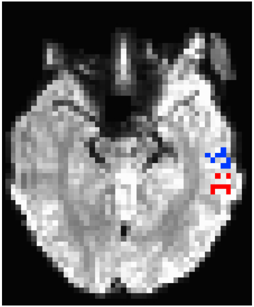
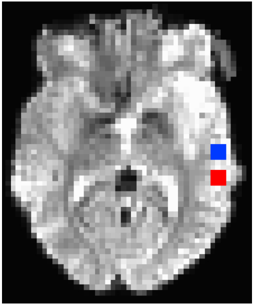
Variational Bayes for fMRI Data: Detecting Scattered Signal



Julia Fisher
University of Arizona
Statistics Consulting
Laboratory

April 20, 2018



Project Goal

Project Goal

In this project, I examine

Project Goal

In this project, I examine

- the efficacy of the Variational Bayes algorithm from Penny, Kiebel, and Friston (2003)

Project Goal

In this project, I examine

- the efficacy of the Variational Bayes algorithm from Penny, Kiebel, and Friston (2003)
- at detecting scattered signal

Project Goal

In this project, I examine

- the efficacy of the Variational Bayes algorithm from Penny, Kiebel, and Friston (2003)
- at detecting scattered signal
- in simulated functional magnetic resonance imaging (fMRI) data.

What are functional neuroimaging data?

What are functional neuroimaging data?

- Functional magnetic resonance imaging data are

What are functional neuroimaging data?

- Functional magnetic resonance imaging data are
  - a series of 3D images

What are functional neuroimaging data?

- Functional magnetic resonance imaging data are
  - a series of 3D images
  - measuring the ratio of oxygenated to deoxygenated blood (hemodynamic response)

What are functional neuroimaging data?

- Functional magnetic resonance imaging data are
  - a series of 3D images
  - measuring the ratio of oxygenated to deoxygenated blood (hemodynamic response)
  - at evenly-spaced points in time.

What are functional neuroimaging data?

- Functional magnetic resonance imaging data are
  - a series of 3D images
  - measuring the ratio of oxygenated to deoxygenated blood (hemodynamic response)
  - at evenly-spaced points in time.
- The raw data are thus $D_{\text{fMRI}} = [\mathrm{Y}_1 \ \mathrm{Y}_2 \ \cdots \ \mathrm{Y}_J]$ where

What are functional neuroimaging data?

- Functional magnetic resonance imaging data are
  - a series of 3D images
  - measuring the ratio of oxygenated to deoxygenated blood (hemodynamic response)
  - at evenly-spaced points in time.
- The raw data are thus $D_{\text{fMRI}} = [\text{Y}_1 \ \text{Y}_2 \ \cdots \ \text{Y}_J]$ where
  - J is the number of voxels in each image.

What are functional neuroimaging data?

- Functional magnetic resonance imaging data are
  - a series of 3D images
  - measuring the ratio of oxygenated to deoxygenated blood (hemodynamic response)
  - at evenly-spaced points in time.
- The raw data are thus $D_{\text{fMRI}} = [\mathrm{Y}_1 \ \mathrm{Y}_2 \ \cdots \ \mathrm{Y}_J]$ where
  - J is the number of voxels in each image.
  - Each $\mathrm{Y}_j = [Y_{1j} \ Y_{2j} \ \cdots \ Y_{\tau j}]^T$.

What are functional neuroimaging data?

- Functional magnetic resonance imaging data are
  - a series of 3D images
  - measuring the ratio of oxygenated to deoxygenated blood (hemodynamic response)
  - at evenly-spaced points in time.
- The raw data are thus $D_{\mathrm{fMRI}} = [\mathrm{Y}_1 \ \mathrm{Y}_2 \ \cdots \ \mathrm{Y}_J]$ where
  - J is the number of voxels in each image.
  - Each $\mathrm{Y}_j = [Y_{1j} \ Y_{2j} \ \cdots \ Y_{\tau j}]^T$.
  - τ is the number of time points at which images are acquired.

What are functional neuroimaging data?

What are functional neuroimaging data?

- Functional magnetic resonance imaging (fMRI) data

What are functional neuroimaging data?

- Functional magnetic resonance imaging (fMRI) data
  - are temporally and spatially autocorrelated and

What are functional neuroimaging data?

- Functional magnetic resonance imaging (fMRI) data
  - are temporally and spatially autocorrelated and
  - are traditionally analyzed with a mass univariate, frequentist, general linear model approach

What are functional neuroimaging data?

- Functional magnetic resonance imaging (fMRI) data
    - are temporally and spatially autocorrelated and
    - are traditionally analyzed with a mass univariate, frequentist, general linear model approach
    - that often incorporates smoothing as a pre-processing step and

What are functional neuroimaging data?

- Functional magnetic resonance imaging (fMRI) data
  - are temporally and spatially autocorrelated and
  - are traditionally analyzed with a mass univariate, frequentist, general linear model approach
  - that often incorporates smoothing as a pre-processing step and
  - something like a cluster-size threshold to control the false signal detection rate.

What are functional neuroimaging data?

- Functional magnetic resonance imaging (fMRI) data
  - are temporally and spatially autocorrelated and
  - are traditionally analyzed with a mass univariate, frequentist, general linear model approach
  - that often incorporates smoothing as a pre-processing step and
  - something like a cluster-size threshold to control the false signal detection rate.
- A natural consequence of these sorts of steps is that traditional analyses are very good at detecting clusters of activation.

What are functional neuroimaging data?

