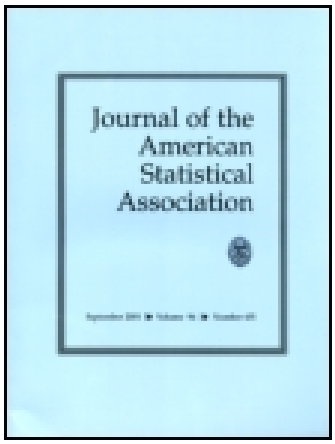
This article was downloaded by: [Carnegie Mellon University]
On: 06 August 2015, At: 20:17
Publisher: Taylor & Francis
Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: 5 Howick Place, London, SW1P 1WG



Journal of the American Statistical Association
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/uasa20

False Discovery Rate Regression: An Application to Neural Synchrony Detection in Primary Visual Cortex

James G. Scott, Ryan C. Kelly, Matthew A. Smith, Pengcheng Zhou & Robert E. Kass

Accepted author version posted online: 07 Jan 2015.Published online: 06 Jul 2015.

CrossMark
Click for updates

To cite this article: James G. Scott, Ryan C. Kelly, Matthew A. Smith, Pengcheng Zhou & Robert E. Kass (2015) False Discovery Rate Regression: An Application to Neural Synchrony Detection in Primary Visual Cortex, Journal of the American Statistical Association, 110:510, 459-471, DOI: 10.1080/01621459.2014.990973

To link to this article: http://dx.doi.org/10.1080/01621459.2014.990973

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the "Content") contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at http://www.tandfonline.com/page/terms-and-conditions

Supplementary materials for this article are available online. Please go to www.tandfonline.com/r/JASA

False Discovery Rate Regression: An Application to Neural Synchrony Detection in Primary Visual Cortex

James G. SCOTT, Ryan C. KELLY, Matthew A. SMITH, Pengcheng ZHOU, and Robert E. KASS

This article introduces false discovery rate regression, a method for incorporating covariate information into large-scale multiple-testing problems. FDR regression estimates a relationship between test-level covariates and the prior probability that a given observation is a signal. It then uses this estimated relationship to inform the outcome of each test in a way that controls the overall false discovery rate at a prespecified level. This poses many subtle issues at the interface between inference and computation, and we investigate several variations of the overall approach. Simulation evidence suggests that: (1) when covariate effects are present, FDR regression improves power for a fixed false-discovery rate; and (2) when covariate effects are absent, the method is robust, in the sense that it does not lead to inflated error rates. We apply the method to neural recordings from primary visual cortex. The goal is to detect pairs of neurons that exhibit fine-time-scale interactions, in the sense that they fire together more often than expected due to chance. Our method detects roughly 50% more synchronous pairs versus a standard FDR-controlling analysis. The companion R package `FDRreg` implements all methods described in the article. Supplementary materials for this article are available online.

KEY WORDS: False discovery rate; Empirical Bayes; Multiple testing; Neuroscience.

Downloaded by [Carnegie Mellon University] at 20:17 06 August 2015

1. INTRODUCTION

1.1 Multiple Testing in the Presence of Covariates

The problem of multiple testing concerns a group of related null hypotheses $h_1, \ldots, h_n$ that are tested simultaneously. In its simplest form, each test yields a summary statistic z_i, and the goal is to decide which of the z_i are signals ($h_i = 1$) and which are null ($h_i = 0$). One classic solution to this problem, known as Bonferroni correction, aims to control the family-wise error rate (FWER): the probability of incorrectly rejecting at least one null hypothesis, assuming that they are all true. An alternative, which has become the dominant approach in many fields, is to control the false discovery rate (FDR): the proportion of false positives among those null hypotheses that are rejected (Benjamini and Hochberg 1995). Regardless of which error rate they aim to control, however, most existing approaches obey a monotonicity property: if test statistic z_i is declared significant, and z_j is more extreme than z_i, then z_j is also declared significant. Yet in many cases, we have auxiliary covariate information about each test statistic, such as location in the brain or distance along a chromosome. If significant test statistics tend to cluster in covariate space, then monotonicity becomes undesirable, and a procedure that takes account of the covariate should perform better. In this article, we introduce a method called *false-discovery-rate regression* (FDRR) that incorporates covariates directly into the multiple-testing problem.

James G. Scott is Associate Professor, McCombs School of Business, University of Texas at Austin, 1 University Station, B6500, Austin, TX 78712 (E-mail: *james.scott@mccombs.utexas.edu*). Ryan C. Kelly, Google, New York, NY 10011 (E-mail: *ryekelly@gmail.com*). Matthew A. Smith is Assistant Professor, Department of Ophthalmology, University of Pittsburgh, Pittsburgh, PA 15213 (E-mail: *smithma@pitt.edu*). Pengcheng Zhou is Graduate Student, Center for the Neural Basis of Cognition, Carnegie Mellon University, Pittsburgh, PA 15213 (E-mail: *zhoupc1988@gmail.com*). Robert E. Kass is Professor, Department of Statistics, Carnegie Mellon University, Pittsburgh, PA 15213 (E-mail: *kass@stat.cmu.edu*). The authors thank the editor, associate editor, and two anonymous referees for their detailed and helpful feedback. JGS was supported by NSF CAREER grant DMS-1255187. MAS was supported by NIH grants EY018894 and EY022928. PZ, RCK, and REK were supported by NIH grant R01MH064537, and PZ by a grant from the R. K. Mellon Foundation.

Color versions of one or more of the figures in the article can be found online at *www.tandfonline.com/r/jasa*.

The method we describe here builds on the two-groups model (Efron et al. 2001), an intuitive framework for controlling the false-discovery rate. In the two-groups model, some small fraction c of the test statistics are assumed to come from an unknown signal population, and the remainder from a known null population. Our proposal is to allow the mixing fraction c to depend upon covariates, and to estimate the form of this dependence from the data. Extensive simulation evidence shows that, by relaxing the monotonicity property in a data-dependent way, FDR regression can improve power while still controlling the global false-discovery rate. The method is implemented in the publicly available R package `FDRreg` (Scott 2014).

Our motivating application is the identification of interactions among many simultaneously recorded neurons, which has become a central issue in computational neuroscience. Specifically, we use FDR regression to detect fine-time-scale neural interactions ("synchrony") among 128 units (either single neurons or multi-unit groups) recorded simultaneously from the primary visual cortex (V1) of a rhesus macaque monkey (Kelly et al. 2010; Kelly and Kass 2012). The experiment from which the data are drawn produced thousands of pairs of neurons, each involving a single null hypothesis of no interaction. In this case, combining all tests into a single FDR-controlling analysis would inappropriately ignore the known spatial and functional relationships among the neurons (e.g., Smith and Kohn 2008). Our approach for false-discovery rate regression avoids this problem: it detects roughly 50% more significant neuron pairs compared with a standard analysis by exploiting the fact that spatially and functionally related neurons are more likely to exhibit synchronous firing.

1.2 The Two-Groups Model

In the two-groups model for multiple testing, one assumes that test statistics $z_1, \ldots, z_n$ arise from the mixture

$$z \sim c \cdot f_1(z) + (1 - c) \cdot f_0(z), \qquad (1)$$

© 2015 American Statistical Association

Downloaded by [Carnegie Mellon University] at 20:17 06 August 2015

where $c \in (0, 1)$, and where f_0 and f_1 respectively describe the null ($h_i = 0$) and alternative ($h_i = 1$) distributions of the test statistics. For each z_i, one then reports the quantity

$$w_i = \mathrm{P}(h_i = 1 \mid z_i) = \frac{c \cdot f_1(z_i)}{c \cdot f_1(z_i) + (1 - c) \cdot f_0(z_i)}. \quad (2)$$

As Efron (2008a) observed, the information contained in w_i provides a tidy methodological unification to the multiple-testing problem. Bayesians may interpret w_i as the posterior probability that z_i is a signal, while frequentists may interpret $1 - w_i$ as a local false-discovery rate. The global false-discovery rate of some set Z_1 of putative signals can then estimated as

$$\mathrm{FDR}(Z_1) \approx \frac{1}{|Z_1|} \sum_{i : z_i \in Z_1} (1 - w_i).$$

Efron et al. (2001) showed that this Bayesian formulation of FDR is biased upward as an estimate of frequentist FDR, and therefore conservative.

An elegant property of the local FDR approach is that it is both frequentist and fully conditional: it yields valid error rates, yet also provides a measure of significance that depends on the precise value of z_i, and not merely its inclusion in a larger set (see Berger 2003). This can be achieved, moreover, at little computational cost. To see how, observe that (2) may be reexpressed in marginalized form as

$$1 - w_i = \frac{(1 - c) \cdot f_0(z_i)}{f(z_i)}, \quad (3)$$

where $f(z) = c \cdot f_1(z_i) + (1 - c) \cdot f_0(z_i)$ is the overall marginal density. Importantly, $f(z)$ can be estimated from the empirical distribution of the test statistics; this is typically smooth, which makes estimating $f(z)$ notably easier than a generic density-estimation problem. Therefore one may compute local FDR using cheap plug-in estimates $\hat{f}(z)$ and $\hat{c}$, and avoid the difficult deconvolution problem that would have to be solved to find $f_1(z)$ explicitly (e.g., Efron et al. 2001; Newton 2002; Martin and Tokdar 2012).

1.3 FDR Regression

Implicit in the two-groups model is the assumption that all tests should be combined into a single analysis with a common mixing weight c in (1). Yet for some datasets, this may be highly dubious. In our analysis of neural recordings, for example, a test statistic z_i is a measure of pairwise synchrony in the firing rates of two neurons recorded from an array of electrodes, and these z_i's exhibit spatial dependence across the array: two nearby neurons are more likely to fire synchronously than are two neurons at a great distance. Similar considerations are likely to arise in many applications.

False-discovery-rate regression addresses this problem through a conceptually simple modification of (1), in which covariates x_i may affect the prior probability that z_i is a signal. In its most general form, the model assumes that

$$\begin{aligned} z_i &\sim c(x_i) \cdot f_1(z_i) + \{1 - c(x_i)\} \cdot f_0(z_i) \\ c(x_i) &= G\{s(x_i)\} \end{aligned} \quad (4)$$

for an unknown regression function $s(x)$ and known link function $G : \mathcal{R} \to (0, 1)$.

This new model poses two main challenges versus the ordinary two-groups model (1). First, we must estimate a regression model for an unobserved binary outcome: whether z_i comes from f_1, and is therefore a signal. Second, because each mixing weight in (4) depends on x_i, there is no longer a common mixture distribution $f(z)$ for all the test statistics. We therefore cannot express the Bayes probabilities in marginalized form (3), and cannot avoid estimating $f_1(z)$ directly.

