Valid Inference for Selected Regions of Interest with Application to fMRI Studies

Amit Meir
Department of Statistics
University of Washington

Yuval Benjamini
Department of Statistics
The Hebrew University

May 15, 2018

Abstract

Functional magnetic resonance imaging (fMRI) studies often try to identify associations between localized brain activity and specific changes in cognition or behavior. In cluster-wise analysis, the researcher tries to simultaneously (a) group neighboring voxels that display a strong response to the experiment to a cluster and (b) estimate the strength of the response in the cluster. However, using the same data for both steps creates a "selective inference" bias, where the observed effect diminishes in validation experiments. Current methods that account for multivariate selection require either a very strong set of assumptions (and have been found to produce inaccurate results), or are non-parametric with relatively low power. Furthermore, all current methods focus on testing and do not give estimates and intervals for non-null effects. Here, we propose to apply recently-developed multivariate selective inference techniques to the problem of cluster-wise analysis in neuroimaging studies. Specifically, we use fast stochastic gradient methods to compute post-selection maximum likelihood estimates and construct efficient post-selection confidence intervals based on a profile likelihood approach. Our methods work well on both simulated data, as well as on realistic datasets based on resting-state fMRI recordings from the 1000 Functional Connectomes Project repository. The point estimates are shown to reduce bias and estimation error compared to unadjusted point estimates and the confidence intervals are shown to obtain close to nominal coverage rates. The computation of the estimators and confidence intervals is implemented in an accompanying software package.

1 Introduction

1.1 Functional MRI Studies

Functional magnetic resonance imaging (fMRI) studies often try to identify associations between localized brain activity and specific changes in cognition or behavior. Functional MRI measures a correlate of neural activity, the blood-oxygen-level dependent (BOLD) contrast, over a dense grid of volumized pixels (voxels) covering the cortex. Many experimental designs can be adapted to run within the fMRI scanner, allowing researchers to assess at each voxel the neural response to variation in the task.

Typically, a linear model would be fit to each voxel separately, summarized by a single coefficient of interest; together, these coefficients form a spatial map of activity (first level analysis). Maps from multiple subjects can be aligned to a template and a cross-subject linear contrast (e.g. a difference of means) on the coefficients can be estimated at each voxel (second level analysis). For either level of analysis, the standard goal is to identify regions within the cortex that display a strong response. For new experimental tasks, this requires (a) identifying regions of the cortex that respond to experimental variation, and (b) estimating the strength of the changes.

In trying to identify locally strong signal in a large stochastic map, researchers need to account for chance effects of noise. Regional inference procedures (Friston et al. (1994), Hayasaka and Nichols (2003)) aggregate neighboring responses to share strength and reduce the number of hypotheses tested. Usually, regional inference is based on two thresholding stages: First, the map of statistics is thresholded at a primary threshold and voxels with statistic values below this threshold are discarded. The remaining voxels are then organized into spatially connected regions. Finally, a summary for each region (typically the number of voxels in the region) is calculated, and a secondary region threshold is set to control the probability of falsely detecting regions. Regions that pass this second thresholding are reported, sorted by their summary statistic. Estimating the effect size of a newly detected region is hard. Model selection - in our case, defining the region of interest - can induce bias in classical estimates if both steps use the same data. Naive implementations that do not account for region selection can greatly overestimate the magnitude of the effect and underestimate the uncertainty in the results, producing results that diminish or even disappear in later experiments.

Both parametric and non-parametric methods have been proposed to set the threshold for cluster size, and an ongoing debate in the community (e.g. Eklund et al. (2016), arguing against parametric methods, and Flandin and Friston (2016), for them) has underscored strengths and weaknesses in both approaches. Briefly, parametric methods use approximations based on Gaussian Random Theory to derive cutoffs for the expected cluster-size under the null. The most commonly used formula (Friston et al., 1994) is valid only for

very high thresholds (that is, the probability under the null of crossing the threshold is small), and requires strong assumptions on the dependence between voxels. Non-parametric approaches are based on permutation analysis of full-brain images within the linear design. The power of these methods to detect signals is limited in small samples and in the presence of one true large cluster, other clusters may be undetectable.

Moreover, there is no current method that can estimate the effect size for regions that were found to be significant. Both approaches are based on the sharp null (zero-effect everywhere), making it hard to characterize the effect of selection on the alternative. Regional null-models proposed in the statistical literature (Heller et al., 2006; Schwartzman et al., 2011), also emphasize hypothesis testing rather than non-null inference. This is disconcerting, because statistically significant but scientifically unimportant effects are common in human-subject studies, where no experimental condition can be perfectly controlled.

