

Asymptotic Properties of Bayes Risk of a General Class of Shrinkage Priors in Multiple Hypothesis Testing Under Sparsity

Prasenjit Ghosh^{*}, Xueying Tang[†], Malay Ghosh[‡], and Arijit Chakrabarti[§]

Abstract. Consider the problem of simultaneous testing for the means of independent normal observations. In this paper, we study some asymptotic optimality properties of certain multiple testing rules induced by a general class of one-group shrinkage priors in a Bayesian decision theoretic framework, where the overall loss is taken as the number of misclassified hypotheses. We assume a two-groups normal mixture model for the data and consider the asymptotic framework adopted in Bogdan et al. (2011) who introduced the notion of asymptotic Bayes optimality under sparsity in the context of multiple testing. The general class of one-group priors under study is rich enough to include, among others, the families of three parameter beta, generalized double Pareto priors, and in particular the horseshoe, the normal–exponential–gamma and the Strawderman–Berger priors. We establish that within our chosen asymptotic framework, the multiple testing rules under study asymptotically attain the risk of the Bayes Oracle up to a multiplicative factor, with the constant in the risk close to the constant in the Oracle risk. This is similar to a result obtained in Datta and Ghosh (2013) for the multiple testing rule based on the horseshoe estimator introduced in Carvalho et al. (2009, 2010). We further show that under very mild assumption on the underlying sparsity parameter, the induced decisions using an empirical Bayes estimate of the corresponding global shrinkage parameter proposed by van der Pas et al. (2014), asymptotically attain the optimal Bayes risk up to the same multiplicative factor. We provide a unifying argument applicable for the general class of priors under study. In the process, we settle a conjecture regarding optimality property of the generalized double Pareto priors made in Datta and Ghosh (2013). Our work also shows that the result in Datta and Ghosh (2013) can be improved further.

MSC 2010 subject classifications: Primary 62C10; secondary 62C10.

Keywords: asymptotic optimality, Bayes Oracle, empirical Bayes, generalized double Pareto, three parameter beta, horseshoe, normal–exponential–gamma, Strawderman–Berger.

^{*}Applied Statistics Unit, Indian Statistical Institute, 203 B.T. Road, Kolkata, Pincode 700108, India, prasenjit_r@isical.ac.in

[†]Department of Statistics, University of Florida, P.O. Box 118545, Gainesville, Florida 32611, USA, xytang@stat.ufl.edu

[‡]Department of Statistics, University of Florida, P.O. Box 118545, Gainesville, Florida 32611, USA, ghoshm@stat.ufl.edu

[§]Applied Statistics Unit, Indian Statistical Institute, 203 B.T. Road, Kolkata, Pincode 700108, India, arc@isical.ac.in

1 Introduction

Multiple hypothesis testing has become a topic of growing importance in statistics, particularly for the analysis of high-dimensional data. Its application extends over various scientific fields such as genomics, bio-informatics, medicine, economics, finance, just to name a few. For example, in microarray experiments, thousands of tests are performed simultaneously to identify the differentially expressed genes, that is, genes whose expression levels are associated with some biological trait of interest. Microarray experiment is just one out of many examples where one needs to analyze sparse high-dimensional data, the main objective being detection of a few signals amidst a large body of noises. Multiple hypothesis testing is one convenient and fruitful approach towards this end. The biggest impetus to research in multiple hypothesis testing came from the classic paper of Benjamini and Hochberg (1995). Since then the topic has received considerable attention from both frequentists and Bayesians.

In this paper, we consider simultaneous testing for means of independent normal observations. Suppose we have m independent observations X_1, \dots, X_m , such that $X_i \sim N(\mu_i, \sigma^2)$, for $i = 1, \dots, m$. The unknown parameters μ_1, \dots, μ_m represent the effects under investigation, while σ^2 is the variance of the random noise. We wish to test $H_{0i} : \mu_i = 0$ against $H_{1i} : \mu_i \neq 0$, for $i = 1, \dots, m$. Our focus is on situations when m is large and the fraction of non-zero μ_i 's is small. For each i , μ_i is assumed to be a random variable whose distribution is determined by the latent binary random variable ν_i , where $\nu_i = 0$ denotes the event that H_{0i} is true while $\nu_i = 1$ corresponds to the event that H_{0i} is false. Here ν_i 's are assumed to be i.i.d. Bernoulli(p) random variables, for some p in $(0, 1)$. Under H_{0i} , $\mu_i = 0$, i.e., $\mu_i \sim \delta_{\{0\}}$, the distribution having mass 1 at 0, while under H_{1i} , $\mu_i \neq 0$ and it is assumed to follow an $N(0, \psi^2)$ distribution with $\psi^2 > 0$. Thus

$$\mu_i \stackrel{i.i.d.}{\sim} (1-p)\delta_{\{0\}} + pN(0, \psi^2), \quad i = 1, \dots, m. \quad (1)$$

The marginal distributions of the X_i 's are then given by the following two-groups model:

$$X_i \stackrel{i.i.d.}{\sim} (1-p)N(0, \sigma^2) + pN(0, \sigma^2 + \psi^2), \quad i = 1, \dots, m. \quad (2)$$

Our testing problem is now equivalent to testing simultaneously

$$H_{0i} : \nu_i = 0 \text{ versus } H_{1i} : \nu_i = 1 \text{ for } i = 1, \dots, m. \quad (3)$$

It is assumed that p , ψ^2 and σ^2 depend on the number of hypotheses m . The parameter p is the theoretical proportion of non-nulls in the population. In sparse situations, where most of the μ_i 's are zero or very small in magnitude, it is natural to assume that p is small and converges to 0 as the number of hypotheses m tends to infinity. The variance component ψ^2 is typically assumed to be large to identify the true signals. Such a model is very natural where one has few potentially large signals among a large pool of noise terms and has been very popular in the literature. See, for example, Mitchell and Beauchamp (1988), for an early use of modeling of this kind in Bayesian variable selection where a uniform prior is used for the absolutely continuous part in place of the normal prior as in (1) above. The two-groups model has the advantage of capturing information across different tests through learning about the common hyperparameters based on information from all the data points. Fully Bayesian approaches towards mul-

multiple testing based on the two-groups model by placing hyperpriors on the underlying model parameters are available in the literature, see, for example, Scott and Berger (2006) and Bogdan et al. (2008). Empirical Bayes approaches using the two-groups formulation have been considered, for example, in Efron (2004, 2008), Storey (2007) and Bogdan et al. (2008), just to name a few. Under model (2) and the usual additive loss function, Bogdan et al. (2011) provided conditions under which the optimal Bayes risk (that is, the risk corresponding to the Bayes rule) can be attained asymptotically under sparsity by a multiple testing procedure as the number of tests grows to infinity. They referred to this property as Asymptotic Bayes Optimality under Sparsity (ABOS). In particular, they showed that the procedures of Benjamini and Hochberg (1995) and Bonferroni attain the ABOS property under mild conditions. The optimal Bayes rule is also referred to as a Bayes Oracle in Bogdan et al. (2008) and Bogdan et al. (2011) and will be discussed in detail further in Section 3.

In contrast to the above two-groups formulation, there are proposals to model the unknown parameters in sparse situations through hierarchical one-group “shrinkage” priors. Such priors can be expressed as scale-mixtures of normals and their use require substantially less computational effort than the two-groups model, especially, in high-dimensional problems as well as in complex parametric frameworks. These priors capture sparsity by assigning large probabilities to means close to zero while at the same time they give non-trivial probabilities to large means. This is achieved by employing two levels of parameters to express the prior variances of the μ_i 's, namely, the “local shrinkage parameters”, which control the degree of shrinkage at the individual levels, and a “global shrinkage parameter”, common for all the μ_i 's to cause an overall shrinking effect. If the mixing density corresponding to the local shrinkage parameters is appropriately heavy tailed, the large observations are left almost unshrunk which is often referred to as the “tail robustness” property. Choice of the global shrinkage parameter varies in different specifications and will be discussed in greater detail in Section 2.

Some early examples of one-group shrinkage priors are the t -prior (Tipping (2001)), the Laplace prior in the context of Bayesian Lasso (Park and Casella (2008) and Hans (2009)) and the family of normal–exponential–gamma priors (Griffin and Brown (2005)). More recently, Carvalho et al. (2009, 2010) introduced a hierarchical Bayesian one-group prior called the horseshoe prior. Various new one-group shrinkage priors have been proposed in the literature and studied since then. Armagan et al. (2011) introduced the class of “three parameter beta normal” mixture priors while the class of “generalized double Pareto” priors was introduced in Armagan et al. (2012). The family of three parameter beta normal mixture priors generalizes some well known shrinkage priors such as the horseshoe, Strawderman–Berger and normal–exponential–gamma priors. See also, Polson and Scott (2011, 2012), Scott (2011) and Griffin and Brown (2010, 2012, 2013), in this context. Many of these one-group priors, including the horseshoe, employ local shrinkage parameters with priors having the aforesaid tail robustness property.

The horseshoe prior has acquired an important place in the literature on “shrinkage” priors and it has been used in estimation as well as in multiple testing and variable selection problems. Carvalho et al. (2010) proposed a new multiple testing procedure for the normal means problem based on the horseshoe prior. They observed through

numerical findings that under sparsity of the true normal means, the procedure based on the horseshoe prior performs closely to the Bayes rule when the true data comes from a two-groups model and the loss of a testing procedure is taken as the number of misclassified hypotheses. Datta and Ghosh (2013) theoretically established this optimality by showing that the ratio of the Bayes risk for this procedure to that of the Bayes Oracle under the two-groups model (2) is within a constant factor asymptotically. Moreover, it was numerically shown in their paper, that priors having exponential or lighter tails, such as the Laplace or the normal prior, fail to achieve such optimality property.

As commented in Carvalho et al. (2009), a carefully chosen two-groups model can be considered a “gold standard” for sparse problems. Therefore, it may be used as a benchmark against which the “shrinkage” priors can be judged. Motivated by this and inspired by the results in Carvalho et al. (2010) and Datta and Ghosh (2013), we want to study in this paper asymptotic optimality properties of multiple testing procedures induced by a very general class of “shrinkage” priors which are heavy tailed and yet handle sparsity well. This class contains the “three parameter beta normal” mixture priors as well as the “generalized double Pareto” priors. We consider multiple testing rules based on these priors and apply them on data generated from a two-groups model. We establish that these rules achieve the same Bayesian optimality property as shown in Datta and Ghosh (2013) for the testing rule based on the horseshoe prior, assuming that the global shrinkage parameter is appropriately chosen based on the theoretical proportion of true alternatives. In case this proportion is unknown, we consider an empirical Bayes version of this test procedure, where the global shrinkage parameter is estimated using the data as in van der Pas et al. (2014). We show that the resulting empirical Bayes testing procedure also attains the optimal Bayes risk asymptotically up to the same multiplicative factor. We also study the performance of such rules on simulated data and our theoretical results are corroborated by the simulations.

The highlight of this paper is a unified treatment of the question of Bayesian optimality in multiple testing under sparsity based on a very general class of one-group priors, taking the same loss function as in Datta and Ghosh (2013). In the process, we not only generalize their results for a very broad class of tail robust shrinkage priors, but also strengthen their optimality result by deriving a sharper asymptotic upper bound to the corresponding Bayes risk. We have a new unifying argument that enables us to establish asymptotic bounds to the risk for this whole class of priors. Datta and Ghosh (2013) conjectured that for the present multiple testing problem, the generalized double Pareto prior should enjoy similar optimality property like the horseshoe prior. We settle this conjecture by showing that the generalized double Pareto is indeed a member of this general class of tail robust priors under consideration. Further, our general technique of proof shows that some of the arguments in Datta and Ghosh (2013) can be simplified.

The organization of this paper is as follows. In Section 2, we describe the general class of one-group priors under study and define the multiple testing procedure based on them. In Section 3, we present our main theoretical results after describing the optimal Bayes rule under the two-groups model and the asymptotic framework under which these theoretical results are derived. The main results of Section 3 crucially depend on some key inequalities involving the posterior distribution of the underlying shrinkage coefficients, that help us deriving important asymptotic bounds to the type I

and type II error probabilities. These inequalities and the bounds on both types of error probabilities are presented in Section 4. Section 5 contains the simulation results followed by a discussion in Section 6. Proofs of the theoretical results are given in the Appendix.

1.1 Notations and definition

Given any two sequences of positive real numbers $\{a_m\}$ and $\{b_m\}$, with $b_m \neq 0$, we write $a_m \sim b_m$ to denote $\lim_{m \rightarrow \infty} a_m/b_m = 1$. For any two sequences of real numbers $\{a_m\}$ and $\{b_m\}$, with $b_m \neq 0$, we write $a_m = O(b_m)$ if $|\frac{a_m}{b_m}| \leq M$ for all m , for some positive real number M independent of m , and $a_m = o(b_m)$ to denote $\lim_{m \rightarrow \infty} a_m/b_m = 0$. Thus $a_m = o(1)$ if $\lim_{m \rightarrow \infty} a_m = 0$. Moreover, given any two positive real valued functions $f(x)$ and $g(x)$, both having a common domain of definition (A, ∞) , $A \geq 0$, we write $f(x) \sim g(x)$ as $x \rightarrow \infty$ to denote $\lim_{x \rightarrow \infty} f(x)/g(x) = 1$.

By a random variable Z we mean a $N(0, 1)$ random variable having cumulative distribution function and probability density function $\Phi(\cdot)$ and $\phi(\cdot)$, respectively.

Definition 1. A positive measurable function L defined over some (A, ∞) , $A \geq 0$, is said to be slowly varying or is said to vary slowly (in Karamata’s sense) if for every fixed $\alpha > 0$, $L(\alpha x) \sim L(x)$ as $x \rightarrow \infty$.

2 The one-group priors and the corresponding induced multiple testing procedures

As mentioned in the introduction, our aim in this paper is to study, through theoretical investigations and simulations, asymptotic risk properties of the multiple testing rules induced by a very broad class of one-group shrinkage priors, when applied to data that come from the two-groups model in (2). The class of one-group priors we study is inspired by a class of priors suggested in Polson and Scott (2011) which can be represented through the following hierarchical formulation:

$$\begin{aligned} \mu_i | (\lambda_i^2, \tau^2, \sigma^2) &\sim N(0, \lambda_i^2 \tau^2 \sigma^2), \text{ independently for } i = 1, \dots, m, \\ \lambda_i^2 &\sim \pi(\lambda_i^2), \text{ independently for } i = 1, \dots, m, \text{ and,} \\ (\tau^2, \sigma^2) &\sim \pi(\tau^2, \sigma^2). \end{aligned}$$

Our specific choices of $\pi(\lambda_i^2)$ and $\pi(\tau^2, \sigma^2)$ are described and explained below. Note that the above hierarchy is in slight variation from that of Polson and Scott (2011) in that we bring the σ earlier in the sequence, while in their formulation the σ comes later through the conditional prior of τ given σ . But both formulations produce the same marginal prior distribution for the μ_i ’s.

The above one-group formulation is often referred to as a global-local scale mixtures of normals. The parameter τ is called the “global” shrinkage parameter, while the parameters λ_i^2 ’s are called the “local” shrinkage parameters. The corresponding posterior

mean of μ_i is given by,

$$E(\mu_i|X_i, \tau, \sigma) = (1 - E(\kappa_i|X_i, \tau, \sigma))X_i, \quad (4)$$

where $\kappa_i = 1/(1 + \lambda_i^2 \tau^2)$ is called the i th shrinkage coefficient. It is observed in Carvalho et al. (2010) and Polson and Scott (2011) that under the two-groups model (2), for large ψ^2 , the posterior mean of μ_i can be approximated as

$$E(\mu_i|X_i, p, \psi, \sigma) \approx \omega_i(X_i)X_i \quad (5)$$

where $\omega_i(X_i)$ denotes the posterior probability that H_{1i} is true. It may be noted further that when $p \approx 0$, most of the ω_i 's are expected to be very close to zero unless X_i is sufficiently large, in which case the corresponding ω_i is expected to be close to 1, provided ψ^2 is large enough. This ensures that the noise observations are mostly shrunk towards zero, while the large X_i 's are left mostly unshrunk. Here the parameter p is responsible for achieving an overall shrinkage, while the large ψ^2 is helpful in discovering the true signals.

Using the above observations, for the one-group model, Polson and Scott (2011) argued that in sparse problems, the global shrinkage parameter τ (whose role is analogous to p in the two-groups prior) should be small and its prior should have substantial mass near zero, whereas the prior for the local shrinkage parameters λ_i^2 should have thick tails. This ensures that the resulting prior for the μ 's is highly peaked near zero but also heavy tailed enough to accommodate large signals. In this sense, the one-group priors can be thought of as approximately similar to a two-groups prior with an appropriately heavy-tailed absolutely continuous part.

Motivated by the preceding discussion and the work of Polson and Scott (2011), we take $\pi(\lambda_i^2)$ to be of the form

$$\pi(\lambda_i^2) = K(\lambda_i^2)^{-a-1}L(\lambda_i^2), \quad (6)$$

in our hierarchical formulation. Here $K > 0$ is the constant of proportionality, a is a positive real number and L is a positive measurable, non-constant, slowly varying function over $(0, \infty)$. It follows from Theorem 1 of Polson and Scott (2011) that the above general class of one-group priors achieves the desired "tail robustness" property in the sense that for any given τ and σ , $E(\mu_i|X_i, \tau, \sigma) \approx X_i$, for large X_i 's. Since $\pi(\lambda_i^2)$ is assumed to be proper, the possibility of $L(\cdot)$ being a constant function is ruled out.

It will be proved in Section 2.2 and Section 2.3 that a very broad class of one-group priors, such as, the generalized double Pareto and the three parameter beta normal mixtures, actually fall inside the general class of shrinkage priors under consideration. It is worth pointing out here that in some one-group formulations, like the original form of the generalized double Pareto in Armagan et al. (2012), the global shrinkage parameter is not explicitly mentioned or equivalently it is kept fixed at 1. In some other cases, like the three parameter beta normal mixtures, a shared global shrinkage parameter is explicitly given. Armagan et al. (2011) opined that it is reasonable to put a prior on the global shrinkage parameter, but this parameter may also be kept fixed at a certain value which reflects the prior knowledge about sparsity *if* such information

is available. In case such prior knowledge is unavailable, one can consider either of the two approaches, namely, (i) a full Bayes approach by placing further hyperprior over τ and (ii) an empirical Bayes approach by learning about τ through the data. In the line of recommendation of Polson and Scott (2011), for a full Bayes treatment of the present multiple testing problem, we consider the following joint prior distribution of (τ, σ) ,

$$\tau \sim C^+(0, 1) \text{ and } \pi(\sigma) \propto \frac{1}{\sigma}, \quad (7)$$

which will be used later in our simulation study. It should be noted here that Gelman (2006) strongly recommended the use of a half-Cauchy (or more generally, a folded non-central- t) distribution as a prior for the global variance component τ in a hierarchical Bayesian formulation. Though, in his original recommendation, he suggested using a half-Cauchy prior $C^+(0, \sigma)$ for τ scaled by the error variance σ^2 , we take $C^+(0, 1)$ since the error variance term σ^2 appears earlier in our hierarchical formulation. We also consider an empirical Bayes approach to be discussed in detail shortly.