Our approach, described in detail in Section 2, is to represent $f_1(z)$ as a location mixture of the null density, here assumed to be a Gaussian distribution:

$$\begin{aligned} f_0(z) &= \mathrm{N}(z \mid \mu, \sigma^2) \\ f_1(z) &= \int_{\mathcal{R}} \mathrm{N}(z \mid \mu + \theta, \sigma^2)\, \pi(\theta)\, d\theta. \end{aligned}$$

Even in the absence of covariates, estimating the mixing density $\pi(\theta)$ is known to be a challenging problem, because Gaussian convolution heavily blurs out any peaks in the prior. We consider two ways of proceeding. The first is an empirical-Bayes method in which an initial plug-in estimate $\hat{\pi}(\theta)$ is fit via predictive recursion (Newton 2002). The regression function is then estimated by an expectation-maximization (EM) algorithm, treating $\hat{\pi}(\theta)$ as fixed. The second is a fully Bayes method in which $\pi(\theta)$ and the regression function $s(x)$ are estimated jointly using Markov chain Monte Carlo. In simulation studies, both methods lead to better power and equally strong protection against false discoveries compared with traditional FDR-controlling approaches.

The rest of the article proceeds as follows. The remainder of Section 1 contains a brief review of the literature on multiple testing. Section 2 describes both empirical-Bayes and fully Bayes methods for fitting the FDR regression model, and draws connections with existing approaches for controlling the false-discovery rate. It also describes how existing methods for fitting an empirical null hypothesis may be combined with the new approach (Section 2.4). Section 3 shows the results of a simulation study that validates the frequentist performance of the method. Section 4 provides background information on the neural synchrony-detection problem. Section 5 shows the results of applying FDR regression to the sychrony-detection dataset. Section 6 contains discussion.

1.4 Connection with Existing Work

Our approach is based on the two-groups model, and therefore in the spirit of much previous work on Bayes and empirical-Bayes multiple testing, including Efron et al. (2001), Johnstone and Silverman (2004), Scott and Berger (2006), Muller, Parmigiani, and Rice (2006), Efron (2008a, 2008b), and Bogdan, Ghosh, and Tokdar (2008). The final reference has a comprehensive bibliography. We will make some of these connections more explicit when they arise in subsequent sections.

Other authors have considered the problem of multiple testing in the presence of correlation (e.g. Clarke and Hall 2009; Fan, Han, and Gu 2012). The focus there is on making the resulting conclusions robust to unknown correlation structure among the test statistics. Because FDR regression explicitly uses covariates to inform the outcome of each test, it is different both in aim and execution from these other approaches.

Downloaded by [Carnegie Mellon University] at 20:17 06 August 2015

On the computational side, we also draw upon a number of recent innovations. Our empirical-Bayes approach uses predictive recursion, a fast and efficient method for estimating a mixing distribution (Newton 2002; Tokdar, Martin, and Ghosh 2009; Martin and Tokdar 2012). Our fully Bayes approach requires drawing posterior samples from a hierarchical logistic-regression model, for which we exploit the Pólya-Gamma data-augmentation scheme introduced by Polson, Scott, and Windle (2013).

There is also a growing body of work on density regression, where an unknown probability distribution is allowed to change flexibly with covariates using nonparametric mixture models (e.g. Dunson, Pillai, and Park 2007). We do not attempt a comprehensive review of this literature, which has different goals than the present application. For example, one of the key issues that arises in multiple testing is the need to limit the flexibility of the model so that the null and alternative hypotheses are identifiable. Ensuring this property is not trivial; see Martin and Tokdar (2012). In density regression, on the other hand, only the overall density is of interest; the mixture components themselves are rarely identifiable.

Our application draws most directly on Kelly et al. (2010) and Kelly and Kass (2012). We review other relevant neuroscience literature in Section 4.

2. FITTING THE FDR REGRESSION MODEL

2.1 An Empirical-Bayes Approach

We use a version of the FDR regression model where

$$\begin{aligned} z_i &\sim c(x_i)\cdot f_1(z_i) + \{1-c(x_i)\}\cdot f_0(z_i) \\ c(x_i) &= \frac{1}{1+\exp\{-s(x_i)\}} \\ f_0(z) &= \mathrm{N}(z \mid \mu, \sigma^2) \\ f_1(z) &= \int_{\mathcal{R}} \mathrm{N}(z \mid \mu+\theta, \sigma^2)\,\pi(\theta)\,d\theta\,. \end{aligned} \tag{5}$$

We have assumed a logistic link and a Gaussian error model, both of which could be modified to suit a different problem. We also assume a linear regression where $s(x) = x^T\beta$, and therefore model nonlinear functions by incorporating a flexible basis set into the covariate vector x. Both μ and σ^2 are initially assumed to be known; in Section 2.4, we describe how to weaken this assumption by estimating an empirical null, in the spirit of Efron (2004).

The unknown parameters of the FDR regression model that must be estimated are the regression coefficients β and the mixing distribution $\pi(\theta)$. This section describes two methods—one empirical-Bayes, one fully Bayes—for doing so. Both methods are implemented in the R package `FDRreg`.

Our empirical-Bayes approach begins with a precomputed plug-in estimate for $\pi(\theta)$ in (5), ignoring the covariates. This is equivalent to assuming that $c(x_i) \equiv c$ for all i, albeit only for the purpose of estimating $\pi(\theta)$. Many methods could be used for this purpose, including finite mixture models. We recommend the predictive-recursion algorithm of Newton (2002) for two reasons: speed, and the strong guarantees of accuracy proven by Tokdar, Martin, and Ghosh (2009). Predictive recursion generates a nonparametric estimate $\hat{\pi}(\theta)$, and therefore an estimate $\hat{f}_1(z)$ for the marginal density under the alternative, after a small number of passes (typically 5–10) through the data. The algorithm itself is similar to stochastic gradient descent, and is reviewed in Appendix 6.

Algorithm 1 EM for FDR regression using a plug-in estimate $\hat{f}_1(z)$. To estimate $\hat{f}_1$, we use predictive recursion (Appendix 6).

Data: Test statistics $z_1, \ldots, z_n$
Input: Densities $f_0(z), f_1(z)$; initial guess $\beta^{(0)}$
Output: Estimated coefficients β and posterior probabilities w_i
While *not converged* **do**

E step: Update $Q(\beta) = E\{l(\beta) \mid \beta^{(t)}\}$ as

$$Q^{(t)}(\beta) = \sum_{i=1}^{n}\left\{w_i^{(t)} x_i^T\beta - \log\left(1+e^{x_i^T\beta}\right)\right\}$$

$$w_i^{(t)} = E(h_i \mid \beta^{(t)}, z_i) = \frac{c(x_i)\cdot f_1(z_i)}{c(x_i)\cdot f_1(z_i) + \{1-c(x_i)\}\cdot f_0(z_i)}$$

$$c(x_i) = \frac{1}{1+\exp\{-x_i^T\beta^{(t)}\}}\,.$$

M step: Update β as

$$\beta^{(t+1)} = \arg\max_{\beta\in\mathcal{R}^d} Q^{(t)}(\beta)$$

using the Newton Raphson method.

end

Upon fixing this empirical-Bayes estimate $\hat{f}_1(z)$, and assuming that $f_0(z)$ is known, we can fit the FDR regression model by expectation-maximization (EM, Dempster, Laird, and Rubin 1977). To carry this out, we introduce binary latent variables h_i such that

$$\begin{aligned} z_i &\sim \begin{cases} f_1(z_i) \text{ if } h_i = 1\,, \\ f_0(z_i) \text{ if } h_i = 0 \end{cases} \\ \mathrm{P}(h_i = 1) &= \frac{1}{1+\exp\{-x_i^T\beta\}}\,. \end{aligned}$$

Marginalizing out each h_i clearly recovers the original model (4). The complete-data log-likelihood for β is

$$l(\beta) = \sum_{i=1}^{n}\left\{h_i x_i^T\beta - \log\left(1+e^{x_i^T\beta}\right)\right\}\,.$$

This is a smooth, concave function of β whose gradient and Hessian matrix are available in closed form. It is therefore easily maximized using standard methods, such as the Newton–Raphson algorithm. Moreover, $l(\beta)$ is linear in h_i, and the conditional expected value for h_i, given β, is just the conditional probability that $h_i = 1$:

$$w_i = E(h_i \mid \beta, z_i) = \frac{c(x_i)\cdot f_1(z_i)}{c(x_i)\cdot f_1(z_i) + \{1-c(x_i)\}\cdot f_0(z_i)}\,. \tag{6}$$

These facts lead to a simple EM algorithm for fitting the model (Algorithm 1; see box).

Thus the overall approach for estimating Model (5) has three steps.

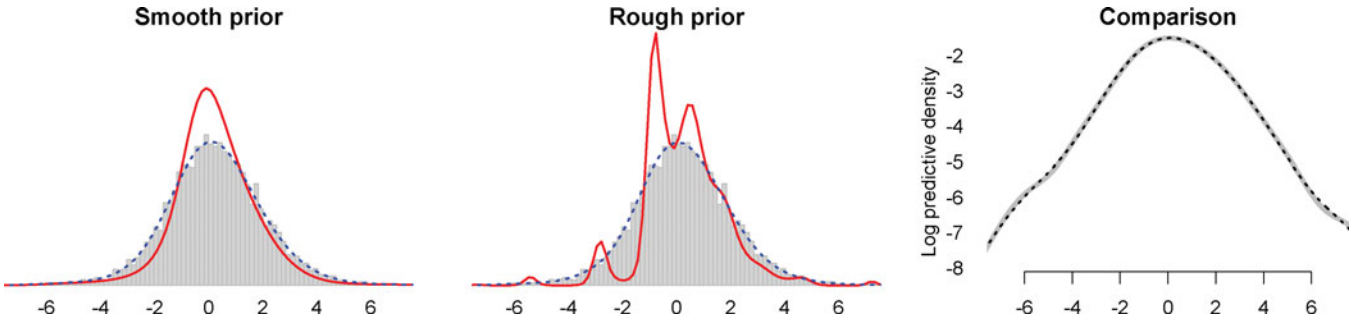


Figure 1. Two deconvolutions, one with a smooth prior (left) and one with a rough prior (middle). The solid red line is the prior, and the dotted blue line is the convolution (predictive density). The right panel shows how similar the two predictive densities are on a log scale. The solid gray line is the convolution of the rough prior, while the dotted black line is the convolution of the smooth prior.

1. Fix μ and σ^2 under the null hypothesis, or estimate an empirical null (see Section 2.4), thereby defining $f_0(z)$.
2. Use predictive recursion to estimate $\pi(\theta)$, and therefore $f_1(z)$, under the two-groups model without covariate effects (see Appendix 6).
3. Use $f_0(z)$ and $f_1(z)$ in Algorithm 1 to estimate w_i and the regression coefficients.