Researchers therefore resort to reporting observed statistics of the regions, such as size or average effect in region. These statistics are prone to selection bias, which may lead in follow up studies to diminishing effects. To see this, define a population vector describing the mean population effect at each voxel $\Theta = (\theta_1, ..., \theta_v)$, which the observed voxel-wise effect $Z = (Z_1, ..., Z_v)$ tries to estimate. (Here, each index in $1, \ldots, v$ refers to a point in the three dimensional grid). Then, for a given region $R \subset V = \{1, ..., v\}$, we can identify the (population) effect size in the region $\bar{\theta}_R = |R|^{-1} \sum_{i \in R} \theta_i$. The idea is that if R was selected without seeing the data, based on a previous experiment, then $\bar{Z}_R$ would approximate $\bar{\theta}_R$ with some added noise ($E[\bar{Z}_R] = \bar{\theta}_R$). However, if R was selected from the data, the observed mean $\bar{Z}_R$ would be (upward) biased when compared to $\bar{\theta}_R$. As an example, if the primary threshold is c, then $\bar{Z}_R \geq c$ regardless of $\bar{\theta}_R$.

1.2 Selective Inference

The field of selective inference is concerned with developing statistical methods that account for selection in inference. While most works focus on constructing valid confidence intervals in selected models (Benjamini and Yekutieli, 2005; Fithian et al., 2014; Lee et al., 2016; Tian et al., 2016; Weinstein et al., 2013), other recent works have focused on estimation (Panigrahi et al., 2016; Reid et al., 2014; Routtenberg and Tong, 2015). In the context of functional MRI studies, Rosenblatt and Benjamini (2014) and Benjamini and Meir (2014) develop inference methods for voxel-wise analysis. Most relevant in our context are the works of Meir and Drton (2017) who develop methods for performing selective maximum likelihood inference in multivariate settings and the work of Benjamini et al. (2016) who construct valid confidence intervals for selected regions in high-throughput assays.

A popular paradigm in selective inference is that of 'conditioning on selection'. In the presence of model selection, a quantity is only estimated if a related random variable satisfies some pre-specified conditions.

Thus, one can capture the added uncertainty induced by model selection through conditioning on the event of selection, in effect constraining the distribution of the data to the region in which it can be observed. In Section 2 we describe how this paradigm can be adapted to performing inference on selected regions in fMRI studies. Specifically we propose a novel conditional profile likelihood approach for constructing post-selection confidence intervals. We discuss the computation of conditional maximum likelihood points estimates (MLE) in Section 3. In Section 4 we demonstrate the validity of the proposed inference methods using real and artificial data. Finally, in Section 5 we conclude with a discussion.

2 Selective Confidence Intervals

Let R be a set of neighboring voxels that pass screening with a pre-specified threshold c. Let $\partial R \subset V\ R$ be the set of voxels that are not in R, but are neighbors of voxels in R. For the region R to be detected, one of two conditions must hold:

$$A_R^+ := \left\{ \begin{array}{ll} Z_i \geq c, & i \in R \ \text{(internal conditions)} \\ Z_i < c, & i \in \partial R \ \text{(external conditions)} \end{array} \right\}, \qquad A_R^- := \left\{ \begin{array}{ll} Z_i \leq -c, & i \in R \ \text{(internal conditions)} \\ Z_i > -c, & i \in \partial R \ \text{(external conditions)} \end{array} \right\}.$$

That is, voxels of interest may be either positively or negatively correlated with an outcome. We denote the union of the two events by $A_R := A_R^+ \cup A_R^-$. Assuming that Z follows a multivariate normal distribution with mean θ and covariance Σ, we can perform inference based on the post-selection distribution of $Z_R|A_R$, which is a truncated multivariate normal distribution.

For hypothesis testing, we utilize an approach similar to that of Benjamini et al. (2016), who construct confidence intervals based on samples from the truncated multivariate normal distribution. Under the sharp local null model ($\Theta_{R\cup\partial R} = \mathbf{0}$), we can sample from $\{\bar{Z}_R|Z \in A_R\}$ and use the empirical sample from the null to estimate the p-value of the observed average. Sampling from the truncated multivariate normal distribution is a well studied task which can be performed efficiently, see for example Kotecha and Djuric (1999). A similar approach can be used to form a confidence interval for $\bar{\theta}_R$.

