Wavelet Thresholding with Bayesian False Discovery Rate Control

Mahlet G. Tadesse,1,* Joseph G. Ibrahim,2 Marina Vannucci,3 and Robert Gentleman4

1Department of Biostatistics and Epidemiology, University of Pennsylvania, Philadelphia, Pennsylvania 19104, U.S.A.
2Department of Biostatistics, University of North Carolina, Chapel Hill, North Carolina 27599, U.S.A.
3Department of Statistics, Texas A&M University, College Station, Texas 77843, U.S.A.
4Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington 98109, U.S.A.
*email: mtadesse@ceb.upenn.edu

Summary. The false discovery rate (FDR) procedure has become a popular method for handling multiplicity in high-dimensional data. The definition of FDR has a natural Bayesian interpretation; it is the expected proportion of null hypotheses mistakenly rejected given a measure of evidence for their truth. In this article, we propose controlling the positive FDR using a Bayesian approach where the rejection rule is based on the posterior probabilities of the null hypotheses. Correspondence between Bayesian and frequentist measures of evidence in hypothesis testing has been studied in several contexts. Here we extend the comparison to multiple testing with control of the FDR and illustrate the procedure with an application to wavelet thresholding. The problem consists of recovering signal from noisy measurements. This involves extracting wavelet coefficients that result from true signal and can be formulated as a multiple hypotheses-testing problem. We use simulated examples to compare the performance of our approach to the Benjamini and Hochberg (1995, *Journal of the Royal Statistical Society, Series B* 57, 289–300) procedure. We also illustrate the method with nuclear magnetic resonance spectral data from human brain.

Key words: False discovery rate; Multiple hypotheses testing; Posterior probability; p-value; Wavelet thresholding.

1. Introduction

Wavelet techniques have become an attractive and efficient tool in function estimation. Suppose we have data \((y_1, \ldots, y_n)\)

\[
y_i = f(t_i) + \varepsilon_i, \quad i = 1, \ldots, n,
\]

where the \(t_i\)'s are equally spaced points and the \(\varepsilon_i\)'s are independent and identically distributed \(\mathcal{N}(0, \sigma^2)\) random variables. The observations \(y_i, i = 1, \ldots, n\), are transformed to the wavelet domain by applying a discrete wavelet transform (DWT). This yields a sequence of empirical wavelet coefficients \(d_{jk}, j = 0, \ldots, J, k = 0, \ldots, 2^j - 1\), where \(J\) is such that \(n = 2^{J+1}\). This relationship between \(n\) and \(J\) is not a real restriction and there are methods that allow the DWT to be applied to any length of data. The empirical coefficients are contaminated with noise and follow a \(\mathcal{N}(\theta_{jk}, \sigma^2)\) density, where the \(\theta_{jk}\)'s correspond to the true coefficients resulting from the expansion of the function \(f\) in terms of an orthonormal basis. Because of the parsimonious representation by wavelets, only a few \(d_{jk}\)'s contain information about the real signal. An estimator \(\hat{f}\) can be obtained by extracting the significant coefficients and applying the inverse DWT to the thresholded coefficients. Common choices of thresholding include hard thresholding and soft thresholding (Donoho and Johnstone, 1994). The former is a “kill” or “keep” rule, where the empirical wavelet coefficients that fall below a threshold are mapped to 0 and the remaining are kept fixed. The latter is a “kill” or “shrink” rule, where the retained coefficients are shrunk toward 0. Hard thresholding generally does better at reproducing peak heights and discontinuities but at some cost in visual smoothness. A comprehensive review of wavelets in statistical modeling can be found in Vidakovic (1999).

The problem of thresholding wavelet coefficients can be viewed as a multiple testing problem (Abramovich and Benjamini, 1996). For each coefficient, the hypothesis \(H_0: \theta_{jk} = 0\) is tested against \(H_1: \theta_{jk} \neq 0\). This raises a problem of multiplicity, where the probability of erroneously declaring significance increases rapidly with the number of tests. The traditional concern in multiple hypotheses testing is to control the family-wise error rate (FWER), that is, the probability of making at least one false rejection among all tests. Numerous methods have been developed to this end (Miller, 1981). Unfortunately, while these procedures reduce the probability of spurious findings, they also severely reduce the probability of identifying real effects. In particular, when testing hundreds of hypotheses, these methods lead to rejection rules that are so stringent that very few, if any, findings are made. The false discovery rate (FDR) procedure developed by Benjamini and Hochberg (1995) controls the proportion of errors among rejected tests and provides a less conservative approach.
The FDR has been the subject of much recent research in the statistical literature. Storey (2002, 2003) introduced a modified version, called the positive false discovery rate (pFDR), which has a natural Bayesian interpretation. Here, we propose a novel Bayesian approach for controlling the pFDR and compare its performance to the Benjamini and Hochberg (1995) procedure in the context of wavelet thresholding. Our procedure has the advantage of allowing available prior or external information to be incorporated into the evaluation of evidence about the hypotheses.