In principle, the estimate for $\pi(\theta)$ could be improved by using the information in the covariates. Despite this, our experiments show that the empirical-Bayes approach is essentially just as effective as using a full Bayes approach to estimate $\pi(\theta)$ and the regression function jointly. This can be explained by the fact that Gaussian deconvolution is such a poorly conditioned problem. For example, consider the simulated data in Figure 1. In the left panel, a simple Gaussian deconvolution of the form $f(y) = \int N(y \mid \theta, 1)\pi(\theta)d\theta$ has been fit to the data in the histogram, assuming that $\pi(\theta)$ is supported on discrete grid. A smoothness penalty has been imposed on $\pi(\theta)$, yielding a smooth deconvolution. In the middle panel, the same model has been fit to the data, but the smoothness penalty has been relaxed, yielding a rougher deconvolution. In each panel, the solid red line shows the estimated $\pi(\theta)$, and the dotted blue line shows the predictive density—that is, the Gaussian convolution of $\pi(\theta)$. The priors themselves are quite different, but their convolutions are nearly indistinguishable, and both fit the histogram well. The right panel shows the two predictive densities of the two priors on a log scale, confirming how similar they are.

This example illustrates a more general point about Gaussian deconvolution. For any finite set of observations from $f_1(z)$ in (5), there is a large "near-equivalence" class of approximately compatible priors $\pi(\theta)$. Because the posterior probability in (6) depends on $\pi(\theta)$ only through $f_1(z)$, any prior in this near-equivalence class will yield nearly the same set of answers. In this sense, our approach based on predictive recursion seems to provide a "good enough" estimate for $\pi(\theta)$, despite ignoring the covariates.

2.2 Empirical Bayes with Marginalization

In Section 2.1, we ignored the covariates to estimate $f_1(z)$. But the most direct generalization of the local-FDR approach of Efron et al. (2001) would be to ignore the covariates and estimate the overall mixture density $f(z)$ instead. We now explain why this is a poor solution to the FDR regression problem. Let us begin with the key insight in Efron et al.'s approach to estimating local FDR: the marginal $f(z)$ is common to all test statistics and can be estimated well using the empirical distribution of the z_i, without explicit deconvolution of the mixture. This motivates a simple empirical-Bayes strategy: (i) compute a nonparametric estimate $\hat{f}(z)$ of the marginal density, along with a likelihood- or moment-based estimate $\hat{c}$ of the mixing fraction; and (ii) plug $\hat{f}$ and $\hat{c}$ into the marginalized form of the posterior probability (3) to get local FDR for each test statistic. One caveat is that $\hat{c}$ must be chosen to ensure that (3) falls on the unit interval for all i. But this is an easy constraint to impose during estimation. See Efron (2012, chap. 5) for further details.

We have already remarked that the posterior probabilities in the FDR regression model (4) do not share a common mixture density $f(z)$, and so cannot be expressed in marginalized form. Nonetheless, it is natural to wonder what happens if we simply ignore this fact, estimate a global $\hat{f}(z)$ from the empirical distribution of the test statistics, and use

$$1 - w_i^{(t)} = \frac{\{1 - c(x_i)\} f_0(z_i)}{\hat{f}(z_i)} \tag{7}$$

in lieu of expression (6) used in Algorithm 1. This has the seemingly desirable feature that it avoids the difficulties of explicit deconvolution. But (7) is not guaranteed to lie on the unit interval, and constraining it to do so is much more complicated than in the no-covariates case (3). Moreover, simply truncating (7) to the unit interval during the course of the estimation procedure leads to very poor answers.

In our simulation studies, we do consider the following ad hoc modification of (7), in an attempt to mimic the original local-FDR approach as closely as possible:

$$1 - w_i^{(t)} = \begin{cases} 1 & \text{if } (1 - \hat{c}) f_0(z_i) \geq \hat{f}(z_i), \\ T_u\left[\frac{\{1-c(x_i)\} f_0(z_i)}{\hat{f}(z_i)}\right] & \text{otherwise,} \end{cases} \tag{8}$$

where $T_u(a)$ is the projection of a to the unit interval, while $\hat{f}(z)$ and $\hat{c}$ are plug-in estimates of the marginal density and the mixing fraction using the no-covariates method described by Efron (2012, chap. 5). In our simulation studies, this modification (despite no longer being a valid EM algorithm) does give stable answers with qualitatively correct covariate effects. But because it zeroes out the posterior probabilities for all z_i within a neighborhood of the origin, it yields heavily biased estimates

Downloaded by [Carnegie Mellon University] at 20:17 06 August 2015

Downloaded by [Carnegie Mellon University] at 20:17 06 August 2015

for β and, in our studies, is less powerful than the method of Section 2.1.

2.3 Full Bayes

From a Bayesian perspective, the hierarchical model

$$\begin{aligned}(z_i \mid \theta_i) &\sim \mathrm{N}(\mu + \theta_i, \sigma^2)\\ (\theta_i \mid h_i) &\sim h_i \cdot \pi(\theta_i) + \{1 - h_i\} \cdot \delta_0\\ \mathrm{P}(h_i = 1) = c(x_i) &= \frac{1}{1 + \exp(-x_i^T \beta)}\,, \end{aligned} \tag{9}$$

together with priors for β and the unknown distribution $\pi(\theta)$, defines a joint posterior distribution over all model parameters. We use a Markov chain Monte Carlo algorithm to sample from this posterior, drawing iteratively from three complete conditional distributions: for the mixing density $\pi(\theta)$; for the latent binary variables h_i that indicate whether z_i is signal or null; and for the regression coefficients β.

An important question is how to parameterize $\pi(\theta)$. In the no-covariates multiple-testing problem, there have been many proposals, including simple parametric families (Scott and Berger 2006; Polson and Scott 2012) and nonparametric priors based on mixtures of Dirichlet processes (Do, Müller, and Tang 2005). In principle, any of these methods could be used. In our analyses, we model $\pi(\theta)$ as a K-component mixture of Gaussians with unknown means, variances, and weights. We choose K via a preliminary run of the EM algorithm for deconvolution mixture models, picking the K that minimizes the Akaike information criterion (AIC). (In simulations, we found that AIC was slightly better than BIC at recovering K for the deconvolution problem, as distinct from the ordinary density-estimation problem.)

The model's chief computational difficulty is the analytically inconvenient form of the conditional posterior distribution for β. The two major issues here are that the response h_i depends nonlinearly on the parameters, and that there is no natural conjugate prior to facilitate posterior computation. These issues are present in all Bayesian analyses of the logit model, and have typically been handled using the Metropolis–Hastings algorithm. A third issue, particular to our setting, is that the binary event h_i is a latent variable, and not actually observed.

We proceed by exploiting the Pólya-Gamma data-augmentation scheme for binomial models recently proposed by Polson, Scott, and Windle (2013). Let x_i be the vector of covariates for test statistic i, including an intercept term. The complete conditional for β depends only $h = \{h_i : i = 1, \ldots, n\}$, and may be written as

$$p(\beta \mid h) \propto p(\beta) \prod_{i=1}^{n} \frac{(e^{x_i^T \beta})^{h_i}}{1 + e^{x_i^T \beta}}\,,$$

where $p(\beta)$ is the prior. By introducing latent variables $\omega_i \sim \mathrm{PG}(1, 0)$, each having a standard Pólya-Gamma distribution, we may reexpress each term in the above product as the marginal of a more convenient joint density:

$$\frac{(e^{x_i^T \beta})^{h_i}}{1 + e^{x_i^T \beta}} \propto e^{\kappa_i x_i^T \beta} \int_0^\infty e^{-\omega_i (x_i^T \beta)^2/2}\; p(\omega_i)\, d\omega_i\,,$$

where $\kappa_i = h_i - 1/2$ and $p(\omega)$ is the density of a PG(1,0) variate. Assuming a normal prior $\beta \sim \mathrm{N}(c, D)$, it can be shown that β has a conditionally Gaussian distribution, given the diagonal matrix $\Omega = \mathrm{diag}(\omega_1, \ldots, \omega_n)$. Moreover, the conditional for each ω_i, given β, is also in the Pólya-Gamma family, and may be efficiently simulated.

Together with standard results on mixture models, the Pólya-Gamma scheme leads to a simple, efficient Gibbs sampler for the fully Bayesian FDR regression model. Further details of the Bayesian method can be found in Appendix 6, including the priors we use, the conditionals needed for sampling, and the default settings implemented in `FDRreg`.

On both simulated and real datasets, we have observed that the empirical-Bayes and fully Bayes approaches give very similar answers for the local false-discovery rates, and thus reach similar conclusions about which cases are significant. The advantage of the Bayesian approach is that it provides a natural way to quantify uncertainty about the regression function $s(x)$ and $\pi(\theta)$ jointly. This is counterbalanced by the additional computational complexity of the fully Bayesian method.

2.4 Using an Empirical Null

The FDR regression model (5) assumes that μ and σ^2 are both known, or can be derived from the distributional theory of the test statistic in question. But as Efron (2004) observed, many datasets are poorly described by this "theoretical null," and it is necessary to estimate an empirical null to reach plausible conclusions. This is a common situation in high-dimensional screening problems, where correlation among the test statistics, along with many other factors, can invalidate the theoretical null.

Estimating both an empirical null f_0 and an empirical alternative f_1 in (1) would initially seem to involve an intractable identifiability problem. For example, one cannot estimate both the signal and noise densities in a fully nonparametric manner, or else the concepts of null and alternative hypotheses become meaningless. This is a variant of the label-switching issue that arises in estimating finite mixture models: without some nontrivial identifying restriction, nothing in the two-groups model (1) encodes the scientific idea that signals are rarer and systematically more extreme than the test statistics seen under the null.

The existing literature suggests two ways of solving the identifiability problem in estimating an empirical null. We refer to these as the zero assumption and the tail assumption.

The zero assumption expresses the idea that essentially all of the test statistics near the origin are noise. Efron (2004) proposed two methods for exploiting this assumption to estimate an empirical Gaussian null: maximum likelihood and central matching. Both approaches share the idea of using the shape of the histogram near zero to fit a local Gaussian distribution. Either can be incorporated into the empirical-Bayes FDR regression method of Section 2.1 as a simple preprocessing step to estimate μ and σ^2. The central-matching approach is used later in our analysis of the neural synchrony data, and is offered as an option in `FDRreg`. This two-stage approach may be slightly less efficient than performing a full Bayesian analysis, but we believe the improved stability is worth the trade-off.