Let $q(\theta_R)$ be the probability measure determined by the multivariate normal density function with mean θ_R and covariance Σ_R and define $\theta_R(a) := \arg\inf_{\bar{\theta}'_R = a} \mathrm{KL}(q(\theta_R)\|q(\theta'_R))$. As a solution for constructing valid post-selection confidence intervals for selected regions, Benjamini et al. (2016) propose to utilize the well known result that

$$P_{\theta_R}\left(\bar{\theta}_R \in \{a \in \mathbb{R} :\ \alpha/2 \leq F_{\theta_R(a)}(\bar{Z}) \leq 1-\alpha/2\}\right) = \alpha.$$

The main obstacle is that $\theta(a)$ is unknown and that a mean value a, on its own, does not fully specify a distribution for Z_R. Benjamini et al. resolve this issue by positing a specific family of alternatives defined by

linear mapping of the mean parameter $\bar{\theta}_R$ to the population vector $\Theta_R = \bar{\theta}_R * \nu$ for a profile vector satisfying $\sum \nu_i = 1$. They specify ν as a function of Σ_R in such a way as to ensure that the resulting confidence interval is contiguous.

Here, we seek to employ a more rigorous profile likelihood approach (Venzon and Moolgavkar, 1988). Define the conditional profile likelihood of $\bar{\theta}_R$ as $\mathcal{L}(\bar{\theta}_R|A_R) := \arg\max_{\theta_R} \mathcal{L}(\bar{\theta}_R, \theta_R|A_R)$. The profile likelihood of $\bar{\theta}_R$ defines a proper distribution for each candidate value of $\bar{\theta}_R$. Let $\hat{\theta}_{R,n}(a) = \arg\max_{\theta_R} \mathcal{L}_n(a, \theta_R|A_R)$. Then, it is clear that given a sequence of constrained estimators that satisfy $\hat{\theta}_{R,n}(a) \to^p \theta_R(a)$ the resulting confidence intervals are consistent, satisfying

$$\lim_{n\to\infty} P_{\theta_R}\left(\bar{\theta}_R \in \{a \in \mathbb{R} :\ \alpha/2 \leq F_{\hat{\theta}_{R,n}(a)}(\bar{Z}) \leq 1 - \alpha/2\}\right) = \alpha.$$

The parameter estimate $\hat{\theta}_R(a)$ is a constrained MLE. In general, the maximization of the conditional likelihood is a non-trivial task. In the next Section, we discuss efficient techniques for computing the conditional MLE.

3 Maximum Likelihood Inference for Selected Regions

3.1 A Stochastic Gradient Method

Assume that Z follows a multivariate normal distribution with mean θ, variance Σ and density function φ and define the (unconstrained) conditional maximum likelihood estimator (MLE) for the signal in a selected region R:

$$\hat{\theta}_R = \arg\max_{\theta_R} \log \varphi(Z_R; \theta_R, \Sigma_R) - \log P_{\theta_R}(A_R). \tag{1}$$

The conditional MLE is given by the maximizer of the unconditional log-likelihood of Z_R penalized by the log of the probability of selecting the region R. This likelihood is notoriously difficult to maximize as the probability of selecting R is a high-dimensional integral. To bypass the difficult likelihood, Meir and Drton (2017) propose to use samples from the distribution of $Z_R|A_R$ and approximate the conditional expectation $E(Z_R|A_R)$, which in turn can be used to approximate the gradient.

Denote by $z(\theta_R)$ a sample from the distribution of Z_R at a parameter value generated according to the distribution determined by θ_R and let $\{\gamma_t\}_{t=1}^{\infty}$ be a deterministic sequence which satisfies $\sum_{i=1}^{\infty} \gamma_i = \infty$, $\sum_{i=1}^{\infty} \gamma^2 < \infty$. Then, the conditional MLE can be computed by taking optimization steps of the form

$$\theta_R^t = \theta_R^{t-1} + \gamma_t \Sigma^{-1} \left(Z_R - z(\theta_R^{t-1})\right). \tag{2}$$

To see why the algorithm defined by (2) converges to the MLE, consider the score function of a truncated multivariate normal distribution $\Sigma^{-1}(Z_R - E(Z_R|A_R))$. In (2), we replaced the conditional expectation of Z_R given A_R with the sample $z(\theta_R)$, resulting in a stochastic gradient method (Ma and Wang, 2015). We compute constrained MLE estimates using a similar, projected stochastic gradient method.

3.2 Inducing Smoothness

Given that the observed realization of fMRI experiments tend to be smooth, we could expect the underlying signal to be smooth as well. However, the conditional MLE does not always yield a smooth estimate. This may become a problem when performing constrained estimation in large clusters, as fixing the mean leaves many degrees of freedom for the estimation routine to fit the data closely using a highly variable and non-smooth parameter estimate.