Many authors have compared frequentist p-values to Bayesian posterior probabilities, finding various degrees of reconciliation depending on the type of tests considered. In the one-sided problem, Casella and Berger (1987) show that reconciliation depending on the type of tests considered. In the one-sided problem, Casella and Berger (1987) show that there is agreement between the two quantities. However, in the two-sided problem, the use of p-values as a measure of evidence has been called into question as they tend to overstate evidence against $H_0$, that is, p-values tend to be much smaller than posterior probabilities (Berger and Delampady, 1987; Berger and Sellke, 1987). Westfall, Johnson, and Utts (1997) examined the correspondence of these measures in the context of multiple tests where the Bonferroni correction is used, and give conditions under which the adjusted posterior probabilities roughly correspond to the Bonferroni adjusted p-values. Our work can be viewed as an extension of these comparisons to the situation where multiple tests are performed with control of the FDR.

This article is organized as follows. Section 2 describes various existing FDR procedures. In Section 3, we present our Bayesian pFDR method. Section 4 compares the performance of our approach to the Benjamini and Hochberg (1995) method in the context of wavelet thresholding using simulated datasets. We also look at the results using the Akaike (AIC) and the Bayesian (BIC) information criteria. In Section 5, we apply the procedures to nuclear magnetic resonance (NMR) spectral data from a brain tissue image. Section 6 concludes the article with a brief discussion.

2. The False Discovery Rate

The outcomes from testing $m$ hypotheses can be classified into four categories, as described in Table 1. The number of hypotheses to be performed, $m$, is known in advance. The numbers of rejected and nonrejected null hypotheses, $R$ and $W$, are observable random variables, while all the other quantities are unknown. $S$ and $U$ correspond respectively to the number of correct rejections and nonrejections. $V$ and $T$ are the incorrect decisions and correspond respectively to the number of rejected null hypotheses (false positives) and nonrejected alternative hypotheses (false negatives).

The FDR procedure attempts to minimize the expected proportion of null hypotheses erroneously rejected. This can be viewed through the random variable $Q = V/R$. However, since taking the expectation of $Q$ is not mathematically sound when $R = 0$, Benjamini and Hochberg (1995) consider the following alternatives:

$$
\begin{align*}
(a) & \quad E \left[ \frac{V}{R} \mid R > 0 \right] \cdot \Pr(R > 0), \\
(b) & \quad E \left[ \frac{V}{R} \mid R > 0 \right], \\
(c) & \quad E \left[ \frac{V}{R} \mid R = r \right] = \frac{E[V \mid R = r]}{r}, \\
(d) & \quad \frac{E[V]}{E[R]}.
\end{align*}
$$

The formulations in (b), (c), and (d) would identically be 1 if all null hypotheses are true and therefore cannot be controlled in the traditional p-value-based framework. Consequently, Benjamini and Hochberg (1995) choose to work with formulation (a) and propose a sequential p-value method. Let $\alpha$ be the desired error-rate level and $p_1, \ldots, p_m$ denote the ordered p-values from the $m$ tests. The procedure rejects $k$ of the smallest p-values, where

$$
k = \max \left\{ j : p_j \leq \frac{j \cdot \alpha}{m} \right\}. \tag{3}
$$

The authors show this guarantees that the FDR $\leq \alpha$. In fact, the FDR is controlled at a lower level, $\alpha\frac{m_0}{m}$. In a later work, Benjamini and Hochberg (2000) propose an adaptive procedure, where the number of null hypotheses, $m_0$, is first estimated and then used in the sequential p-value procedure, so that

$$
k = \max \left\{ j : p_j \leq \frac{j \cdot \alpha}{m} \right\}. \tag{4}
$$

These procedures and all the others discussed in this article assume independent hypotheses. We just mention that Benjamini and Yekutieli (2001) have proposed an FDR procedure for general forms of dependence, where the criterion of rejection is modified to

$$
k = \max \left\{ j : p_j \leq \frac{j \cdot \alpha}{m} \sum_{i=1}^{m} \frac{1}{j} \right\}. \tag{5}
$$

Storey (2003) argues that formulation (b) in equation (2), which he calls the positive false discovery rate (pFDR), is a more appropriate measure because investigators are not concerned with controlling the rate of false detections when no test is significant. He proposes to fix the rejection region and estimate the corresponding pFDR. Let $T$ be the specified rejection region and $T_1, \ldots, T_m$ be independent and identically distributed test statistics for $m$ identical hypotheses. The pFDR can be written as a posterior probability

$$
pFDR(\Gamma) = E \left[ \frac{V}{R} \mid R > 0 \right] = \Pr(H = 0 \mid T \in \Gamma)
\times \frac{\pi_0 \cdot \Pr(T \in \Gamma \mid H = 0)}{\pi_0 \cdot \Pr(T \in \Gamma \mid H = 0) + (1 - \pi_0) \cdot \Pr(T \in \Gamma \mid H = 1)}. \tag{6}
$$