We describe below the multiple testing rules considered in this paper. We first consider two rules (defined in (8) and (10) below) for which asymptotic optimality results have been derived theoretically. For this we assume σ^2 to be known and equal to 1. For the first rule, we treat τ as a tuning parameter to be chosen freely depending on the value of p , while the second one is based on an empirical Bayes estimate of τ . Note that a comparison between the expressions in (4) and (5) for the posterior mean of μ_i , together with the previous discussion, suggest that the posterior shrinkage weights $E(1 - \kappa_i | X_i, \tau)$ based on tail robust shrinkage priors, should behave like the posterior inclusion probability $\omega_i(X_i)$ in the two-groups model. Using this observation, Carvalho et al. (2010) proposed a natural classification rule based on the posterior shrinkage weights under a symmetric 0–1 loss for the horseshoe prior. Borrowing the same idea, we consider the following multiple testing procedure based on our chosen class of tail-robust one-group shrinkage priors, given by:

$$\text{reject } H_{0i} \text{ if } 1 - E(\kappa_i | X_i, \tau) > 0.5, \quad i = 1, \dots, m. \quad (8)$$

As mentioned in the introduction, Datta and Ghosh (2013) considered the multiple testing rule defined in (8) based on the horseshoe prior. They showed that it asymptotically attains the optimal Bayes risk up to a multiplicative factor. It will be seen later that the Oracle optimality property of the decision rule in (8) based on our general class of one-group priors, critically depends on appropriate choice of τ depending on p . This plays a significant role in the limiting value of the type II error measure and in controlling the rate of the overall contribution from type I error in the risk function. This is similar to the observations made in Datta and Ghosh (2013) for the above multiple testing rule based on the horseshoe prior.

In a recent article, van der Pas et al. (2014) considered the problem of estimating an m -dimensional multivariate normal mean vector which is sparse in the nearly black sense, that is, the number of non-zero entries is of a smaller order than m as $m \rightarrow \infty$. They modeled the mean vector through the horseshoe prior and estimated it by the corresponding posterior mean, namely, the horseshoe estimator. They showed that for

suitably chosen τ depending on the proportion of non-zero elements of the mean vector, the horseshoe estimator asymptotically attains the corresponding minimax l_2 risk, possibly up to a multiplicative constant, and the corresponding posterior distribution contracts at this optimal rate. But in practice p is usually unknown. A natural approach in such situations is to learn about τ from the data and then plug this choice into the corresponding posterior mean. When p is unknown, van der Pas et al. (2014) proposed a natural estimator of τ and showed that the horseshoe estimator based on this estimate, attains the corresponding minimax l_2 risk up to some multiplicative factor. Inspired by this, we consider the following estimator of τ due to van der Pas et al. (2014) in case p is unknown:

$$\hat{\tau} = \max \left\{ \frac{1}{m}, \frac{1}{c_2 m} \sum_{j=1}^m 1\{|X_j| > \sqrt{c_1 \log m}\} \right\} \quad (9)$$

where $c_1 \geq 2$ and $c_2 \geq 1$ are some predetermined finite positive constants. Note that the above estimator of τ is truncated below by $\frac{1}{m}$ and hence it is not susceptible to collapsing to zero, which is a major concern for the use of such empirical Bayes approaches as mentioned in Carvalho et al. (2009), Scott and Berger (2010), Bogdan et al. (2008) and Datta and Ghosh (2013). We refer to van der Pas et al. (2014) for a detailed discussion on this point. Let $E(1 - \kappa_i | X_i, \hat{\tau})$ denote the posterior shrinkage weight $E(1 - \kappa_i | X_i, \tau)$ evaluated at $\tau = \hat{\tau}$. We consider the following empirical Bayes procedure based on $E(1 - \kappa_i | X_i, \hat{\tau})$, $i = 1, \dots, m$, given by:

$$\text{reject } H_{0i} \text{ if } 1 - E(\kappa_i | X_i, \hat{\tau}) > 0.5, \quad i = 1, \dots, m. \quad (10)$$

For the simulations we consider two cases, firstly, a full Bayes treatment with (τ, σ) given the joint prior distribution as in (7), and the corresponding rule is defined as

$$\text{reject } H_{0i} \text{ if } 1 - E(\kappa_i | X_1, \dots, X_m) > 0.5, \quad i = 1, \dots, m, \quad (11)$$

where $1 - E(\kappa_i | X_1, \dots, X_m)$ denotes the i th posterior shrinkage weight after integrating $E(1 - \kappa_i | X_i, \tau, \sigma)$ with respect to the joint posterior density of (τ, σ) . We also consider the empirical Bayes decisions as in (10) for the simulation study where we fix $\sigma^2 = 1$. We apply the above decision rules in (11) or (10) induced by these priors in the multiple testing problem (3), where the true data are generated from the two-groups mixture model (2) described before. We show in this paper, through theoretical analysis and simulations, that the aforesaid decision rules enjoy similar optimality property as shown for the horseshoe prior in Datta and Ghosh (2013).

2.1 Some well known one-group shrinkage priors

In this section, we demonstrate that some popular shrinkage priors actually fall within the general class of one-group priors considered in this paper. This follows from observing that the mixing density $\pi(\lambda_i^2)$ corresponding to the local shrinkage parameter λ_i^2 can be expressed in the form (6) where $L(\cdot)$ is a slowly varying function over $(0, \infty)$. This in turn can be shown by proving that the corresponding $L(t)$ converges to a finite positive limit as t goes to infinity. We also show that for each of these priors the corresponding

$L(\cdot)$ is uniformly bounded by some finite positive constant. The boundedness property of $L(\cdot)$ is important, as it makes the proofs of the theoretical results of this paper much simpler. This will become clear in Section 4 of this paper.

2.2 Three parameter beta normal mixtures

Let us consider the following global-local scale mixture formulation of one-group priors:

$$\begin{aligned} \mu_i \mid \lambda_i^2, \tau^2, \sigma^2 &\sim N(0, \lambda_i^2 \tau^2 \sigma^2) \text{ independently for } i = 1, \dots, m, \\ \lambda_i^2 &\sim \pi(\lambda_i^2) \text{ independently for } i = 1, \dots, m, \\ (\tau^2, \sigma^2) &\sim \pi(\tau^2, \sigma^2) \end{aligned}$$

with

$$\pi(\lambda_i^2) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} (\lambda_i^2)^{\alpha-1} (1 + \lambda_i^2)^{-(\alpha+\beta)} \tag{12}$$

for $\alpha > 0, \beta > 0$. The mixing density given in (12), in fact, corresponds to an inverted-beta density (or, beta density of the second kind) with parameters α and β . The prior density corresponding to the shrinkage coefficients $\kappa_i = \frac{1}{1+\lambda_i^2 \tau^2}$ is then given by

$$\pi(\kappa_i) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} (\tau^2)^\beta \kappa_i^{\beta-1} (1 - \kappa_i)^{\alpha-1} \left\{ 1 - (1 - \tau^2)\kappa_i \right\}^{-(\alpha+\beta)}$$

which corresponds to an $TPB(\alpha, \beta, \tau^2)$ density. Therefore, the above hierarchical one-group formulation can alternatively be represented as

$$\begin{aligned} \mu_i \mid \kappa_i, \sigma^2 &\sim N(0, (\kappa_i^{-1} - 1)\sigma^2) \text{ independently for } i = 1, \dots, m, \\ \kappa_i &\sim TPB(\alpha, \beta, \tau^2) \text{ independently for } i = 1, \dots, m. \end{aligned}$$

This gives the three parameter beta normal mixture priors introduced by Armagan et al. (2011) and is denoted by $TPBN(\alpha, \beta, \tau^2 \sigma^2)$. The TPBN family of priors is rich enough to generalize some well known shrinkage priors, such as the horseshoe prior with $\alpha = \frac{1}{2}, \beta = \frac{1}{2}$, the Strawderman–Berger prior with $\alpha = 1, \beta = \frac{1}{2}$ and $\tau^2 = 1$ and the normal–exponential–gamma priors with $\alpha = 1, \beta > 0$.

Note that the prior in (12) can also be written as

$$\pi(\lambda_i^2) = K(\lambda_i^2)^{-\beta-1} L(\lambda_i^2)$$

where $L(\lambda_i^2) = (1 + \frac{1}{\lambda_i^2})^{-(\alpha+\beta)}$ and $K = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}$. Clearly, $\lim_{\lambda^2 \rightarrow \infty} L(\lambda^2) = 1$, thereby implying that the TPBN family of priors falls within our general class of global-scale mixture normals. Also, note that $\sup_{t \in (0, \infty)} L(t) = 1$, which shows that the associated function $L(\cdot)$ is bounded as mentioned earlier.

2.3 Generalized double Pareto priors

Let us consider the hierarchical one-group global-local scale mixture formulation as described at the beginning of Section 2, where $\pi(\lambda_i^2)$ is defined as

$$\begin{aligned}\lambda_i^2 | \gamma_i &\sim \text{Exponential}\left(\frac{\gamma_i^2}{2}\right) \text{ independently for } i = 1, \dots, m, \\ \gamma_i | \alpha, \beta &\sim \text{Gamma}(\alpha, \beta) \text{ independently for } i = 1, \dots, m,\end{aligned}$$

for some fixed $\alpha > 0$ and $\beta > 0$. It follows that $\mu_i | (\tau, \sigma)$ has the density

$$\pi(\mu_i | \tau, \sigma) = \frac{1}{2\tau\sigma\beta/\alpha} \left(1 + \frac{|\mu_i|}{\alpha \cdot \tau\sigma\beta/\alpha}\right)^{-(1+\alpha)}. \quad (13)$$

The density in (13) above, corresponds to a generalized double Pareto density with shape parameter α and scale parameter $\xi = \tau\sigma\beta/\alpha > 0$ and is denoted by $GDP(\alpha, \xi)$. Equivalently, it may also be interpreted as the density of a $GDP(\alpha, \beta/\alpha)$ random variable multiplied by $\tau\sigma$. When $\alpha = 1$ and $\beta = 1$, a $GDP(\alpha, \beta/\alpha)$ distribution is known as the standard double Pareto distribution. We refer to this hierarchical global-local scale mixture formulation with $\pi(\lambda_i^2)$ defined as above, as the generalized double Pareto prior introduced by Armagan et al. (2012). For simulations in Section 5, in our hierarchical global-local scale mixture formulation, when we talk about the standard double Pareto prior, we mean that $\lambda_i^2 \sim GDP(1, 1)$, and we mix further with respect to the joint density of (τ, σ) for a full Bayes treatment or use an empirical Bayes estimate of τ taking σ^2 to be fixed, as mentioned before.

Now we demonstrate that the generalized double Pareto prior falls within our chosen class of tail robust shrinkage priors. Towards this end, we first observe that the mixing density $\pi(\lambda_i^2)$ corresponding to the generalized double Pareto prior can be written as

$$\pi(\lambda_i^2) = \frac{\beta^\alpha}{2\Gamma(\alpha)} \int_0^\infty e^{-(\frac{\gamma_i^2 \lambda_i^2}{2} + \beta \gamma_i)} \gamma_i^{\alpha+1} d\gamma_i. \quad (14)$$

Note that using Fubini's Theorem one has $\int_0^\infty \pi(\lambda_i^2) d\lambda_i^2 = 1$, so that the density given in (14) is proper. Now, using the change of variable $u = \lambda_i^2 \gamma_i^2 / 2$ in the integral on the right hand side of (14), we obtain

$$\begin{aligned}\pi(\lambda_i^2) &= \frac{\beta^\alpha (\lambda_i^2)^{-\frac{\alpha}{2}-1}}{2^{1-\frac{\alpha}{2}} \Gamma(\alpha)} \int_0^\infty e^{-\beta \sqrt{\frac{2u}{\lambda_i^2}}} e^{-u} u^{(\frac{\alpha}{2}+1)-1} du \\ &= K (\lambda_i^2)^{-\frac{\alpha}{2}-1} L(\lambda_i^2), \text{ say,}\end{aligned}$$

where $L(\lambda_i^2) = 2^{\frac{\alpha}{2}-1} \int_0^\infty e^{-\beta \sqrt{\frac{2u}{\lambda_i^2}}} e^{-u} u^{(\frac{\alpha}{2}+1)-1} du$ and $K = \frac{\beta^\alpha}{\Gamma(\alpha)}$.

Now applying Lebesgue's Dominated Convergence Theorem, we obtain

$$\lim_{\lambda_i^2 \rightarrow \infty} L(\lambda_i^2) = 2^{\frac{\alpha}{2}-1} \int_0^\infty e^{-u} u^{(\frac{\alpha}{2}+1)-1} du = 2^{\frac{\alpha}{2}-1} \Gamma\left(\frac{\alpha}{2} + 1\right) > 0,$$

which means that $L(\cdot)$ defined above slowly varies over $(0, \infty)$. This also shows that the mixing density given in (14) can be expressed in the form given by (6) with $L(\cdot)$ as above and $a = \alpha/2$. Thus, the generalized double Pareto prior falls within our general class of tail-robust shrinkage priors. Moreover, using the monotone convergence theorem, it follows that $\sup_{t \in (0, \infty)} L(t) = 2^{\frac{\alpha}{2}-1} \Gamma(\frac{\alpha}{2} + 1)$, which means the function $L(\cdot)$ defined above is bounded as well.

3 Asymptotic framework and the main results

In this section, we present our major theoretical results about asymptotic optimality of the multiple testing rules (8) and (10) under study. In Section 3.1, first we describe the decision theoretic setting and the optimal Bayes rule under this setting. We then describe the asymptotic framework under which our theoretical results are derived. Section 3.2 presents the main theoretical results of this paper involving asymptotic bounds to the Bayes risk of the induced decisions (8) and (10) under study. The Oracle optimality properties of these decision rules up to $O(1)$ then follow immediately.

3.1 Optimal Bayes rule and the asymptotic framework

Suppose X_1, \dots, X_m are independently distributed according to the two-groups model (2), with $\sigma^2 = 1$. We are interested in the multiple testing problem (3). We assume a symmetric 0–1 loss for each individual test and the total loss of a multiple testing procedure is assumed to be the sum of the individual losses incurred in each test. Letting t_{1i} and t_{2i} denote the probabilities of type I and type II errors of the i th test, respectively, the Bayes risk of a multiple testing procedure under the two-groups model (2) is given by

$$R = \sum_{i=1}^m \{(1-p)t_{1i} + pt_{2i}\}. \tag{15}$$

It was shown in Bogdan et al. (2008) and Bogdan et al. (2011) that the multiple testing rule which minimizes the Bayes Risk in (15) is the test which, for each $i = 1, \dots, m$, rejects H_{0i} if

$$\frac{f(x_i|\nu_i = 1)}{f(x_i|\nu_i = 0)} > \frac{1-p}{p}, \text{ i.e., } X_i^2 > c^2,$$

where $f(x_i|\nu_i = 1)$ denotes the marginal density of X_i under H_{1i} while $f(x_i|\nu_i = 0)$ denotes that under H_{0i} and $c^2 \equiv c_{\psi, f}^2 = \frac{1+\psi^2}{\psi^2}(\log(1 + \psi^2) + 2\log(f))$, with $f = \frac{1-p}{p}$. The above rule is called Bayes Oracle since it makes use of the unknown parameters ψ and p , and hence is not attainable in finite samples. By introducing two new parameters $u = \psi^2$ and $v = uf^2$, the above threshold becomes

$$c^2 \equiv c_{u,v}^2 = (1 + \frac{1}{u})(\log v + \log(1 + \frac{1}{u})).$$

Bogdan et al. (2011) considered the following asymptotic scheme:

Assumption 1. The sequence of vectors (ψ_m, p_m) satisfies the following conditions:

1. $p_m \rightarrow 0$ as $m \rightarrow \infty$.
2. $u_m = \psi_m^2 \rightarrow \infty$ as $m \rightarrow \infty$.
3. $v_m = u_m f^2 = \psi_m^2 \left(\frac{1-p_m}{p_m}\right)^2 \rightarrow \infty$ as $m \rightarrow \infty$.
4. $\frac{\log v_m}{u_m} \rightarrow C \in (0, \infty)$ as $m \rightarrow \infty$.

Under Assumption 1, Bogdan et al. (2011) obtained the following asymptotic expressions of type I and type II error probabilities of the Bayes Oracle given by

$$t_1^{BO} = e^{-C/2} \sqrt{\frac{2}{\pi v \log v}} (1 + o(1)) \text{ and} \quad (16)$$

$$t_2^{BO} = (2\Phi(\sqrt{C}) - 1)(1 + o(1)), \quad (17)$$

and the corresponding optimal Bayes risk was given by

$$R_{Opt}^{BO} = m((1-p)t_1^{BO} + pt_2^{BO}) = mp(2\Phi(\sqrt{C}) - 1)(1 + o(1)). \quad (18)$$

In (16)–(18) above, the $o(1)$ terms tend to zero as $m \rightarrow \infty$.

We want to study asymptotic optimality properties of the multiple testing rules (8) and (10), induced by our general class of one-group tail robust shrinkage priors when applied to data generated from the two-groups model (2), where the hyperparameters (ψ_m, p_m) of the two-groups model satisfy Assumption 1. For simplicity of notation, henceforth we drop the subscript m from p_m , τ_m^2 and ψ_m^2 . For the sake of completeness, we describe below the one-group prior specification for our theoretical analysis:

$$\left. \begin{aligned} \mu_i | (\lambda_i^2, \tau^2) &\stackrel{ind}{\sim} N(0, \lambda_i^2 \tau^2), \text{ for } i = 1, \dots, m, \\ \lambda_i^2 &\stackrel{ind}{\sim} \pi(\lambda_i^2) = K(\lambda_i^2)^{-a-1} L(\lambda_i^2), \text{ for } i = 1, \dots, m, \end{aligned} \right\} \quad (19)$$

where $a > 0$, $K > 0$ and L is a non-constant slowly varying function over $(0, \infty)$. Under (19), the shrinkage coefficients $\kappa_i = 1/(1 + \lambda_i^2 \tau^2)$'s are independently distributed given $(X_1, \dots, X_m, \tau^2)$, with the posterior of κ_i only depending on (X_i, τ^2) and given by

$$\pi(\kappa_i | X_i, \tau) \propto \kappa_i^{a+\frac{1}{2}-1} (1 - \kappa_i)^{-a-1} L\left(\frac{1}{\tau^2} \left(\frac{1}{\kappa_i} - 1\right)\right) e^{-\frac{\kappa_i X_i^2}{2}}, \quad \kappa_i \in (0, 1).$$

3.2 Main theoretical results

In this section, we present in Theorem 1 and Theorem 2 the main theoretical findings of this paper. Theorem 1 gives asymptotic upper and lower bounds to the Bayes risk of the multiple testing procedure (8) under study, when the global shrinkage parameter τ is treated as a tuning parameter, while Theorem 2 gives asymptotic upper bound to the Bayes risk of the empirical Bayes procedure defined in (10). Proofs of Theorem 1 and Theorem 2 are based on some asymptotic bounds for the corresponding type I and

type II error probabilities of the individual decisions in (8) and (10), which, in turn, depend on a set of concentration and moment inequalities. We present these inequalities and the asymptotic bounds on both kinds of error probabilities in Section 4 of this paper. Proofs of Theorem 1 and Theorem 2 are given in the Appendix.