As an alternative, Martin and Tokdar (2012) proposed the tail assumption: the idea that the signals have heavier tails than the null distribution. The authors provided a lengthy mathe-

Downloaded by [Carnegie Mellon University] at 20:17 06 August 2015

matical discussion of the identifiability issue, with their main theorem establishing that f_0 and f_1 are jointly identifiable if: (1) the null test statistics come from a normal distribution with unknown mean and variance, and (2) the signal test statistics are an unknown, compactly supported location mixture of the null distribution. Their discussion emphasizes two key intuitive requirements for identifiability: that the tail weight of the null distribution is known, and that the alternative is heavier tailed than the null.

An important theoretical limitation of both approaches is that they restrict the null hypothesis to be a Gaussian distribution, albeit one with unknown mean and variance. Ideally, we would be able to estimate other properties of the empirical null beyond its first two moments, such as its skewness or rate of tail decay. But doing so in full generality—even in the much simpler no-covariates version of the two-groups model—does not seem possible using currently available techniques, and may not be possible even in principle. An important open research question that has gone unaddressed, both here and in the wider literature, is: what are the weakest assumptions on f_0 and f_1 necessary to identify the null? Martin and Tokdar (2012) provided the weakest set of sufficient conditions we are aware of. (Here we mean "weakest" in a scientific rather than mathematical sense—neither the zero assumption nor the tail assumption implies the other.) Moreover, it seems unlikely that their conditions could be drastically weakened. To take an extreme example, if the null hypothesis were allowed to have an arbitrary unknown tail weight, then nothing would rule out a Cauchy distribution. In this case, we could not reject the null even for a z-statistic in the hundreds or thousands. In cases where there are serious doubts that the Gaussian null assumption is broadly sensible, one approach would be to conduct a smaller pilot experiment in which the null is known to hold, so that the problem of identification is absent.

It is reasonable to conjecture that both the zero assumption and the tail assumption could be generalized to handle the case where the null is any known location-scale family. Martin and Tokdar (2012) gave the specific example of a Student's-t null with fixed, known degrees of freedom. But even if true, this generalization could be exploited only by a data analyst with prior knowledge of the functional form of the null distribution. In the rare case where such prior knowledge is available, we recommend that it be incorporated into the analysis via a suitable modification of (4).

In our motivating example, each test statistic z_i arises from a maximum-likelihood estimate in a generalized-linear model with a modest number of parameters, fit to hundreds or thousands of neural spikes. The large-sample properties of the MLE in generalized linear models suggest that our z_i will be very close to Gaussian under the null, and so we are comfortable with this version of the model. At the same time, we are sensitive to the possibility of under- or over-dispersion versus a standard normal, for the reasons discussed by Efron (2008a) and the references contained therein. We therefore fit an empirical null using the central-matching method (see Section 5).

3. SIMULATIONS

This section presents the results of a simulation study that confirms the advantage of false-discovery-rate regression in problems where signals cluster in covariate space. We simulated datasets having two covariates $x_i = (x_{i1}, x_{i2})$, with each test statistic z_i drawn according to the covariate-dependent mixture model (5). We considered five choices for $s(x)$:

(A) $s(x) = -3 + 1.5x_1 + 1.5x_2$
(B) $s(x) = -3.25 + 3.5x_1^2 - 3.5x_2^2$
(C) $s(x) = -1.5(x_1 - 0.5)^2 - 5|x_2|$
(D) $s(x) = -4.25 + 2x_1^2 + 2x_2^2 - 2x_1x_2$
(E) $s(x) = -3$

These choices are shown in the left four panels of Figure 2. Function A is linear; Functions B and C are nonlinear but

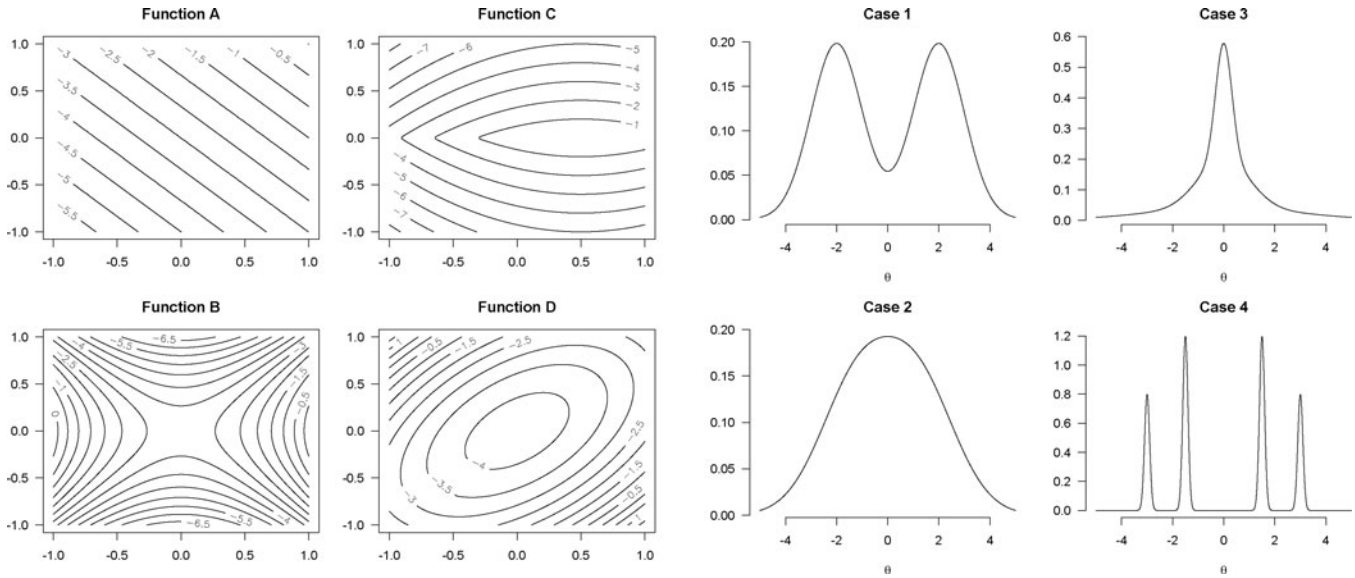


Figure 2. Settings for the simulation study. Left four panels: choices for the bivariate regression function $s(x_1, x_2)$. The contours show the prior log odds that a test statistic in that part of covariate space will be a signal. Function E, not shown, is a flat function: $s(x) = -3$. Right four panels: choices for $\pi(\theta)$.

Downloaded by [Carnegie Mellon University] at 20:17 06 August 2015

additive in x_1 and x_2; Function D is neither linear nor additive. Function E, not shown, is the flat function $s(x) = -3$. This is included to understand the behavior of FDR regression when it is inappropriately applied to a dataset with no covariate effects. The parameters for each function were chosen so that between 6% and 10% of the z_i were drawn from the nonnull signal population $f_1(z)$.

We also considered four choices for $\pi(\theta)$, all discrete mixtures of Gaussians $\mathrm{N}(\mu, \tau^2)$:

1. $\pi(\theta) = 0.48 \cdot \mathrm{N}(-2, 1) + 0.04 \cdot \mathrm{N}(0, 16) + 0.48 \cdot \mathrm{N}(2, 1)$,
2. $\pi(\theta) = 0.4 \cdot \mathrm{N}(-1.25, 2) + 0.2 \cdot \mathrm{N}(0, 4) + 0.4 \cdot \mathrm{N}(1.25, 2)$,
3. $\pi(\theta) = 0.3 \cdot \mathrm{N}(0, 0.1) + 0.4 \cdot \mathrm{N}(0, 1) + 0.3 \cdot \mathrm{N}(0, 9)$,
4. $\pi(\theta) = 0.2 \cdot \mathrm{N}(-3, 0.01) + 0.3 \cdot \mathrm{N}(-1.5, 0.01) + 0.3 \cdot \mathrm{N}(1.5, 0.01) + 0.2 \cdot \mathrm{N}(3, 0.01)$.

These choices for $\pi(\theta)$ are shown in the right four panels. Choices 1 and 4 have most of the nonnull signals separated from zero, and are thus easier problems overall. Choices 2 and 3 have most of the signals near zero, and are thus harder problems overall.

For each of the 20 possible combinations for $s(x)$ and $\pi(\theta)$ listed above, we simulated 100 datasets of $n = 10000$ test statistics. Each design point x_i was drawn uniformly from the unit cube, $[0, 1]^2$. To each simulated dataset, we applied the following methods.

BH: the Benjamini–Hochberg procedure (Benjamini and Hochberg 1995).

2G: the two-groups model described in Chapter 5 of Efron (2012).

EB: Empirical Bayes FDR regression using predictive recursion (Section 2.1).

EBm: Empirical Bayes FDR regression with ad hoc marginalization (Section 2.2).

FB: Fully Bayes FDR regression using Pólya-Gamma data-augmentation (Section 2.3).

For the EB, EBm, and FB methods, we fit a nonlinear additive model by expanding each covariate in a B-spline basis with five equally spaced knots. Because additivity is assumed, these methods cannot recover Function D exactly, making this a useful robustness check. The theoretical N(0, 1) null was assumed in all cases. All methods are implemented in the R package `FDRreg`, and the R script used for the study is available as a supplemental file to the article.

In each case, we selected a set of nonnull signals by attempting to control the global false discovery rate at 10% using each method (see, e.g., Morris et al. 2008) We then calculated both the realized false discovery rate and the realized true positive rate (TPR) by comparing the selected test statistics to the truth. Recall that the TPR is defined to be the number of true signals discovered, as a fraction of the total number of true signals.