To obtain smooth post-selection estimates, we can make use of an appropriate penalty function. Specifically, we use a Tikohonov ℓ_2 penalty which penalizes the spatial first and second differences in the estimated mean, solving the following optimization problem:

$$\hat{\theta}_R = \arg\max_{\theta_R} \log \varphi(Z_R; \theta_R, \Sigma) - \log P_{\theta_R}(A_R) - \lambda_1 \|\Gamma_1 \theta_R\|_2^2 - \lambda_2 \|\Gamma_2 \theta_R\|_2^2, \tag{3}$$

where Γ_1 is a matrix constructed in such a way so $\Gamma_1 \theta_R$ gives a vector of the first order differences in θ_R and similarly $\Gamma_2 \theta_R$ is the vector of second order differences.

The problem remains of specifying the amount of smoothness to be induced. If we are willing to assume that both the noise and the signal in our data are smooth, then we can let the data guide us in determining the degree of smoothness to induce in our estimates. In particular, we specify the regularization parameters $\lambda_1, \lambda_2 \in \mathbb{R}^+$ in such a way as to ensure that $\|\Gamma_1 \theta_R\|_2^2 \leq \lambda \|\Gamma_1 Z_R\|_2^2$ and $\|\Gamma_2 \theta_R\|_2^2 \leq \lambda \|\Gamma_2 Z_R\|_2^2$, where $\lambda \propto \hat{\bar{\theta}}_R / \bar{Z}_R$. Thus, we estimate a signal that has a degree of smoothness proportional to the smoothness of the observed data linearly shrunk to have the same mean as the estimated mean of the parameter. In our simulations we use $\lambda = 2\hat{\bar{\theta}}_R / \bar{Z}_R$.

3.3 Estimating the External Parameters

The stochastic gradient algorithm proposed in Section 3.1 calls for approximating $E(Z_R|A_R)$ with a sample from the distribution of $Z_R|A_R$. Because A_R induces a truncation on both Z_R and $Z_{\partial R}$, one cannot sample from $Z_R|A_R$ without also sampling $Z_{\partial R}$. In order to sample from $Z_{\partial R}$, we must have an estimate for $\theta_{\partial R}$. Despite it being a nuisance parameter, the role of $\theta_{\partial R}$ cannot be discounted as its misspecification may result in a distorted maximum likelihood estimates and invalid hypothesis tests. Furthermore, $\theta_{\partial R}$ cannot

be estimated directly via maximum likelihood as the upper truncation often leads to an undesirable inflation of point estimates (Meir and Drton, 2017).

In order to overcome this issue and obtain reasonable estimate of $\theta_{\partial R}$, we make use of our implicit smoothness assumption on the form of the true signal. For a voxel $i \in \partial R$ we find $i' = \arg\min_{j \in R} d(i,j)$, the nearest voxel to i in R, and assign a plug-in estimate for θ_i setting $\hat{\theta}_i = \hat{\theta}_{i'}$. In simulations where the underlying signal is smooth and known, we find that this neighbor plug-in method works almost as well as plugging-in the true parameter values for the boundary.

4 Simulations

With candidate methods for performing inference in selected regions on hand, we turn to evaluating them on relevant data sets. We conduct experiments in two data regimes: the first is based on multivariate normal data with exponentially decaying correlation. The second is based on observed and preprocessed functional fMRI data from multiple subjects.

4.1 Methods

4.1.1 Simulated data

In the first set of experiments, we model the signal as a smooth unimodal process and the noise as an additive multivariate normal process. We generate an observation vector $Z = (Z)_i$, where i is a multi-index for a voxel in a three-dimensional cube of dimensions $11 \times 11 \times 9$. At voxel i, we set $Z_i = \theta_i + \varepsilon_i$, for θ_i the signal at voxel i and ε_i the noise. The noise component ε is sampled from a multivariate normal distribution with mean zero and covariance $\Sigma_{ij} = \rho^{\|i-j\|_2^2}$. For estimating coverage levels, we generate a smooth signal process: first we select a multi index i^* in the cube at random, and set $\theta_i = s \exp((i-i^*)'\Omega^{-1}(i-i^*))$ where $\Omega_{k,l} = 2\boldsymbol{I}\{k=l\} + 0.6\boldsymbol{I}\{k \neq l\}$ and c is a constant that can be thought of as determining the signal to noise ratio (snr) in the data. We vary the simulation parameters over $s \in \{0,1,2,3,4\}$ and $\rho \in \{0.5, 0.75\}$.