Theorem 1. *Suppose X_1, \dots, X_m are i.i.d. observations having the two-groups normal mixture distribution in (2) with $\sigma^2 = 1$, and we wish to test the m hypotheses $H_{0i} : \nu_i = 0$ vs $H_{1i} : \nu_i = 1$, for $i = 1, \dots, m$, simultaneously, using the decision rule (8) induced by the one-group priors (19). Suppose Assumption 1 is satisfied by the sequence of parameters (ψ^2, p) . Further assume that $\tau \rightarrow 0$ as $m \rightarrow \infty$ such that $\lim_{m \rightarrow \infty} \tau/p \in (0, \infty)$, and $\pi(\lambda_i^2)$ is such that*

$$(I) \quad \frac{1}{2} < a < 1,$$

$$(II) \quad a = \frac{1}{2} \text{ and } L(t)/\sqrt{\log(t)} \rightarrow 0 \text{ as } t \rightarrow \infty.$$

Then, as $m \rightarrow \infty$, the Bayes risk of the multiple testing rules in (8), denoted R_{OG} , satisfies

$$mp[2\Phi(\sqrt{2a}\sqrt{C}) - 1](1 + o(1)) \leq R_{OG} \leq mp\left[2\Phi\left(\sqrt{\frac{2aC}{\eta(1-\delta)}}\right) - 1\right](1 + o(1)) \quad (20)$$

for every fixed $\eta \in (0, \frac{1}{2})$ and $\delta \in (0, 1)$. The $o(1)$ terms above are not necessarily the same, tend to zero as $m \rightarrow \infty$, and depend on the choice of $\eta \in (0, \frac{1}{2})$ and $\delta \in (0, 1)$.

As a consequence of Theorem 1, for a very large class of priors covered by (I) or (II) of the stated theorem, the ratio of the Bayes risk of the induced decisions in (8) to that of the Bayes Oracle (see (18) in Section 3.1) is asymptotically bounded by

$$\frac{2\Phi(\sqrt{2a}\sqrt{C}) - 1}{2\Phi(\sqrt{C}) - 1}(1 + o(1)) \leq \frac{R_{OG}}{R_{Opt}^{BO}} \leq \frac{2\Phi(\sqrt{2a/(\eta(1-\delta))}\sqrt{C}) - 1}{2\Phi(\sqrt{C}) - 1}(1 + o(1)), \quad (21)$$

as $m \rightarrow \infty$, for every fixed $\eta \in (0, \frac{1}{2})$ and every fixed $\delta \in (0, 1)$. That is,

$$R_{OG} = O(R_{Opt}^{BO}) \text{ as } m \rightarrow \infty.$$

For small values of C and appropriately chosen $\eta \in (0, \frac{1}{2})$ and $\delta \in (0, 1)$, the ratios in (21) given above can be made close to 1. Therefore, we see that, in sparse situations when the global shrinkage parameter τ is asymptotically of the order of the proportion of true alternatives p , the decision rules (8), imposed by a very broad class of tail robust one-group priors satisfying (I) or (II) of Theorem 1, asymptotically attain the optimal Bayes risk up to a multiplicative constant, the constant being close to 1. It may be seen that the condition (II) of Theorem 1 is satisfied if, in the prior on the local shrinkage parameter in (19), one has $a = \frac{1}{2}$ and $L(\cdot)$ is, say, uniformly bounded or $\lim_{t \rightarrow \infty} L(t) \in (0, \infty)$. It has already been shown in Section 2 that the horseshoe prior, the Strawderman–Berger prior, members from the families of normal–exponential–gamma priors and generalized double Pareto priors with appropriate choice of (α, β) , satisfy these conditions.

The theoretical results of the forthcoming sections of this paper suggest that, for the above Oracle optimality property to be true, the optimal choice of τ should be such that it is asymptotically of the order of p , that is, $\frac{\tau}{p}$ has a finite, positive limit as the number of tests m grows to infinity. It will be shown further that there are other choices of τ depending on p , for which the desired Oracle optimality up to $O(1)$ may no longer be true. These will be discussed later in a greater detail in Section 4.2 of this paper.

The next theorem gives an asymptotic upper bound for the Bayes risk of the empirical Bayes procedure defined in (10) under the asymptotic framework of Bogdan et al. (2011) together with the assumption that $p \equiv p_m \propto m^{-\epsilon}$ for $0 < \epsilon < 1$. As a consequence, the Oracle optimality property of the empirical Bayes procedure (10) follows immediately. Note that the condition $p \propto m^{-\epsilon}$, where $0 < \epsilon < 1$, is very mild in nature and covers most of the cases of theoretical and practical interest.

Theorem 2. *Suppose X_1, \dots, X_m are i.i.d. observations having the two-groups mixture distribution described in (2) with $\sigma^2 = 1$, and we wish to test the m hypotheses $H_{0i} : \nu_i = 0$ vs $H_{1i} : \nu_i = 1$, $i = 1, \dots, m$, simultaneously, using the decision rule (10) induced by the one-group priors (19). Suppose Assumption 1 is satisfied by (ψ^2, p) with $p \propto m^{-\epsilon}$, for some $0 < \epsilon < 1$. Further assume that in the prior $\pi(\lambda_i^2)$ for the local shrinkage parameter λ_i^2 in (19) satisfies:*

$$(I) \quad \frac{1}{2} < a < 1, \text{ or,}$$

$$(II) \quad a = \frac{1}{2} \text{ and } L(t)/\sqrt{\log(t)} \rightarrow 0 \text{ as } t \rightarrow \infty.$$

Then, the Bayes risk of the multiple testing rules in (10), denoted R_{OG}^{EB} , is bounded above by

$$R_{OG}^{EB} \leq mp \left[2\Phi \left(\sqrt{\frac{2aC}{\eta(1-\delta)}} \right) - 1 \right] (1 + o(1)) \text{ as } m \rightarrow \infty, \quad (22)$$

for every fixed $\eta \in (0, \frac{1}{2})$ and $\delta \in (0, 1)$, where the $o(1)$ term above tends to zero as $m \rightarrow \infty$ and depends on the choice of $\eta \in (0, \frac{1}{2})$ and $\delta \in (0, 1)$.

Now using Theorem 2 it follows immediately that

$$R_{OG}^{EB} = O(R_{Opt}^{BO}) \text{ as } m \rightarrow \infty.$$

As before, for small values of C , and appropriately chosen $\eta \in (0, \frac{1}{2})$ and $\delta \in (0, 1)$, the ratio of risk R_{OG}^{EB}/R_{Opt}^{BO} can be made close to 1.

Using the techniques employed for deriving asymptotic upper bounds for the type I and type II error probabilities of the empirical Bayes decisions in (10), one can show easily that the empirical Bayes estimate $\hat{\tau}$ defined in (9) consistently estimates the unknown degree of sparsity p up to some multiplicative factor. This will be made more precise in Remark 4. As mentioned already that the desired Bayesian optimality property as presented in Theorem 1 holds when τ is asymptotically of the order of p , which seems to be the optimal choice of τ in case p is known. This perhaps explains the good performance of our proposed empirical Bayes procedure using the estimate $\hat{\tau}$ and gives a strong theoretical support in favor of using such a plug-in estimate of τ .

3.3 A comparison with the work of Datta and Ghosh (2013)

A careful inspection of the proof of Theorem 3.4 of Datta and Ghosh (2013) reveals the following. Under Assumption 1, when $\lim_{m \rightarrow \infty} \tau/p \in (0, \infty)$, the Bayes risk of the decision rules (8) induced by the horseshoe prior, denoted $R_{HS(DG)}$, satisfies

$$R_{HS(DG)} \leq mp \left[2\Phi \left(\sqrt{\frac{2C}{\eta(1-\delta)}} \right) - 1 \right] (1 + o(1)) \text{ as } m \rightarrow \infty, \tag{23}$$

for every fixed $\eta \in (0, \frac{1}{2})$ and $\delta \in (0, 1)$. A comparison between the upper bounds in (20) and (23) shows that our results not only generalize the theoretical finding concerning the asymptotic Bayes optimality of the horseshoe prior, but at the same time, sharpen the upper bound to the Bayes risk of the induced decisions under study, for $\frac{1}{2} \leq a < 1$, across the general class of priors given in (19), and satisfying conditions (I) or (II) of Theorem 1, including the horseshoe, in particular.

Although a few ideas employed in the proofs of this paper are similar to those in Datta and Ghosh (2013), our arguments heavily hinge upon appropriate use of properties of slowly varying functions. It will be observed later in this paper that application of well-known properties of slowly varying functions often leads to exact asymptotic orders of certain integrals, without the need to depend mainly on using algebraic upper and lower bounds which can be improved further. In fact, using this technique, we obtain a sharper asymptotic bound to the probability of type II errors and hence on the overall risk (in Theorem 1) as compared to that in Datta and Ghosh (2013). See Remark 2 in this context.

4 Some key inequalities and bounds on probabilities of type I and type II errors

In Section 4.1, we present some concentration and moment inequalities involving the posterior distributions of the shrinkage coefficients κ_i 's. These inequalities are essential for deriving asymptotic bounds for probabilities of type I and type II errors of the multiple testing procedures (8) and (10) under study, presented in Section 4.2 and Section 4.3, respectively. Proofs of all these results are given in the Appendix.

4.1 Concentration and moment inequalities

Before presenting the theoretical results of this section, let us first briefly describe how they can be useful in studying the error probabilities of two kinds. Let t_{1i} and t_{2i} respectively denote the probabilities of type I and type II errors of the i th individual decision in (8). Then, by definition, $t_{1i} = \Pr(E(1 - \kappa_i|X_i, \tau) > \frac{1}{2}|H_{0i} \text{ is true})$ and $t_{2i} = \Pr(E(\kappa_i|X_i, \tau) > \frac{1}{2}|H_{1i} \text{ is true})$. It seems that finding the exact asymptotic orders of t_{1i} and t_{2i} is infeasible. Therefore, one convenient and fruitful approach to study their asymptotic behaviors is to find non-trivial asymptotic bounds for them. One way of accomplishing this is to obtain appropriate bounds for either of $E(1 - \kappa_i|X_i, \tau)$ and

$\Pr(\kappa_i > \eta | X_i, \tau)$ (since $E(\kappa_i | X_i, \tau)$ can be bounded above by $\eta + \Pr(\kappa_i > \eta | X_i, \tau)$, for any $\eta \in (0, 1)$), followed by some judicious applications of these bounds.

The following theorem is our first step towards this and gives the first concentration inequality involving the posterior distribution of κ_i 's. Using this theorem, one can derive an upper bound to $E(1 - \kappa_i | X_i, \tau)$ in a very simple way in case the function L in (19) is bounded above, as indicated in Remark 1 below.

Theorem 3. *Suppose $X_i \sim N(\mu_i, 1)$ independently for $i = 1, \dots, m$. Consider the one-group prior given in (19) and let $\kappa_i = \frac{1}{1 + \lambda_i^2 \tau^2}$. Then, for any fixed $\epsilon \in (0, 1)$ and any fixed $\tau > 0$,*

$$\Pr(\kappa_i < \epsilon | X_i, \tau) \leq K e^{\frac{x_i^2}{2}} \left\{ \int_{\frac{1}{\tau^2}(\frac{1}{\epsilon} - 1)}^{\infty} t^{-a-1} L(t) dt \right\} (1 + o(1)),$$

where the $o(1)$ term above is independent of both the index i and the data point X_i , but depends on τ in such a way that $\lim_{\tau \rightarrow 0} o(1) = 0$.

Corollary 1. *Suppose the function $L(\cdot)$ in (19) is uniformly bounded above by some constant $M > 0$. Then under the assumptions of Theorem 3, for any fixed $\epsilon \in (0, 1)$ and any fixed $\tau > 0$,*

$$\Pr(\kappa_i < \epsilon | X_i, \tau) \leq \frac{KM}{a} \epsilon^a (1 - \epsilon)^{-a} e^{\frac{x_i^2}{2}} \tau^{2a} (1 + o(1)),$$

where the $o(1)$ term above is independent of both the index i and the data point X_i , but depends on τ in such a way that $\lim_{\tau \rightarrow 0} o(1) = 0$.

Remark 1. In case the function $L(\cdot)$ is bounded above by some $M > 0$, then using Corollary 1 one can readily obtain the following upper bound on $E(1 - \kappa_i | X_i, \tau)$:

$$E(1 - \kappa_i | X_i, \tau) = \int_0^1 \Pr(\kappa_i < \epsilon | X_i, \tau) d\epsilon \leq \frac{KM}{a(1-a)} e^{\frac{x_i^2}{2}} \tau^{2a} (1 + o(1)). \quad (24)$$

It has already been shown in Section 2.2 and Section 2.3 that for many of the commonly used shrinkage priors including the horseshoe, the corresponding $L(\cdot)$ is bounded above by some constant M . Using the upper bound from Theorem 3 makes the task of finding an upper bound for $E(1 - \kappa_i | X_i, \tau)$ very simple in such cases. Finding an upper bound for $E(1 - \kappa_i | X_i, \tau)$ in case of a general $L(\cdot)$, as given in Theorem 4 below, is quite non-trivial and requires pretty delicate arguments based on properties of slowly varying functions.

Theorem 4. *Consider the set up of Theorem 3, with the prior on the local shrinkage parameter as in (19) with $a \in (0, 1)$. Then, for every fixed $\tau < 1$,*

$$E(1 - \kappa_i | X_i, \tau) \leq \frac{A_0 K}{a(1-a)} e^{\frac{x_i^2}{2}} \tau^{2a} L\left(\frac{1}{\tau^2}\right) (1 + o(1)), \quad (25)$$

where the $o(1)$ term above is independent of both the index i and the data point X_i , but depends on τ in such a way that $\lim_{\tau \rightarrow 0} o(1) = 0$. Here $A_0 \geq 1$ is a constant depending on L , such that, $L(\cdot)$ is bounded in every compact subset of $[A_0, \infty)$.

The next theorem gives the second concentration inequality of this paper involving the term $\Pr(\kappa_i > \eta | X_i, \tau)$.

Theorem 5. *Under the setup of Theorem 3, for any fixed $\tau > 0$, and each fixed $\eta \in (0, 1)$ and $\delta \in (0, 1)$,*

$$\Pr(\kappa_i > \eta | X_i, \tau) \leq \frac{H(a, \eta, \delta) e^{-\frac{\eta(1-\delta)X_i^2}{2}}}{\tau^{2a} \Delta(\tau^2, \eta, \delta)}, \text{ uniformly in } X_i \in \mathbb{R},$$

$$\text{where } \Delta(\tau^2, \eta, \delta) = \xi(\tau^2, \eta, \delta) L\left(\frac{1}{\tau^2} \left(\frac{1}{\eta\delta} - 1\right)\right),$$

$$\xi(\tau^2, \eta, \delta) = \frac{\int_{\frac{1}{\tau^2} \left(\frac{1}{\eta\delta} - 1\right)}^{\infty} t^{-(a+\frac{1}{2}+1)} L(t) dt}{\left(a + \frac{1}{2}\right)^{-1} \left(\frac{1}{\tau^2} \left(\frac{1}{\eta\delta} - 1\right)\right)^{-(a+\frac{1}{2})} L\left(\frac{1}{\tau^2} \left(\frac{1}{\eta\delta} - 1\right)\right)}, \text{ and}$$

$$H(a, \eta, \delta) = \frac{\left(a + \frac{1}{2}\right)(1 - \eta\delta)^a}{K(\eta\delta)^{\left(a+\frac{1}{2}\right)}}.$$

Remark 2. It is to be observed in this context that, for $0 < a < 1$, the upper bound in Theorem 5 of the present article is of a smaller order compared to that derived in Theorem 3.2 of Datta and Ghosh (2013). In particular, using properties of slowly varying functions (see the Appendix), it can be easily established that the ratio of the former to the latter tends to zero as $\tau \rightarrow 0$. The sharper asymptotic bound in Theorem 5 results in a sharper asymptotic upper bound to the probability of type II error, and hence on the overall risk R_{OG} (in Theorem 1) of the procedure (8) as compared to that in Datta and Ghosh (2013).

Several important features of the posterior distribution of the shrinkage coefficients κ_i 's based on our general class of tail robust shrinkage priors, now become clear from Theorem 3 through Theorem 5. These are listed in Corollary 2 – Corollary 5 given below. While Corollary 2 and Corollary 3 are derived using Theorem 3 and Theorem 4, respectively, the rest follow from Theorem 5. Proofs of these results are trivial and hence are omitted. It should however be remembered that these corollaries have no direct use in proving the main theoretical results of this paper.

Corollary 2. *Under the assumptions of Theorem 3, $\Pr(\kappa_i \geq \epsilon | X_i, \tau) \rightarrow 1$ as $\tau \rightarrow 0$ for any fixed $\epsilon \in (0, 1)$ uniformly in $X_i \in \mathbb{R}$.*

Thus, for each fixed $x \in \mathbb{R}$, the posterior distribution of κ_i 's, based on the tail robust priors under consideration, tend to concentrate near 1 for small values of τ .

Corollary 3. *Under the assumptions of Theorem 3, $E(1 - \kappa_i | X_i, \tau) \rightarrow 0$ as $\tau \rightarrow 0$ for any fixed $\epsilon \in (0, 1)$ uniformly in $X_i \in \mathbb{R}$.*

Corollary 3 above says that for small values of τ , noise observations will be squelched towards the origin by the kind of one-group priors considered in this paper.

Corollary 4. Under the assumptions of Theorem 5, $\Pr(\kappa_i \leq \eta | X_i, \tau) \rightarrow 1$ as $X_i \rightarrow \infty$, for any fixed $\tau > 0$ and every fixed $\eta \in (0, 1)$.

Corollary 5. Under the assumptions of Theorem 5, $E(1 - \kappa_i | X_i, \tau) \rightarrow 1$ as $X_i \rightarrow \infty$, for any fixed $\tau > 0$.

Corollary 5 above shows that, for each of the heavy tailed shrinkage priors under consideration, even if the global variance component τ is very small, the amount of posterior shrinkage will be negligibly small for large X_i 's, thus leaving the large observations almost unshrunk.

4.2 Asymptotic bounds on probabilities of type I and type II errors when τ is treated as a tuning parameter

Theorem 6 and Theorem 7 below give asymptotic upper bounds to the probability of type I error (t_{1i}) and the probability of type II error (t_{2i}) of the i th decision in (8), respectively, while Theorem 8 and Theorem 9 respectively give asymptotic lower bounds for t_{1i} and t_{2i} . As mentioned before, these results lead to the asymptotic bounds on the Bayes risk (R_{OG}) of the multiple testing procedure in (8).

Theorem 6. Suppose X_1, \dots, X_m are i.i.d. observations having the two-groups mixture distribution described in (2) with $\sigma^2 = 1$ and suppose Assumption 1 is satisfied by (ψ^2, p) . Suppose one is testing $H_{0i} : \nu_i = 0$ vs $H_{1i} : \nu_i = 1$ using the decision rule (8) induced by the general class of one-group shrinkage priors (19) where $a \in (0, 1)$ in $\pi(\lambda_i^2)$. Suppose $\tau = \tau_m \rightarrow 0$ as $m \rightarrow \infty$. Then the probability t_{1i} of type I error of the i th decision in (8) satisfies

$$t_1 \equiv t_{1i} \leq \frac{1}{\sqrt{\pi a}} \cdot \frac{2A_0 K}{a(1-a)} \cdot \frac{\tau^{2a} L(\frac{1}{\tau^2})}{\sqrt{\log(\frac{1}{\tau^2})}} (1 + o(1)) \text{ as } m \rightarrow \infty,$$

where the $o(1)$ term above does not depend on i and tends to zero as $m \rightarrow \infty$. The constant A_0 has already been defined in Theorem 4.