- Functional magnetic resonance imaging (fMRI) data
  - are temporally and spatially autocorrelated and
  - are traditionally analyzed with a mass univariate, frequentist, general linear model approach
  - that often incorporates smoothing as a pre-processing step and
  - something like a cluster-size threshold to control the false signal detection rate.
- A natural consequence of these sorts of steps is that traditional analyses are very good at detecting clusters of activation.
- They are not designed to detect scattered (or patterns of) signal.

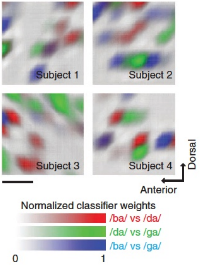
What is scattered signal?

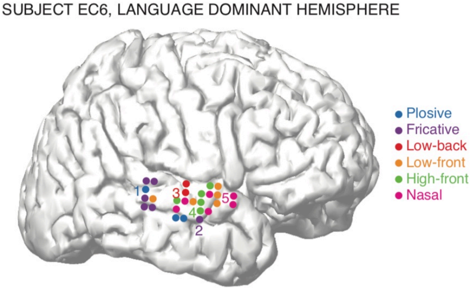
What is scattered signal?

Evidence from some invasive neuroimaging studies indicates that some activation occurs in scattered patterns, not large clusters.

What is scattered signal?

Evidence from some invasive neuroimaging studies indicates that some activation occurs in scattered patterns, not large clusters.





Left image is from Chang et al. (2010, pg. 1431) and right from Mesgarani et al. (2014, supplemental material).

How do you detect scattered signal?

How do you detect scattered signal?

- Much hope has been placed in classification analyses (k-Nearest Neighbors, support vector machines, etc.).

How do you detect scattered signal?

- Much hope has been placed in classification analyses (k-Nearest Neighbors, support vector machines, etc.).
- However, many of these approaches have a variety of pitfalls, and

How do you detect scattered signal?

- Much hope has been placed in classification analyses (k-Nearest Neighbors, support vector machines, etc.).
- However, many of these approaches have a variety of pitfalls, and
- personal work exploring the efficacy of a variety of these methods has not revealed one that reliably detects scattered signal.

Why take a Bayesian approach?

Why take a Bayesian approach?

Bayesian methods...

Why take a Bayesian approach?

Bayesian methods...

- are highly flexible.

Why take a Bayesian approach?

Bayesian methods...

- are highly flexible.
- can easily incorporate multiple sources of information (e.g. prior results, structural neurological information, etc.).

Why take a Bayesian approach?

Bayesian methods...

- are highly flexible.
- can easily incorporate multiple sources of information (e.g. prior results, structural neurological information, etc.).
- can more easily incorporate spatial information into the primary analysis.

Penny, Kiebel, and Friston (2003)

Penny, Kiebel, and Friston (2003)

- In order to better understand which aspects of a Bayesian approach would most effectively allow detection of scattered signal, I decided to explore some existing methodologies.

Penny, Kiebel, and Friston (2003)

- In order to better understand which aspects of a Bayesian approach would most effectively allow detection of scattered signal, I decided to explore some existing methodologies.
- Moreover, while quite complex Bayesian approaches to fMRI analyses exist in the literature, it felt appropriate to begin with a more simple approach.

Penny, Kiebel, and Friston (2003)

- In order to better understand which aspects of a Bayesian approach would most effectively allow detection of scattered signal, I decided to explore some existing methodologies.
- Moreover, while quite complex Bayesian approaches to fMRI analyses exist in the literature, it felt appropriate to begin with a more simple approach.
- Penny, Kiebel, and Friston (2003) propose analyzing fMRI data with a voxelwise, autoregressive model estimated by Variational Bayes approximation.

The Data

The Data

- Five $10 \times 10 \times 10$ voxel datasets in each of three signal-to-noise ratio conditions (2, 5, 10) were simulated.

The Data

- Five $10 \times 10 \times 10$ voxel datasets in each of three signal-to-noise ratio conditions (2, 5, 10) were simulated.
- The data were simulated to have spatial and temporal autocorrelation.

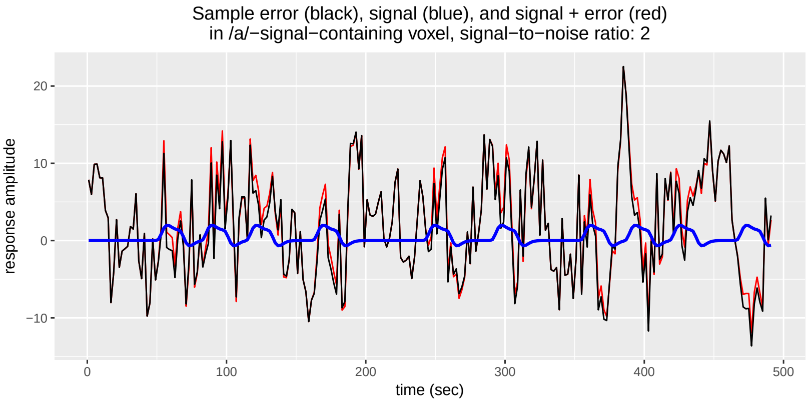
The Data

- Five 10 $\times$ 10 $\times$ 10 voxel datasets in each of three signal-to-noise ratio conditions (2, 5, 10) were simulated.
- The data were simulated to have spatial and temporal autocorrelation.
- Two types of signal were simulated (/a/ and /i/).