Table 1

<table>
<thead>
<tr>
<th>Summary of outcomes in multiple hypotheses testing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maintain null</td>
</tr>
<tr>
<td>----------------</td>
</tr>
<tr>
<td>Null true</td>
</tr>
<tr>
<td>Null false</td>
</tr>
<tr>
<td>$W$</td>
</tr>
</tbody>
</table>
where \( \pi_0 \) is the prior probability that a null hypothesis is true. Storey (2002) expresses the pFDR in terms of p-value-based rejections. All rejections are then in the interval \([0, \gamma]\), for some \( \gamma \geq 0 \), and equation (6) can be written as

\[
pFDR(\gamma) = \frac{\pi_0 \cdot \Pr(P \leq \gamma \mid H = 0)}{\Pr(P \leq \gamma)},
\]

where \( P \) is the random p-value resulting from any test and \( P \mid H = 0 \sim \text{Uniform}(0, 1) \). The other quantities in pFDR are estimated from the data.

Bayesian FDR approaches have also been proposed. Here, the rejection rule is defined in terms of posterior probabilities of the null. Rosner and Vidakovic (2000) order the posterior probabilities of the null hypotheses, \( p(1), \ldots, p(m) \), and reject the first \( k \), where

\[
k = \max \left\{ j : \frac{1}{j} \sum_{i=0}^{j-1} i \cdot P_j(i) < \alpha \right\},
\]

and \( P_j(i) \) is the coefficient of \( z^i \) in the generating polynomial \( \phi_j(z) = \prod_{i=0}^{j} (q(i) + p(i)z) = \sum_{i=0}^{j} P_j(i)z^i \) with \( q(i) = 1 - p(i) \).

Newton et al. (2003) also consider using the ranked posterior probabilities of the null hypotheses. Unlike the approach we propose here, their method does not control the average rate of false detections. Instead, it uses a data-dependent estimate of the rate of false positives. The procedure rejects all hypotheses with posterior probabilities less than \( \kappa \), where \( \kappa \) is the largest value such that \( \frac{1}{C(\kappa)} \leq \alpha \) with

\[
J(\kappa) = \{ j : p_j \leq \kappa \} \quad \text{and} \quad C(\kappa) = \sum_j p_j I\{ p_j \leq \kappa \}.
\]

3. Bayesian Positive False Discovery Rate

3.1 Definition

Our Bayesian pFDR approach is also based on the posterior probabilities of the null hypotheses. We adopt Storey’s (2003) definition of the FDR,

\[
E \left[ \frac{V}{R} \right] \mid R > 0.
\]

Suppose we observe data \( X = (X_1, \ldots, X_m) \) on \( m \) independent variables and for each \( X_j \) test identical hypotheses, \( H_j = 0 \) versus \( H_j = 1 \). The decision to keep or reject a null hypothesis is based on its posterior probability

\[
p_j = \Pr(H_j = 0 \mid X_j) = \frac{\Pr(X_j \mid H_j = 0) \cdot \Pr(H_j = 0)}{\Pr(H_j = 0) \cdot \Pr(X_j \mid H_j = 0) + [1 - \Pr(H_j = 0)] \cdot \Pr(X_j \mid H_j = 1)},
\]

where \( \Pr(H_j = 0) \) is the prior probability that the null is true and \( \Pr(X_j \mid H_j = i) \) is obtained by integrating over the parameter space

\[
\Pr(X_j \mid H_j = i) = \int_{\theta_i} \Pr(X_j \mid \theta, H_j = i) \Pr(\theta \mid H_j = i) \, d\theta, \quad i = 0, 1
\]

where \( \Pr(X_j \mid \theta, H_j = i) \) is the likelihood function and \( \Pr(\theta \mid H_j = i) \) is the prior density of the set of parameters \( \theta \) under \( H_j = i \). If external information on the parameters is available, it may be incorporated in the prior specification. Otherwise, \( \Pr(\theta \mid H_j = i) \) may be chosen relatively flat in the region where the likelihood is large. In some cases, such as exponential family distributions with conjugate priors, the integral in (10) can be evaluated analytically. In situations where it is intractable one can resort to numerical methods or Monte Carlo approximations (Kass and Raftery, 1995).