4.1.2 fMRI data

In the second set of experiments we generate controlled two-group comparison of functional MRI data. We obtained three-dimensional time-series data (T=119 timepoints) for N = 85 subjects from the Cambridge Buckner dataset in the 1000 Functional Connectomes repository (Biswal et al., 2010). The data was collected in resting state, meaning that no explicit cognitive task was performed. All subjects recordings were registered to a standard cortical template (55366 voxels, embedded within a $54 \times 64 \times 50$ box), and spatially

smoothed at a bandwidth of 4mm. Following the protocol of Eklund et al. (2016), for each subject the data were processed as if an osculating stimulus sequence was presented. The result was a t-statistic map for each subject representing the estimated t coefficient at each voxel, for which the underlying value of the parameter is 0. We cropped the t maps to a $11 \times 11 \times 9$ cube within the cortex.

In each experiment, we sample without replacement two groups of n subjects each (A and B), and form a map of group-average differences. Letting $\tilde{\epsilon}^{(j)}$ be the three-dimensional t-map of the j'th subject, we set $\varepsilon = \frac{1}{n}\sum_{j\in A}\tilde{\epsilon}^{(j)} - \frac{1}{n}\sum_{j'\in B}\tilde{\epsilon}^{(j')}$. ε is further scaled so that the standard deviation would be 1 on average. For coverage analysis, we added a known signal as before. We vary the simulation parameters over $s \in \{0, 1, 2, 3, 4\}$ and $n \in \{8, 16\}$.

4.1.3 Algorithm

Given a simulated dataset, we follow the oft prescribed two-step inference method of applying a hard threshold to the observed voxel intensities and grouping the voxels which pass the threshold into clusters. We use two different thresholds corresponding to univariate p-values of 0.01 and 0.001. We then use the 22-point method to group the voxels into regions, meaning that the neighbors of a voxel are the voxels that it shares a side or an edge with, but not a corner. Note that we used the true covariance of the noise in the algorithm and did not try to estimate it. For each detected region we compute a regularized conditional MLE as defined in (3) with regularization parameter $\lambda \approx 2\hat{\bar{\theta}}_R/\bar{Z}_R$. We also report results for the naive estimate, given by the mean of the observed voxels $\bar{Z}_R$.

4.2 Results

The results for the the simulated and real data are very similar, that despite the fact that the fMRI data can't necessarily be expected to follow a normal distribution. The comparison between the root mean squared error (RMSE) of the naive and conditional estimators, defined as $\|\hat{\bar{\theta}}_R - \bar{\theta}_R\|_2$, is depicted in Figure 1. In each simulation we often detect more than one region, and so in order to get an estimate for typical error in a simulation we first compute the mean of the errors in each artificial data set weighted by the sizes of the regions, and then average over the replications. The conditional MLE exhibits lower RMSE in all experiments. In Figure 2 we plot the bias, as defined by $\text{sign}(\bar{Z}_R)(\hat{\bar{\theta}}_R - \bar{\theta}_R)$ for the naive and conditional estimators. The conditional estimator mitigates much of the bias induced by the selection process, exhibiting an upward bias of up to 1 when the signal to noise ratio is low and a downward bias of up to about -1 when the signal ratio is high. As could be expected naive estimates exhibits a significant amount of upward bias in all simulation settings.

In Figure 3 we plot the coverage rates for the proposed conditional confidence intervals. The selection adjusted confidence intervals obtain close to nominal coverage rates in all simulation settings. That, in contrast to existing parametric tests for the significance of selected regions which are known to be sensitive to the degree of smoothness of the data and the threshold used in the first screening step (Flandin and Friston, 2016).

We plot the power to detect true signals in the right panel of Figure 4. In each simulation instance we only check whether the largest signal was successfully detected. The cluster that has the largest signal is in most cases the largest cluster detected and is often centered around i^*. We get the most power when the threshold used for initial screening is set low and the noise is more smooth. In the experiments with fMRI data, we also have slightly better power when the initial screening is performed with a less stringent threshold. This can be expected, because using a lower threshold conserves more information for the conditional inference step (Fithian et al., 2014).

5 Conclusion

In this work we developed a set of specialized computational technique for performing inference on selected regions of interest in the context of Functional MRI studies. The proposed techniques are based on ideas stemming from the field of selective inference. In the context of the selective inference literature, our work introduces two novel concepts. The first, is the construction of profile likelihood confidence interval through the computation of constrained conditional maximum likelihood estimates. The second, is that of maximizing a penalized conditional likelihood in order to obtain better behaved point estimates.

Central motivations for this project were the works of Vul et al. (2009) who pointed out the problem of many false findings in fMRI studies and Eklund et al. (2016) who found existing parametric methods for performing inference in selected regions lacking. In contrast to these existing parametric methods, our methods only assume knowledge of the selection rule and an asymptotic normality of the observed residuals. Because we only examine one region at a time, we do not need to make restrictive assumptions regarding the covariance structure in the entire brain image and are able to use standard covariance estimates computed just in the selected regions. Additionally, conditioning on selection allows us to safely discard small regions, a practice that may exacerbate the selection bias if selection is not explicitly accounted for.