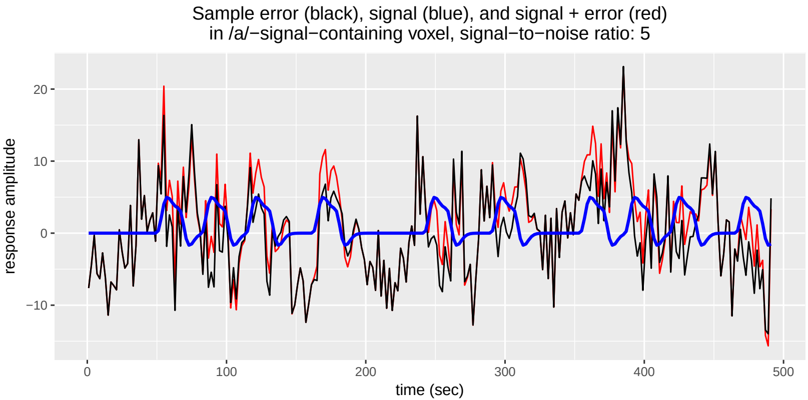
The Data

- Five $10 \times 10 \times 10$ voxel datasets in each of three signal-to-noise ratio conditions (2, 5, 10) were simulated.
- The data were simulated to have spatial and temporal autocorrelation.
- Two types of signal were simulated (/a/ and /i/).
- Each type was added to 27 randomly-chosen voxels.

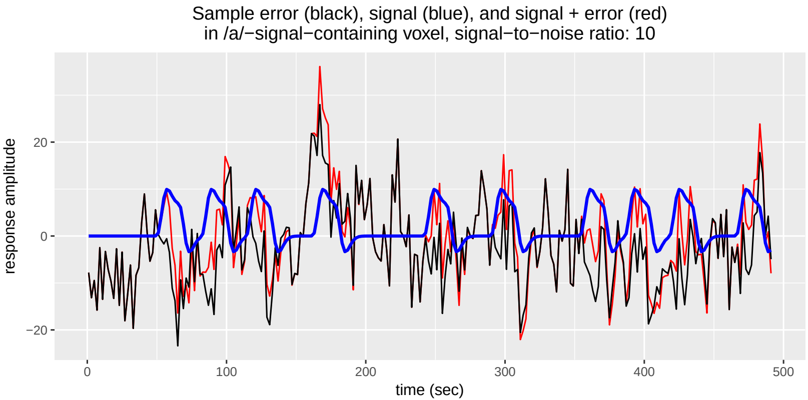
Signal-to-Noise Ratio: 2



Signal-to-Noise Ratio: 5



Signal-to-Noise Ratio: 10



The Model

The Model

- Recall that $\mathrm{Y}_j = [Y_{1j} \; Y_{2j} \; \cdots \; Y_{\tau j}]^T$ is the timeseries from voxel j.

The Model

- Recall that $\mathrm{Y}_j = [Y_{1j} \;\; Y_{2j} \;\; \cdots \;\; Y_{\tau j}]^T$ is the timeseries from voxel j.
- We model Y_j with an GLM-AR(2) model:

The Model

- Recall that $\mathrm{Y}_j = [Y_{1j} \; Y_{2j} \; \cdots \; Y_{\tau j}]^T$ is the timeseries from voxel j.
- We model Y_j with an GLM-AR(2) model:

$$\mathrm{Y}_j = \mathrm{X}\beta + \mathrm{E}_j$$

The Model

- Recall that $\mathsf{Y}_j = [Y_{1j} \ Y_{2j} \ \cdots \ Y_{\tau j}]^T$ is the timeseries from voxel j.
- We model Y_j with an GLM-AR(2) model:

$$\mathsf{Y}_j = \mathsf{X}\beta + \mathsf{E}_j$$

$$\mathsf{E}_j = \tilde{E}_j\alpha + \mathsf{Z}_j$$

The Model

- Recall that $\mathsf{Y}_j = [Y_{1j}\ \ Y_{2j}\ \ \cdots\ \ Y_{\tau j}]^T$ is the timeseries from voxel j.
- We model Y_j with an GLM-AR(2) model:

$$\mathsf{Y}_j = \mathsf{X}\beta + \mathsf{E}_j$$

$$\mathsf{E}_j = \tilde{E}_j\alpha + \mathsf{Z}_j$$

- $\tilde{E}_j$ is an embedded error matrix.

The Model

- Recall that $\mathsf{Y}_j = [Y_{1j} \; Y_{2j} \; \cdots \; Y_{\tau j}]^T$ is the timeseries from voxel j.
- We model Y_j with an GLM-AR(2) model:

$$\mathsf{Y}_j = \mathsf{X}\beta + \mathsf{E}_j$$

$$\mathsf{E}_j = \tilde{E}_j\alpha + \mathsf{Z}_j$$

- $\tilde{E}_j$ is an embedded error matrix.
- $\mathsf{Z}_j \overset{\text{iid}}{\sim} \mathsf{N}(0, \; \lambda^{-1}I)$

The Model

- Recall that $\mathsf{Y}_j = [Y_{1j} \; Y_{2j} \; \cdots \; Y_{\tau j}]^T$ is the timeseries from voxel j.
- We model Y_j with an GLM-AR(2) model:

$$\mathsf{Y}_j = \mathsf{X}\beta + \mathsf{E}_j$$

$$\mathsf{E}_j = \tilde{E}_j\alpha + \mathsf{Z}_j$$

- $\tilde{E}_j$ is an embedded error matrix.
- $\mathsf{Z}_j \overset{\text{iid}}{\sim} \mathsf{N}(0,\, \lambda^{-1}I)$
- The design matrix X contains an intercept column, a column for /a/ signal, and a column for /i/ signal.

Priors

Priors

- We need priors on the following parameters:

Priors

- We need priors on the following parameters:
    - β: coefficients of primary interest

Priors

- We need priors on the following parameters:
  - β: coefficients of primary interest
  - α: autoregression coefficients

Priors

- We need priors on the following parameters:
  - β: coefficients of primary interest
  - α: autoregression coefficients
  - λ: noise precision

Priors

- We need priors on the following parameters:
  - β: coefficients of primary interest
  - α: autoregression coefficients
  - λ: noise precision
- We collectively call these θ.