The magnitude of the posterior probability reflects the strength of the support for the null hypothesis. Thus, \( H_j = 0 \) is rejected if \( p_j \leq \gamma \), where \( \gamma \) is the cut-off for rejection. Let \( P_j \) be the random variable defined by the posterior probability of the \( j \)th null hypothesis and let \( \Gamma \) be the rejection region. The pFDR defined in (9) is then equal to

\[
E \left[ \frac{V}{R} \right] \mid R > 0 = \sum_{r=1}^{m} E \left[ \frac{V}{r} \right] \cdot \Pr(R = r \mid R > 0)
\]

\[
= \sum_{r=1}^{m} \sum_{j \in S_r} \sum_{j=r}^{m} \Pr(P_{j,}, \ldots, P_{j,} \in \Gamma \mid R > 0)
\]

\[
\times \Pr(P_{j,}, \ldots, P_{j,} \notin \Gamma \mid R > 0)
\]

\[
= \sum_{r=1}^{m} \frac{1}{r} \sum_{j \in S_r} \left\{ \prod_{i=1}^{r} \Pr(P_{j,} \leq \gamma) \cdot \prod_{i=j+1}^{m} \Pr(P_{j,} > \gamma) \right\}
\]

\[
\cdot \Pr(R > 0)
\]

where \( S_r \) consists of all distinct subsets of size \( r \), \( S_r = \{ j : \{ j_1, \ldots, j_r \} \} \), and \( \{ j_{r+1}, \ldots, j_m \} \) is the set of hypotheses not included in \( S_r \). The last equality follows from the independence of the \( X_j \)’s. Calculating (12) is computationally prohibitive as it involves all \( (2^m - 1) \) possible scenarios of rejection. This derivation can be facilitated by assuming that all hypotheses \( H_j (j = 1, \ldots, m) \) have equal prior probability of coming from the null, and by specifying identical prior densities for the model parameters under \( H_j = i (i = 0, 1) \). The \( P_j \)’s are then independent and identically distributed and we can omit the subscript \( j \) in the notation. For readability, let \( H_0 \) and \( H_1 \) denote the null and alternative hypotheses, respectively. We then have \( V \mid R = r \sim \text{binomial}(r, \Pr(H_0 \mid P \leq \gamma)) \) and equation (12) reduces to

\[
bFDR(\gamma) = \sum_{r=1}^{m} \frac{1}{r} \Pr(H_0 \mid P \leq \gamma) \times \Pr(R = r \mid R > 0)
\]

\[
= \frac{1}{\Pr(R > 0)} \cdot \Pr(H_0 \mid P \leq \gamma)
\]

\[
= \frac{\Pr(P \leq \gamma \mid H_0) \cdot \Pr(H_0)}{\Pr(P \leq \gamma)}.
\]

The assumption of identical priors for all hypotheses is limiting, especially in a Bayesian procedure where a major
incentive is to incorporate additional information. In practice, observations believed to behave similarly may be grouped together, and each data subset can be assigned different priors and examined separately. We illustrate this in Section 5 with the NMR spectral data.

3.2 Estimation of the Bayesian Positive False Discovery Rate

We note that equation (13) has the same form as (7). However, the two quantities are inherently different because they use different measures of evidence. In (13), we work with posterior probabilities of the null rather than p-values. In Storey’s approach, $P \mid H_0$ is Uniform $(0, 1)$ since the p-values are tail probabilities computed under the null hypothesis. The posterior probabilities of the null, however, do not have such interpretation, and it is not possible to obtain an analytic expression for their posterior densities, $g(p \mid H_i)(i = 0, 1)$. We use Monte Carlo simulation to evaluate $Pr(P \leq \gamma \mid H_i)$. We generate $Y_i^{(t)}(t = 1, \ldots, T)$ from $f(x \mid \hat{\theta}_i)$, where $\hat{\theta}_i$ are the parameter estimates under $H_i$. For each $Y_i^{(t)}$ we calculate the posterior probability of the null hypothesis, $p_i^{(t)}$. An estimate of the desired tail area is then given by

$$\int_0^\gamma g(p \mid H_i) dp \approx \frac{1}{T} \sum_{t=1}^T I\{p_i^{(t)} \leq \gamma\}.$$ 