Theorem 7. Consider the set-up of Theorem 6 but allow the parameter a to be any positive real number in the definition of the prior $\pi(\lambda_i^2)$ of the local shrinkage parameter in (19). Assume further that $\tau = \tau_m \rightarrow 0$ as $m \rightarrow \infty$ in such a way that $\lim_{m \rightarrow \infty} \frac{\tau}{p} \in (0, \infty)$. Then the probability t_{2i} of type II error of the i th decision in (8) satisfies

$$t_2 \equiv t_{2i} \leq \left[2\Phi\left(\sqrt{\frac{2aC}{\eta(1-\delta)}}\right) - 1 \right] (1 + o(1)) \text{ as } m \rightarrow \infty,$$

for every fixed $\eta \in (0, \frac{1}{2})$ and $\delta \in (0, 1)$. Here the $o(1)$ term above depends on $\eta \in (0, \frac{1}{2})$ and $\delta \in (0, 1)$ and is independent of i , and tends to zero as $m \rightarrow \infty$.

Remark 3. We record here that, instead of $\lim_{m \rightarrow \infty} \tau/p \in (0, \infty)$, if we assume $\log \tau \sim \log p$, keeping the other conditions unaltered, the proof of Theorem 7 goes through. Consequently, the upper bound on the Bayes risk R_{OG} in Theorem 1, which has been

derived by combining the results of Theorem 6 and Theorem 7, also holds when $\log \tau \sim \log p$, under conditions specified in (I). A similar inspection shows that under conditions specified in (II), the upper bound of Theorem 1 holds when the conditions $\log \tau \sim \log p$ and $\tau = O(p)$ both hold as $m \rightarrow \infty$.

Theorem 8. *Consider the set-up of Theorem 6. Let us fix any $0 < \eta < 1/2$ and any $0 < \delta < 1$. Then the probability t_{1i} of type I error of the i th decision in (8) satisfies*

$$t_1 \equiv t_{1i} \geq \frac{(\frac{1}{2} - \eta)/\sqrt{\pi a}}{H(a, \eta, \delta)} \cdot \frac{\tau^{\frac{2\alpha}{\eta(1-\delta)}} L(\frac{1}{\tau^2})}{\sqrt{\log(\frac{1}{\tau^2})}} (1 + o(1)) \text{ as } m \rightarrow \infty,$$

where the $o(1)$ term above depends on $\eta \in (0, \frac{1}{2})$ and $\delta \in (0, 1)$ and is independent of i , and tends to zero as $m \rightarrow \infty$. The constant $H(a, \eta, \delta)$ has already been defined in Theorem 5.

Theorem 9. *Consider the set-up of Theorem 7. Then the probability t_{2i} of type II error of the i th decision in (8) satisfies*

$$t_2 \equiv t_{2i} \geq (2\Phi(\sqrt{2a\sqrt{C}}) - 1)(1 + o(1)) \text{ as } m \rightarrow \infty,$$

where the $o(1)$ term above does not depend on i and tends to zero as $m \rightarrow \infty$.

Some important observations regarding an appropriate choice of τ now follow as consequences of Theorem 6 – Theorem 9. Note that, the type I and type II error probabilities of the i th decision in (8), that is, t_{1i} and t_{2i} , do not depend on i and their common values are given by t_1 and t_2 , respectively. See the proofs of Theorem 6 and Theorem 7 in the Appendix for an explanation of this fact. Thus, the Bayes risk of the decision rules in (8) is given by $R_{OG} = mp(\frac{1-p}{p}t_1 + t_2)$ (using (15)). Suppose now $\tau \rightarrow 0$ at such a rate that

$$\frac{(1-p)\tau^{\frac{2\alpha}{\eta(1-\delta)}} L(\frac{1}{\tau^2})}{p\sqrt{\log(\frac{1}{\tau^2})}} \rightarrow \infty \text{ as } m \rightarrow \infty. \tag{26}$$

Then combining Theorem 8 and Theorem 9 together, it follows that $\frac{1-p}{p}t_1 + t_2 \rightarrow \infty$ as $m \rightarrow \infty$. Consequently, $R_{OG}/R_{Opt}^{BO} \rightarrow \infty$ as $m \rightarrow \infty$. Consider, for example, the horseshoe or the standard double Pareto prior. For each of these priors, one has $a = 0.5$ and the corresponding $L(\cdot)$ has a finite positive limit at infinity as already shown before. Let us take $\tau = p^\alpha$, for $\alpha > 0$. Now, for $0 < \alpha < \frac{1}{2}$, one can always choose some $\eta \in (0, \frac{1}{2})$ and some $\delta \in (0, 1)$, such that $0 < \alpha < \eta(1-\delta) < \frac{1}{2}$. As a result, (26) holds and we have $R_{OG}/R_{Opt}^{BO} \rightarrow \infty$ as $m \rightarrow \infty$, when $0 < \alpha < \frac{1}{2}$. Thus the desired Bayesian optimality property up to $O(1)$ no longer holds in such situations. However, from the derived lower bounds, we cannot yet conclude the same if $\alpha \geq \frac{1}{2}$ since, in that case, the quantity in (26) tends to zero as $m \rightarrow \infty$. But, the upper bound for $\frac{1-p}{p}t_1 + t_2$, as obtained by combining Theorem 6 and Theorem 7, tends to infinity as $m \rightarrow \infty$, when $\frac{1}{2} \leq \alpha < 1$. This indicates, though not conclusively, that, for such tail robust priors, $\tau = p^\alpha$ with $0 < \alpha < 1$ is not likely to be a good choice.

It should be noted further that the proofs of Theorem 7 and Theorem 9 work even if we take $\tau = p^\alpha$, for $\alpha \geq 1$ (see Appendix), and the constants in the corresponding asymptotic bounds should be replaced by $(2\Phi(\sqrt{2a\alpha C}/(\eta(1-\delta))) - 1)$ and $(2\Phi(\sqrt{2a\alpha C}) - 1)$, respectively. Each of these bounds increases with an increase in α , and so does the corresponding bound on the overall Bayes risk R_{OG} . Thus, R_{OG} tends to be away from the optimal Bayes risk R_{Opt}^{BO} in (18), for values of $\alpha > 1$, thereby implying that $\tau = p^\alpha$ is also not a good choice for $\alpha > 1$. Note that all these arguments remain valid even if we assume τ/p^α has a finite positive limit as $m \rightarrow \infty$. Thus our results give a partial indication that $\tau = p$ (or, $\lim_{m \rightarrow \infty} \tau/p \in (0, \infty)$) should be the optimal choice of τ . Similar observations were also made by van der Pas et al. (2014) for the horseshoe prior when $p = o(1)$. They found that for optimal contraction of the corresponding posterior distribution around the truth as well as the horseshoe estimator, the optimal choice for τ should be $\tau = p$ (or, up to some logarithmic factor of it). Moreover, they showed that when $\tau = p^\alpha$, the posterior distribution based on the horseshoe prior, contracts around the horseshoe estimator at a sub-optimal rate in the squared l_2 sense if $0 < \alpha < 1$, while it contracts too quickly to yield an adequate measure of uncertainty when $\alpha > 1$. Our results are, therefore, in a partial agreement with that of van der Pas et al. (2014) regarding the optimal choice of τ , when p is assumed to be known.

4.3 Asymptotic bounds on probabilities of type I and type II errors for the empirical Bayes procedure

In Theorem 10 and Theorem 11 below, we present asymptotic upper bounds to the probabilities of type I and type II errors of the individual decisions corresponding to the empirical Bayes procedure in (10). Proofs of these theorems are significantly different from those of Theorem 6 and Theorem 7, when τ is treated as a tuning parameter. This is so because for each i , $E(1 - \kappa_i | X_i, \hat{\tau})$ depends on the data (X_1, \dots, X_m) through $\hat{\tau} \equiv \hat{\tau}(X_1, \dots, X_m)$ and X_i in a complicated manner. Therefore, in order to avoid dealing with $E(1 - \kappa_i | X_i, \hat{\tau})$ directly, we employ some nontrivial and novel arguments as follows. First, we introduce a dummy parameter that turns out to be asymptotically of the order of p , and then using this new parameter, we divide the range of $\hat{\tau}$ into two mutually disjoint parts. For one part, as noted by van der Pas et al. (2014), we use the fact that, for each fixed $x \in \mathbb{R}$, $E(1 - \kappa_i | x, \tau)$ is non-decreasing in τ , and then we use the results of Theorem 6 and Theorem 7, while for the other part, we exploit the structure of $\hat{\tau}$, together with the independence among the X_i 's.

Theorem 10. *Suppose X_1, \dots, X_m are i.i.d. observations having the two-groups mixture distribution described in (2) with $\sigma^2 = 1$, and we wish to test the m hypotheses $H_{0i} : \nu_i = 0$ vs $H_{1i} : \nu_i = 1$, $i = 1, \dots, m$, simultaneously, using the decision rule (10) induced by the one-group priors (19) where $a \in (0, 1)$ in $\pi(\lambda_i^2)$. Suppose Assumption 1 is satisfied by (ψ^2, p) with $p \propto m^{-\epsilon}$, for some $0 < \epsilon < 1$. Then, the probability \tilde{t}_{1i} of type I error of the i th induced decision in (10) satisfies*

$$\tilde{t}_{1i} \leq B_1^* \frac{\alpha_m^{2a} L(\frac{1}{\alpha_m^2})}{\sqrt{\log(\frac{1}{\alpha_m^2})}} (1 + o(1)) + \frac{1/\sqrt{\pi}}{m^{c_1/2} \sqrt{\log m}} + e^{-2(2 \log 2 - 1)\beta m p(1+o(1))} \text{ as } m \rightarrow \infty,$$

where the $o(1)$ terms appearing above are independent of i , and tend to zero as $m \rightarrow \infty$. Here B_1^* and β are some finite positive constants, each being independent of m , while $\alpha_m = \Pr(|X_1| > \sqrt{c_1 \log m})$ depends on m .

Theorem 11. *Let us consider the set-up of Theorem 10. Then the probability \tilde{t}_{2i} of type II error of the i th decision in (10) satisfies*

$$\tilde{t}_{2i} \leq \left[2\Phi\left(\sqrt{\frac{2aC}{\eta(1-\delta)}}\right) - 1 \right] (1 + o(1)) \text{ as } m \rightarrow \infty,$$

for every fixed $\eta \in (0, \frac{1}{2})$ and $\delta \in (0, 1)$. Here the $o(1)$ term tends to zero as $m \rightarrow \infty$ and is independent of i , but depends on the choices of $\eta \in (0, \frac{1}{2})$ and $\delta \in (0, 1)$.

Remark 4. Using the architecture of the proofs of Theorem 10 and Theorem 11, one can show now that

$$\frac{\hat{\tau}}{p} \xrightarrow{p} 2\beta \text{ as } m \rightarrow \infty, \tag{27}$$

where the above probability convergence is taken with respect to the joint distribution of X_i 's defined through (2). Since $\alpha_m \sim 2\beta p$ and $\beta = 1 - \Phi(c_1 C / (2\epsilon)) > 0$ (see the proof of Theorem 10), it will be enough to show that, given any $\delta_0 > 0$, $\Pr(|\frac{\hat{\tau}}{\alpha_m} - 1| > \delta_0) = o(1)$. This follows quite easily using the techniques used for proving Theorem 10 and Theorem 11. Thus, (27) says that the estimator $\hat{\tau}$ will be asymptotically of the order of the proportion of true alternatives p . Note that when $c_1 = 2$ and C/ϵ is small, $2\beta \approx 1$. In that case, $\hat{\tau}$ is expected to be very close to the unknown degree of sparsity p with a large probability.

5 Simulations

In this section, we present and interpret the results obtained in our simulation study. The simulation study has several objectives. The first objective is to motivate the use of the decision rules based on the global-local tail robust priors when data actually come from a two-groups model. Secondly, we want to study empirically the suitability of these priors for handling sparsity as well as their robustness in handling large signals. Thirdly, we want to study the role of τ as a global shrinkage parameter. The most important objective is to compare the simulation averages of the proportion of misclassified hypotheses (as estimate of the misclassification probability) of these testing rules with that of the Bayes Oracle for the two-groups problem to understand how closely these rules actually perform vis-a-vis the Oracle.

We present in this section the numerical results obtained by using the horseshoe prior, the standard double Pareto prior, the Strawderman–Berger prior, and the normal–exponential–gamma prior (with $\alpha = 1, \beta = 0.6$) when the data actually come from a two-groups model. We consider a fully Bayesian approach as well as an empirical Bayes approach. For the fully Bayesian approach, (τ, σ) is assigned a hyperprior given by

$$\tau \sim C^+(0, 1) \text{ and } \pi(\sigma) \propto \frac{1}{\sigma},$$

and the marginal prior for μ_i 's is obtained by mixing further with respect to this joint prior distribution of (τ, σ) . For the empirical Bayes approach, σ is taken to be equal to 1, and we use the procedure in (10), where we take $c_1 = 2$ and $c_2 = 1$ in the definition of $\hat{\tau}$ in (9).

Our simulation data are generated as follows. For each fixed $p \in (0, 1)$, we draw $m = 200$ independent observations X_1, \dots, X_m using the two groups model (2), with $\psi_m = \sqrt{2 \log m} = 3.26$ and $\sigma^2 = 1$. For estimating the misclassification probability, the process is replicated 1000 times and simulation averages of misclassification proportions are taken as estimates of the misclassification probabilities of the different multiple testing procedures under study. Our results lend support to our theoretical findings and also justification for our theoretical study presented earlier in Section 3 and Section 4.

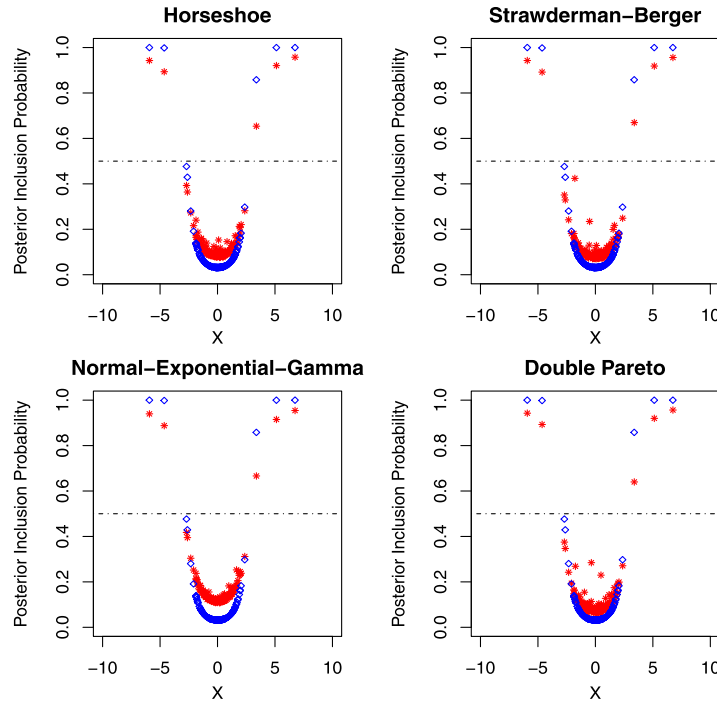


Figure 1: Comparison between the posterior inclusion probabilities and the posterior shrinkage weights $1 - E(\kappa_i | X_1, \dots, X_m)$ when $p = 0.10$.

Taking $p = 0.10$, we plot in Figure 1, the theoretical posterior inclusion probabilities $w_i(X_i) = P(\nu_i = 1 | X_i)$ for the two-groups model (2) given by

$$\omega_i(X_i) = \pi(\nu_i = 1 | X_i) = \left\{ \left(\frac{1-p}{p} \right) \sqrt{1 + \psi^2} e^{-\frac{X_i^2}{2} \frac{\psi^2}{1+\psi^2}} + 1 \right\}^{-1},$$

along with the shrinkage weights $(1 - E(\kappa_i | X_1, \dots, X_m))$ corresponding to the four one-group shrinkage priors mentioned above against the data. The blue dots in the figure

denote the theoretical posterior inclusion probabilities while the red dots correspond to the shrinkage weights $(1 - E(\kappa_i|X_1, \dots, X_m))$. The figures clearly show the proximity of the two quantities for small values of the sparsity parameter p for each of the four shrinkage priors mentioned above. This fact and the theoretical observations made in Section 2 justify the use of $(1 - E(\kappa_i|X_1, \dots, X_m))$ as an approximation to the corresponding posterior inclusion probabilities $\omega_i(X_i)$ in sparse situations and thus motivates the use of decision rules based on $(1 - E(\kappa_i|X_1, \dots, X_m))$ using one-group tail robust shrinkage priors.

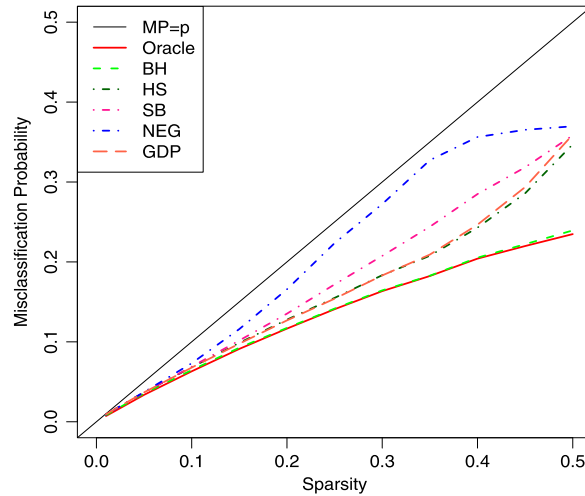


Figure 2: Estimated misclassification probabilities for the full Bayes approach.

Figure 2 shows the (estimated) misclassification probability (MP) plots of the decision rule (11) corresponding to the four priors under consideration along with those of the Bayes Oracle and the Benjamini–Hochberg rule against values of p lying in the set $\{0.01, 0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45, 0.5\}$. As mentioned before, these are obtained as average values of misclassification proportions of the decision rule (11) and that corresponding to the Bayes Oracle (defined in Section 3.1). The Bayes Oracle serves as the lower bound to the MP whereas the line $MP = p$ corresponds to the situation when we reject all null hypotheses without looking into the data. It is clear from Figure 2 that when the sparsity parameter p is small, the MP plots corresponding to the four priors under consideration almost coincide with that of the Bayes Oracle which is in conformity with the theoretical results of the present article. While the MP plots of the horseshoe and the standard double Pareto prior are nearly identical, the MP plot corresponding to the the Strawderman–Berger prior is in close proximity. The same is true for the MP plot for the normal–exponential–gamma plot, for small values of p . When p is larger, say above 0.4, performance of each of these priors become inferior compared to the Bayes Oracle. We also plot the MP for the Benjamini–Hochberg rule, for $\alpha = 1/\log m = 0.1887$. Bogdan et al. (2011) theoretically established that for such choices of α , the corresponding Benjamini–Hochberg rule becomes ABOS in the present

set up. This is also corroborated by Figure 2 where we see that the Benjamini–Hochberg rule achieves practically the same MP as the Bayes Oracle. Similar phenomenon can also be observed for the empirical Bayes procedure (10) in Figure 3. However, Figure 3 also shows that the normal–exponential–gamma prior (with $\alpha = 1$ and $\beta = 0.6$) performs significantly better in terms of the overall MP in the empirical Bayes approach compared to its full Bayes counterpart in Figure 2.