Priors

- We need priors on the following parameters:
  - β: coefficients of primary interest
  - α: autoregression coefficients
  - λ: noise precision
- We collectively call these θ.
- I follow Penny, Kiebel, and Friston (2003)'s prior specifications:

Priors

- We need priors on the following parameters:
  - β: coefficients of primary interest
  - α: autoregression coefficients
  - λ: noise precision
- We collectively call these θ.
- I follow Penny, Kiebel, and Friston (2003)'s prior specifications:
  - $p(\beta|b) = \mathrm{N}(0, b^{-1}I)$, where $b = 10^{-6}$

Priors

- We need priors on the following parameters:
  - β: coefficients of primary interest
  - α: autoregression coefficients
  - λ: noise precision
- We collectively call these θ.
- I follow Penny, Kiebel, and Friston (2003)'s prior specifications:
  - $p(\beta|b) = \mathrm{N}(0, b^{-1}I)$, where $b = 10^{-6}$
  - $p(\alpha|a) = \mathrm{N}(0, a^{-1}I)$, where $a = 10^{-3}$

Priors

- We need priors on the following parameters:
  - β: coefficients of primary interest
  - α: autoregression coefficients
  - λ: noise precision
- We collectively call these θ.
- I follow Penny, Kiebel, and Friston (2003)'s prior specifications:
  - $p(\beta|b) = \mathrm{N}(0, b^{-1}I)$, where $b = 10^{-6}$
  - $p(\alpha|a) = \mathrm{N}(0, a^{-1}I)$, where $a = 10^{-3}$
  - $p(\lambda) = \mathrm{Gamma}(\ell_1, \ell_2)$, where $\ell_1 = 1000$ and $\ell_2 = 0.001$

What is Variational Bayes?

What is Variational Bayes?

- Variational Bayes is a technique used to approximate the posterior distribution of parameters given the data.

What is Variational Bayes?

- Variational Bayes is a technique used to approximate the posterior distribution of parameters given the data.
- It is often less computationally intensive than finding the true posterior.

What is Variational Bayes?

- Variational Bayes is a technique used to approximate the posterior distribution of parameters given the data.
- It is often less computationally intensive than finding the true posterior.
- However, there are costs to using Variational Bayes:

What is Variational Bayes?

- Variational Bayes is a technique used to approximate the posterior distribution of parameters given the data.
- It is often less computationally intensive than finding the true posterior.
- However, there are costs to using Variational Bayes:
  - It often underestimates variances.

What is Variational Bayes?

- Variational Bayes is a technique used to approximate the posterior distribution of parameters given the data.
- It is often less computationally intensive than finding the true posterior.
- However, there are costs to using Variational Bayes:
  - It often underestimates variances.
  - It also often does a poor job at data correlation structure estimation.

What is Variational Bayes?

What is Variational Bayes?

Variational Bayes comes out of a recognition that $\log p(Y)$ can be expressed as follows, where $q(\theta|Y)$ is a generic probability density:

What is Variational Bayes?

Variational Bayes comes out of a recognition that $\log p(Y)$ can be expressed as follows, where $q(\theta|Y)$ is a generic probability density:

$$\log p(Y) = \int q(\theta|Y) \log p(Y) d\theta$$

What is Variational Bayes?

Variational Bayes comes out of a recognition that $\log p(Y)$ can be expressed as follows, where $q(\theta|Y)$ is a generic probability density:

$$\begin{aligned}\log p(Y) &= \int q(\theta|Y) \log p(Y) d\theta \\ &= \int q(\theta|Y) \log \frac{p(\theta, Y)}{p(\theta|Y)} d\theta\end{aligned}$$

What is Variational Bayes?

Variational Bayes comes out of a recognition that $\log p(Y)$ can be expressed as follows, where $q(\theta|Y)$ is a generic probability density:

$$
\begin{aligned}
\log p(Y) &= \int q(\theta|Y) \log p(Y) d\theta \\
&= \int q(\theta|Y) \log \frac{p(\theta, Y)}{p(\theta|Y)} d\theta \\
&= \int q(\theta|Y) \log \frac{q(\theta|Y) p(\theta, Y)}{p(\theta|Y) q(\theta|Y)} d\theta
\end{aligned}
$$

What is Variational Bayes?

Variational Bayes comes out of a recognition that $\log p(Y)$ can be expressed as follows, where $q(\theta|Y)$ is a generic probability density:

$$
\begin{aligned}
\log p(Y) &= \int q(\theta|Y) \log p(Y) d\theta \\
&= \int q(\theta|Y) \log \frac{p(\theta, Y)}{p(\theta|Y)} d\theta \\
&= \int q(\theta|Y) \log \frac{q(\theta|Y) p(\theta, Y)}{p(\theta|Y) q(\theta|Y)} d\theta \\
&= \int q(\theta|Y) \log \frac{q(\theta|Y)}{p(\theta|Y)} d\theta + \int q(\theta|Y) \log \frac{p(\theta, Y)}{q(\theta|Y)} d\theta
\end{aligned}
$$

What is Variational Bayes?

Variational Bayes comes out of a recognition that $\log p(Y)$ can be expressed as follows, where $q(\theta|Y)$ is a generic probability density:

$$\begin{aligned}
\log p(Y) &= \int q(\theta|Y) \log p(Y) d\theta \\
&= \int q(\theta|Y) \log \frac{p(\theta, Y)}{p(\theta|Y)} d\theta \\
&= \int q(\theta|Y) \log \frac{q(\theta|Y) p(\theta, Y)}{p(\theta|Y) q(\theta|Y)} d\theta \\
&= \int q(\theta|Y) \log \frac{q(\theta|Y)}{p(\theta|Y)} d\theta + \int q(\theta|Y) \log \frac{p(\theta, Y)}{q(\theta|Y)} d\theta \\
&= \text{Kullback-Leibler divergence} + \text{negative variational free energy}
\end{aligned}$$

What is Variational Bayes?