The results, shown in Table 1 and Figure 3, support several conclusions. First, the FDR regression methods (EB, EBm, and FB) effectively control the false discovery rate at the nominal level (10%). When covariate effects are present, the empirical-Bayes method violates the nominal level about as often as the other methods. It does so with greater frequency only when covariate effects are absent (function E). Even then, the

Table 1. Results of the simulation study

		False discovery rate (%)					True positive rate (%)				
$\pi(\theta)$	$s(x)$	BH	2G	EBm	EB	FB	BH	2G	EBm	EB	FB
1	A	8.9	9.0	9.2	9.7	*10.7	22.5	22.4	24.0	30.1	31.0
	B	9.5	9.5	9.4	9.7	*11.1	21.8	21.7	23.7	32.5	34.2
	C	9.5	9.5	9.4	9.5	10.2	22.8	22.4	24.2	33.3	34.3
	D	9.3	9.3	9.5	9.7	*11.2	22.3	22.0	23.7	29.2	30.4
	E	9.4	9.0	9.7	*11.0	10.2	18.0	17.4	18.1	18.7	18.0
2	A	9.2	9.1	9.3	9.7	10.2	13.5	13.4	14.0	18.0	18.6
	B	8.6	8.8	8.7	9.2	10.4	13.0	13.1	13.7	19.0	20.2
	C	9.2	9.3	9.4	9.3	10.3	13.7	13.6	14.3	19.9	20.9
	D	9.3	9.3	9.5	9.7	10.6	13.8	13.7	14.4	17.7	18.3
	E	9.6	8.8	9.6	*10.8	9.3	11.3	10.9	11.3	11.7	11.1
3	A	8.9	*10.6	9.4	9.0	8.6	9.4	9.7	9.7	11.1	11.0
	B	9.0	10.4	9.5	8.7	8.9	9.3	9.6	9.5	11.6	11.8
	C	8.6	10.0	9.0	8.4	7.9	9.6	9.9	9.8	11.8	11.7
	D	9.2	*10.7	9.8	9.3	9.0	9.8	10.0	10.0	11.3	11.2
	E	10.0	*10.9	10.5	*11.3	9.3	8.7	8.7	8.6	8.7	8.4
4	A	9.0	9.1	9.1	10.3	*10.9	21.8	22.0	23.7	30.8	31.6
	B	9.4	9.4	9.4	10.3	*11.1	21.8	21.7	24.1	33.8	34.7
	C	9.2	9.5	9.6	10.1	*10.4	22.5	22.7	24.6	34.7	35.2
	D	9.0	9.3	9.4	9.9	*11.1	22.3	22.5	24.1	30.3	31.1
	E	9.9	9.3	10.1	*10.9	10.0	17.0	16.2	17.0	17.5	06.7

NOTE: The rows show different configurations for $\pi(\theta)$ and $s(x)$; see Figure 2. The columns show the realized false discovery rate and true positive rate for the five different procedures listed in Section 3. The rates are shown as percentages, with results averaged over 100 simulated datasets. In all cases, the nominal FDR was controlled at the 10% level. FDR entries marked with a star are significantly larger than the nominal level of 10%, as judged by a one-sided t-test ($p < 0.05$).

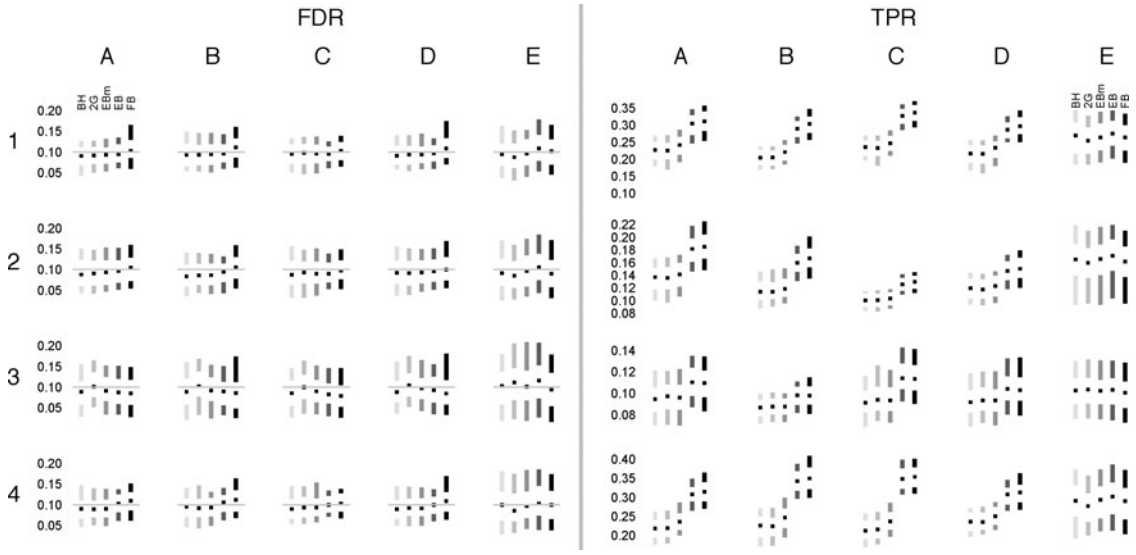


Figure 3. Boxplots of false discovery rate (FDR, left 20 panels) and true positive rate (TPR, right 20 panels) for the simulation study. The rows are the different true priors $\pi(\theta)$, and the columns are the different true regression function $s(x_1, x_2)$. Within each panel, the boxplots show the median, interquartile range (white bar), and range (gray bar) for each method across the 100 simulated datasets. Within each panel, the methods are arranged from left to right as: Benjamini–Hochberg (BH), the two-groups model without covariates (2G), empirical Bayes with marginalization (EBm, Section 2.2), empirical Bayes (EB, Section 2.1), and fully Bayes (FB, Section 2.3).

Downloaded by [Carnegie Mellon University] at 20:17 06 August 2015

average FDR across the different simulated datasets is only slightly higher than the nominal level (e.g., 11% versus 10%). To put this in context, 1% is much smaller than the variability of the realized FDR from the Benjamini–Hochberg method, which ranges between 5% and 15% across the datasets within a given scenario. See the left 20 panels of Figure 3.

When covariate effects are present, FDR regression has better power than existing methods for a fixed level of desired FDR control. The amount of improvement depends on the situation. For priors 1 and 4, the improvement was substantial: usually between 40–50% more discoveries, which is similar to what we see in the real-data illustration of the next section. For priors 2 and 3, the power gains of FDR regression were more modest, but still noticeable. These broad trends were consistent across the different functions.

The empirical-Bayes and fully Bayes methods perform very similarly overall. The only noticeable difference is that, when covariate effects are absent (Function E), the empirical Bayes method violates the nominal FDR level slightly more often than the fully Bayes method. We do not understand why this is so, but as the boxplots in Figure 3 show, the effect is small in absolute terms. They also show that the performance of the empirical-Bayes method in these cases is similar in this respect to the two-groups model without covariates (labeled 2G in the plots).

4. DETECTING NEURAL SYNCHRONY

4.1 Background

The ability to record dozens or hundreds of neural spike trains simultaneously has posed many new challenges for data analysis in neuroscience (Brown, Kass, and Mitra 2004; Buzsaki 2004; Aertsen 2010; Stevenson and Kording 2011). Among these, the problem of identifying neural interactions has, since the advent of multi-unit recording, been recognized as centrally important (Perkel, Gerstein, and Moore 1967). Neural interactions may occur on sub-behavioral timescales, where two neurons may fire repeatedly within a few milliseconds of each other. It has been proposed that such fine-timescale synchrony is crucial for binding visual objects (see Gray 1999; Shadlen and Movshon 1999, for opposing views), enhancing the strength of communication between groups of neurons (Niebur, Hsiao, and Johnson 2002), and coordinating the activity of multiple brain regions (Fries 2009; Saalmann and Kastner 2011). It has also been argued that the disruption of synchrony may play a role in cognitive dysfunctions and brain disorders (Uhlhaas et al. 2009).

Furthermore, there is growing recognition that synchrony and other forms of correlated spiking have an impact on population coding (Averbeck, Latham, and Pouget 2006) and decoding (Graf et al. 2011). The proposed roles of neural synchrony in numerous computational processes and models of coding and decoding, combined with the knowledge that the amount of synchrony can depend on stimulus identity and strength (Kohn and Smith 2005) as well as the neuronal separation (Smith and Kohn 2008), make it particularly important that we have effective tools for measuring synchrony and determining how it varies under different experimental paradigms.

Rigorous statistical detection of synchrony in the activity of two neurons requires formulation of a statistical model that specifies the stochastic behavior of the two neural spike trains under the assumption that they are conditionally independent, given some suitable statistics or covariates (Harrison, Amarasingham, and Kass 2013). When n neural spike trains are recorded there $N = \binom{n}{2}$ null hypotheses to be tested, which raises the problem of multiplicity. In the face of this difficulty, the false-discovery rate could be controlled using the Benjamini-Hochberg procedure Benjamini and Hochberg (1995), combining all N test statistic into a single analysis. But this omnibus approach ignores potentially useful information about the spatial and functional relationships among individual neuron pairs. We therefore use false discovery rate regression to incorporate these covariates

Downloaded by [Carnegie Mellon University] at 20:17 06 August 2015

into an investigation of synchrony in the primary visual cortex (V1). Specifically, we analyzed data from V1 neurons recorded from an anesthetized monkey in response to visual stimuli consisting of drifting sinusoidal gratings, that is, light whose luminance was governed by a sine wave that moved along an axis having a particular orientation. Details of the experiment and recording technique may be found in Kelly et al. (2007). Drifting gratings are known to drive many V1 neurons to fire at a rate that depends on orientation. Thus, many V1 neurons will have a maximal firing rate for a grating in a particular orientation (which is the same as an orientation rotated by 180 degrees) and a minimal firing rate when the grating orientation is rotated by 90 degrees. For a given neuron, a plot of average firing rate against angle of orientation produces what is known as the neuron's "tuning curve." Spike trains from 128 neurons were recorded in response to gratings in 98 equally spaced orientations, across 125 complete replications of the experiment (125 trials). Here we analyze data from the first 3 sec of each 30-second trial. The 128 neurons generated 8,128 pairs, and thus 8,128 tests of synchrony. We applied the model in Equation (4) to examine the way the probability of synchrony $c(x)$ for a pair of neurons depends on two covariates: the distance between the neurons and the correlation of their tuning curves (i.e., the Pearson correlation between the two vectors of length 98 that contain average firing rate as a function of orientation). The idea is that when neurons are close together, or have similar tuning curves, they may be more likely to share inputs and thus more likely to produce synchronous spikes, compared to the number predicted under conditional independence. Our analysis, reported below, substantiates the observation of Smith and Kohn (2008) that the probability of fine time-scale synchrony for pairs of V1 neurons tends to decrease with the distance between the two neurons and increase with the magnitude of tuning-curve correlation.