In order to illustrate these concepts and show the distribution of the posterior probabilities of the null and that of the p-values, we simulate a dataset with $m = 1000$ variables. Let $x_{ij}$ be the measurement for sample $i$ $(i = 1, \ldots, n)$ and variable $j$ $(j = 1, \ldots, m)$. The data are generated such that

$$x_{ij} \sim N(0, 1) + 2 \cdot I(j > 500),$$

and for each $j$, we test $H_0: \mu = 0$ versus $H_1: \mu \neq 0$. We use the following priors to calculate $p(H_0 \mid X)$:

$$\begin{align*}
\mu &\mid H_0 \sim \delta(0) \\
\mu &\mid \sigma^2, H_1 \sim N(0, \sigma^2) \\
\sigma^2 &\sim IG(r, s)
\end{align*}$$

where $\delta(0)$ is a point mass at 0. The hyperparameters $r, s$ are chosen to yield noninformative priors. We consider varying the prior probabilities for $H_0$. The plots in Figure 1 show the distribution of $P = Pr(H_0 \mid X)$ under the null, $g(p \mid H_0)$, and its marginal distribution, $g(p)$. We see that they depend on the choice of prior. A weak prior in the truth of the null yields smaller posterior probabilities. For example, when $Pr(H_0) = 0.1$ we have $Pr(H_0 \mid X)$ in the range $[0, 0.5]$. As $Pr(H_0)$ increases, $g(p \mid H_0)$ becomes more skewed toward 1, whereas $g(p)$ becomes a symmetric distribution on $[0, 1]$. We also calculate the p-values for each of the variables and the distribution of this measure is shown in Figure 2. We note that the p-values do indeed follow a uniform distribution under the null hypothesis.

The steps to implement the Bayesian pFDR procedure are as given in Section 3.3.

3.3 Algorithm to Estimate the Bayesian False Discovery Rate

Let $(X_1, \ldots, X_m)$ be $m$ independent variables with density $f(x \mid \theta, H_i)$ under hypothesis $H_i$ $(i = 0, 1)$ and let $[0, \gamma]$ be the rejection region.

1. Compute the posterior probability of the null, $p_j$, for each test, $j = 1, \ldots, m$.

2. Obtain estimates for the posterior means of the parameters under the null and alternative hypotheses:

   (a) Calculate the posterior estimates, $\hat{\theta}_j \mid H_1 = E[\theta \mid X_j, H_1]$, for each variable $j$.
   (b) Let $n_0 = \sum_{j=1}^m I(p_j > \gamma)$ and $n_1 = m - n_0$.
   (c) Define $\hat{\theta}_0 = \frac{1}{n_0} \sum_{j:p_j > \gamma} \hat{\theta}_j \mid H_0$ and $\hat{\theta}_1 = \frac{1}{n_1} \sum_{j:p_j > \gamma} \hat{\theta}_j \mid H_1$.

3. Generate samples $Y_i^{(t)}$ from $f(x \mid \hat{\theta}_i, H_i), t = 1, \ldots, T$.

4. For each $Y_i^{(t)}$, calculate the corresponding posterior probability of the null, $p_i^{(t)}$.

5. Estimate $bFDR(\gamma)$ with

$$\widehat{bFDR}(\gamma) = \frac{Pr(H_0) \cdot \widehat{Pr}(P \leq \gamma \mid H_0)}{Pr(P \leq \gamma)},$$

where

$$\widehat{Pr}(P \leq \gamma \mid H_i) = \frac{1}{T} \sum_{t=1}^T I\{p_i^{(t)} \leq \gamma\}, \quad i = 0, 1,$$

and

$$\widehat{Pr}(P \leq \gamma) = Pr(H_0) \cdot \widehat{Pr}(P \leq \gamma \mid H_0) + Pr(H_1) \cdot \widehat{Pr}(P \leq \gamma \mid H_1).$$

4. Simulation Study

We assess the performance of our procedure in the context of wavelet thresholding using four benchmark datasets: “Heavy Sine,” “Blocks,” “Bumps,” and “Doppler” (Donoho and Johnstone, 1994). For each function, 500 noisy datasets were generated by corrupting the true signal with independent random noise $\varepsilon_i \sim N(0, \sigma^2)$ at 1024 points uniformly spaced on $[0, 1]$. The values of $\sigma$ were taken to correspond to a signal-to-noise ratio $\delta = sd(f) / \sigma = 3, 5, \text{and } 7$. The goal is to remove the noise and recover the true functions. We applied the DWT implemented in the Matlab toolbox WavBox (by Carl Taswell) using Daubechies minimum phase wavelets with four vanishing moments (Daubechies, 1992).

We compare the reconstructions obtained by applying the inverse wavelet transform to the coefficients thresholded by the Bayesian pFDR and the Benjamini and Hochberg (1995) approaches. Bayesian methods provide a natural way to obtain shrunken estimates through the posterior means of the wavelet coefficients and we make use of these. Our approach is therefore a soft thresholding procedure. On the other hand, the Benjamini–Hochberg approach provides a hard thresholding method (Abramovich and Benjamini, 1996).