Variational Bayes comes out of a recognition that $\log p(Y)$ can be expressed as follows, where $q(\theta|Y)$ is a generic probability density:

$$\begin{aligned}
\log p(Y) &= \int q(\theta|Y) \log p(Y) d\theta \\
&= \int q(\theta|Y) \log \frac{p(\theta, Y)}{p(\theta|Y)} d\theta \\
&= \int q(\theta|Y) \log \frac{q(\theta|Y) p(\theta, Y)}{p(\theta|Y) q(\theta|Y)} d\theta \\
&= \int q(\theta|Y) \log \frac{q(\theta|Y)}{p(\theta|Y)} d\theta + \int q(\theta|Y) \log \frac{p(\theta, Y)}{q(\theta|Y)} d\theta \\
&= \text{Kullback-Leibler divergence} + \text{negative variational free energy} \\
&= KL + F
\end{aligned}$$

What is Variational Bayes?

What is Variational Bayes?

- The quantity $\log p(Y)$ is fixed.

What is Variational Bayes?

- The quantity $\log p(Y)$ is fixed.
- The Kullback-Leibler (KL) divergence is always greater than or equal to zero.

What is Variational Bayes?

- The quantity $\log p(Y)$ is fixed.
- The Kullback-Leibler (KL) divergence is always greater than or equal to zero.
- The KL divergence equals zero when $q(\theta|Y) = p(\theta|Y)$.

What is Variational Bayes?

- The quantity $\log p(Y)$ is fixed.
- The Kullback-Leibler (KL) divergence is always greater than or equal to zero.
- The KL divergence equals zero when $q(\theta|Y) = p(\theta|Y)$.
- By minimizing the KL divergence, we can get $q(\theta|Y)$ as close as possible to $p(\theta|Y)$.

What is Variational Bayes?

- The quantity $\log p(Y)$ is fixed.
- The Kullback-Leibler (KL) divergence is always greater than or equal to zero.
- The KL divergence equals zero when $q(\theta|Y) = p(\theta|Y)$.
- By minimizing the KL divergence, we can get $q(\theta|Y)$ as close as possible to $p(\theta|Y)$.
- Minimizing the KL divergence is the same as maximizing the negative variational free energy (F).

What is Variational Bayes?

What is Variational Bayes?

- With Variational Bayes, you impose restrictions on $q(\theta|Y)$.

What is Variational Bayes?

- With Variational Bayes, you impose restrictions on $q(\theta|Y)$.
- Penny, Kiebel, and Friston (2003) impose the limitation that the approximate posterior distributions of β, α, and λ be independent:

What is Variational Bayes?

- With Variational Bayes, you impose restrictions on $q(\theta|Y)$.
- Penny, Kiebel, and Friston (2003) impose the limitation that the approximate posterior distributions of β, α, and λ be independent:

$$q(\theta|Y) = q(\beta|Y)q(\alpha|Y)q(\lambda|Y)$$

What is Variational Bayes?

- With Variational Bayes, you impose restrictions on $q(\theta|Y)$.
- Penny, Kiebel, and Friston (2003) impose the limitation that the approximate posterior distributions of β, α, and λ be independent:

$$q(\theta|Y) = q(\beta|Y)q(\alpha|Y)q(\lambda|Y)$$

- Because of this restriction, the maximum value of F is attained when the individual approximate posteriors are of specific parametric forms ($\beta|Y$ is normal, $\alpha|Y$ is normal, $\lambda|Y$ is gamma).

What is Variational Bayes?

- With Variational Bayes, you impose restrictions on $q(\theta|Y)$.
- Penny, Kiebel, and Friston (2003) impose the limitation that the approximate posterior distributions of β, α, and λ be independent:

$$q(\theta|Y) = q(\beta|Y)q(\alpha|Y)q(\lambda|Y)$$

- Because of this restriction, the maximum value of F is attained when the individual approximate posteriors are of specific parametric forms ($\beta|Y$ is normal, $\alpha|Y$ is normal, $\lambda|Y$ is gamma).
- The formula to calculate the parameters of each posterior distribution involves the other posteriors. Thus, iterated application of the formulas is used until a previous value of F and a current value are within a preset tolerance.

Results

- Five datasets (i.e. "participants") are simulated for each of the three signal-to-noise ratio conditions (2, 5, 10).

Results

- Five datasets (i.e. "participants") are simulated for each of the three signal-to-noise ratio conditions (2, 5, 10).
- For each dataset, the approximate posteriors are estimated separately.

Results

- Five datasets (i.e. "participants") are simulated for each of the three signal-to-noise ratio conditions (2, 5, 10).
- For each dataset, the approximate posteriors are estimated separately.
- For each dataset, the probability that the coefficient associated with /a/ signal is greater than 0 is calculated.