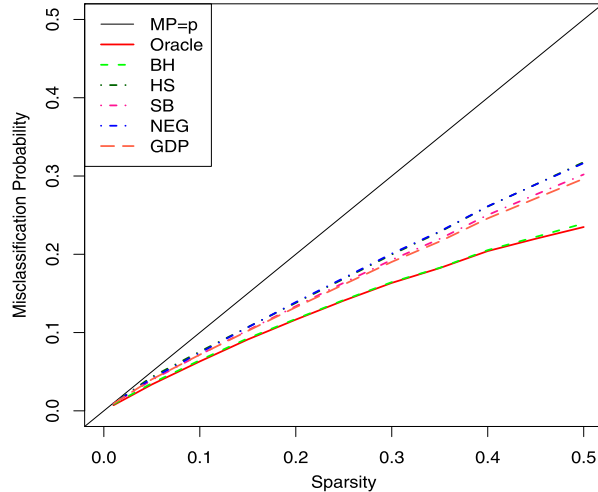


Figure 3: Estimated misclassification probabilities for the empirical Bayes approach.

Shrinkage properties corresponding to the four tail-robust priors under consideration along with the Laplace and the half-normal priors having exponential tails are demonstrated through Figure 4. Here we plot the posterior expectations $E(\mu_i | X_1, \dots, X_m)$ against different values (X_i) of the observations. Figure 4 clearly shows that the noise observations are shrunk towards zero efficiently while the big signals are left mostly unshrunk by each of the four tail robust shrinkage priors under consideration. That is, for each of the four tail robust priors under consideration, the amount of posterior shrinkage (that is, the difference between the 45° straight line and the posterior expectation curve) is negligibly small for observations lying on the two extremes of the horizontal axis, while the difference becomes more prominent for observations near the origin. This corroborates our theoretical findings given in Corollary 2 – Corollary 5 regarding the nature of the posterior distribution of the shrinkage coefficients κ_i 's based on the general class of one-group priors under study. However, this is clearly not the case for the normal and the Laplace priors since these priors shrink even the large signals by some non-diminishing amounts which is evident from Figure 4. Similar observations were also made in Datta and Ghosh (2013) about the Laplace and the half-normal priors.

Finally, we demonstrate in Figure 5 how the global shrinkage parameter τ adapts to the sparsity level of the data in the full Bayes approach. We draw box-plots of the posterior draws for τ across different levels of sparsity for each of the four priors in focus. It may be observed that box-plots are highly concentrated near zero for small values

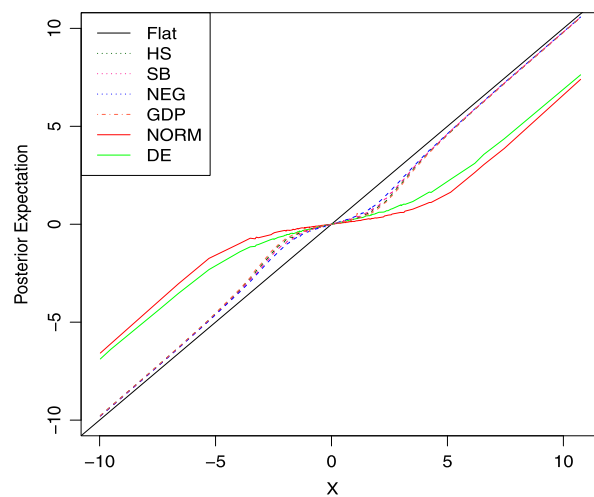


Figure 4: Posterior Mean $E(\mu|X)$ versus X plot for $p = 0.25$.

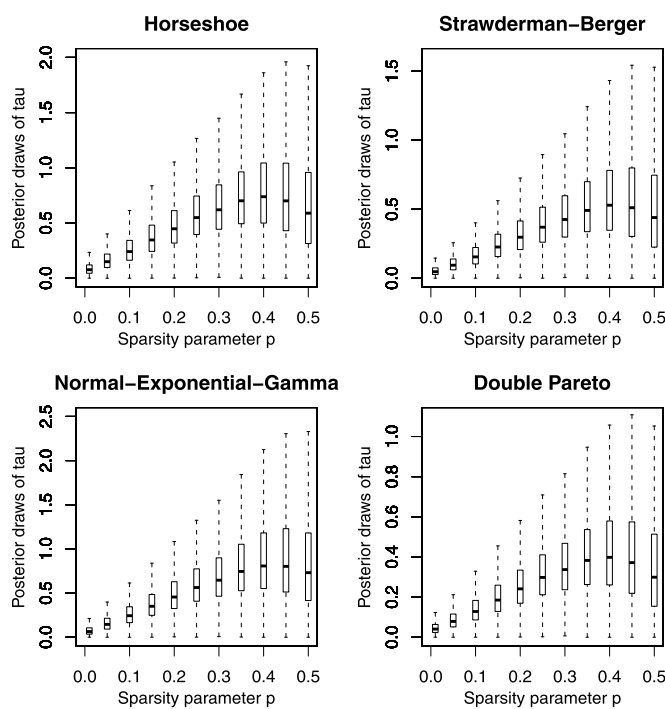


Figure 5: Posterior draws of τ for the horseshoe, the Strawderman-Berger, the normal-exponential-gamma and the standard double Pareto priors across different levels of sparsity.

of p while as p gets large, the median of the posterior samples of τ tends to increase systematically. Thus, the overall sparsity level of the data is reflected well through the posterior distribution of the global shrinkage parameter τ . Moreover, the ratio of the posterior draws of τ to p seems to lie within a bounded factor as can be seen from Figure 5. Similar phenomenon was observed in Datta and Ghosh (2013) in the context of the horseshoe prior. This indicates how global shared parameters can control error rates in multiple testing by estimating the overall sparsity level, as already discussed in Scott and Berger (2006) and Carvalho et al. (2010).

6 Discussion

We have considered in this paper multiple hypothesis testing under sparsity in a decision theoretic framework. Global-local shrinkage priors are used towards this end. We have proved an Oracle property of the resulting decision rules similar to those of Datta and Ghosh (2013). We have also considered an empirical Bayes version of the induced decisions by using an estimate of the global shrinkage parameter τ and shown its Oracle optimality property under very mild restriction over the sparsity parameter p . One of the salient features of our work is that we have provided unified results for a very general class of shrinkage priors including some of the commonly used priors such as the horseshoe prior, generalized double Pareto priors, normal-exponential-gamma priors, and many others. As a special case of our general result, we have strengthened the optimality result for the horseshoe prior as considered in Datta and Ghosh (2013). Further, we have settled a conjecture of these authors related to generalized double Pareto priors. Our technique of proof shows that some of the arguments of Datta and Ghosh (2013) can be simplified. Moreover, in the theoretical treatment of this paper, we have exploited properties of slowly varying functions to obtain a general unifying argument that works across a large class of one-group priors for investigating their theoretical properties, which to the best of our knowledge, has not been done before and can be very useful in other contexts.

The Oracle optimality property of the multiple testing rules studied in this paper assumes σ^2 to be known and treats τ as a tuning parameter or it is estimated from the data. A natural question is whether these decision rules retain this optimality property if we use a full Bayes approach by assigning a hyperprior to (τ, σ) . Thus one would like to know whether the decision rules in (11) enjoy similar Bayesian optimality property. Our simulation results indicate that this is indeed the case, though giving a formal theoretical justification for the same is difficult. To elucidate this point, let us first take a look at the posterior means of μ_i 's based on which the full Bayes decision rules in (11) are defined. For each $i = 1, \dots, m$, it is given by

$$\begin{aligned} E(\mu_i | X_1, \dots, X_m) &= \int_0^\infty \int_0^\infty E(\mu_i | X_1, \dots, X_m, \tau, \sigma) \pi(\tau, \sigma | X_1, \dots, X_m) d\tau d\sigma \\ &= (1 - E(\kappa_i | X_1, \dots, X_m)) X_i \end{aligned}$$

where $E(\kappa_i | X_1, \dots, X_m) = E_{(\tau, \sigma | X_1, \dots, X_m)} E(\kappa_i | X_1, \dots, X_m, \tau, \sigma)$. Here, the posterior distribution of κ_i depends on the entire dataset and not only on X_i , which is the case

when τ is treated as a tuning parameter and σ^2 is assumed to be known. Also, for each i , the posterior distribution $\pi(\kappa_i|X_1, \dots, X_m)$ is analytically quite intractable in this case. Therefore, finding estimates or asymptotic bounds on the type I and type II error measures directly using this posterior distribution does not look feasible. By looking at the expression for $E(\kappa_i|X_1, \dots, X_m)$ given above, one may however wish to explore for possibilities of using the concentration and moment inequalities based on $E(\kappa_i|X_1, \dots, X_m, \tau, \sigma)$ first (as given in Section 4.2), and then finding suitable bounds for $E(\kappa_i|X_1, \dots, X_m)$ using $\pi(\tau, \sigma|X_1, \dots, X_m)$. But that also seems very non-trivial as we do not have any proper handle over $\pi(\tau, \sigma|X_1, \dots, X_m)$. The problem does not become any easier even if we assume σ^2 to be known and put a hyperprior for τ only. It seems that one needs to invoke significantly new techniques and arguments for a full Bayes treatment of the type of one-group priors studied in this paper. We leave this as an interesting problem for future research.

We expect that Oracle optimality properties like those studied in this paper, should now easily be obtained even if one considers $\pi(\lambda_i^2) \sim (\lambda_i^2)^{-a-1}L(\lambda_i^2)$ as $\lambda_i^2 \rightarrow \infty$, as originally considered in Theorem 1 of Polson and Scott (2011), by employing the techniques given in this paper together with the fact that the ratio of these two functions belongs to a small neighborhood of 1 not containing the origin, for all sufficiently large λ_i^2 . It may be commented at this point that we have provided a set of sufficient conditions that are quite general in nature under which such Bayesian optimality results hold true. However, characterizing priors of this kind having such optimality property is quite challenging and remains an interesting and open problem till date.

Our results bear the potential of application also in variable selection. In particular, we want to examine how to extend the proposed decision rule for selection of regression parameters and study optimality of such decision rules in that context. It has been mentioned earlier that van der Pas et al. (2014) considered the problem of estimation of a sparse multivariate normal mean vector using the horseshoe estimator. They showed among other things that the horseshoe estimator achieves (up to a multiplicative factor) the minimax squared error risk and the corresponding posterior distribution contracts around the truth at the minimax rate. A similar optimality result was proved for an empirical Bayes version of the horseshoe estimator. The thing to be noted is that the “optimal” choice of τ that makes all the good results come through is given by the theoretical proportion of non-zero entries p in the mean vector (up to a logarithmic factor), and choices of τ of the order of p^α for $\alpha > 1$ and $\alpha < 1$ were shown to be sub-optimal for such purposes. It is interesting to note, as already argued in Section 4.2, that a choice of τ which is asymptotically of the order of p seems to be the optimal choice of τ in our case, when p is assumed to be known. Therefore, the choice of τ in the vicinity of p seems to be optimal for the two most important inferential problems. In a recent technical report, Ghosh and Chakrabarti (2015) adopted the framework of van der Pas et al. (2014) and proved generalizations of their results using Bayes estimators coming from a very general class of one-group tail robust priors and taking τ proportional to the fraction p . Proofs of these results crucially exploit the inequalities given in Theorem 4 and Theorem 5 of the present article. It is also worth noting that the empirical Bayes estimate of τ that turned out to be useful in the paper of van der Pas et al. (2014) also turns out to be equally handy in the present multiple testing problem. We have already

commented that a possible reason for this is the fact that the empirical Bayes estimate $\hat{\tau}$ proposed by van der Pas et al. (2014) is asymptotically of the order of p under the present two-groups formulation.

Like Datta and Ghosh (2013) or van der Pas et al. (2014), our theoretical results mostly treat the global shrinkage parameter τ as a tuning parameter. As explained before, a fully Bayesian approach using a hyperprior on τ seems difficult to handle. The difficulty in analyzing the posterior distribution for a full Bayes analysis corresponding to horseshoe-type priors with heavy tails has also been addressed in Bhattacharya et al. (2012, 2014) and Pati et al. (2014). The authors of these articles studied the contraction properties of global-local shrinkage priors and proved their sub-optimal properties in cases where the tails of the distributions decay at an exponential or faster rate. However, they conjectured that heavy tailed priors, such as, the horseshoe, the generalized double Pareto, and the normal-exponential-gamma priors, should have optimal posterior contraction properties. This has already been established in van der Pas et al. (2014) for the horseshoe prior and subsequently by Ghosh and Chakrabarti (2015) for a broader class of one-group tail-robust priors which is rich enough to include, among others, the horseshoe, the generalized double Pareto and the normal-exponential-gamma priors in particular. Results of the last two articles, however, assume τ to be a tuning parameter. On the other hand, Bhattacharya et al. (2012, 2014) and Pati et al. (2014) used full Bayes approaches for studying concentration probabilities around sparse vectors and rate of convergence of the corresponding posterior distributions based on their proposed class of shrinkage priors, referred to as Dirichlet-Laplace (DL)-type priors. Such priors result in a prior distribution for the individual means which is highly peaked near the origin and has thick tails. However, the DL-type priors are fundamentally different from the global-local scale mixture representation of Polson and Scott (2011) and hence cannot be covered by our general class of tail robust one-group priors. For example, Bhattacharya et al. (2014) used a set of scalars $(\phi_1\tau, \dots, \phi_m\tau)$ instead of a single global shrinkage component τ and assigned a *Dirichlet* prior to (ϕ_1, \dots, ϕ_m) , with a common Dirichlet concentration parameter. Moreover, the asymptotics of contraction for the DL-type priors are quite different from those derived in this paper, and hence do not seem to apply to our present asymptotic framework, at least, as far as we can see.

Recently, there has been a growing interest in identifying the “rare and weak” signals and also in estimating the proportion of true signals using a two-groups formulation slightly different from the one considered in this paper. See, for example, Ingster (1997), Donoho and Jin (2004), Meinshausen and Rice (2006), Cai et al. (2007) and Cai and Jin (2010), in this context. As suggested by one of the referees, another interesting and related problem to consider is when both the mean and the variance of the signal density vary with the number of tests m under such two-groups framework. For example, people have considered the case when the mean tends to infinity at a slower rate, but the variance remains bounded as m goes to infinity. However, in order to look at the feasibility of any one-group formulation, one needs to first carefully understand what kind of shrinkage priors, if any, would possibly be a good approximation to the two-groups formulation in such contexts. We hope to address these issues elsewhere in future.

Appendix

A.1 Some important properties of slowly varying functions

The theoretical results derived in this paper depend heavily up on some fundamental properties of general slowly varying functions. These are listed as Lemmas 1 – 4 given below. Interested readers are referred to the classic text of Bingham et al. (1987) for a detailed treatment of these results.

Lemma 1. *If L is any slowly varying function and $\alpha < -1$, then*

$$\frac{\int_x^\infty t^\alpha L(t) dt}{x^{\alpha+1} L(x)} \sim \frac{-1}{\alpha + 1} \text{ as } x \rightarrow \infty.$$

Lemma 2. *If L is any slowly varying function then there exists $A_0 > 0$ such that L is locally bounded in $[A_0, \infty)$, that is, L is bounded in all compact subsets of $[A_0, \infty)$.*

Lemma 3. *If L is any slowly varying function, A_0 is so large such that L is locally bounded in $[A_0, \infty)$ and $\alpha > -1$, then*

$$\frac{\int_{A_0}^x t^\alpha L(t) dt}{x^{\alpha+1} L(x)} \sim \frac{1}{1 + \alpha} \text{ as } x \rightarrow \infty.$$

Lemma 4. *If L is any slowly varying function then*

(i) $\lim_{x \rightarrow \infty} \frac{\log L(x)}{\log(x)} = 0,$

(ii) L^β is slowly varying for all $\beta \in \mathbb{R}$, and,

(iii) $\lim_{x \rightarrow \infty} x^{-\alpha} L(x) = 0,$ for all $\alpha > 0.$

A.2 Proofs

This section contains the proofs of all the major theoretical results of this paper. But before proving the main theoretical results given in Section 3, namely, Theorem 1 and Theorem 2, we shall first present the proofs of the results of Section 4 since these results are essential for deriving Theorem 1 and Theorem 2. But above all, let us first prove the following lemma, which we shall use often for proving rest of the results of this paper.