In validating our methodologies, we applied them to real and artificial data. In both settings, our confidence intervals achieved the desired nominal coverage rates. In addition, we demonstrated that the conditional maximum likelihood estimators exhibit less bias and less estimation error compared to the naive unadjusted estimates. Estimating the covariance poses an additional challenge to our method, and we hope

results showing that a somewhat inflated covariance matrix estimate maintains the coverage with little loss of power (Benjamini et al., 2016) would persist for this data. The question of whether our methods yield a significant gain in power compared to the commonly used non-parametric methods remains to be investigated.

There also remain some open theoretical questions regarding our computational methods. Meir and Drton (2017) outline some conditions under which the stochastic optimization routine proposed here converges and the conditional MLE is consistent. However, some of these assumptions are violated by the thresholding rule utilized here. It is our conjecture that the smoothness penalty imposed on the conditional likelihood mitigates the potential undesirable behaviors of the conditional MLE, though we have yet to provide rigorous proof to that effect and we hope to do so in the near future.

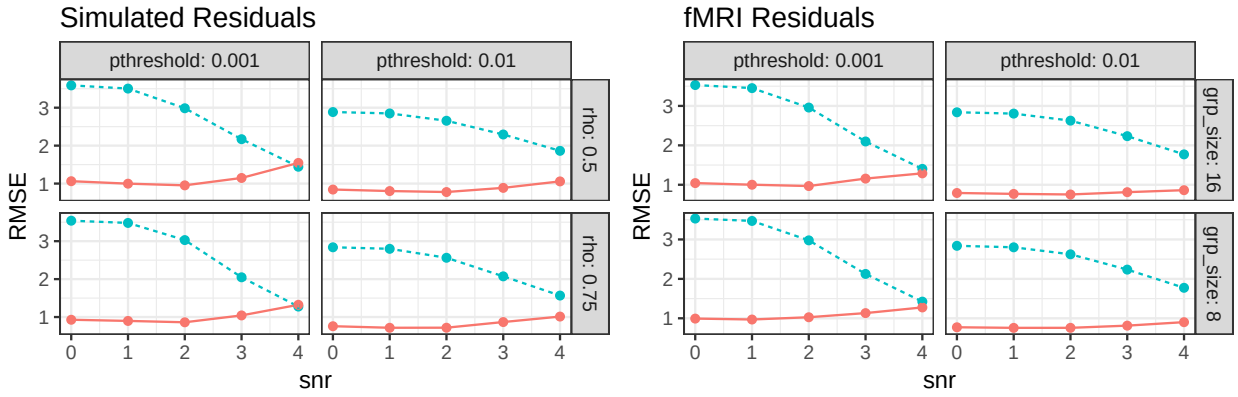

Figure 1: **Root mean squared error for estimating the mean signals in selected regions.** We plot the mean squared error for two types of estimates, the conditional MLE (solid red line) and the naive unadjusted estimate $\bar{Z}_R$ (dashed blue line).

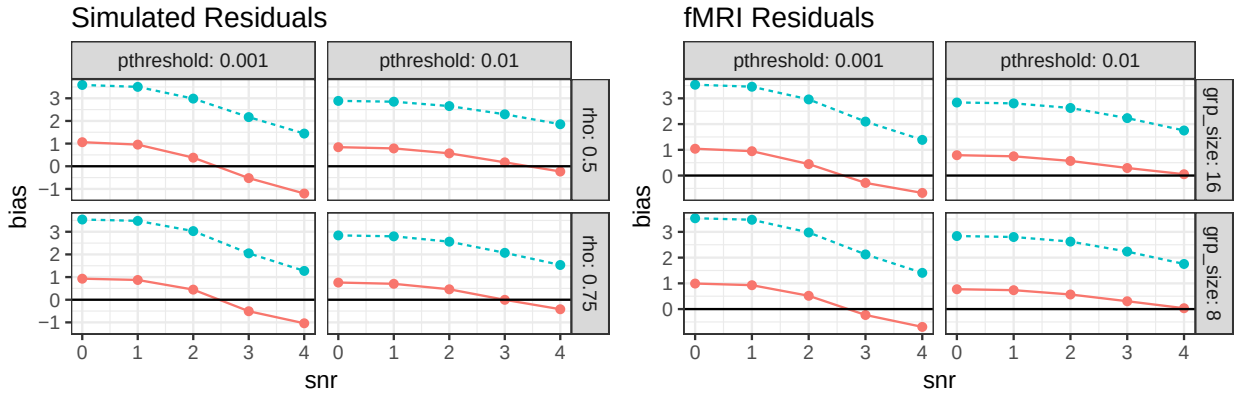

Figure 2: **Bias in estimating the mean signals in selected regions.** We plot the bias for two types of estimates, the conditional MLE (solid red line) and the naive unadjusted estimate $\bar{Z}_R$ (dashed blue line). While the naive estimators exhibits a significant upward bias in all simulation settings the conditional MLE mitigates most of the bias induced by the selection process.