For each wavelet coefficient, $d_{jk} = \theta_{jk} + \varepsilon_{jk}$, we test $H_0: \theta_{jk} = 0$ against $H_1: \theta_{jk} \neq 0$ to identify the coefficients to be thresholded. The p-values corresponding to testing each $\theta_{jk}$ are given by

$$2 \cdot \left[1 - \Phi\left(|d_{jk}| / \hat{\sigma}\right)\right],$$

where $\hat{\sigma}$ is the estimated standard deviation.
Bayesian False Discovery Rate Control

Figure 1. Distribution of $\Pr(H = 0 \mid X)$ for different choices of prior.

where $\hat{\sigma} = \text{MAD}(d_j)$, the median absolute deviation of the wavelet coefficients at the finest level adjusted by $1/0.6745$.

In the Bayesian approach, we specify a prior model similar to the one considered by Clyde, Parmigiani, and Vidakovic (1998). For the coefficients $\theta_{jk}$, we take a point mass prior at 0 under the null and we consider a common normal prior under the alternative hypothesis:

\[
\begin{align*}
\theta_{jk} \mid \sigma^2, H_0 &\sim \delta(0) \\
\theta_{jk} \mid \sigma^2, H_1 &\sim N(0, c \sigma^2) \\
\sigma^2 &\sim IG(r, s) \\
H_1 \mid \alpha &\sim \text{Bernoulli}(\alpha). 
\end{align*}
\]  

(18)

We set $c = 100$ to accommodate the possibility of large coefficients, $s = \hat{\sigma}^2$ and $r = 2$, the smallest integer such that $\text{E}[\sigma^2] = s/r - 1$ is defined. The hyperparameter $\alpha$ corresponds to the proportion of coefficients expected a priori to be non-negligible and depends on the smoothness of the true signal. Donoho and Johnstone (1994) offered a universal threshold $\lambda = \sigma(2 \log n)^{1/2}$ to separate signal from noise. We elicit $\alpha$ as the proportion of coefficients that pass this threshold

\[
\alpha = \frac{\sum_{j=1}^{j' - 1} \sum_{k=0}^{j - 1} I\{|d_{jk}| > \lambda\}}{n - 2}.
\]

(19)

The Benjamini–Hochberg method fixes the error rate whereas our Bayesian approach fixes the rejection region. Since the former is based on p-values and the latter is based on posterior probabilities of the null hypotheses, fixing a common rejection region and estimating the corresponding FDRs would not be appropriate. We therefore make a small modification to the Bayesian procedure and estimate the largest cut-off value that yields the specified error rate,

\[
\gamma = \max\{\gamma_j : \hat{\gamma}_F(\gamma_j) \leq \alpha\},
\]

(20)

where $\hat{\gamma}_F(\gamma_j)$ is given by equation (16). Thus, with our Bayesian procedure, all hypotheses with posterior probabilities of the null \( \leq \gamma \) are rejected, and with the Benjamini–Hochberg method the smallest p-values chosen according to (3) are rejected. This should put the two methods on equal ground for comparison. Under both procedures, we select the significant coefficients $d_{jk}$ to be retained while controlling the FDR at level $\alpha = 0.05$.

In the Bayesian approach, the nonrejected coefficients are estimated by their posterior means, which provide shrinkage. In the Benjamini–Hochberg method, these coefficients are fixed to their empirical values. An estimator $\hat{f}$ for the true function is then obtained by applying the inverse DWT to the thresholded coefficients. The goodness of fit of each estimator is measured by the mean-square error (MSE)

\[
\text{MSE} = n^{-1} \sum_{i=1}^{n} (\hat{f}_i - f_i)^2.
\]

(21)
Table 2 reports, for each of the four functions, the average MSE and the standard error of the MSE over the 500 replications. For all cases considered, both procedures have comparable goodness-of-fit measures. Figures 3–6 show the true signals and the noisy signals with $\delta = 5$ along with the reconstructions obtained after thresholding the coefficients. In these figures, we have superimposed the estimators, which are shown in solid lines, and the true functions, which appear in dotted lines. We see that both methods do a very good job at removing the noise.

A comprehensive comparison of various wavelet-thresholding and wavelet-shrinkage methods is given by Antoniadis, Bigot, and Sapatinas (2001). Here, we also present the results obtained using AIC (Akaike, 1973) and BIC (Schwarz, 1978), two commonly used model-selection criteria. These procedures maximize a penalized log likelihood with different penalty terms. In a regression model with orthogonal regressors, AIC amounts to selecting variables with $z$-score greater than $(2)^{1/2}$, and BIC corresponds to including variables with $|z| > (\log n)^{1/2}$ (George and Foster, 2000). We note that the BIC yields comparable results to the FDR procedures, whereas the AIC performs quite poorly.

5. Application to NMR Spectral Data

We now illustrate the Bayesian FDR procedure with wavelet thresholding applied to nuclear magnetic resonance (NMR) data where we use adaptive priors, that is, priors that incorporate information on the resolution level of the wavelet coefficients, in the spirit of Chipman, Kolaczyk, and McCulloch (1997). The data were recorded from a single voxel of a three-dimensional image of a normal brain (Figure 8a). The goal is to “denoise” the data while preserving the location and size of the peaks in the underlying signals. These are used to determine the molecular properties of the tissue.