4.2 Data Preprocessing

Our analysis takes advantage of a recently developed technique for measuring synchrony across binned spike trains (where the time bins are small, such as 5 milliseconds). For a pair of neurons labeled 1 and 2, we calculate

$$\hat{\xi} = \frac{\text{number of bins in which both neurons spike}}{\sum_t \hat{P}(\text{neuron 1 spikes at } t \mid D_t^{(1)}) \cdot \hat{P}(\text{neuron 2 spikes at } t \mid D_t^{(2)})}, \tag{10}$$

where $D_t^{(1)}$ and $D_t^{(2)}$ refer to relevant conditioning information for the firing activity of neurons 1 and 2 at time t, and the sum is over all time bins across all experimental trials (Kelly and Kass 2012). The denominator of (10) is an estimate of the number of joint spikes that would be expected, by chance, if $D_t^{(j)}$ characterized the spiking activity of neuron j (with $j = 1, 2$) and, apart from these background effects, the neurons were independent. When $\hat{\xi} \approx 1$, or $\log \hat{\xi} \approx 0$, the conclusion would be that the number of observed synchronous spikes is consistent with the prediction of synchronous spiking under independence, given the background effects $D^{(j)}$. Note that this conditioning information $D_t^{(j)}$ is intended to capture effects—including tuning curve information—on each neuron separately, whereas the covariates that enter the FDRR model (4) operate pairwise. Thus, for example, two independent neurons having similar tuning curves would both be driven to fire more rapidly by a grating stimulus in a particular orientation, and would therefore be likely to produce more synchronous spikes by chance than a pair of independent neurons with dissimilar tuning curves. As another example of conditioning information, under anesthesia there are pronounced periods during which most recorded neurons increase their firing rate (Brown, Lydic, and Schiff 2010). These waves of increased network activity have much lower frequency than many other physiological wave-like neural behaviors, and are called "slow waves." One would expect slow-wave activity to account for considerable synchrony, even if, conditionally on the slow-wave activity, a pair of neurons were independent. The statistic $\hat{\xi}$ in formula (10) is a maximum-likelihood estimator in the continuous-time framework discussed by Kass, Kelly, and Loh (2011), specifically their equation (22). The purpose of that framework, and of (10), is to describe the way synchrony might depend on background information. For example, Kass, Kelly, and Loh (2011) contrasted results from two pairs of V1 neurons. Both pairs exhibited highly statistically significantly enhanced synchrony above that predicted by stimulus effects (tuning curves) alone. However, the two pairs were very different with respect to the relationship of synchrony to slow-wave network activity. In one pair, when background information characterizing the presence of slow-wave network activity was used in (10), the enhanced synchrony vanished, with $\log \hat{\xi}_1 = 0.06 \pm 0.15$. In the other pair, it persisted with $\log \hat{\xi}_2 = 0.82 \pm 0.23$, indicating the number of synchronous spikes was more than double the number predicted by slow-wave network activity together with trial-averaged firing rate.

The set of 8,128 $\log \hat{\xi}$ coefficients analyzed here, together with their standard errors, was created with a model that differed in two ways from that used by Kass, Kelly, and Loh (2011). First, to better capture slow-wave network effects, in place of a linear model based on single count variable (for neurons i and j, Kass et al. used the total number of spikes within the past 50 ms time among all neurons other than neurons i and j) a general nonparametric function of the count was fitted, using splines. Second, a nonparametric function capturing spike history effects was used. This allows non-Poisson variability, which is important in many contexts (Kass, Eden, and Brown 2014), and exploratory analysis indicated that it is consequential for this dataset as well.

5. ANALYSIS AND RESULTS

As an illustration of the method, we apply false discovery rate regression to search for evidence of enhanced synchrony in a three-second window of recordings on these 128 V1 neurons. We emphasize that what we refer to here as "findings" or "discoveries" are necessarily tentative. As in many genomics data-analysis pipelines, additional follow-up work is clearly necessary to verify any individual discovery arising from an FDR-controlling analysis of a large-scale screening experiment. Nonetheless, because of their clear covariate effects, the V1 neural recordings provide a good illustration of the FDR regression method.

We use the subscript i to index a pair of neurons, to maintain notational consistency with the rest of the article. Let $y_i = \log \hat{\xi}_i$ denote the observed synchrony statistic for the ith pair of neurons being tested. This comes from Formula (10), after conditioning

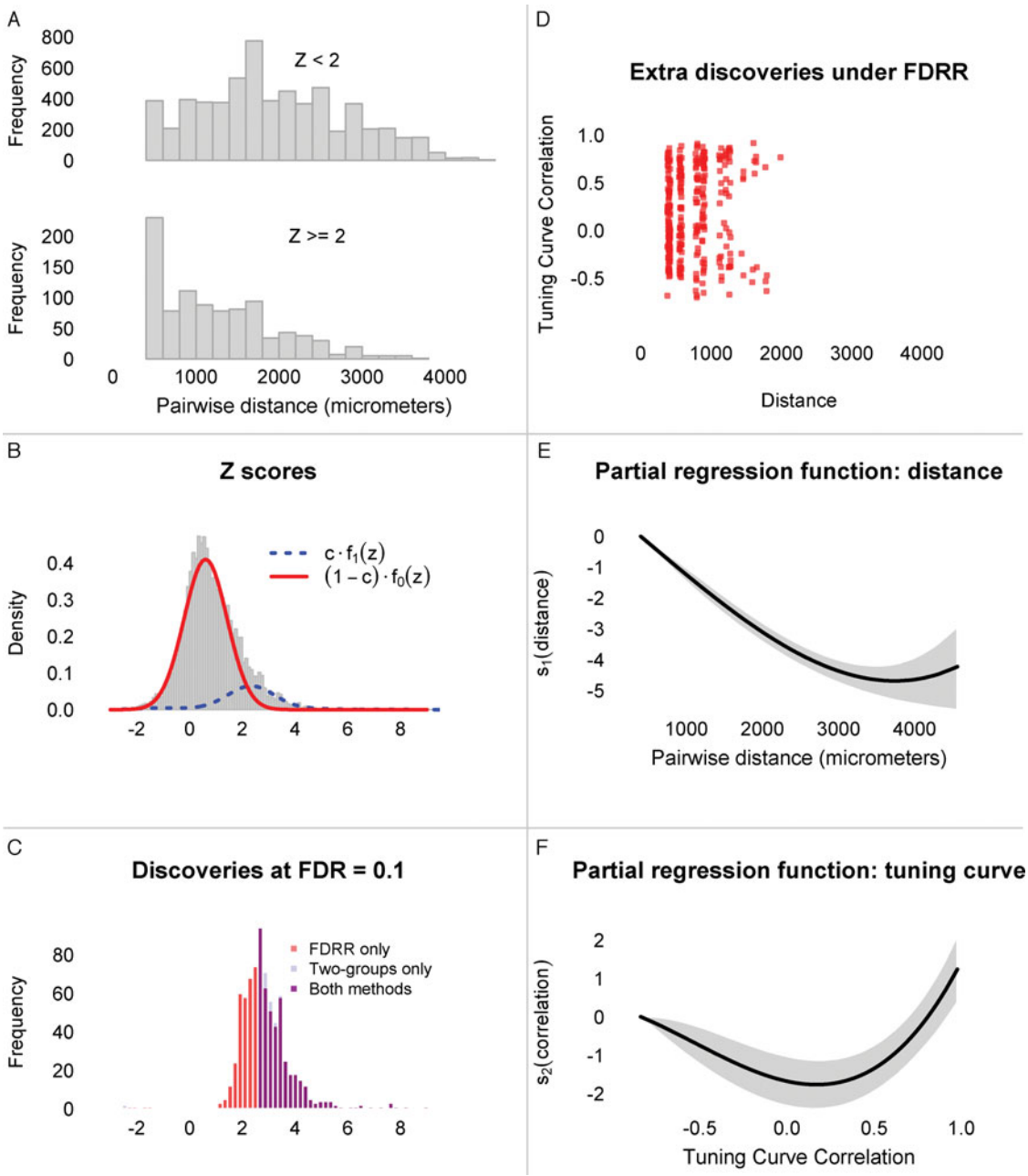


Figure 4. Panel A shows histograms of inter-neuron distances for pairs with synchrony z-score less than 2, versus those with z-score larger than 2. Panel B shows the empirical null density $f_0(z)$, together with the signal density $f_1(z)$ estimated by predictive recursion, superimposed upon the histogram of the raw z-scores. Panel C compares discoveries at the 10% FDR level using the ordinary two-groups local FDR model, versus those under the FDR regression model. Panel D shows that the extra discoveries made by FDR regression (red points) tend to concentrate at short inter-neuron distances compared with the rest of the neuron pairs. Panels E and F show the estimated partial regression functions for prior log-odds of being a signal versus distance and tuning curve correlation. The black lines are the estimates, and the gray areas show 95% posterior credible intervals arising from the full Bayes analysis.

on slow-wave activity. Let s_i denote the estimated standard error for $\log \hat{\zeta}_i$, which is obtained from a parametric bootstrap procedure, following Kass, Kelly, and Loh (2011) and Kelly and Kass (2012). We define $z_i = y_i/s_i$ as our test statistic, and assume that the z_i arise from Model (5). The pairs where $z_i \sim f_0(z)$ correspond to the null hypothesis of conditional independence, given slow-wave network activity. As previously mentioned, there are two relevant covariates: (1) inter-neuron distance, measured in micrometers; and (2) tuning-curve correlation (r_i).

Panel A of Figure 4 provides some initial exploratory evidence for a substantial distance effect. It shows two histograms of z-scores: one for neuron pairs where $z_i < 2$ (suggesting no synchrony enhancement), and another for neuron pairs where $z_i \geq 2$ (suggesting possible synchrony enhancement). It is clear from the figure that nearby neuron pairs are much more likely to have $z_i \geq 2$, versus neuron pairs at a longer distance. This motivates the use of covariate-dependent prior probabilities in (5).

Downloaded by [Carnegie Mellon University] at 20:17 06 August 2015

Downloaded by [Carnegie Mellon University] at 20:17 06 August 2015

To fit the FDR regression model, we first estimated an empirical null $f_0(z)$, as described in Section 2.4. This was necessary because the empirical distribution of z-scores was poorly described by a standard normal density. We used the maximum-likelihood method from Efron (2004), which yielded $\mu = 0.61$ and $\sigma = 0.81$. This suggested underdispersion and a positive bias versus the theoretical N(0, 1) null. Fixing μ and σ at these estimated values, we then ran predictive recursion to estimate $f_1(z)$, as described in Section 2.1. Panel B of Figure 4 shows the estimates for $f_0(z)$ (solid red line) and $f_1(z)$ (dashed blue line), scaled by the empirical-Bayes estimate of the mixing fraction c in the two-groups model (1), and superimposed on the histogram of z-scores. The alternative hypothesis appears to be dominated by cases where $z_i > 0$.

Having computed estimates for $f_0(z)$ and $f_1(z)$, we then used the empirical-Bayes method of Section 2.1 to estimate the FDR regression model by expectation–maximization. We assumed that the prior log odds of synchrony ($\theta_i \neq 0$) could be described by an additive model involving distance and tuning-curve correlation:

$$s(x) = \beta_0 + s_1(\text{distance}) + s_2(\text{correlation}).$$

The partial regression functions were modeled by expanding each covariate in a B-spline basis, with the degrees of freedom chosen to minimize AIC. To regularize the estimates, we used $N(0, 1)$ priors on the spline coefficients. The partial regression functions are identified only up to additive constants. To identify them, we estimated an overall intercept β_0, and fixed s_1 and s_2 to be zero at their left-most endpoints. As a robustness check, we also ran the full Bayes method, which does not require a precomputed estimate for $f_1(z)$. We focus mainly on results for the empirical-Bayes approach, but the fully Bayes estimates of local FDR were very similar, and we use the full Bayes method to construct confidence bands for the underlying regression function.