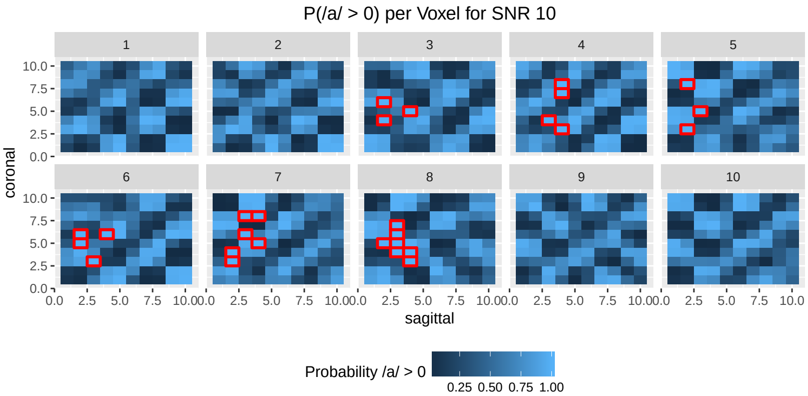
Results

- Five datasets (i.e. "participants") are simulated for each of the three signal-to-noise ratio conditions (2, 5, 10).
- For each dataset, the approximate posteriors are estimated separately.
- For each dataset, the probability that the coefficient associated with /a/ signal is greater than 0 is calculated.
- The same is done for /i/.

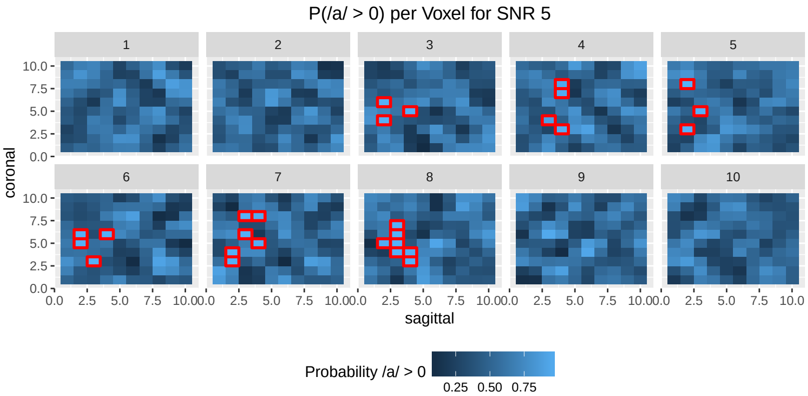
Results

- Five datasets (i.e. "participants") are simulated for each of the three signal-to-noise ratio conditions (2, 5, 10).
- For each dataset, the approximate posteriors are estimated separately.
- For each dataset, the probability that the coefficient associated with /a/ signal is greater than 0 is calculated.
- The same is done for /i/.
- Results from the two conditions are similar; the next plots thus just depict P(/a/ coefficient $>$ 0).

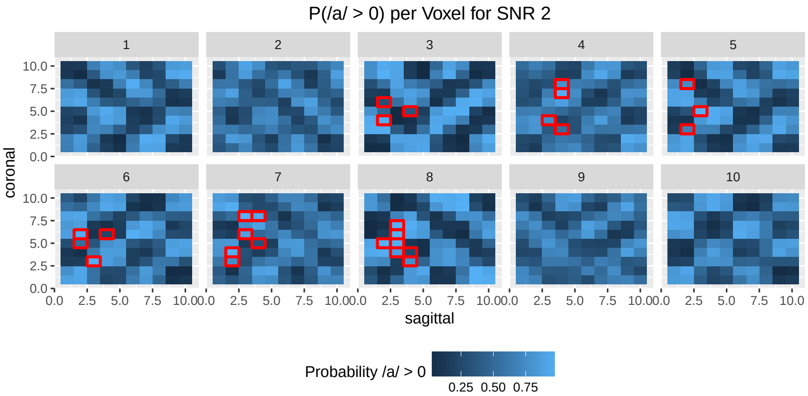
Signal-to-Noise Ratio: 10



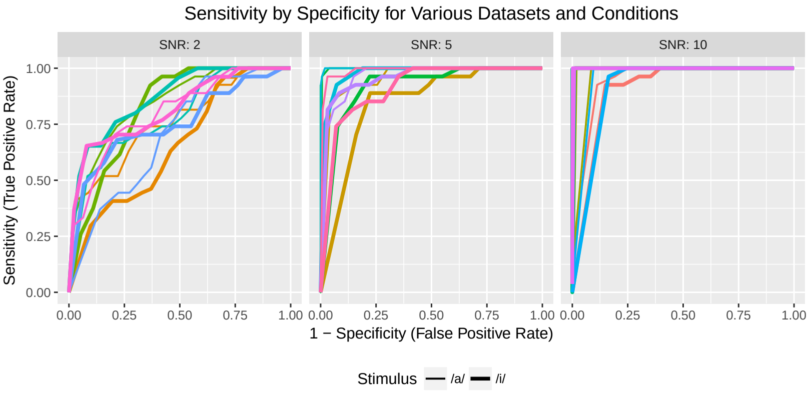
Signal-to-Noise Ratio: 5



Signal-to-Noise Ratio: 2



Receiver Operating Characteristic Curves



Conclusions

- With higher signal-to-noise ratios, Penny, Kiebel, and Friston's (2003) algorithm an do an excellent job at detecting scattered signal and only that signal (if the right threshold for "activated" voxels is chosen).

Conclusions

- With higher signal-to-noise ratios, Penny, Kiebel, and Friston's (2003) algorithm an do an excellent job at detecting scattered signal and only that signal (if the right threshold for "activated" voxels is chosen).
- For the more realistic low signal-to-noise condition, the algorithm did not perform very well.

Conclusions

- With higher signal-to-noise ratios, Penny, Kiebel, and Friston's (2003) algorithm an do an excellent job at detecting scattered signal and only that signal (if the right threshold for "activated" voxels is chosen).
- For the more realistic low signal-to-noise condition, the algorithm did not perform very well.
- More algorithms need to be explored.

