\big) \le g\big(\zeta(w_2)\big).$$

Next, using the relation $\zeta^{\kappa-1}g(\zeta) \asymp \tilde{m}(w)$ from Lemma S-23, one obtains $g(\xi(r(w_1, t))) \lesssim \zeta(w_2)^{1-\kappa}\tilde{m}(w_2) \lesssim \zeta(w_1)^{1-\kappa}\tilde{m}(w_2)$ so that

$$\begin{aligned}E_{\theta_0} V_\ell(w_1) &\le Ct(n-\sigma_0)w_1\tilde{m}(w_2)\zeta(w_1)^{-\kappa} \\ &\le c^* t(n-\sigma_0)w_2\tilde{m}(w_2)\zeta(w_1)^{-\kappa}\end{aligned}$$

for a constant $c^* = c^*(\upsilon, g) > 0$. On the other hand, by definition of S_ℓ,

$$E_{\theta_0} S_\ell(w_2) = \sum_{i:\theta_{0,i}\neq 0} \overline{\Phi}\big(\xi\big(r(w_2, t)\big) - \theta_{0,i}\big) + \overline{\Phi}\big(\xi\big(r(w_2, t)\big) + \theta_{0,i}\big).$$

Lemma S-17 now enables to bound from below the two terms in the previous display in terms of $\zeta(w_2)$, and further restricting the sum to the set of indices $\mathcal{C}_0(w_2, K_1)$ defined by (75) with the same choice of K_1 leads to

$$E_{\theta_0} S_\ell(w_2) \ge Ct \sum_{i \in \mathcal{C}_0(w_2, K_1)} \overline{\Phi}\big(\zeta(w_2) - |\theta_{0,i}|\big).$$

Apart from the Ct term in factor, it is the same bound as for q-values; see (76). Hence, using the bound obtained above, for n large enough and $c_* = c_*(\upsilon, g) > 0$,

$$E_{\theta_0} S_\ell(w_2) \geq c_* t(n - \sigma_0) w_2 \tilde{m}(w_2).$$

Combining the previous bounds leads to

$$E_{\theta_0}\left[\frac{V_\ell(\hat{w})}{V_\ell(\hat{w}) + S_\ell(\hat{w}) \vee 1} \mathbf{1}\{w_2 \leq \hat{w} \leq w_1\}\right] \leq e^{-c_* t(n-\sigma_0)w_2\tilde{m}(w_2)} + 12\frac{c^*}{c_*}\frac{1}{\zeta(w_1)^\kappa}.$$

As in (77), we have $(n - \sigma_0)w_2\tilde{m}(w_2) \geq C'(n - \sigma_0)w_0\tilde{m}(w_0) \geq C'M$. One concludes that, in Case 2, for some constants $d_1 = d_1(\upsilon, g), d_2 = d_2(\upsilon, g) > 0$ and taking $\nu = 1/2$, setting $\zeta(w_1) = \zeta_1$,

$$\begin{aligned} \mathrm{FDR}\big(\theta_0, \varphi^{\ell\text{-val}}(t; \hat{w}, g)\big) &\leq d_2\zeta_1^{-\kappa} + e^{-C'Mc_*t} + 2e^{-CM} \\ &\leq d_2\zeta_1^{-\kappa} + 3e^{-d_1 Mt}. \end{aligned} \tag{80}$$

8.3.3. *Combining Cases* 1 *and* 2. For q-values, for $\nu = 1/2$ and $t \leq 3/4$ we get by combining (72) and (79)

$$\mathrm{FDR}\big(\theta_0, \varphi^{q\text{-val}}(t; \hat{w}, g)\big) \leq \max\big\{10Mt + e^{-C_0 M}, c_2 t + 3e^{-c_1 M}\big\}.$$

Taking $M = (C_0 \wedge c_1)^{-1} \log(1/t)$ gives the upper bound

$$\mathrm{FDR}\big(\theta_0, \varphi^{q\text{-val}}(t; \hat{w}, g)\big) \leq \max\big\{C't\log(1/t) + e^{-\log(1/t)}, c_2 t + 3e^{-\log 1/t}\big\},$$

which is smaller than $Ct\log(1/t)$, giving the result for q-values.