We controlled the (Bayesian) false discovery rate at the 10% level, and compared the resulting discoveries under the FDR regression model to those under the ordinary two-groups (local FDR) model without covariate effects. The regression model yielded roughly 50% more discoveries compared to the two-groups model, 763 versus 489. Panels C and D of Figure 4 show that these extra discoveries tend to be at the borderline of statistical significance ($z_i \approx 2$), but heavily concentrated at short distances, where the prior odds of a significant z-score are much higher.

Panels E and F of Figure 4 show the estimated partial regression functions s_1 and s_2, together with 95% posterior credible intervals derived from the fully Bayesian posterior distribution of the spline coefficients. The distance effect suggested by Panel A is confirmed by the confidence bands of the partial regression function for distance in Panel E. Tuning-curve correlation also appears to play a role in the prior odds of synchrony, with its effect roughly (though not exactly) symmetric about zero. To provide intuition about the magnitude of the covariate effects, we compare two sets of pairs.

- A neuron pair at distance 2433 micrometers (the 75th percentile), and with tuning-curve correlation of 0.12 (the median), was estimated to have a 2.3% prior probability of being a nonnull signal. A neuron pair with the same tuning curve correlation but separated by only 1200 micrometers (the 25th percentile) was estimated to have a 16.5% prior probability of being a nonnull signal.
- A neuron pair at distance 1789 micrometers (the median), and with tuning-curve correlation of 0, was estimated to have a 6.9% prior probability of being a nonnull signal. Another neuron pair at the same distance of 1789 micrometers, but with tuning-curve correlation of 0.5, was estimated to have a 10% prior probability of being a nonnull signal. A third pair at the same distance and tuning-curve correlation of 0.75 was estimated to have a 24% prior probability of being a nonnull signal.

6. FINAL REMARKS

Our FDR regression model preserves the spirit of the unified Bayes/frequentist approach of the two-groups model (1) while incorporating test-level covariates that, in our motivating example, describe the physical and functional relationships among neurons. Our results show that involving these covariates directly in the multiple-testing model has the potential to improve inferences about fine-time-scale neural interactions. While we consider our findings to be preliminary, the distance and tuning-curve effects suggested by our analysis are easily interpretable and support the previous analyses of Smith and Kohn (2008).

The neural-recordings dataset we have analyzed here is typical of many found in today's pressing scientific problems, in that it exhibits two important statistical features: the need to adjust for simultaneous inference, and the presence of spatial information, or some other nontrivial covariate structure. Previous attempts to handle this structure have typically involved separate analyses on subsets of the data, such as the "front-versus-back of brain" split considered by Efron (2008b). When there is an obvious subset structure in the data, such an approach may be appealing. Yet it requires case-by-case judgments, and opens the door to further multiplicity issues regarding the choice of subsets. Our results show that false-discovery-rate regression can avoid these difficulties, without compromising on the global error rate, by incorporating covariates directly into the testing problem. It is therefore suited to the increasingly common situation in which test statistics should not be considered exchangeable.

APPENDIX A: PREDICTIVE RECURSION

Predictive recursion is used to estimate the mixing distribution $\pi(\theta)$ in the following formulation of the two-groups model without covariates:

$$\begin{aligned} z_i &\sim c \cdot f_1(z_i) + (1-c) \cdot f_0(z_i) \\ f_0(z) &= \mathrm{N}(z \mid \mu, \sigma^2) \\ f_1(z) &= \int_{\mathcal{R}} \mathrm{N}(z \mid \mu + \theta, \sigma^2)\, \pi(\theta). \end{aligned} \tag{A.1}$$

An equivalent formulation is

$$\begin{aligned} z_i &\sim \mathrm{N}(\mu + \theta_i, \sigma^2) \\ \theta_i &\sim \Psi, \quad \Psi = \tilde{\pi}_1(\theta) + \pi_0 \delta_0, \end{aligned}$$

where Ψ is absolutely continuous with respect to the dominating measure ν defined as the sum of Lebesgue measure on $\mathcal{R}$ and a point mass

Downloaded by [Carnegie Mellon University] at 20:17 06 August 2015

Algorithm 2 Predictive recursion for estimating $\pi(\theta)$ and c in Model (A.1). The subdensity $\tilde{\pi}_1(\theta) = c\pi(\theta)$ is approximated on a discrete grid, and integrals with respect to $\tilde{\pi}_1(\theta)$ are calculated by the trapezoid rule.

Data: Test statistics $z_1, \ldots, z_n$

Input: Null model $\mathrm{N}(\mu, \sigma^2)$; weights $\gamma^{[i]}$; initial guess $\Psi^{[0]} = \tilde{\pi}_1^{[0]}(\theta) + \pi_0^{[0]}\delta_0$ having a continuous subdensity $\tilde{\pi}_1^{[0]}(\theta)$ and a Dirac measure at zero of mass $\pi_0^{[0]}$.

for: $i = 1, \ldots, n$ **do**

$$m_0^{[i]} = \pi_0^{[i-1]} \cdot \mathrm{N}(z_i \mid \mu, \sigma^2)$$

$$f_1^{[i]}(\theta) = \mathrm{N}(z_i \mid \mu + \theta, \sigma^2)\, \tilde{\pi}_1^{[i-1]}(\theta) \quad \text{(discrete grid)}$$

$$m_1^{[i]} = \int_{\mathcal{R}} f_1^{[i]}(\theta)\, d\theta \qquad \text{(trapezoid rule)}$$

$$\pi_0^{[i]} = (1 - \gamma^{[i]}) \cdot \pi_0^{[i-1]} + \gamma^{[i]} \cdot \left(\frac{m_0^{[i]}}{m_0^{[i]} + m_1^{[i]}} \right)$$

$$\tilde{\pi}_1^{[i]}(\theta) = (1 - \gamma^{[i]}) \cdot \tilde{\pi}_1^{[i-1]}(\theta) + \gamma^{[i]} \cdot \left(\frac{f_1^{[i]}(\theta)}{m_0^{[i]} + m_1^{[i]}} \right)$$

end

Output: Estimates $c = 1 - \pi_0^{[n]}$ and $\pi(\theta) = \pi_1^{[n]}(\theta)/c$.

at 0. Here $\tilde{\pi}_1(\theta) = c \cdot \pi(\theta)$ is a subdensity corresponding to signals, and $\pi_0 = 1 - c$ is the mass at zero corresponding to nulls.

Predictive recursion (Newton 2002) is a stochastic algorithm for estimating Ψ, or for any mixing density with respect to an arbitrary dominating measure ν, from observations $z_1, \ldots, z_n$. Assume that μ and σ^2 are fixed. Begin with a guess $\Psi^{[0]}$ and a sequence of weights $\gamma^{[i]} \in (0, 1)$. For $i = 1, 2, \ldots, n$, recursively compute the update

$$m^{[i-1]}(z_i) = \int_{\mathcal{R}} \mathrm{N}(z_i \mid \mu + u, \sigma^2)\, \Psi^{[i-1]}(du) \tag{A.2}$$

$$\Psi^{[i]}(du) = (1 - \gamma^{[i]})\Psi^{[i-1]}(du) + \gamma^{[i]} \cdot \left\{ \frac{\mathrm{N}(z_i \mid \mu + u, \sigma^2)\Psi^{[i-1]}(du)}{m^{[i-1]}(z_i)} \right\}. \tag{A.3}$$

The final update,

$$\Psi^{[n]} = \tilde{\pi}_1^{[n]}(\theta) + \pi_0^{[n]}\delta_0 = c^{[n]} \cdot \pi^{[n]}(\theta) + (1 - c^{[n]}) \cdot \delta_0\,,$$

provides estimates for c and the mixing density $\pi(\theta)$. In practice, the continuous component $\pi(\theta)$ is approximated on a discrete grid of points, and the integral in (A.2) is computed using the trapezoid rule over this grid.

The key advantages of predictive recursion are its speed and its flexibility. Moreover, Tokdar, Martin, and Ghosh (2009) derived conditions on the weights $\gamma^{[i]}$ that lead to almost-sure weak convergence of the PR estimate to the true mixing distribution. They also show that, when the mixture model is mis-specified, the final estimate converges in total variation to the mixing density that minimizes the Kullback-Leibler divergence to the truth. The conditions on the weights $\gamma^{[i]}$ necessary to ensure these results are satisfied by $\gamma^{[i]} = (i + 1)^{-a}$, $a \in (2/3, 1)$. We use the default value $a = 0.67$ recommended by Tokdar, Martin, and Ghosh (2009).

Algorithm 2 describes the steps of predictive recursion in detail, including a clear separation of the continuous and discrete components of the mixture distribution Ψ. In our implementation, we pass through the data 10 times, randomizing the sweep order in each pass. This yields stable estimates that are relatively insensitive to the order in which the data points are processed, and is consistent with the practice of other authors who have studied predictive recursion (Newton 2002; Tokdar, Martin, and Ghosh 2009; Martin and Tokdar 2012).

APPENDIX B: DETAILS OF THE FULLY BAYES METHOD

Our implementation of the fully Bayes FDR regression model in (9) assumes that $\pi(\theta)$ is a K-component discrete mixture of Gaussians, parameterized by a set of component weights η_k, means μ_k and variances τ_k^2.

Let h_i be the binary indicator of whether z_i is signal or noise, let β denote the regression vector, and let

$$c(x_i) = \frac{1}{1 + e^{-x_i^T \beta}}\,.$$

We assume the conditionally conjugate priors

$$\begin{aligned} \beta &\sim N(b_0, B_0) \\ \mu_k &\sim N(0, v_\mu) \\ \tau_k^2 &\sim \mathrm{IG}(a/2, b/2) \\ (\eta_1, \ldots, \eta_K) &\sim \mathrm{Dirichlet}(\alpha)\,. \end{aligned}$$

Under these priors, the full conditionals needed to implement a Gibbs sampler are as follows. To lighten the notation, a dash (—) is used to denote "all variables not otherwise named." Our simulation studies use the prior parameters $b_0 = 0$, $B_0 = 100I$, $v_\mu = 100$, $a = b = 1$, and $\alpha = (1, \ldots, 1)$.