Lemma 5. *Let $L : (0, \infty) \rightarrow (0, \infty)$ be a measurable function and a be a real number such that $\int_0^\infty t^{-a-1} L(t) dt = K^{-1}, K \in (0, \infty)$. Then, assuming $\tau \rightarrow 0,$*

$$\int_0^1 u^{a+\frac{1}{2}-1} (1-u)^{-a-1} L\left(\frac{1}{\tau^2} \left(\frac{1}{u} - 1\right)\right) du = K^{-1} \tau^{-2a} (1 + o(1)),$$

where the $o(1)$ term above depends on τ in such a way that $\lim_{\tau \rightarrow 0} o(1) = 0.$

Proof. Let $J = \int_0^1 u^{a+\frac{1}{2}-1}(1-u)^{-a-1}L(\frac{1}{\tau^2}(\frac{1}{u}-1))du$. Since the integrand in J is non-negative, using the change of variable $t = \frac{1}{\tau^2}(\frac{1}{u}-1)$, one has

$$J = (\tau^2)^{-a} \int_0^\infty (1+t\tau^2)^{-\frac{1}{2}}t^{-a-1}L(t)dt.$$

Since $t^{-a-1}L(t)$ is assumed to be integrable, the proof follows immediately using Lebesgue’s Dominated Convergence Theorem, where the $o(1)$ term is such that $\lim_{\tau \rightarrow 0} o(1) = 0$. □

Proof of Theorem 3. Fix any $\epsilon \in (0, 1)$. Then by definition,

$$\begin{aligned} \Pr(\kappa_i < \epsilon | X_i, \tau) &= \frac{\int_0^\epsilon \kappa_i^{a+\frac{1}{2}-1}(1-\kappa_i)^{-a-1}L(\frac{1}{\tau^2}(\frac{1}{\kappa_i}-1))e^{-\frac{\kappa_i X_i^2}{2}}d\kappa_i}{\int_0^1 \kappa_i^{a+\frac{1}{2}-1}(1-\kappa_i)^{-a-1}L(\frac{1}{\tau^2}(\frac{1}{\kappa_i}-1))e^{-\frac{\kappa_i X_i^2}{2}}d\kappa_i} \\ &\leq \frac{e^{-\frac{X_i^2}{2}} \int_0^\epsilon \kappa_i^{a+\frac{1}{2}-1}(1-\kappa_i)^{-a-1}L(\frac{1}{\tau^2}(\frac{1}{\kappa_i}-1))d\kappa_i}{\int_0^1 \kappa_i^{a+\frac{1}{2}-1}(1-\kappa_i)^{-a-1}L(\frac{1}{\tau^2}(\frac{1}{\kappa_i}-1))d\kappa_i} \end{aligned}$$

which follows from the fact that $e^{-\frac{X_i^2}{2}} \leq e^{-\frac{\kappa_i X_i^2}{2}} \leq 1$ for every $\kappa_i \in (0, 1)$. Now using the change of variable $t = \frac{1}{\tau^2}(\frac{1}{\kappa_i}-1)$ to the numerator of the right hand side of the above inequality and applying Lemma 5 to the corresponding denominator, we obtain

$$\Pr(\kappa_i < \epsilon | X_i, \tau) \leq K e^{-\frac{X_i^2}{2}} \left\{ \int_{\frac{1}{\tau^2}(\frac{1}{\epsilon}-1)}^\infty t^{-a-1}L(t)dt \right\} (1 + o(1)).$$

Here the $o(1)$ term is independent of X_i and tends to zero as $\tau \rightarrow 0$. This completes the proof of Theorem 3. □

Proof of Theorem 4. Since L is slowly varying and $a \in (0, 1)$, by Lemma 3, there exists some $A_0 > 0$ such that L is locally bounded on $[A_0, \infty)$ and

$$\lim_{z \rightarrow \infty} \frac{\int_{A_0}^z t^{-a}L(t)dt}{z^{1-a}L(z)} = \frac{1}{1-a}. \tag{28}$$

Without any loss of generality, one may assume that $A_0 \geq 1$. Now observe that from the definition of $E(1 - \kappa_i | X_i, \tau)$, it directly follows that

$$\begin{aligned} E(1 - \kappa_i | X_i, \tau) &= \frac{\int_0^1 \kappa_i^{a+\frac{1}{2}-1}(1-\kappa_i)^{-a}L(\frac{1}{\tau^2}(\frac{1}{\kappa_i}-1))e^{-\frac{\kappa_i X_i^2}{2}}d\kappa_i}{\int_0^1 \kappa_i^{a+\frac{1}{2}-1}(1-\kappa_i)^{-a-1}L(\frac{1}{\tau^2}(\frac{1}{\kappa_i}-1))e^{-\frac{\kappa_i X_i^2}{2}}d\kappa_i} \\ &\leq \frac{e^{-\frac{X_i^2}{2}} \int_0^1 \kappa_i^{a+\frac{1}{2}-1}(1-\kappa_i)^{-a}L(\frac{1}{\tau^2}(\frac{1}{\kappa_i}-1))d\kappa_i}{\int_0^1 \kappa_i^{a+\frac{1}{2}-1}(1-\kappa_i)^{-a-1}L(\frac{1}{\tau^2}(\frac{1}{\kappa_i}-1))d\kappa_i}. \end{aligned} \tag{29}$$

Then, using the change of variable $t = \frac{1}{\tau^2}(\frac{1}{\kappa_i} - 1)$ to the numerator of (29) and applying Lemma 5 to its denominator, we obtain the following:

$$\begin{aligned} E(1 - \kappa_i | X_i, \tau) &\leq \frac{e^{\frac{x_i^2}{2}}(\tau^2)^{-a+1} \int_0^\infty (1+t\tau^2)^{-\frac{3}{2}} t^{-a} L(t) dt}{K^{-1}\tau^{-2a}(1+o(1))} \\ &= Ke^{\frac{x_i^2}{2}} \tau^2 \int_0^\infty (1+t\tau^2)^{-\frac{3}{2}} t^{-a} L(t) dt (1+o(1)) \\ &= KJ_\tau e^{\frac{x_i^2}{2}} (1+o(1)), \text{ say,} \end{aligned} \tag{30}$$

where $J_\tau = \tau^2 \int_0^\infty (1+t\tau^2)^{-\frac{3}{2}} t^{-a} L(t) dt$ and the $o(1)$ term in (30) does not depend on X_i nor on the index i , and tends to zero as $\tau \rightarrow 0$.

Now we observe that for any $\tau < 1$, we can split J_τ as

$$J_\tau = \left(\int_0^{A_0} + \int_{A_0}^{\frac{A_0}{\tau^2}} + \int_{\frac{A_0}{\tau^2}}^\infty \right) \frac{t\tau^2}{(1+t\tau^2)^{\frac{3}{2}}} t^{-a-1} L(t) dt = J_{1\tau} + J_{2\tau} + J_{3\tau}, \text{ say.}$$

First note that

$$J_{1\tau} \leq \int_0^{A_0} \frac{A_0\tau^2}{(1+t\tau^2)^{\frac{3}{2}}} t^{-a-1} L(t) dt \leq A_0\tau^2 K^{-1}. \tag{31}$$

Next, we have

$$J_{2\tau} \leq \frac{A_0^{1-a}\tau^{2a}}{1-a} L\left(\frac{A_0}{\tau^2}\right) (1+o(1)) \leq \frac{A_0\tau^{2a}}{1-a} L\left(\frac{1}{\tau^2}\right) (1+o(1)), \tag{32}$$

where the above inequality in (32) comes from using (28) and then the slowly varying property of L , together with the fact that $A_0 \geq 1$ and $a \in (0, 1)$.

Finally, using Lemma 1 and the slowly varying property of L , we have

$$J_{3\tau} \leq \frac{A_0^{-a}\tau^{2a}}{a} L\left(\frac{A_0}{\tau^2}\right) (1+o(1)) \leq \frac{A_0\tau^{2a}}{a} L\left(\frac{1}{\tau^2}\right) (1+o(1)). \tag{33}$$

Using (31), (32) and (33) it follows that

$$\begin{aligned} J_\tau &\leq \frac{A_0\tau^{2a} L\left(\frac{1}{\tau^2}\right)}{a(1-a)} \left[K^{-1} \frac{\left(\frac{1}{\tau^2}\right)^{a-1}}{L\left(\frac{1}{\tau^2}\right)} + a(1+o(1)) + (1-a)(1+o(1)) \right] \\ &= \frac{A_0\tau^{2a} L\left(\frac{1}{\tau^2}\right)}{a(1-a)} (1+o(1)). \end{aligned} \tag{34}$$

The equality in (34) follows as $\lim_{\tau \rightarrow 0} \left(\frac{1}{\tau^2}\right)^{a-1} / L\left(\frac{1}{\tau^2}\right) = 0$, which, in turn, follows from Proposition part (iii) of Lemma 4. Theorem 4 now follows from (30) and (34). \square

Proof of Theorem 5. Fix $\eta \in (0, 1)$ and $\delta \in (0, 1)$. Then by definition,

$$\Pr(\kappa_i > \eta | X_i, \tau) = \frac{\int_\eta^1 \kappa_i^{a+\frac{1}{2}-1} (1-\kappa_i)^{-a-1} L\left(\frac{1}{\tau^2} \left(\frac{1}{\kappa_i} - 1\right)\right) e^{-\frac{\kappa_i X_i^2}{2}} d\kappa_i}{\int_0^1 \kappa_i^{a+\frac{1}{2}-1} (1-\kappa_i)^{-a-1} L\left(\frac{1}{\tau^2} \left(\frac{1}{\kappa_i} - 1\right)\right) e^{-\frac{\kappa_i X_i^2}{2}} d\kappa_i}$$

$$\leq \frac{e^{-\frac{\eta(1-\delta)X_i^2}{2}} \int_{\eta}^1 \kappa_i^{a+\frac{1}{2}-1} (1-\kappa_i)^{-a-1} L\left(\frac{1}{\tau^2}\left(\frac{1}{\kappa_i}-1\right)\right) d\kappa_i}{\int_0^{\eta\delta} \kappa_i^{a+\frac{1}{2}-1} (1-\kappa_i)^{-a-1} L\left(\frac{1}{\tau^2}\left(\frac{1}{\kappa_i}-1\right)\right) d\kappa_i}.$$

Now applying the change of variable $t = \frac{1}{\tau^2}\left(\frac{1}{\kappa_i}-1\right)$ to both the numerator and denominator on the right hand side of the preceding inequality, we obtain

$$\begin{aligned} \Pr(\kappa_i > \eta | X_i, \tau) &\leq \frac{e^{-\frac{\eta(1-\delta)X_i^2}{2}} (\tau^2)^{-a} \int_0^{\frac{1}{\tau^2}\left(\frac{1}{\eta}-1\right)} (1+t\tau^2)^{-\frac{1}{2}} t^{-a-1} L(t) dt}{(\tau^2)^{-a} \int_{\frac{1}{\tau^2}\left(\frac{1}{\eta\delta}-1\right)}^{\infty} (1+t\tau^2)^{-\frac{1}{2}} t^{-a-1} L(t) dt} \\ &\leq \frac{K^{-1} e^{-\frac{\eta(1-\delta)X_i^2}{2}}}{\sqrt{\frac{1-\eta\delta}{\tau^2}} \int_{\frac{1}{\tau^2}\left(\frac{1}{\eta\delta}-1\right)}^{\infty} t^{-(a+\frac{1}{2}+1)} L(t) dt} \\ &= \frac{H(a, \eta, \delta) e^{-\frac{\eta(1-\delta)X_i^2}{2}}}{\tau^{2a} \Delta(\tau^2, \eta, \delta)}, \end{aligned}$$

where we use the fact that $\int_0^{\frac{1}{\tau^2}\left(\frac{1}{\eta}-1\right)} \frac{t^{-a-1} L(t)}{\sqrt{1+t\tau^2}} dt \leq \int_0^{\infty} t^{-a-1} L(t) dt = K^{-1}$ and $\frac{t\tau^2}{1+t\tau^2} \geq 1-\eta\delta$ for each $t \geq \frac{1}{\tau^2}\left(\frac{1}{\eta\delta}-1\right)$ as $t \mapsto t\tau^2/(1+t\tau^2)$ is an increasing function of t for any fixed $\tau^2 > 0$. This completes the proof of Theorem 5. \square

Proof of Theorem 6. First note that the form of the posterior density of κ_i given (X_i, τ^2) as a function of (X_i, τ^2) is the same for each $i = 1, \dots, m$. Also under H_{0i} , the distribution of X_i does not depend on i . Therefore the probability $t_{1i} = \Pr(E(1-\kappa_i | X_i, \tau) > \frac{1}{2} | H_{0i} \text{ is true})$ of type I error for the i th test is the same for each i , and we denote it by t_1 . Using Theorem 4, for any $\tau < 1$, the event $\{E(1-\kappa_i | X_i, \tau) > \frac{1}{2}\}$ implies the following event:

$$\left\{ X_i^2 > 2a \log\left(\frac{1}{\tau^2}\right) - 2 \log L\left(\frac{1}{\tau^2}\right) - 2 \log\left(\frac{2A_0K}{a(1-a)}\right) - 2 \log(1+o(1)) \right\},$$

where the $o(1)$ term tends to zero as $\tau \rightarrow 0$ and is independent of X_i . Note that $2a \log\left(\frac{1}{\tau^2}\right) \rightarrow \infty$ as $\tau \rightarrow 0$, $\log(1+o(1)) \rightarrow 0$ as $\tau \rightarrow 0$ and $2A_0K/(a(1-a))$ is a positive constant. Also, using part (i) of Lemma 4, $\lim_{\tau \rightarrow 0} \log(L(\frac{1}{\tau^2}))/\log(\frac{1}{\tau^2}) = 0$. Therefore, we have, for all sufficiently small $\tau < 1$,

$$t_1 \leq \Pr \left(X_i^2 > 2 \left\{ a \log\left(\frac{1}{\tau^2}\right) + \log\left(\frac{1}{L\left(\frac{1}{\tau^2}\right)}\right) + \log\left(\frac{a(1-a)}{2A_0K}\right) \right\} \middle| H_{0i} \text{ is true} \right) (1+o(1)), \tag{35}$$

where $\lim_{\tau \rightarrow 0} o(1) = 0$ and this term does not depend on i , since under H_{0i} , the distribution of X_i is $N(0, 1)$ independently of i . Also note that for all sufficiently small $\tau < 1$,

$$\left\{ a \log\left(\frac{1}{\tau^2}\right) + \log\left(\frac{1}{L\left(\frac{1}{\tau^2}\right)}\right) + \log\left(\frac{a(1-a)}{2A_0K}\right) \right\} > 0. \tag{36}$$

Using (35) and (36), and the fact that $1 - \Phi(t) < \frac{\phi(t)}{t}$, for $t > 0$ we have, as $\tau \rightarrow 0$,

$$\begin{aligned} t_1 &\leq \Pr \left(|Z| > \sqrt{2a \log \left(\frac{1}{\tau^2} \right) + 2 \log \left(\frac{1}{L \left(\frac{1}{\tau^2} \right)} \right) + 2 \log \left(\frac{a(1-a)}{2A_0K} \right)} \right) (1 + o(1)) \\ &\leq 2 \cdot \frac{\phi \left(\sqrt{2a \log \left(\frac{1}{\tau^2} \right) + 2 \log \left(\frac{1}{L \left(\frac{1}{\tau^2} \right)} \right) + 2 \log \left(\frac{a(1-a)}{2A_0K} \right)} \right)}{\sqrt{2a \log \left(\frac{1}{\tau^2} \right) + 2 \log \left(\frac{1}{L \left(\frac{1}{\tau^2} \right)} \right) + 2 \log \left(\frac{a(1-a)}{2A_0K} \right)}} (1 + o(1)) \\ &= \frac{1}{\sqrt{\pi a}} \cdot \frac{2A_0K}{a(1-a)} \cdot \frac{\tau^{2a} L \left(\frac{1}{\tau^2} \right)}{\sqrt{\log \left(\frac{1}{\tau^2} \right)}} (1 + o(1)). \end{aligned}$$

In the above the $o(1)$ term tends to zero as $\tau \rightarrow 0$, and clearly it does not depend on i . The last equality follows using the functional form of $\phi(\cdot)$ in the numerator and by getting an asymptotic expression for the denominator using the observations made earlier about the relative magnitudes of the different terms as $\tau \rightarrow 0$. Since as $m \rightarrow \infty$, $\tau = \tau_m \rightarrow 0$, all the limiting statements used in the theorem also hold when $m \rightarrow \infty$. \square

Proof of Theorem 7. Let us fix any $\eta \in (0, \frac{1}{2})$ and any $\delta \in (0, 1)$. Using the inequality

$$\kappa_i \leq 1\{\eta < \kappa_i \leq 1\} + \eta,$$

we get the following:

$$E(\kappa_i | X_i, \tau) \leq \Pr(\kappa_i > \eta | X_i, \tau) + \eta. \tag{37}$$

Equation (37) coupled with Theorem 5, implies that for every $X_i \in \mathbb{R}$ we have

$$\left\{ E(\kappa_i | X_i, \tau) > \frac{1}{2} \right\} \subseteq \left\{ \frac{H(a, \eta, \delta) e^{-\frac{\eta(1-\delta)X_i^2}{2}}}{\tau^{2a} \Delta(\tau^2, \eta, \delta)} > \frac{1}{2} - \eta \right\}. \tag{38}$$

Using exactly similar argument as used in Theorem 6 about equality of t_{1i} for $i = 1, \dots, m$, but now noting that under H_{1i} , the distribution of X_i does not depend on i , it follows that $t_{2i} = t_2$ for some t_2 for $i = 1, \dots, m$, where t_{2i} denotes the probability of type II error of the i th test. Now observe that for sufficiently large m , $\tau = \tau_m < 1$ and hence $\log(\frac{1}{\tau^2}) \neq 0$. Therefore, using (38), we have,

$$\begin{aligned} t_2 &= \Pr \left(E(\kappa_i | X_i, \tau) > \frac{1}{2} \mid H_{1i} \text{ is true} \right) \\ &\leq \Pr \left(\frac{H(a, \eta, \delta) e^{-\frac{\eta(1-\delta)X_i^2}{2}}}{\tau^{2a} \Delta(\tau^2, \eta, \delta)} > \frac{1}{2} - \eta \mid H_{1i} \text{ is true} \right) \\ &= \Pr \left(X_i^2 < \frac{2}{\eta(1-\delta)} \left(a \log \left(\frac{1}{\tau^2} \right) + \log \left(\frac{1}{\Delta(\tau^2, \eta, \delta)} \right) + \log \left(\frac{H(a, \eta, \delta)}{\frac{1}{2} - \eta} \right) \right) \mid H_{1i} \text{ is true} \right) \\ &= \Pr \left(X_i^2 < \frac{2a}{\eta(1-\delta)} \log \left(\frac{1}{\tau^2} \right) (1 + o(1)) \mid H_{1i} \text{ is true} \right) \text{ as } m \rightarrow \infty. \end{aligned} \tag{39}$$

To prove the equality in (39), we first observe that $\log(\frac{1}{\tau^2}) \rightarrow \infty$ as $m \rightarrow \infty$ and that $\log(\frac{H(a,\eta,\delta)}{\frac{1}{2}-\eta})$ is a bounded quantity. Furthermore, one can show that

$$\frac{\log(\frac{1}{\Delta(\tau^2,\eta,\delta)})}{\log(\frac{1}{\tau^2})} \rightarrow 0 \text{ as } \tau \rightarrow 0 \text{ and hence as } m \rightarrow \infty. \tag{40}$$

Combination of these facts prove the last equality in (39). For the time being let us assume (40), and the proof of this will be given at the end.

To complete the proof of the theorem let us proceed as follows. Note that under $H_{1i}, X_i \sim N(0, 1 + \psi^2)$. Therefore, by (39) and the fact that $\lim_{m \rightarrow \infty} \frac{\psi^2}{1+\psi^2} = 1$ under Assumption 1, we have

$$t_2 \leq \Pr \left(|Z| < \sqrt{\frac{2a}{\eta(1-\delta)}} \sqrt{\frac{\log(\frac{1}{\tau^2})}{\psi^2}} (1 + o(1)) \right) \text{ as } m \rightarrow \infty. \tag{41}$$

Now under the assumption $\lim_{m \rightarrow \infty} \tau/p \in (0, \infty)$, it follows from Assumption 1 that, $\log(\frac{1}{\tau^2})/\psi^2 \rightarrow C \in (0, \infty)$ as $m \rightarrow \infty$. Together with (41), this shows

$$\begin{aligned} t_2 &\leq \Pr \left(|Z| < \sqrt{\frac{2aC}{\eta(1-\delta)}} (1 + o(1)) \right) \text{ as } m \rightarrow \infty \\ &= \Pr \left(|Z| < \sqrt{\frac{2aC}{\eta(1-\delta)}} (1 + o(1)) \right) \text{ as } m \rightarrow \infty \\ &= \left[2\Phi \left(\sqrt{\frac{2aC}{\eta(1-\delta)}} \right) - 1 \right] (1 + o(1)) \text{ as } m \rightarrow \infty. \end{aligned}$$

It is clear from the proof that the $o(1)$ terms do not depend on i . This completes the proof of the theorem, modulo the proof of (40) which is given below.