In the ℓ-values case, with $\zeta_1 \leq \zeta_0$ and setting $\nu = 1/2$, we get by combining (71) and (80)

$$\begin{aligned} \mathrm{FDR}\big(\theta_0, \varphi^{\ell\text{-val}}(t; \hat{w}, g)\big) &\leq \max\big\{CM\zeta_0^{-\kappa}t + e^{-C_0 M}, d_2\zeta_1^{-\kappa} + 3e^{-d_1 tM}\big\} \\ &\leq d_3\big\{M\zeta_1^{-\kappa}t + \zeta_1^{-\kappa} + e^{-d_4 tM}\big\}. \end{aligned}$$

The announced bound is obtained upon setting $M = t^{-1}d_4^{-1}\log(\zeta_1^\kappa)$ and noting that $\zeta_1^2 \lesssim \log(1/w_1) \lesssim \log n$ and $\zeta_1^2 \gtrsim \log(1/w_1) \gtrsim \log n$ by using Lemmas S-1, S-2 to bound w_1 and Lemma S-14 to bound $\zeta(w_1)$. This concludes the proof of Theorem 1 for ℓ-values and Theorem 2 for q-values.

8.3.4. *Proof for* `EBayesq`.0. First, notice that

$$\begin{aligned} \mathrm{FDR}\big(\theta_0, \varphi^{q\text{-val}.0}(t; \hat{w}, g)\big) &= E_{\theta_0}\left[\frac{\sum_{i=1}^n \mathbf{1}\{\theta_{0,i} = 0\}\varphi^{q\text{-val}.0}(t; \hat{w}, g)}{1 \vee \sum_{i=1}^n \varphi^{q\text{-val}.0}(t; \hat{w}, g)}\right] \\ &= E_{\theta_0}\left[\frac{\sum_{i=1}^n \mathbf{1}\{\theta_{0,i} = 0\}\varphi^{q\text{-val}}(t; \hat{w}, g)}{1 \vee \sum_{i=1}^n \varphi^{q\text{-val}}(t; \hat{w}, g)}\mathbf{1}\{\hat{w} > \omega_n\}\right], \end{aligned} \tag{81}$$

by definition of algorithm `EBayesq`.0. The strategy of proof is similar to the q-value case. Let us take M in the definition (66) of w_0 equal to L_n from the statement of Theorem 2 (see (33)) and suppose $L_n \in [1, \log n]$. Let us show for n large enough,

$$\omega_n \geq w_0. \tag{82}$$

As $\zeta(w_0) \leq \zeta(1/n) \leq \sqrt{2.1\log n}$ for n large enough by Lemmas S-1, S-14,

$$\omega_n = \frac{L_n}{n\overline{G}(\sqrt{2.1\log n})} \geq \frac{L_n}{n\overline{G}(\zeta(1/n))} \geq \frac{L_n}{n\overline{G}(\zeta(w_0))}.$$

Now, by using Lemma S-23, for n large enough,

$$\frac{L_n}{n\overline{G}(\zeta(w_0))} \geq 0.9\frac{2L_n}{n\tilde{m}(w_0)} \geq \frac{L_n}{n\tilde{m}(w_0)} = w_0,$$

leading to (82). Next, on the one hand, in Case 1 the FDR is bounded by

$$\mathrm{FDR}(\theta_0, \varphi^{q\text{-val.0}}(t; \hat{w}, g)) \leq P_{\theta_0}(\hat{w} > \omega_n) \leq P_{\theta_0}(\hat{w} > w_0).$$

By using (69), the last display is at most $e^{-C_0\nu^2 L_n}$. On the other hand, in Case 2 we simply use that by (81),

$$\mathrm{FDR}(\theta_0, \varphi^{q\text{-val.0}}(t; \hat{w}, g)) \leq \mathrm{FDR}(\theta_0, \varphi^{q\text{-val}}(t; \hat{w}, g)) \leq c_2 t + 3e^{-c_1 L_n},$$

which concludes the proof.