To update h_i, note that, from standard results on Gaussian mixtures, the conditional predictive density under the alternative is

$$f_1(z_i \mid —) = \sum_{k=1}^{K} \mathrm{N}(z_i \mid \mu_k, \tau_k^2 + \sigma^2)\,.$$

Thus to draw h_i, we sample from the Bernoulli distribution

$$(h_i \mid —) \sim \begin{cases} 1 \text{ with probability } w_i \\ 0 \text{ otherwise,} \end{cases}$$

where

$$w_i = \frac{c(x_i) \cdot f_1(z_i \mid —)}{c(x_i) \cdot f_1(z_i \mid —) + \{1 - c(x_i)\} \cdot f_0(z_i)}\,.$$

Conditionally on h_i, the regression coefficients β can be updated in two stages using the Polya-Gamma latent-variable scheme. First we draw auxiliary variables ω_i from a Polya-Gamma distribution as

$$(\omega_i \mid —) \sim \mathrm{PG}(1, x_i^T \beta)\,,$$

using the method of Polson, Scott, and Windle (2013), and implemented by Windle, Polson, and Scott (2014). Let $\Omega = \mathrm{diag}(\omega_1, \ldots, \omega_n)$ and $\kappa = (h_1 - 1/2, \ldots, h_n - 1/2)$. We use these to update β as

$$(\beta \mid —) \sim \mathrm{N}(m_\beta, V_\beta)\,,$$

where

$$\begin{aligned} V_\beta^{-1} &= X^T \Omega X + B_0^{-1} \\ m_\beta &= V_\beta^{-1}(X^T \kappa + B_0^{-1} b_0)\,. \end{aligned}$$

Given the z_i corresponding to signals ($h_i = 1$), the mixture-model weights, means, and variances involve straightforward conjugate updates, and are described in many standard textbooks on Bayesian analysis. Thus we do not include them here; see, for example, Chapter 22 of Gelman et al. (2013).

Downloaded by [Carnegie Mellon University] at 20:17 06 August 2015

SUPPLEMENTARY MATERIALS

The supplementary materials include the R script that was used for the study. Please also refer to *https://github.com/jgscott/FDRreg/*.

[Received February 2013. Revised February 2014.]

REFERENCES

Aertsen, A. (2010), " Foreword, " in *Analysis of Parallel Spike Trains*, eds. S. Grün and S. Rotter, New York: Springer. [466]

Averbeck, B. B., Latham, P. E., and Pouget, A. (2006), "Neural Correlations, Population Coding and Computation," *Nature Reviews Neuroscience*, 7, 358–366. [466]

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, 289–300. [459,465,466]

Berger, J. O. (2003), "Could Fisher, Jeffreys, and Neyman have Agreed on Testing?," *Statistical Science*, 18, 1–32. [460]

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,* (Vol. 1), Institute of Mathematical Statistics, pp. 211–230. [460]

Brown, E., Lydic, R., and Schiff, N. (2010), "General Anesthesia, Sleep, and Coma," *New England Journal of Medicine*, 363, 2638–2650. [467]

Brown, E. N., Kass, R. E., and Mitra, P. P. (2004), "Multiple Neural Spike Train Data Analysis: State-of-the-art and Future Challenges," *Nature Neuroscience*, 7, 456–461. [466]

Buzsaki, G. (2004), "Large-scale Recording of Neuronal Ensembles," *Nature Neuroscience*, 7, 446–451. [466]

Clarke, S., and Hall, P. (2009), "Robustness of Multiple Testing Procedures Against Dependence," *The Annals of Statistics*, 37, 332–358. [460]

Dempster, A., Laird, N., and Rubin, D. (1977), "Maximum Likelihood From Incomplete Data Via the EM Algorithm" (with discussion), *Journal of the Royal Statistical Society,* Series B, 39, 1–38. [461]

Do, K.-A., Müller, P., and Tang, F. (2005), "A Bayesian Mixture Model for Differential Gene Expression," *Journal of the Royal Statistical Society,* Series C, 54, 627–644. [463]

Dunson, D. B., Pillai, N. S., and Park, J.-H. (2007), "Bayesian Density Regression," *Journal of the Royal Statistical Society,* Series B, 69, 163–183. [461]

Efron, B. (2004), "Large-Scale Simultaneous Hypothesis Testing: The Choice of a Null Hypothesis," *Journal of the American Statistical Association*, 99. [461,463,469]

——— (2008a), "Microarrays, Empirical Bayes and the Two-groups Model" (with discussion), *Statistical Science*, 1, 1–22. [460,464]

——— (2008b), "Simultaneous Inference: When Should Hypothesis Testing Problems be Combined?," *The Annals of Applied Statistics*, 2, 197–223. [460,469]

——— (2012), *Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction*, Institute of Mathematical Statistics Monographs, Cambridge, UK: Cambridge University Press. [462,465]

Efron, B., Tibshirani, R., Storey, J., and Tusher, V. (2001), "Empirical Bayes Analysis of a Microarray Experiment," *Journal of American Statistical Association*, 96, 1151–1160. [459,460,462]

Fan, J., Han, X., and Gu, W. (2012), "Estimating False Discovery Proportion Under Arbitrary Covariance Dependence" (with discussion), *Journal of the American Statistical Association*, 107, 1019–1035. [460]

Fries, P. (2009), "Neuronal Gamma-band Synchronization as a Fundamental Process in Cortical Computation," *Annual Review of Neuroscience*, 32. [466]

Gelman, A., Carlin, J., Stern, H., Dunson, D. B., Vehtari, A., and Rubin, D. (2013), *Bayesian Data Analysis.* (3rd ed.), Chapman and Hall/CRC. [470]

Graf, A. B. A., Kohn, A., Jazayeri, M., and Movshon, J. A. (2011), "Decoding the Activity of Neuronal Populations in Macaque Primary Visual Cortex," *Nature Neuroscience*, 14, 239–245. [466]

Gray, C. (1999), "The Temporal Correlation Hypothesis of Visual Feature Integration: Still Alive and Well," *Neuron*, 24. [466]

Harrison, M., Amarasingham, A., and Kass, R. (2013), "Statistical Identification of Synchronous Spiking," in *Spike Timing: Mechanisms and Function*, eds. P. D. Lorenzo and J. Victor, pp. 77–120. [466]

Johnstone, I., and Silverman, B. W. (2004), "Needles and Straw in Haystacks: Empirical-Bayes Estimates of Possibly Sparse Sequences," *The Annals of Statistics*, 32, 1594–1649. [460]

Kass, R. E., Kelly, R. C., and Loh, W.-L. (2011), "Assessment of Synchrony in Multiple Neural Spike Trains Using Loglinear Point Process Models," *The Annals of Applied Statistics*, 5, 1262–1292. [467]

Kass, R. E., Eden, U., and Brown, E. (2014), *Analysis of Neural Data*, New York: Springer. [467]

Kelly, R., and Kass, R. (2012), "A Framework for Evaluating Pairwise and Multiway Synchrony Among Stimulus-driven Neurons," *Neural Computation*, 24, 2007–2032. [459,461,467]

Kelly, R., Smith, M., Samonds, J., Kohn, A., Bonds, A., Movshon, J., and Lee, T. (2007), "Comparison of Recordings From Microelectrode Arrays and Single Electrodes in the Visual Cortex," *Journal of Neuroscience*, 27, 261–264. [467]

Kelly, R. C., Smith, M. A., Kass, R. E., and Lee, T. (2010), "Local Field Potentials Indicate Network State and Account for Neuronal Response Variability," *Journal of Computational Neuroscience*, 29, 567–579. [459,461]

Kohn, A., and Smith, M. A. (2005), "Stimulus Dependence of Neuronal Correlation in Primary Visual Cortex of the Macaque," *Journal of Neuroscience*, 25, 3661–3673. [466]

Martin, R., and Tokdar, S. (2012), "A Nonparametric Empirical Bayes Framework for Large-scale Multiple Testing," *Biostatistics*, 13, 427–439. [460,461,463,464,470]

Morris, J., Brown, P., C., H. R., Baggerly, K., and Coombes, K. (2008), "Bayesian Analysis of Mass Spectrometry Proteomic Data Using Wavelet-based Functional Mixed Models," *Biometrics*, 64, 479–489. [465]

Muller, P., Parmigiani, G., and Rice, K. (2006), "FDR and Bayesian Multiple Comparisons Rules," in *Proceedings of the 8th Valencia World Meeting on Bayesian Statistics*. Oxford, UK: Oxford University Press. [460]

Newton, M. A. (2002), "A Nonparametric Recursive Estimator of the Mixing Distribution," *Sankhya,* Series A, 64, 306–322. [460,461,470]

Niebur, E., Hsiao, S., and Johnson, K. (2002), "Synchrony: a Neuronal Mechanism for Attentional Selection?," *Current Opinion in Neurobiology*, 12, 190–194. [466]

Perkel, D. H., Gerstein, G. L., and Moore, G. P. (1967), "Neuronal Spike Trains and Stochastic Point Processes i. The Single Spike Train," *Biophysical Journal*, 7, 391–418. [466]

Polson, N. G., and Scott, J. G. (2012), "Good, Great, or Lucky? Screening for Firms With Sustained Superior Performance Using Heavy-Tailed Priors," *The Annals of Applied Statistics*, 6, 161–185. [463]

Polson, N. G., Scott, J. G., and Windle, J. (2013), "Bayesian Inference for Logistic Models Using Polya-Gamma Latent Variables," *Journal of the American Statistical Association*, 108, 1339–1349. [461,463,470]

Saalmann, Y., and Kastner, S. (2011), "Cognitive and Perceptual Functions of the Visual Thalamus," *Neuron*, 71, 209–223. [466]

Scott, J. G. (2014), *FDRreg: False Discovery Rate Regression*. Available at *https://github.com/jgscott/FDRreg*. R package version 0.2. [459]

Scott, J. G., and Berger, J. O. (2006), "An Exploration of Aspects of Bayesian Multiple Testing," *Journal of Statistical Planning and Inference*, 136, 2144–2162. [460,463]

Shadlen, M., and Movshon, J. (1999), "Synchrony Unbound: A Critical Evaluation of the Temporal Binding Hypothesis," *Neuron*, 24, 67–77, 111–125. [466]

Smith, M., and Kohn, A. (2008), "Spatial and Temporal Scales of Neuronal Correlation in Primary Visual Cortex," *Journal of Neuroscience*, 21, 12591–12603. [459,466,469]

Stevenson, I. H., and Kording, K. P. (2011), "How Advances in Neural Recording Affect Data Analysis," *Nature Neuroscience*, 14, 139–142. [466]

Tokdar, S., Martin, R., and Ghosh, J. (2009), "Consistency of a Recursive Estimate of Mixing Distributions," *The Annals of Statistics*, 37, 2502–2522. [461,470]

Uhlhaas, P., Pipa, G., Lima, B., Melloni, L., Neuenschwander, S., Nikolic, D., and Singer, W. (2009), "Neural Synchrony in Cortical Networks: History, Concept and Current Status," in *Frontiers in Integrative Neuroscience*, Vol. 3. [466]

Windle, J., Polson, N. G., and Scott, J. G. (2014), *BayesLogit: Bayesian Logistic Regression*, Available at *http://cran.r-project.org/web/packages/BayesLogit/index.html*. R package version 0.4. [470]