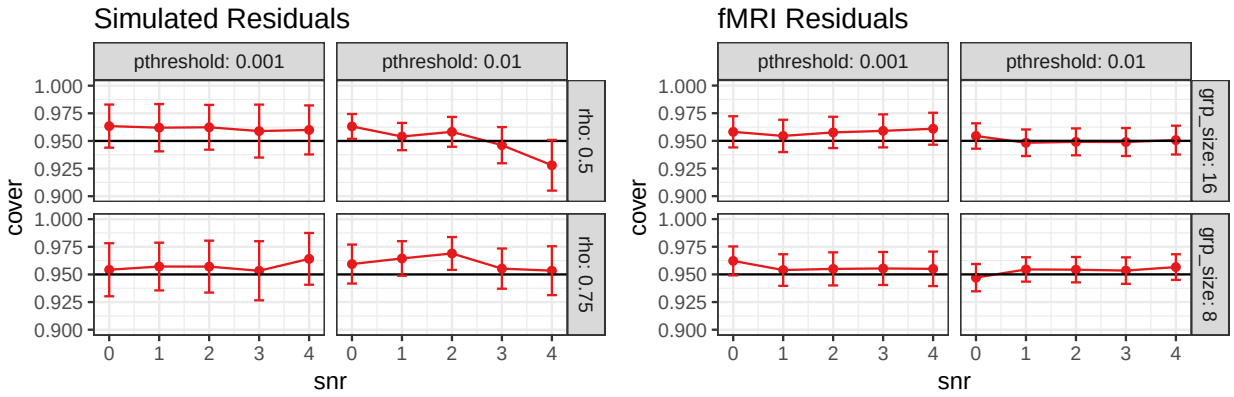

Figure 3: **Coverage rates of selective confidence intervals.** We accompany the estimate of the coverage rate in each simulation setting with a 95% Bonferroni confidence intervals . The selection adjusted confidence intervals obtain close to the nominal coverage rate of 95% in almost all simulation settings. The only simulation setting in which the coverage rate was below the nominal one was when a low threshold was used with non-smooth simulated data. However, this observed deviation is not significant and it may be due to Monte-Carlo error.

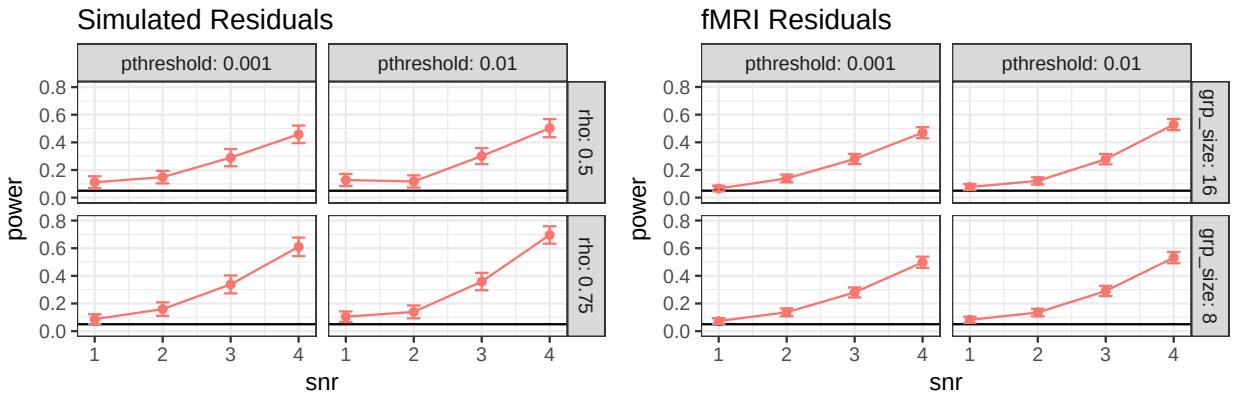

Figure 4: **Power for detecting true signal using post-selection tests.** We accompany the estimate of the power in each simulation setting with a 95% Bonferroni adjusted confidence intervals. The power is measured as the power to detect the signal in the selected region where the signal was largest in each simulated dataset. The region with the most signal also tends to be the largest region. When the signal to noise ratio is low, the selective test is quite under powered, and the power to detect true signal increases with the ratio of signal to noise.