Acknowledgments. The authors would like to thank the Associate Editor and the referees for their helpful and constructive comments.

This work was supported in part by grants of the ANR, projects ANR-16-CE40-0019 (SansSouci) and ANR-17-CE40-0001 (BASICS).

SUPPLEMENTARY MATERIAL

Supplement to "On spike and slab empirical Bayes multiple testing" (DOI: 10.1214/19-AOS1897SUPP; .pdf). This supplement contains all remaining proofs, additional details on some of the procedures introduced in the main text, and complementary numerical experiments.

REFERENCES

[1] ABRAMOVICH, F. and ANGELINI, C. (2006). Bayesian maximum *a posteriori* multiple testing procedure. *Sankhyā* **68** 436–460. MR2322194

[2] ABRAMOVICH, F., BENJAMINI, Y., DONOHO, D. L. and JOHNSTONE, I. M. (2006). Adapting to unknown sparsity by controlling the false discovery rate. *Ann. Statist.* **34** 584–653. MR2281879 https://doi.org/10.1214/009053606000000074

[3] ABRAMOVICH, F., GRINSHTEIN, V. and PENSKY, M. (2007). On optimality of Bayesian testimation in the normal means problem. *Ann. Statist.* **35** 2261–2286. MR2363971 https://doi.org/10.1214/009053607000000226

[4] ARIAS-CASTRO, E. and CHEN, S. (2017). Distribution-free multiple testing. *Electron. J. Stat.* **11** 1983–2001. MR3651021 https://doi.org/10.1214/17-EJS1277

[5] BELITSER, E. and GHOSAL, S. (2019). Empirical Bayes oracle uncertainty quantification for regression. *Ann. Statist.* To appear.

[6] BELITSER, E. and NURUSHEV, N. (2019). Needles and straw in a haystack: Robust empirical Bayes confidence for possibly sparse sequences. *Bernoulli*. To appear.

[7] BENJAMINI, Y. and HOCHBERG, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser. B* **57** 289–300. MR1325392

[8] BENJAMINI, Y., KRIEGER, A. M. and YEKUTIELI, D. (2006). Adaptive linear step-up procedures that control the false discovery rate. *Biometrika* **93** 491–507. MR2261438 https://doi.org/10.1093/biomet/93.3.491

[9] BENJAMINI, Y. and YEKUTIELI, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Ann. Statist.* **29** 1165–1188. MR1869245 https://doi.org/10.1214/aos/1013699998

[10] BOGDAN, M., CHAKRABARTI, A., FROMMLET, F. and GHOSH, J. K. (2011). Asymptotic Bayes-optimality under sparsity of some multiple testing procedures. *Ann. Statist.* **39** 1551–1579. MR2850212 https://doi.org/10.1214/10-AOS869

[11] BOGDAN, M., GHOSH, J. K. and TOKDAR, S. T. (2008). A comparison of the Benjamini–Hochberg procedure with some Bayesian rules for multiple testing. In *Beyond Parametrics in Interdisciplinary Research*: *Festschrift in Honor of Professor Pranab K. Sen. Inst. Math. Stat.* (*IMS*) *Collect.* **1** 211–230. IMS, Beachwood, OH. MR2462208 https://doi.org/10.1214/193940307000000158

[12] BOGDAN, M., VAN DEN BERG, E., SABATTI, C., SU, W. and CANDÈS, E. J. (2015). SLOPE—Adaptive variable selection via convex optimization. *Ann. Appl. Stat.* **9** 1103–1140. MR3418717 https://doi.org/10.1214/15-AOAS842

[13] CAI, T. T. and SUN, W. (2009). Simultaneous testing of grouped hypotheses: Finding needles in multiple haystacks. *J. Amer. Statist. Assoc.* **104** 1467–1481. MR2597000 https://doi.org/10.1198/jasa.2009.tm08415

[14] CAI, T. T., SUN, W. and WANG, W. (2019). Covariate-assisted ranking and screening for large-scale two-sample inference. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **81** 187–234. MR3928141

[15] CAO, J., XIE, X.-J., ZHANG, S., WHITEHURST, A. and WHITE, M. A. (2009). Bayesian optimal discovery procedure for simultaneous significance testing. *BMC Bioinform.* **10** Article ID 5.