Conclusions

- With higher signal-to-noise ratios, Penny, Kiebel, and Friston's (2003) algorithm an do an excellent job at detecting scattered signal and only that signal (if the right threshold for "activated" voxels is chosen).
- For the more realistic low signal-to-noise condition, the algorithm did not perform very well.
- More algorithms need to be explored.
- (Carefully!) capitalizing on spatial information may help.

Many thanks to...

- Ed Bedrick
- Dean Billheimer
- All of you!

References

Chang, E. F., Rieger, J. W., Johnson, K., Berger, M. S., Barbaro, N. M., and Knight, R. T. (2010). Categorical speech representation in human superior temporal gyrus. Nature Neuroscience, 13, 1428 - 1433.

Cover, T. M., and Thomas, J. A. (1991). Elements of Information Theory. Wiley, New York.

Mesgarani, N., Cheung, C., Johnson, K., and Chang, E. F. (2014). Phonetic feature encoding in human superior temporal gyrus. Science, 343, 1006 - 1010.

Penny, W., Kiebel, S., and Friston, K. (2003). Variational Bayesian inference for fMRI time series. NeuroImage, 19, 727 - 741.

Plant, R. E. (2012). Spatial Data Analysis in Ecology and Agriculture Using R. CRC Press, Boca Raton, FL.

Zhang, L., Guindani, M., Versace, F., Engelmann, J. M., and Vannucci, M. (2016). A Spatiotemporal Nonparametric Bayesian Model of Multi-Subject fMRI Data. The Annals of Applied Statistics, 10, 638 - 666.

The Noise Base

The Noise Base

- Each simulated dataset was a $10 \times 10 \times 10$ grid of voxels.

The Noise Base

- Each simulated dataset was a $10 \times 10 \times 10$ grid of voxels.
- To approximate the temporally autocorrelated noise present in fMRI data, 1000 independent timecourses were simulated from the model $e_t = \tilde{e}_t \alpha + z_t$, where

The Noise Base

- Each simulated dataset was a $10 \times 10 \times 10$ grid of voxels.
- To approximate the temporally autocorrelated noise present in fMRI data, 1000 independent timecourses were simulated from the model $e_t = \tilde{e}_t\alpha + z_t$, where
  - $\alpha^T = [0.3, 0.4]$

The Noise Base

- Each simulated dataset was a $10 \times 10 \times 10$ grid of voxels.
- To approximate the temporally autocorrelated noise present in fMRI data, 1000 independent timecourses were simulated from the model $e_t = \tilde{e}_t \alpha + z_t$, where
  - $\alpha^T = [0.3, 0.4]$
  - $\tilde{e}_t^T = [e_{t-1}, e_{t-2}]$

The Noise Base

- Each simulated dataset was a $10 \times 10 \times 10$ grid of voxels.
- To approximate the temporally autocorrelated noise present in fMRI data, 1000 independent timecourses were simulated from the model $e_t = \tilde{e}_t \alpha + z_t$, where
  - $\alpha^T = [0.3, 0.4]$
  - $\tilde{e}_t^T = [e_{t-1}, e_{t-2}]$
  - $z_t \overset{\text{iid}}{\sim} N(0, 1)$ for all t

The Noise Base

- Each simulated dataset was a $10 \times 10 \times 10$ grid of voxels.
- To approximate the temporally autocorrelated noise present in fMRI data, 1000 independent timecourses were simulated from the model $e_t = \tilde{e}_t \alpha + z_t$, where
  - $\alpha^T = [0.3, 0.4]$
  - $\tilde{e}_t^T = [e_{t-1}, e_{t-2}]$
  - $z_t \overset{\text{iid}}{\sim} N(0, 1)$ for all t
  - $t \in [1, 2, ..., 246]$

The Noise Base

The Noise Base

- To introduce spatial autocorrelation to the data, the simultaneous autoregressive model $B = \rho WB + \epsilon$ was employed:

The Noise Base

- To introduce spatial autocorrelation to the data, the simultaneous autoregressive model $B = \rho W B + \epsilon$ was employed:
  - W is a neighbor matrix with elements $w_{ij} = 1$ if voxels i and j are neighbors and 0 otherwise.

The Noise Base

- To introduce spatial autocorrelation to the data, the simultaneous autoregressive model $B = \rho WB + \epsilon$ was employed:
  - W is a neighbor matrix with elements $w_{ij} = 1$ if voxels i and j are neighbors and 0 otherwise.
  - $\rho = 0.3$ is the spatial autocorrelation coefficient.

The Noise Base

- To introduce spatial autocorrelation to the data, the simultaneous autoregressive model $B = \rho WB + \epsilon$ was employed:
  - W is a neighbor matrix with elements $w_{ij} = 1$ if voxels i and j are neighbors and 0 otherwise.
  - $\rho = 0.3$ is the spatial autocorrelation coefficient.
  - ϵ is the 1000×246 matrix of temporally autocorrelated errors for each of the 1000 voxels created in the previous step.

The Noise Base

- To introduce spatial autocorrelation to the data, the simultaneous autoregressive model $B = \rho WB + \epsilon$ was employed:
  - W is a neighbor matrix with elements $w_{ij} = 1$ if voxels i and j are neighbors and 0 otherwise.
  - $\rho = 0.3$ is the spatial autocorrelation coefficient.
  - ϵ is the 1000×246 matrix of temporally autocorrelated errors for each of the 1000 voxels created in the previous step.
  - B is the final matrix of spatially and temporally autocorrelated data.