We first consider identical priors for all coefficients as in Section 4. Our second set of priors incorporates the

<table>
<thead>
<tr>
<th>Function</th>
<th>$\delta$</th>
<th>AIC Ave. (MSE)</th>
<th>AIC s.e. (MSE)</th>
<th>BH Ave. (MSE)</th>
<th>BH s.e. (MSE)</th>
<th>bFDR Ave. (MSE)</th>
<th>bFDR s.e. (MSE)</th>
<th>BIC Ave. (MSE)</th>
<th>BIC s.e. (MSE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heavy Sine</td>
<td>3</td>
<td>0.0810</td>
<td>0.0216</td>
<td>0.1204</td>
<td>0.0318</td>
<td>0.5907</td>
<td>0.0483</td>
<td>0.1200</td>
<td>0.0275</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.0298</td>
<td>0.0073</td>
<td>0.0277</td>
<td>0.0070</td>
<td>0.0214</td>
<td>0.0159</td>
<td>0.0453</td>
<td>0.0097</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>0.0174</td>
<td>0.0037</td>
<td>0.0164</td>
<td>0.0034</td>
<td>0.1101</td>
<td>0.0085</td>
<td>0.0251</td>
<td>0.0049</td>
</tr>
<tr>
<td>Blocks</td>
<td>3</td>
<td>0.1281</td>
<td>0.0138</td>
<td>0.1281</td>
<td>0.0129</td>
<td>0.2616</td>
<td>0.0192</td>
<td>0.1274</td>
<td>0.0134</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.0491</td>
<td>0.0052</td>
<td>0.0493</td>
<td>0.0050</td>
<td>0.0949</td>
<td>0.0071</td>
<td>0.0491</td>
<td>0.0051</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>0.0269</td>
<td>0.0026</td>
<td>0.0269</td>
<td>0.0025</td>
<td>0.0482</td>
<td>0.0033</td>
<td>0.0267</td>
<td>0.0025</td>
</tr>
<tr>
<td>Bumps</td>
<td>3</td>
<td>0.0183</td>
<td>0.0017</td>
<td>0.0199</td>
<td>0.0020</td>
<td>0.0323</td>
<td>0.0023</td>
<td>0.0180</td>
<td>0.0017</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.0065</td>
<td>$6.5 \times 10^{-4}$</td>
<td>0.0064</td>
<td>$6.3 \times 10^{-4}$</td>
<td>0.0115</td>
<td>$8.1 \times 10^{-4}$</td>
<td>0.0065</td>
<td>$6.5 \times 10^{-4}$</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>0.0032</td>
<td>$3.3 \times 10^{-4}$</td>
<td>0.0033</td>
<td>$3.4 \times 10^{-4}$</td>
<td>0.0058</td>
<td>$4.2 \times 10^{-4}$</td>
<td>0.0031</td>
<td>$3.3 \times 10^{-4}$</td>
</tr>
<tr>
<td>Doppler</td>
<td>3</td>
<td>0.0018</td>
<td>$2.6 \times 10^{-4}$</td>
<td>0.0018</td>
<td>$2.6 \times 10^{-4}$</td>
<td>0.0058</td>
<td>$4.3 \times 10^{-4}$</td>
<td>0.0020</td>
<td>$2.7 \times 10^{-4}$</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>$7.7 \times 10^{-4}$</td>
<td>$9.0 \times 10^{-5}$</td>
<td>$7.4 \times 10^{-4}$</td>
<td>$8.8 \times 10^{-5}$</td>
<td>0.0021</td>
<td>$1.5 \times 10^{-4}$</td>
<td>$7.9 \times 10^{-4}$</td>
<td>$1.0 \times 10^{-4}$</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>$4.1 \times 10^{-4}$</td>
<td>$5.0 \times 10^{-5}$</td>
<td>$4.1 \times 10^{-4}$</td>
<td>$5.2 \times 10^{-5}$</td>
<td>0.0011</td>
<td>$7.3 \times 10^{-5}$</td>
<td>$4.3 \times 10^{-4}$</td>
<td>$5.6 \times 10^{-5}$</td>
</tr>
</tbody>
</table>
Figure 3. Heavy Sine function.

Figure 4. Blocks function.
Figure 5. Bumps function.