Since $L(\cdot)$ is slowly varying and $a > 0$, from Lemma 1 it follows that, for every fixed $\eta \in (0, 1)$ and $\delta \in (0, 1)$,

$$\lim_{\tau \rightarrow 0} \xi(\tau^2, \eta, \delta) = \lim_{\tau \rightarrow 0} \frac{\int_{\frac{1}{\tau^2}(\frac{1}{\eta\delta}-1)}^{\infty} t^{-(a+\frac{1}{2}+1)} L(t) dt}{(a + \frac{1}{2})^{-1} (\frac{1}{\tau^2}(\frac{1}{\eta\delta} - 1))^{-(a+\frac{1}{2})} L(\frac{1}{\tau^2}(\frac{1}{\eta\delta} - 1))} = 1, \tag{42}$$

where $\xi(\tau^2, \eta, \delta)$ is defined in the statement of Theorem 5. Again, using Lemma 3, we obtain

$$\lim_{\tau \rightarrow 0} \frac{\log L(\frac{1}{\tau^2}(\frac{1}{\eta\delta} - 1))}{\log \frac{1}{\tau^2}} = 0, \tag{43}$$

for every fixed $\eta \in (0, 1)$ and every fixed $\delta \in (0, 1)$, since $L(\cdot)$ is slowly varying. (42) and (43), together with the definition of $\Delta(\tau^2, \eta, \delta)$, lead to (40) immediately. \square

Proof of Theorem 8. By definition, the probability of type I error for the i th decision in (8) is given by

$$t_1 = \Pr(E(1 - \kappa_i|X_i, \tau) > \frac{1}{2} | H_{0i} \text{ is true})$$

where t_1 does not depend on i as already shown in the proof of Theorem 6 before.

Now, using (37) and Theorem 5, for every fixed $\eta \in (0, 1/2)$ and every fixed $\delta \in (0, 1)$, we have

$$E(\kappa_i|X_i, \tau) \leq \eta + \frac{H(a, \eta, \delta)e^{-\frac{\eta(1-\delta)X_i^2}{2}}}{\tau^{2a}\Delta(\tau^2, \eta, \delta)}$$

whence it follows that

$$\left\{ E(1 - \kappa_i|X_i, \tau) > \frac{1}{2} \right\} \supseteq \left\{ \frac{H(a, \eta, \delta)e^{-\frac{\eta(1-\delta)X_i^2}{2}}}{\tau^{2a}\Delta(\tau^2, \eta, \delta)} < \frac{1}{2} - \eta \right\}. \tag{44}$$

Therefore, using the definition of t_1 and (44), we obtain the following:

$$\begin{aligned} t_1 &\geq \Pr\left(\frac{H(a, \eta, \delta)e^{-\frac{\eta(1-\delta)X_i^2}{2}}}{\tau^{2a}\Delta(\tau^2, \eta, \delta)} < \frac{1}{2} - \eta | H_{0i} \text{ is true}\right) \\ &= \Pr\left(X_i^2 > \frac{2}{\eta(1-\delta)}\left\{a \log\left(\frac{1}{\tau^2}\right) + \log\left(\frac{1}{\Delta(\tau^2, \eta, \delta)}\right) + \log\left(\frac{H(a, \eta, \delta)}{\frac{1}{2} - \eta}\right)\right\} | H_{0i} \text{ is true}\right). \end{aligned}$$

Then using (40) coupled with the arguments as in the proof of Theorem 6, we obtain

$$t_1 \geq \frac{(\frac{1}{2} - \eta)/\sqrt{\pi a}}{H(a, \eta, \delta)} \cdot \frac{\tau^{\frac{2a}{\eta(1-\delta)}}}{\sqrt{\log(\frac{1}{\tau^2})}} \Delta(\tau^2, \eta, \delta)(1 + o(1)) \text{ as } m \rightarrow \infty,$$

where the $o(1)$ term above does not depend on i , and tends to zero as $m \rightarrow \infty$.

Since $\Delta(\tau^2, \eta, \delta) \sim L(\frac{1}{\tau^2})$ as $\tau \rightarrow 0$ and hence as $m \rightarrow \infty$, the stated result follows immediately. This completes the proof of Theorem 8. \square

Proof of Theorem 9. By definition, the probability of type II error for the i th decision in (8) is given by

$$t_2 = \Pr(E(1 - \kappa_i|X_i, \tau) \leq \frac{1}{2} | H_{1i} \text{ is true})$$

where t_2 does not depend on i as already shown in the proof of Theorem 7 before.

Note that, by our assumption, $\tau \rightarrow 0$ as $m \rightarrow \infty$. Therefore, using Theorem 4, for all sufficiently large m , we have

$$\left\{ \frac{A_0 K}{a(1-a)} e^{\frac{x_i^2}{2}} \tau^{2a} L\left(\frac{1}{\tau^2}\right)(1 + o(1)) \leq \frac{1}{2} \right\} \subseteq \left\{ E(1 - \kappa_i|X_i, \tau) \leq \frac{1}{2} \right\},$$

where the $o(1)$ term tends to zero as $m \rightarrow \infty$, and does not depend on i or X_i . Hence, for all such m ,

$$\begin{aligned} t_2 &= \Pr \left(E(1 - \kappa_i | X_i, \tau) \leq \frac{1}{2} \mid H_{1i} \text{ is true} \right) \\ &\geq \Pr \left(\frac{A_0 K}{a(1-a)} e^{\frac{X_i^2}{2}} \tau^{2a} L\left(\frac{1}{\tau^2}\right) (1 + o(1)) \leq \frac{1}{2} \mid H_{1i} \text{ is true} \right) \\ &\geq \Pr \left(X_i^2 \leq 2 \left\{ \log\left(\frac{1}{\tau^{2a}}\right) + \log\left(\frac{1}{L\left(\frac{1}{\tau^2}\right)}\right) \right. \right. \\ &\quad \left. \left. + \log\left(\frac{a(1-a)}{2A_0 K}\right) + \log(1 + o(1)) \right\} \mid H_{1i} \text{ is true} \right). \end{aligned}$$

Since $\lim_{\tau \rightarrow 0} \log\left(\frac{1}{L\left(\frac{1}{\tau^2}\right)}\right) / \log\left(\frac{1}{\tau^2}\right) = 0$ and $\log\left(\frac{1}{\tau^2}\right) \rightarrow \infty$ as $\tau \rightarrow 0$, we have, for all sufficiently large m ,

$$2 \left\{ \log\left(\frac{1}{\tau^{2a}}\right) + \log\left(\frac{1}{L\left(\frac{1}{\tau^2}\right)}\right) + \log\left(\frac{a(1-a)}{2A_0 K}\right) + \log(1 + o(1)) \right\} = 2a \log\left(\frac{1}{\tau^2}\right) (1 + o(1)) > 0.$$

Since $X_i \sim N(0, 1 + \psi^2)$ under H_{1i} and $\psi^2 \rightarrow \infty$, we have,

$$\begin{aligned} t_2 &\geq \Pr \left(X_i^2 \leq 2a \log\left(\frac{1}{\tau^2}\right) (1 + o(1)) \mid H_{1i} \text{ is true} \right) \\ &= \Pr \left(|Z| \leq \sqrt{2a} \sqrt{\frac{\log\left(\frac{1}{\tau^2}\right)}{\psi^2}} (1 + o(1)) \right) \\ &= 2(\Phi(\sqrt{2a}\sqrt{C}) - 1)(1 + o(1)). \end{aligned}$$

The second inequality in the above chain of inequalities follows using Assumption 1 and the fact that $\lim_{m \rightarrow \infty} \tau/p \in (0, \infty)$. This completes the proof of Theorem 9. \square

Proof of Theorem 10. We fix any $c_1 \geq 2$ and $c_2 \geq 1$ in the definition of $\hat{\tau}$ in (9).

Now, using the facts $X_1 | \nu_1 = 0 \sim N(0, 1)$ and $X_1 | \nu_1 = 1 \sim N(0, 1 + \psi^2)$, we have

$$\begin{aligned} \alpha_m &= \Pr \left(|X_1| > \sqrt{c_1 \log m} \right) \\ &= 2 \left[(1-p) \Pr \left(Z > \sqrt{c_1 \log m} \right) + p \Pr \left(Z > \sqrt{\frac{c_1 \log m}{1 + \psi^2}} \right) \right]. \end{aligned} \quad (45)$$

Observe that under the assumption $p \equiv p_m \propto m^{-\epsilon}$, $0 < \epsilon < 1$ and Assumption 1, one has $c_1 \log m / (1 + \psi^2) \rightarrow c_1 C / (2\epsilon)$, where the constant $C \in (0, \infty)$ has already been defined in Assumption 1. Hence, $\Pr(Z > \sqrt{c_1 \log m / (1 + \psi^2)}) = \beta(1 + o(1))$, where $\beta = 1 - \Phi(c_1 C / (2\epsilon))$. Clearly, $0 < \beta < 1/2$. On the other hand, applying Mill's ratio, we have $p^{-1} \Pr(Z > \sqrt{c_1 \log m}) \sim \frac{m^{-c_1/2} p^{-1}}{\sqrt{2\pi c_1 \log m}} \rightarrow 0$ as $m \rightarrow \infty$. Therefore, using these observations in (45) we obtain $\alpha_m = 2\beta p(1 + o(1))$.

Now, using the definition of type I error of the i th decision in (10), we have

$$\begin{aligned} \tilde{t}_{1i} &= \Pr(E(1 - \kappa_i|X_i, \hat{\tau}) > \frac{1}{2} | H_{0i} \text{ is true}) \\ &= \Pr(E(1 - \kappa_i|X_i, \hat{\tau}) > \frac{1}{2}, \hat{\tau} \leq 2\alpha_m | H_{0i} \text{ is true}) \\ &\quad + \Pr(E(1 - \kappa_i|X_i, \hat{\tau}) > \frac{1}{2}, \hat{\tau} > 2\alpha_m | H_{0i} \text{ is true}). \end{aligned} \tag{46}$$

As noted in van der Pas et al. (2014), for each fixed x , $E(1 - \kappa_i|x, \tau)$ is non-decreasing in τ . Therefore, $E(1 - \kappa_i|X_i, \hat{\tau}) \leq E(1 - \kappa_i|X_i, 2\alpha_m)$ whenever $\hat{\tau} \leq 2\alpha_m$. Thus, we have

$$\begin{aligned} &\Pr(E(1 - \kappa_i|X_i, \hat{\tau}) > \frac{1}{2}, \hat{\tau} \leq 2\alpha_m | H_{0i} \text{ is true}) \\ &\leq \Pr(E(1 - \kappa_i|X_i, 2\alpha_m) > \frac{1}{2} | H_{0i} \text{ is true}) \\ &\leq B_1^* \frac{\alpha_m^{2a} L(\frac{1}{\alpha_m^2})}{\sqrt{\log(\frac{1}{\alpha_m^2})}} (1 + o(1)) \text{ as } m \rightarrow \infty, \end{aligned} \tag{47}$$

for some finite positive constant B_1^* , independent of m . The last step of the above chain of inequalities follows using the same arguments as in the proof of Theorem 6 and then applying the slowly varying property of L .

Let $\hat{\tau}_1 \equiv \frac{1}{m}$ and $\hat{\tau}_2 = \frac{1}{c_2 m} \sum_{j=1}^m 1\{|X_j| > \sqrt{c_1 \log m}\}$. Thus, $\hat{\tau} = \max\{\hat{\tau}_1, \hat{\tau}_2\}$.

Now observe that, since $\alpha_m = 2\beta p(1 + o(1))$ and $p \propto m^{-\epsilon}$, for $0 < \epsilon < 1$, we have, $1/m < 2\alpha_m$ for all large m . Therefore, $\Pr(\hat{\tau}_1 > 2\alpha_m | H_{0i} \text{ is true}) = 0$ for all sufficiently large m . Using this observation and the fact that $\{\hat{\tau} > 2\alpha_m\} \subseteq \{\hat{\tau}_1 > 2\alpha_m\} \cup \{\hat{\tau}_2 > 2\alpha_m\}$, we obtain for all sufficiently large m the following:

$$\begin{aligned} &\Pr(E(1 - \kappa_i|X_i, \hat{\tau}) > \frac{1}{2}, \hat{\tau} > 2\alpha_m | H_{0i} \text{ is true}) \\ &\leq \Pr(\hat{\tau} > 2\alpha_m | H_{0i} \text{ is true}) \\ &\leq \Pr(\hat{\tau}_1 > 2\alpha_m | H_{0i} \text{ is true}) + \Pr(\hat{\tau}_2 > 2\alpha_m | H_{0i} \text{ is true}) \\ &= \Pr(\hat{\tau}_2 > 2\alpha_m | H_{0i} \text{ is true}) \\ &\leq \Pr(|X_i| > \sqrt{c_1 \log m} | H_{0i} \text{ is true}) + \Pr(\hat{\tau}_2 > 2\alpha_m, |X_i| \leq \sqrt{c_1 \log m} | H_{0i} \text{ is true}) \\ &\leq \frac{1/\sqrt{\pi}}{m^{c_1/2} \sqrt{\log m}} + \Pr(\hat{\tau}_2 > 2\alpha_m, |X_i| \leq \sqrt{c_1 \log m} | H_{0i} \text{ is true}), \end{aligned} \tag{48}$$

where we use the facts $X_i \sim N(0, 1)$ under H_{0i} , $1 - \Phi(t) < \frac{\phi(t)}{t}$ for $t > 0$, and $2/c_1 \leq 1$, for the last step in (48).

Note that $\hat{\tau}_2 = \frac{1}{c_2 m} \sum_{j(\neq i)=1}^m 1\{|X_j| > \sqrt{c_1 \log m}\}$ over the set $\{|X_i| \leq \sqrt{c_1 \log m}\}$. Therefore it follows that

$$\Pr\left(\hat{\tau}_2 > 2\alpha_m, |X_i| \leq \sqrt{c_1 \log m} \mid H_{0i} \text{ is true}\right)$$

$$\begin{aligned}
&\leq \Pr\left(\frac{1}{c_2 m} \sum_{j(\neq i)=1}^m 1\{|X_j| > \sqrt{c_1 \log m}\} > 2\alpha_m \mid H_{0i} \text{ is true}\right) \\
&= \Pr\left(\frac{1}{c_2 m} \sum_{j(\neq i)=1}^m 1\{|X_j| > \sqrt{c_1 \log m}\} > 2\alpha_m\right), \tag{49}
\end{aligned}$$

where the last step in (49) follows from the fact that the distribution of the remaining X_j 's do not depend on that of X_i . Therefore, we have

$$\begin{aligned}
&\Pr\left(\frac{1}{c_2 m} \sum_{j(\neq i)=1}^m 1\{|X_j| > \sqrt{c_1 \log m}\} > 2\alpha_m\right) \\
&= \Pr\left(\frac{1}{m-1} \sum_{j(\neq i)=1}^m 1\{|X_j| > \sqrt{c_1 \log m}\} > \frac{2c_2 m}{m-1} \alpha_m\right) \\
&\leq \Pr\left(\frac{1}{m-1} \sum_{j(\neq i)=1}^m 1\{|X_j| > \sqrt{c_1 \log m}\} \geq 2\alpha_m\right) \left[\text{since } \frac{c_2 m}{m-1} > 1\right]. \tag{50}
\end{aligned}$$

Note that $1\{|X_j| > \sqrt{2 \log m}\} \stackrel{i.i.d.}{\sim} \text{Bernoulli}(\alpha_m)$ for $j \in \{1, \dots, m\} \setminus \{i\}$. Also $0 < \alpha_m < 2\alpha_m < 1$ for all sufficiently large m . Therefore, applying Hoeffding's inequality (Hoeffding (1963)), we obtain, for all sufficiently large m ,

$$\Pr\left(\frac{1}{m-1} \sum_{j(\neq i)=1}^m 1\{|X_j| > \sqrt{c_1 \log m}\} > 2\alpha_m\right) \leq e^{-(m-1)D(2\alpha_m, \alpha_m)} \tag{51}$$

where $D(2\alpha_m, \alpha_m) = 2\alpha_m \log 2 + (1 - 2\alpha_m) \log\left(\frac{1-2\alpha_m}{1-\alpha_m}\right)$.

Recall that $\log\left(\frac{1}{1-x}\right)/x \rightarrow 1$ as $x \downarrow 0$. Therefore, since $\alpha_m \rightarrow 0$ as $m \rightarrow \infty$, one can write $D(2\alpha_m, \alpha_m)$ as

$$\begin{aligned}
D(2\alpha_m, \alpha_m) &= 2 \log 2 \cdot \alpha_m - (1 - 2\alpha_m) \frac{\alpha_m}{1 - \alpha_m} (1 + o(1)) \\
&= (2 \log 2 - 1) \alpha_m (1 + o(1)). \tag{52}
\end{aligned}$$

Since $\alpha_m \sim 2\beta p$, we have, $(m-1)D(2\alpha_m, \alpha_m) = 2(2 \log 2 - 1)\beta m p(1 + o(1))$ as $m \rightarrow \infty$. Therefore, by combining equations (46)–(52), we finally obtain, for each $i = 1, \dots, m$,

$$\tilde{t}_{1i} \leq B_1^* \frac{\alpha_m^{2a} L\left(\frac{1}{\alpha_m^2}\right)}{\sqrt{\log\left(\frac{1}{\alpha_m^2}\right)}} (1 + o(1)) + \frac{1/\sqrt{\pi}}{m^{c_1/2} \sqrt{\log m}} + e^{-2(2 \log 2 - 1)\beta m p(1 + o(1))}$$

provided m is sufficiently large. Before we conclude our arguments, it should be noted that the $o(1)$ terms appearing in the present proof tend to zero as $m \rightarrow \infty$ and are the same for any $i = 1, \dots, m$. This completes the proof of Theorem 10. \square

Proof of Theorem 11. We fix any $c_1 \geq 2$ and $c_2 \geq 1$ in the definition of $\hat{\tau}$ in (9). Let us choose any fixed $\gamma \in (0, \frac{1}{c_2})$.