[16] CASTILLO, I. and MISMER, R. (2018). Empirical Bayes analysis of spike and slab posterior distributions. *Electron. J. Stat.* **12** 3953–4001. MR3885271 https://doi.org/10.1214/18-EJS1494

[17] CASTILLO, I. and ROQUAIN, E. (2020). Supplement to "On spike and slab empirical Bayes multiple testing." https://doi.org/10.1214/19-AOS1897SUPP.

[18] CASTILLO, I. and SZABÓ, B. (2019). Spike and slab empirical Bayes sparse credible sets. *Bernoulli*. To appear.

[19] CASTILLO, I. and VAN DER VAART, A. (2012). Needles and straw in a haystack: Posterior concentration for possibly sparse sequences. *Ann. Statist.* **40** 2069–2101. MR3059077 https://doi.org/10.1214/12-AOS1029

[20] DONOHO, D. L., JOHNSTONE, I. M., HOCH, J. C. and STERN, A. S. (1992). Maximum entropy and the nearly black object. *J. Roy. Statist. Soc. Ser. B* **54** 41–81. MR1157714

[21] EFRON, B. (2007). Size, power and false discovery rates. *Ann. Statist.* **35** 1351–1377. MR2351089 https://doi.org/10.1214/009053606000001460

[22] EFRON, B. (2008). Microarrays, empirical Bayes and the two-groups model. *Statist. Sci.* **23** 1–22. MR2431866 https://doi.org/10.1214/07-STS236

[23] EFRON, B., TIBSHIRANI, R., STOREY, J. D. and TUSHER, V. (2001). Empirical Bayes analysis of a microarray experiment. *J. Amer. Statist. Assoc.* **96** 1151–1160. MR1946571 https://doi.org/10.1198/016214501753382129

[24] FINNER, H., DICKHAUS, T. and ROTERS, M. (2007). Dependency and false discovery rate: Asymptotics. *Ann. Statist.* **35** 1432–1455. MR2351092 https://doi.org/10.1214/009053607000000046

[25] GEORGE, E. I. and FOSTER, D. P. (2000). Calibration and empirical Bayes variable selection. *Biometrika* **87** 731–747. MR1813972 https://doi.org/10.1093/biomet/87.4.731

[26] GERARD, D. and STEPHENS, M. (2018). Empirical Bayes shrinkage and false discovery rate estimation, allowing for unwanted variation. *Biostatistics* To appear.

[27] GHOSAL, S., GHOSH, J. K. and VAN DER VAART, A. W. (2000). Convergence rates of posterior distributions. *Ann. Statist.* **28** 500–531. MR1790007 https://doi.org/10.1214/aos/1016218228

[28] GUINDANI, M., MÜLLER, P. and ZHANG, S. (2009). A Bayesian discovery procedure. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **71** 905–925. MR2750250 https://doi.org/10.1111/j.1467-9868.2009.00714.x

[29] JIANG, W. and ZHANG, C.-H. (2009). General maximum likelihood empirical Bayes estimation of normal means. *Ann. Statist.* **37** 1647–1684. MR2533467 https://doi.org/10.1214/08-AOS638

[30] JOHNSTONE, I. M. and SILVERMAN, B. W. (2004). Needles and straw in haystacks: Empirical Bayes estimates of possibly sparse sequences. *Ann. Statist.* **32** 1594–1649. MR2089135 https://doi.org/10.1214/009053604000000030

[31] JOHNSTONE, I. M. and SILVERMAN, B. W. (2005). EbayesThresh: R programs for empirical Bayes thresholding. *J. Stat. Softw.* **12** Issue 8.

[32] MARTIN, R. and TOKDAR, S. (2012). A nonparametric empirical Bayes framework for large-scale significance testing. *Biostatistics* **13** 427–439.