Figure 6. Doppler function.
information on the resolution level. The hypotheses at the
different levels \( j \) are tested separately using the following
priors:

\[
\begin{align*}
\theta_{\beta} | \sigma^2, H_0 & \sim \delta(0) \\
\theta_{\beta} | \sigma^2, H_1 & \sim \mathcal{N}(0, h_j \sigma^2) \\
\sigma^2 & \sim \text{IG}(r, s) \\
H_1 | \alpha_j & \sim \text{Bernoulli}(\alpha_j),
\end{align*}
\] (22)

where the hyperparameters are specified to capture our belief
that most of the wavelet coefficients at the finest resolution
level result from random noise. We take \( H_j = 10 \times 2^{(J-j)} \) and
\( \alpha_j = 1 - (j/J + 1)^{-2} \). The other hyperparameters are defined
as in equation (18). Under this setting, we chose to include
the coefficients at the coarsest levels, \( j \leq 3 \).

The effect of the different FDR procedures on the wavelet
coefficients can be seen in Figure 7, where the thresholded
estimates are plotted against the empirical coefficients. In all
cases, the coefficients deemed negligible are set to 0. With the
Benjamini–Hochberg approach the remaining ones are kept at
their empirical values whereas the Bayesian procedure shrinks
them toward 0.

The relative performance of the three estimators \( \hat{f} \) can be
assessed from the reconstructions in Figure 8. The one re-
sulting from the IDWT of the coefficients thresholded by
the Benjamini–Hochberg approach is given in Figure 8b.
Figure 8c and 8d shows the underlying signal reconstructed
using the coefficient estimates from the Bayesian FDR pro-
cedure with the two different prior models. We note that the
Benjamini–Hochberg procedure retains the sharpness of the
peaks but the smooth parts of the curve are still a bit noisy.
The Bayesian approach with constant priors effectively re-
moves the noise but attenuates the peaks. The Bayesian ap-
proach with adaptive priors, on the other hand, retains the
perceived sharp structures of the spectrum without much at-
tenuation and does the most complete job of “denoising” the
data.

6. Discussion

We have provided a Bayesian approach for controlling the
FDR and have applied it to the problem of wavelet thresh-
olding. The Bayesian FDR is based on posterior probabilities
of the null hypothesis. This measure of evidence is quite at-
tractive as it allows inference on the validity of a hypothesis
using the actual data as well as external information. The pro-
posed method controls the expected proportion of coefficients
mistakenly included in the reconstruction of the true signal.
It is a soft thresholding procedure, which sets the rejected
coefficients to 0 and shrinks those that are kept. Simulated
examples were used to illustrate the method and to compare
its performance to the Benjamini and Hochberg (1995) FDR
procedure. We also presented an application to NMR spectral
data using adaptive priors.

We conclude the article by putting our work in perspective
with other Bayesian thresholding and shrinkage procedures.
Vidakovic (1998) considered Bayesian hypothesis testing to
address this problem. He proposed a hard thresholding rule
via Bayes factors, where the empirical wavelet coefficients
are kept the same if the null hypothesis is rejected and are
mapped to zero otherwise. Outside the hypothesis-testing

![Figure 7.](image-url)
setting, Chipman et al. (1997) proposed an adaptive shrinkage estimator by modeling the wavelet coefficients as a mixture of two normal distributions with known variance and specifying level-dependent hyperparameters. This approach shrinks the wavelet coefficient estimates but does not set any of them to zero. Clyde et al. (1998) and Abramovich, Sapatinas, and Silverman (1998) used a scale mixture of a normal distribution and a point mass at zero. Their method allows both thresholding and shrinkage of the wavelet coefficient estimates, as does our procedure.

Acknowledgements

Tadesse was supported by NCI grant CA90301; Ibrahim was partially supported by NIH grants CA 70101 and CA 74015; and Vannucci was supported by NSF CAREER award DMS-0093208.

Figure 8. NMR spectrum and different reconstructions. (a) NMR spectrum. (b) Reconstruction using Benjamini and Hochberg’s thresholding. (c) Reconstruction using Bayesian procedure with constant priors. (d) Reconstruction using Bayesian approach with adaptive priors.

Résumé

La procédure du «False Discovery Rate» (FDR) est aujourd’hui une méthode largement répandue pour gérer les questions de multiplicité liées à des données multidimensionnelles. La définition du FDR trouve une interprétation naturelle dans le cadre bayésien, à savoir la proportion attendue d’hypothèses nulles rejetées à tort, à seuil de significativité donné. Dans cet article, nous proposons de contrôler le FDR positif par une approche bayésienne dont la règle de rejet est basée sur les probabilités postérieures des hypothèses nulles. Le lien entre les degrés de significativité bayésiens et frequentistes a été étudié dans plusieurs contextes. Ici, nous élargissons la comparaison aux tests multiples avec contrôle du FDR, et illustrons la procédure par une application au seuillage en ondelettes. Le problème consiste à récupérer un signal à partir de mesures bruitées. Ceci implique d’extraire des coefficients d’ondelettes résultant du vrai signal, et peut être formulé comme un problème de test d’hypothèses...

References