The Noise Base

- To introduce spatial autocorrelation to the data, the simultaneous autoregressive model $B = \rho WB + \epsilon$ was employed:
  - W is a neighbor matrix with elements $w_{ij} = 1$ if voxels i and j are neighbors and 0 otherwise.
  - $\rho = 0.3$ is the spatial autocorrelation coefficient.
  - ϵ is the 1000×246 matrix of temporally autocorrelated errors for each of the 1000 voxels created in the previous step.
  - B is the final matrix of spatially and temporally autocorrelated data.
  - $B = (I - 0.3W)^{-1}\epsilon$

The Signal

The Signal

- Simulated experiment:

The Signal

- Simulated experiment:
  - Two types of stimuli (say the vowels /a/ and /i/) presented in 16-second blocks

The Signal

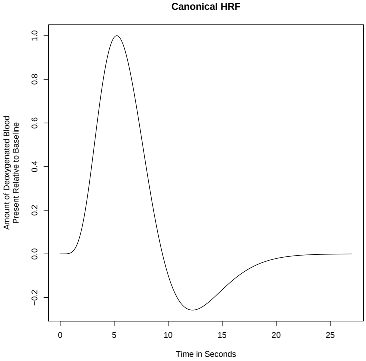
- Simulated experiment:
  - Two types of stimuli (say the vowels /a/ and /i/) presented in 16-second blocks
  - Ten blocks of each stimulus type interspersed with silent blocks

The Signal

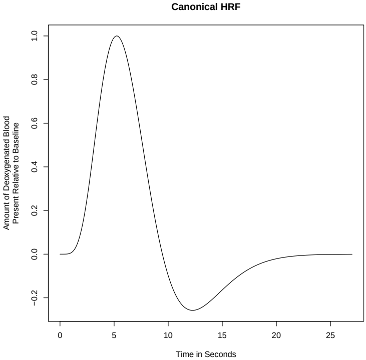
- Simulated experiment:
  - Two types of stimuli (say the vowels /a/ and /i/) presented in 16-second blocks
  - Ten blocks of each stimulus type interspersed with silent blocks
- When a stimulus is presented to a part of the brain that cares about it, the ratio of oxygenated to deoxygenated blood has a (fairly) predictable response — the hemodynamic response function (hrf).

The Signal

The Signal

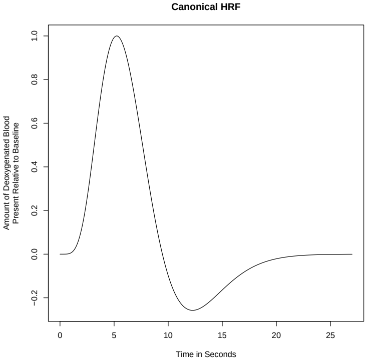

Canonical HRF
Amount of Deoxygenated Blood Present Relative to Baseline
−0.2
0.0
0.2
0.4
0.6
0.8
1.0
0
5
10
15
20
25
Time in Seconds

The Signal



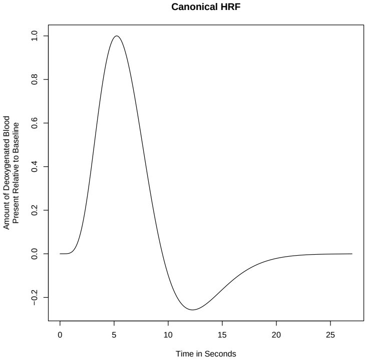
- Convolving the hrf with a stimulus indicator function produces simulated hemodynamic timecourses.

The Signal



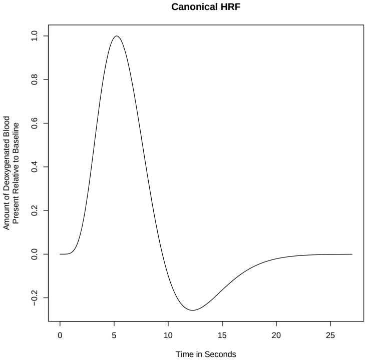
- Convolving the hrf with a stimulus indicator function produces simulated hemodynamic timecourses.
- The simulated timecourses are scaled to have maxima of 2, 5, or 10 (the signal-to-noise ratio conditions).

The Signal



- Convolving the hrf with a stimulus indicator function produces simulated hemodynamic timecourses.
- The simulated timecourses are scaled to have maxima of 2, 5, or 10 (the signal-to-noise ratio conditions).
- 27 voxels per dataset are randomly chosen to receive /a/ signal. Likewise for /i/.

The Signal



- Convolving the hrf with a stimulus indicator function produces simulated hemodynamic timecourses.
- The simulated timecourses are scaled to have maxima of 2, 5, or 10 (the signal-to-noise ratio conditions).
- 27 voxels per dataset are randomly chosen to receive /a/ signal. Likewise for /i/.
- 5 datasets are simulated per condition.

The Design Matrix

The Design Matrix

- Convolving the hrf with an indicator function of when stimuli occur produces predicted hemodynamic timecourses.

The Design Matrix

- Convolving the hrf with an indicator function of when stimuli occur produces predicted hemodynamic timecourses.
- One predicted hemodynamic timecourse is produced for each stimulus and included in the design matrix of the linear model.

The Design Matrix

- Convolving the hrf with an indicator function of when stimuli occur produces predicted hemodynamic timecourses.
- One predicted hemodynamic timecourse is produced for each stimulus and included in the design matrix of the linear model.
- The design matrix X for this work has two predicted hemodynamic timecourses (one for /a/ and one for /i/) and a column vector of ones.

The Design Matrix

- Convolving the hrf with an indicator function of when stimuli occur produces predicted hemodynamic timecourses.
- One predicted hemodynamic timecourse is produced for each stimulus and included in the design matrix of the linear model.
- The design matrix X for this work has two predicted hemodynamic timecourses (one for /a/ and one for /i/) and a column vector of ones.
- The $\hat{\beta}$ associated with a given predictor represents the extent to which a given voxel responded to that predictor.