[33] MITCHELL, T. J. and BEAUCHAMP, J. J. (1988). Bayesian variable selection in linear regression. *J. Amer. Statist. Assoc.* **83** 1023–1036. MR0997578

[34] MÜLLER, P., PARMIGIANI, G., ROBERT, C. and ROUSSEAU, J. (2004). Optimal sample size for multiple testing: The case of gene expression microarrays. *J. Amer. Statist. Assoc.* **99** 990–1001. MR2109489 https://doi.org/10.1198/016214504000001646

[35] NEUVIAL, P. and ROQUAIN, E. (2012). On false discovery rate thresholding for classification under sparsity. *Ann. Statist.* **40** 2572–2600. MR3097613 https://doi.org/10.1214/12-AOS1042

[36] RUDIN, W. (1976). *Principles of Mathematical Analysis*, 3rd ed. *International Series in Pure and Applied Mathematics*. McGraw-Hill, New York. MR0385023

[37] SALOMOND, J.-B. (2017). Risk quantification for the thresholding rule for multiple testing using Gaussian scale mixtures. Preprint. Available at arXiv:1711.08705.

[38] SARKAR, S. K. (2007). Stepup procedures controlling generalized FWER and generalized FDR. *Ann. Statist.* **35** 2405–2420. MR2382652 https://doi.org/10.1214/009053607000000398

[39] SARKAR, S. K., ZHOU, T. and GHOSH, D. (2008). A general decision theoretic formulation of procedures controlling FDR and FNR from a Bayesian perspective. *Statist*. *Sinica* **18** 925–945. MR2440399

[40] SCOTT, J. G. and BERGER, J. O. (2010). Bayes and empirical-Bayes multiplicity adjustment in the variable-selection problem. *Ann*. *Statist*. **38** 2587–2619. MR2722450 https://doi.org/10.1214/10-AOS792

[41] SHEN, X. and WASSERMAN, L. (2001). Rates of convergence of posterior distributions. *Ann*. *Statist*. **29** 687–714. MR1865337 https://doi.org/10.1214/aos/1009210686

[42] STEPHENS, M. (2017). False discovery rates: A new deal. *Biostatistics* **18** 275–294. MR3824755 https://doi.org/10.1093/biostatistics/kxw041

[43] STOREY, J. D. (2003). The positive false discovery rate: A Bayesian interpretation and the q-value. *Ann*. *Statist*. **31** 2013–2035. MR2036398 https://doi.org/10.1214/aos/1074290335

[44] SU, W. and CANDÈS, E. (2016). SLOPE is adaptive to unknown sparsity and asymptotically minimax. *Ann*. *Statist*. **44** 1038–1068. MR3485953 https://doi.org/10.1214/15-AOS1397

[45] SUN, W. and CAI, T. T. (2007). Oracle and adaptive compound decision rules for false discovery rate control. *J*. *Amer*. *Statist*. *Assoc*. **102** 901–912. MR2411657 https://doi.org/10.1198/016214507000000545

[46] SUN, W. and CAI, T. T. (2009). Large-scale multiple testing under dependence. *J*. *R*. *Stat*. *Soc*. *Ser*. *B*. *Stat*. *Methodol*. **71** 393–424. MR2649603 https://doi.org/10.1111/j.1467-9868.2008.00694.x

[47] VAN DER PAS, S., SZABÓ, B. and VAN DER VAART, A. (2017). Adaptive posterior contraction rates for the horseshoe. *Electron*. *J*. *Stat*. **11** 3196–3225. MR3705450 https://doi.org/10.1214/17-EJS1316

[48] VAN DER PAS, S., SZABÓ, B. and VAN DER VAART, A. (2017). Uncertainty quantification for the horseshoe (with discussion). *Bayesian Anal*. **12** 1221–1274. MR3724985 https://doi.org/10.1214/17-BA1065

[49] VAN DER PAS, S. L., KLEIJN, B. J. K. and VAN DER VAART, A. W. (2014). The horseshoe estimator: Posterior concentration around nearly black vectors. *Electron*. *J*. *Stat*. **8** 2585–2618. MR3285877 https://doi.org/10.1214/14-EJS962