References

Benjamini, Y. and Meir, A. (2014). Selective correlations-the conditional estimators. *arXiv preprint arXiv:1412.3242*.

Benjamini, Y., Taylor, J., and Irizarry, R. A. (2016). Selection corrected statistical inference for region detection with high-throughput assays. *bioRxiv*, page 082321.

Benjamini, Y. and Yekutieli, D. (2005). False discovery rate-adjusted multiple confidence intervals for selected parameters. *J. Amer. Statist. Assoc.*, 100(469):71–93. With comments and a rejoinder by the authors.

Biswal, B. B., Mennes, M., Zuo, X.-N., Gohel, S., Kelly, C., Smith, S. M., Beckmann, C. F., Adelstein, J. S., Buckner, R. L., Colcombe, S., et al. (2010). Toward discovery science of human brain function. *Proceedings of the National Academy of Sciences*, 107(10):4734–4739.

Eklund, A., Nichols, T. E., and Knutsson, H. (2016). Cluster failure: why fmri inferences for spatial extent have inflated false-positive rates. *Proceedings of the National Academy of Sciences*, page 201602413.

Fithian, W., Sun, D., and Taylor, J. (2014). Optimal inference after model selection. *arXiv preprint arXiv:1410.2597*.

Flandin, G. and Friston, K. J. (2016). Analysis of family-wise error rates in statistical parametric mapping using random field theory. *arXiv preprint arXiv:1606.08199*.

Friston, K. J., Worsley, K. J., Frackowiak, R., Mazziotta, J. C., and Evans, A. C. (1994). Assessing the significance of focal activations using their spatial extent. *Human brain mapping*, 1(3):210–220.

Hayasaka, S. and Nichols, T. E. (2003). Validating cluster size inference: random field and permutation methods. *Neuroimage*, 20(4):2343–2356.

Heller, R., Stanley, D., Yekutieli, D., Rubin, N., and Benjamini, Y. (2006). Cluster-based analysis of fmri data. *NeuroImage*, 33(2):599–608.

Kotecha, J. H. and Djuric, P. M. (1999). Gibbs sampling approach for generation of truncated multivariate gaussian random variables. In *Acoustics, Speech, and Signal Processing, 1999. Proceedings., 1999 IEEE International Conference on*, volume 3, pages 1757–1760. IEEE.

Lee, J. D., Sun, D. L., Sun, Y., and Taylor, J. E. (2016). Exact post-selection inference, with application to the lasso. *Ann. Statist.*, 44(3):907–927.

Ma, X. and Wang, X. (2015). Convergence analysis of contrastive divergence algorithm based on gradient method with errors. *Math. Probl. Eng.*, pages Art. ID 350102, 9.

Meir, A. and Drton, M. (2017). Tractable post-selection maximum likelihood inference for the lasso. *arXiv preprint*.

Panigrahi, S., Taylor, J., and Weinstein, A. (2016). Bayesian post-selection inference in the linear model. *arXiv preprint arXiv:1605.08824*.

Reid, S., Taylor, J., and Tibshirani, R. (2014). Post-selection point and interval estimation of signal sizes in gaussian samples. *arXiv preprint arXiv:1405.3340*.

Rosenblatt, J. D. and Benjamini, Y. (2014). Selective correlations; not voodoo. *Neuroimage*, 103:401–410.

Routtenberg, T. and Tong, L. (2015). Estimation after parameter selection: Performance analysis and estimation methods. *arXiv preprint arXiv:1503.02045*.

Schwartzman, A., Gavrilov, Y., and Adler, R. J. (2011). Multiple testing of local maxima for detection of peaks in 1D. *Ann. Statist.*, 39(6):3290–3319.

Tian, X., Bi, N., and Taylor, J. (2016). Magic: a general, powerful and tractable method for selective inference. *arXiv preprint arXiv:1607.02630*.

Venzon, D. J. and Moolgavkar, S. H. (1988). A method for computing profile-likelihood-based confidence intervals. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 37(1):87–94.

Vul, E., Harris, C., Winkielman, P., and Pashler, H. (2009). Puzzlingly high correlations in fmri studies of emotion, personality, and social cognition. *Perspectives on psychological science*, 4(3):274–290.

Weinstein, A., Fithian, W., and Benjamini, Y. (2013). Selection adjusted confidence intervals with more power to determine the sign. *J. Amer. Statist. Assoc.*, 108(501):165–176.