Now, by the definition of type II error of the i th decision in (10), we have

$$\begin{aligned} \tilde{t}_{2i} &= \Pr \left(E(\kappa_i|X_i, \hat{\tau}) \geq \frac{1}{2} \mid H_{1i} \text{ is true} \right) \\ &= \Pr \left(E(\kappa_i|X_i, \hat{\tau}) \geq \frac{1}{2}, \hat{\tau} \leq \gamma\alpha_m \mid H_{1i} \text{ is true} \right) \\ &\quad + \Pr \left(E(\kappa_i|X_i, \hat{\tau}) \geq \frac{1}{2}, \hat{\tau} > \gamma\alpha_m \mid H_{1i} \text{ is true} \right). \end{aligned} \tag{53}$$

Recall that for each fixed $x \in \mathbb{R}$, $E(\kappa_i|x, \tau)$ is decreasing in τ . Therefore, one has $E(\kappa_i|X_i, \hat{\tau}) \leq E(\kappa_i|X_i, \gamma\alpha_m)$ whenever $\hat{\tau} > \gamma\alpha_m$, whence it follows that

$$\left\{ E(\kappa_i|X_i, \hat{\tau}) \geq \frac{1}{2}, \hat{\tau} > \gamma\alpha_m \right\} \subseteq \left\{ E(\kappa_i|X_i, \gamma\alpha_m) \geq \frac{1}{2} \right\}.$$

Therefore, applying the same set of arguments used in the proof of Theorem 7 and using the fact that under H_{1i} , $X_i \sim N(0, 1 + \psi^2)$, we obtain for every fixed $0 < \eta < 1/2$ and every fixed $0 < \delta < 1$, the following:

$$\begin{aligned} &\Pr \left(E(\kappa_i|X_i, \hat{\tau}) \geq \frac{1}{2}, \hat{\tau} > \gamma\alpha_m \mid H_{1i} \text{ is true} \right) \\ &\leq \Pr \left(E(\kappa_i|X_i, \gamma\alpha_m) \geq \frac{1}{2} \mid H_{1i} \text{ is true} \right) \\ &= \Pr \left(|Z| \leq \sqrt{\frac{2a}{\eta(1-\delta)}} \sqrt{\frac{2 \log \left(\frac{1}{\gamma\alpha_m} \right)}{1 + \psi^2}} (1 + o(1)) \right). \end{aligned} \tag{54}$$

Since $\alpha_m \sim 2\beta p$, using Assumption 1 it follows that

$$\frac{2 \log \left(\frac{1}{\gamma\alpha_m} \right)}{1 + \psi^2} = \frac{-2 \log p}{\psi^2} (1 + o(1)) = C(1 + o(1)). \tag{55}$$

Therefore, from (54) and (55), for every fixed $0 < \eta < 1/2$ and every fixed $0 < \delta < 1$, we obtain

$$\begin{aligned} &\Pr \left(E(\kappa_i|X_i, \hat{\tau}) \geq \frac{1}{2}, \hat{\tau} > \gamma\alpha_m \mid H_{1i} \text{ is true} \right) \\ &\leq \Pr \left(|Z| \leq \sqrt{\frac{2a}{\eta(1-\delta)}} \sqrt{C} \right) (1 + o(1)) \\ &= \left[2\Phi \left(\sqrt{\frac{2a}{\eta(1-\delta)}} \sqrt{C} \right) - 1 \right] (1 + o(1)). \end{aligned} \tag{56}$$

Our aim is to show now that the first term on the right hand side of (53) goes to 0 as $m \rightarrow \infty$. From the definition of $\hat{\tau}$, it follows that $\hat{\tau} \geq \frac{1}{c_2^2 m} \sum_{j(\neq i)=1}^m 1\{|X_j| > \sqrt{c_1 \log m}\}$. Using this observation and noting that the distribution of the remaining X_j 's do not depend on the distribution of X_i (because of independence), we obtain the following:

$$\Pr \left(E(\kappa_i|X_i, \hat{\tau}) \geq \frac{1}{2}, \hat{\tau} \leq \gamma\alpha_m \mid H_{1i} \text{ is true} \right)$$

$$\begin{aligned}
&\leq \Pr(\hat{\tau} \leq \gamma\alpha_m | H_{1i} \text{ is true}) \\
&\leq \Pr\left(\frac{1}{c_2 m} \sum_{j(\neq i)=1}^m 1\{|X_j| > \sqrt{c_1 \log m}\} \leq \gamma\alpha_m\right) \\
&= \Pr\left(-\left(\frac{1}{m-1} \sum_{j(\neq i)=1}^m 1\{|X_j| > \sqrt{c_1 \log m}\} - \alpha_m\right) \geq \left(1 - \frac{c_2\gamma m}{m-1}\right)\alpha_m\right).
\end{aligned}$$

Note that $1 - \frac{c_2\gamma m}{m-1} \rightarrow 1 - c_2\gamma$ as $m \rightarrow \infty$ and $1 - c_2\gamma > 0$ since $0 < \gamma < \frac{1}{c_2}$. Therefore, $1 - \frac{c_2\gamma m}{m-1} > 0$ for all sufficiently large m .

Therefore, using the preceding arguments and then applying the Markov's inequality, we obtain, for all sufficiently large m , the following:

$$\begin{aligned}
&\Pr\left(E(\kappa_i | X_i, \hat{\tau}) \geq \frac{1}{2}, \hat{\tau} \leq \gamma\alpha_m | H_{1i} \text{ is true}\right) \\
&\leq \Pr\left(-\left(\frac{1}{m-1} \sum_{j(\neq i)=1}^m 1\{|X_j| > \sqrt{c_1 \log m}\} - \alpha_m\right) \geq \left(1 - \frac{c_2\gamma m}{m-1}\right)\alpha_m\right) \\
&\leq \Pr\left(\left|\frac{1}{m-1} \sum_{j(\neq i)=1}^m 1\{|X_j| > \sqrt{c_1 \log m}\} - \alpha_m\right| \geq \left(1 - \frac{c_2\gamma m}{m-1}\right)\alpha_m\right) \\
&\leq \frac{\text{Var}(1\{|X_1| > \sqrt{c_1 \log m}\})}{(m-1)\left(1 - \frac{c_2\gamma m}{m-1}\right)^2 \alpha_m^2} \\
&= \frac{(1 - c_2\gamma)^{-2}(1 - \alpha_m)}{m\alpha_m} (1 + o(1)) \rightarrow 0 \text{ as } m \rightarrow \infty,
\end{aligned}$$

whence we have

$$\Pr\left(E(\kappa_i | X_i, \hat{\tau}) \geq \frac{1}{2}, \hat{\tau} \leq \gamma\alpha_m | H_{1i} \text{ is true}\right) = o(1) \text{ as } m \rightarrow \infty. \quad (57)$$

Combining (53), (56) and (57), it therefore follows that, for each $i = 1, \dots, m$, we have

$$\tilde{t}_{2i} \leq \left[2\Phi\left(\sqrt{\frac{2aC}{\eta(1-\delta)}}\right) - 1\right] (1 + o(1)) \text{ as } m \rightarrow \infty.$$

Note that the $o(1)$ terms appearing in this proof tend to zero as $m \rightarrow \infty$ and are the same for any $i = 1, \dots, m$. This completes the proof of Theorem 11. \square

Proof of Theorem 1. We shall prove only the upper bound here. The corresponding proof for the lower bound will follow analogously. First recall from (15) the general form of the Bayes Risk of a multiple testing rule under our chosen loss. Now since in our case $t_{1i} = t_1$ and $t_{2i} = t_2$ for all $i = 1, \dots, m$, we have

$$R_{OG} = m((1-p)t_1 + pt_2)$$

$$= mp\left(\frac{1-p}{p}t_1 + t_2\right).$$

To prove the result, it suffices to show that under both the situations (I) and (II), $\frac{(1-p)}{p}t_1 \rightarrow 0$ as $m \rightarrow \infty$. We first use the fact that $1-p \leq 1$. Then using the upper bound for t_1 obtained in Theorem 6, we have

$$\frac{(1-p)}{p}t_1 \leq \frac{1}{\sqrt{\pi a}} \cdot \frac{2A_0K}{a(1-a)} \cdot \frac{\tau}{p} \left(\frac{1}{\tau^2}\right)^{-(a-1/2)} \frac{L(\frac{1}{\tau^2})}{\sqrt{\log(\frac{1}{\tau^2})}}(1+o(1)) \text{ as } m \rightarrow \infty. \quad (58)$$

The proof under case (I) follows from the facts that $a > 1/2$, $\lim_{m \rightarrow \infty} \tau/p \in (0, \infty)$ and hence $\lim_{m \rightarrow \infty} \tau = 0$ and by part (iii) of Lemma 4, $\lim_{x \rightarrow \infty} x^{-\beta}L(x) = 0$ for any $\beta > 0$ as $L(\cdot)$ is slowly varying. Proof for the case (II) is simple using (58). \square

Proof of Theorem 2. Recall that

$$R_{OG}^{EB} = \sum_{i=1}^m \{(1-p)\tilde{t}_{1i} + p\tilde{t}_{2i}\} = p \sum_{i=1}^m \left\{ \frac{1-p}{p}\tilde{t}_{1i} + \tilde{t}_{2i} \right\}. \quad (59)$$

For each i , the upper bound to \tilde{t}_{2i} as in Theorem 11, is independent of i . Therefore, for all sufficiently large m ,

$$\sum_{i=1}^m \tilde{t}_{2i} \leq m \left[2\Phi\left(\sqrt{\frac{2aC}{\eta(1-\delta)}}\right) - 1 \right] (1+o(1)). \quad (60)$$

Therefore, to complete the proof of the theorem, it will be enough to show

$$\sum_{i=1}^m \frac{1-p}{p}\tilde{t}_{1i} = o(m) \text{ as } m \rightarrow \infty. \quad (61)$$

For this, we first note that, for each $i = 1, \dots, m$, the upper bound for \tilde{t}_{1i} , as obtained in Theorem 10, is independent of i . Using this and noting that $1-p < 1$, we obtain

$$\frac{1}{m} \sum_{i=1}^m \frac{1-p}{p}\tilde{t}_{1i} \leq B_1^* \frac{\alpha_m^{2a} L(\frac{1}{\alpha_m^2})}{p\sqrt{\log(\frac{1}{\alpha_m^2})}}(1+o(1)) + \frac{1/\sqrt{\pi}}{m^{c_1/2}p\sqrt{\log m}} + \frac{1}{p}e^{-2(2\log 2-1)\beta mp(1+o(1))}, \quad (62)$$

for all sufficiently large m , where B_1^* and β are independent of m and have already been defined in Theorem 10.

Since $p \rightarrow 0$ as $m \rightarrow \infty$ and $\alpha_m \sim 2\beta p$, the first term on the right hand side of (62) can be shown to go to zero as $m \rightarrow \infty$ under case (I), exactly as in the proof of Theorem 1, whereas for case (II), it goes to zero as $m \rightarrow \infty$ using the conditions of the theorem. Also, since $p \propto m^{-\epsilon}$ for $0 < \epsilon < 1$, $mp \rightarrow \infty$ as $m \rightarrow \infty$ and $\log p = o(mp)$. This implies that $\frac{1/\sqrt{\pi}}{m^{c_1/2}p\sqrt{\log m}}$ and $\frac{1}{p}e^{-2(2\log 2-1)\beta mp(1+o(1))} = e^{-2(2\log 2-1)\beta mp(1+o(1))-\log p}$ both tend to zero as $m \rightarrow \infty$. These observations, together with (62), imply that (61) holds, which on combining with (59) and (60) completes the proof of Theorem 2. \square

References

- Armagan, A., Dunson, D. B., and Clyde, M. (2011). “Generalized Beta Mixtures of Gaussians”. In: Shawe-Taylor, J., Zemel, R. S., Bartlett, P. L., Pereira, F. C. N., and Weinberger, K. Q. (eds.), *Advances in Neural Information Processing Systems*, volume 24, 523–531. 755, 758, 761
- Armagan, A., Dunson, D. B., and Lee, J. (2012). “Generalized Double Pareto Shrinkage”. *Statistica Sinica*, 23(1): 119–143. MR3076161. 755, 758, 762
- Benjamini, Y. and Hochberg, Y. (1995). “Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing”. *Journal of the Royal Statistical Society, Series B*, 57(1): 289–300. MR1325392. 754, 755
- Bhattacharya, A., Pati, D., Pillai, N., and Dunson, D. B. (2012). “Bayesian Shrinkage”. arXiv:1212.6088v1. 780
- (2014). “Dirichlet-Laplace priors for optimal shrinkage”. arXiv:1401.5398v1. 780
- Bingham, N. H., Goldie, C. M., and Teugels, J. L. (1987). *Regular Variation*. Cambridge, Great Britain: Encyclopedia of Mathematics and Its Applications, University Press. MR0898871. doi: <http://dx.doi.org/10.1017/CB09780511721434>. 781
- Bogdan, M., Chakrabarti, A., Frommlet, F., and Ghosh, J. K. (2011). “Asymptotic Bayes-Optimality under Sparsity of Some Multiple Testing Procedures”. *The Annals of Statistics*, 39(3): 1551–1579. MR2850212. doi: <http://dx.doi.org/10.1214/10-AOS869>. 753, 755, 763, 764, 766, 775
- 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*, volume 1, 211–230. Beachwood, OH: IMS Collections, IMS. MR2462208. doi: <http://dx.doi.org/10.1214/193940307000000158>. 755, 760, 763
- Cai, T. T. and Jin, J. (2010). “Optimal Rates of Convergence for Estimating the Null Density and Proportion of Nonnull Effects in Large-Scale Multiple Comparisons”. *The Annals of Statistics*, 38(1): 100–145. MR2589318. doi: <http://dx.doi.org/10.1214/09-AOS696>. 780
- Cai, T. T., Jin, J., and Low, M. J. (2007). “Estimation and confidence sets for sparse normal mixtures”. *The Annals of Statistics*, 35(6): 2421–2449. MR2382653. doi: <http://dx.doi.org/10.1214/009053607000000334>. 780
- Carvalho, C., Polson, N., and Scott, J. (2009). “Handling sparsity via the horseshoe”. *Journal of Machine Learning Research W&CP*, 5: 73–80. 753, 755, 756, 760
- (2010). “The horseshoe estimator for sparse signals”. *Biometrika*, 97(2): 465–480. MR2650751. doi: <http://dx.doi.org/10.1093/biomet/asq017>. 753, 755, 756, 758, 759, 778
- Datta, J. and Ghosh, J. K. (2013). “Asymptotic Properties of Bayes Risk for the Horseshoe Prior”. *Bayesian Analysis*, 8(1): 111–132. MR3036256. doi: <http://dx.doi.org/10.1214/13-BA805>. 753, 756, 759, 760, 767, 769, 776, 778, 780

- Donoho, D. and Jin, J. (2004). “Higher criticism for detecting sparse heterogeneous mixtures”. *The Annals of Statistics*, 32(3): 962–994. MR2065195. doi: <http://dx.doi.org/10.1214/009053604000000265>. 780
- Efron, B. (2004). “Large-scale simultaneous hypothesis testing: The choice of a null hypothesis”. *Journal of the American Statistical Association*, 99(465): 96–104. MR2054289. doi: <http://dx.doi.org/10.1198/016214504000000089>. 755
- (2008). “Microarrays, Empirical Bayes and the two-groups Model”. *Statistical Science*, 23(1): 1–22. MR2431866. doi: <http://dx.doi.org/10.1214/07-STS236>. 755
- Gelman, A. (2006). “Prior distributions for variance parameters in hierarchical models”. *Bayesian Analysis*, 1(3): 515–533. MR2221284. 759
- Ghosh, P. and Chakrabarti, A. (2015). “Posterior Concentration Properties of a General Class of Shrinkage Priors around Nearly Black Vectors”. arXiv:1412.8161v4. 779, 780
- Griffin, J. E. and Brown, P. J. (2005). “Alternative prior distributions for variable selection with very many more variables than observations”. Technical report, University of Warwick. 755
- (2010). “Inference with normal–gamma prior distributions in regression problems”. *Bayesian Analysis*, 5(1): 171–188. MR2596440. doi: <http://dx.doi.org/10.1214/10-BA507>. 755
- (2012). “Structuring shrinkage: some correlated priors for regression”. *Biometrika*, 99(2): 481–487. MR2931267. doi: <http://dx.doi.org/10.1093/biomet/asr082>. 755
- (2013). “Some priors for sparse regression modeling”. *Bayesian Analysis*, 8(3): 691–702. MR3102230. doi: <http://dx.doi.org/10.1214/13-BA827>. 755
- Hans, C. (2009). “Bayesian lasso regression”. *Biometrika*, 96(4): 835–845. MR2564494. doi: <http://dx.doi.org/10.1093/biomet/asp047>. 755
- Hoeffding, W. (1963). “Probability inequalities for sums of bounded random variables”. *Journal of the American Statistical Association*, 58: 13–30. MR0144363. 790
- Ingster, Y. I. (1997). “Some problems of hypothesis testing leading to infinitely divisible distributions”. *Mathematical Methods of Statistics*, 6(1): 47–69. MR1456646. 780
- Meinshausen, N. and Rice, J. (2006). “Estimating the proportion of false null hypotheses among a large number of independently tested hypotheses”. *The Annals of Statistics*, 34(1): 373–393. MR2275246. doi: <http://dx.doi.org/10.1214/009053605000000741>. 780
- Mitchell, T. and Beauchamp, J. (1988). “Bayesian variable selection in linear regression (with discussion)”. *Journal of the American Statistical Association*, 83(404): 1023–1036. MR0997578. 754
- Park, T. and Casella, G. (2008). “The Bayesian lasso”. *Journal of the American Statistical Association*, 103(482): 681–686. MR2524001. doi: <http://dx.doi.org/10.1198/016214508000000337>. 755

- Pati, D., Bhattacharya, A., Pillai, N., and Dunson, D. (2014). “Posterior contraction in sparse Bayesian factor models for massive covariance matrices”. *The Annals of Statistics*, 42(3): 1102–1130. MR3210997. doi: <http://dx.doi.org/10.1214/14-AOS1215>. 780
- Polson, N. G. and Scott, J. G. (2011). “Shrink Globally, Act Locally: Sparse Bayesian Regularization and Prediction”. In: *Bayesian Statistics 9, Proceedings of the 9th Valencia International Meeting*, 501–538. Oxford University Press. MR3204017. doi: <http://dx.doi.org/10.1093/acprof:oso/9780199694587.003.0017>. 755, 757, 758, 759, 779, 780
- (2012). “On the Half-Cauchy Prior for a Global Scale parameter”. *Bayesian Analysis*, 7(2): 1–16. MR3000018. doi: <http://dx.doi.org/10.1214/12-BA730>. 755
- Scott, J. and Berger, J. O. (2006). “An exploration of aspects of Bayesian multiple testing”. *Journal of Statistical Planning and Inference*, 136(7): 2144–2162. MR2235051. doi: <http://dx.doi.org/10.1016/j.jspi.2005.08.031>. 755, 778
- (2010). “Bayes and empirical-Bayes multiplicity adjustment in the variable-selection problem”. *The Annals of Statistics*, 38(5): 2587–2619. MR2722450. doi: <http://dx.doi.org/10.1214/10-AOS792>. 760
- Scott, J. G. (2011). “Bayesian estimation of intensity surfaces on the sphere via needlet shrinkage and selection”. *Bayesian Analysis*, 6(2): 307–327. MR2806246. doi: <http://dx.doi.org/10.1214/11-BA611>. 755
- 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. MR2323757. doi: <http://dx.doi.org/10.1111/j.1467-9868.2007.005592.x>. 755
- Tipping, M. (2001). “Sparse Bayesian learning and the Relevance Vector Machine”. *Journal of Machine Learning Research*, 1: 211–244. MR1875838. doi: <http://dx.doi.org/10.1162/15324430152748236>. 755
- 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”. *Electronic Journal of Statistics*, 8: 2585–2618. MR3285877. doi: <http://dx.doi.org/10.1214/14-EJS962>. 753, 756, 759, 760, 772, 779, 780, 789, 796

Acknowledgments

We would like to thank the Editor-in-Chief, the Associate Editor and three anonymous referees for their thoughtful comments which have lead to substantial improvement of this paper. We would also like to thank Dr. Sourabh Bhattacharya of ISI Kolkata for some useful tips on Markov Chain Monte Carlo simulations. We also thank Professor Jayanta K. Ghosh for letting us aware of the recent work of van der Pas et al. (2014) on posterior contraction of the horseshoe prior. Research of Malay Ghosh was partially supported by National Science Foundation (NSF) Grants DMS-1007494 and SES-1026165.