

# A Global Bayes Factor for Observations on an Infinite-Dimensional Hilbert Space, Applied to Signal Detection in fMRI

**Khalil Shafie**

University of Northern Colorado

**Siamak Noorbaloochi**

University of Minnesota

**Mohammad Reza Faridrohani**

Shahid Beheshti University

**Hossein Moradi Rekabdarkolaei**

South Dakota State University

---

## Abstract

Functional Magnetic Resonance Imaging (fMRI) is a fundamental tool in advancing our understanding of the brain's functionality. Recently, a series of Bayesian approaches have been suggested to test for the voxel activation in different brain regions. In this paper, we propose a novel definition for the global Bayes factor to test for activation using the Radon-Nikodym derivative. Our proposed method extends the definition of Bayes factor to an infinite dimensional Hilbert space. Using this extended definition, a Bayesian testing procedure is introduced for signal detection in noisy images when both signal and noise are considered as an element of an infinite dimensional Hilbert space. This new approach is illustrated through a real data analysis to find activated areas of Brain in an fMRI data.

*Keywords:* Bayes factor, functional magnetic resonance imaging, Hilbert space, Radon-Nikodym derivative.

---

## 1. Introduction

Functional Magnetic Resonance Imaging (fMRI) measures the metabolic changes that occur within the brain in response to a stimulus. It is a fundamental tool in advancing our understanding of human brain functionality. fMRI is a sequence of magnetic resonance images (MRI). Each image contains a number of uniformly spaced volume elements, called pixels or voxels given the dimension of the image, that partition the brain into equal size boxes. The intensity of each pixel or voxel represents the spatial distribution of the nuclear spin density in that area. Changes in brain hemodynamics are due to the neuronal activity. These changes impact the local intensity of the MR signal. Therefore, changes in the pixel or voxel intensity across time can be used to infer when and where a particular activity happened. In last two decades, there have been many studies for finding activation in brain images obtained from fMRI. Lazar (2008) provided a comprehensive summary on statistical analysis of fMRI data. However, the drawback of traditional methods is the discretization of a continuous image in their analysis. Furthermore, these approaches make inference on a continuous object using

tools that is originally introduced for discrete data. In current work, we propose a method that directly works with continuous subjects by assuming them to be infinite dimensional random variables.

A common statistical inference approach for analytic (continuous) statistical processes is based on the random fields theory. Level crossings (Friston, Frith, Liddle, and Frackowiak 1991; Friston, Holmes, Worsley, Poline, Frith, and Frackowiak 1994) and differential topology (Worsley, Evans, Marrett, and Neelin 1992) are the two examples of such approach. Recently, a series of Bayesian approaches have been suggested to test for brain activation (Penny and Friston 2003; Flandin and Penny 2007). These detection methods are limited to cases where the signal is observed for at most countable points and the alternative is also a simple hypothesis test. Common Bayesian non-decision theoretic tools for hypotheses testing are posterior probability and Bayes factors (Poor 2013). Makni, Idier, Vincent, Thirion, Dehaene-Lambertz, and Ciuciu (2008) developed a Bayesian framework to conciliate both detection and estimation issues. Kim, Smyth, and Stern (2006) used Bayesian nonparametric methods to automatically select the number of activated clusters in an image. Rohani, Shafie, and Noorbaloochi (2006) proposed a Bayesian procedure to detect signals existing within noisy images when the image is modeled as a scale space random field. In this paper, we extend the scale space results to a general setting using the Radon-Nikodym derivative. Furthermore, we extend the definition of the Bayes factor for testing the point null hypothesis. Using this extended definition, we introduce a Bayesian testing procedure for signal detection in noisy images when both signal and noise are considered as an element of an infinite dimensional Hilbert space. Our proposed method avoids errors that can happen in the discretization process and provides a more informative result.

The rest of the paper is organized as follows. Section 2 briefly reviews the definition of the Bayes factor for testing a sharp null hypothesis against its complement for infinite dimensional abstract spaces. Section 3 briefly reviews the classical results of Gaussian measures on Hilbert spaces, the equivalence and orthogonality of measures, and the derivation of the Radon-Nikodym derivatives. Section 4 applies the Bayes factor to Gaussian modeling of images as Hilbert valued random objects and the Bayesian method of signal detection is explained in detail. We applied our proposed method to two models in Section 5. Finally, Section 6 concludes the paper with a short discussion.

## 2. Bayesian testing of a sharp (point) null hypothesis

The common sharp (point) hypothesis testing conducts the test  $H_0 : \theta = \theta_0$  against  $H_1 : \theta \neq \theta_0$  (Lee 1997). Bayesian testing procedure for this procedure for a finite dimensions setting starts with a random vector,  $X$ , with density  $f(x|\theta)$ . The unobserved conditioning parameter  $\theta$  is an unknown element of the parameter space  $\Theta$  which is a subset of  $\mathbb{R}^n$ . Using the two stage Jeffreys prior (Jeffreys 1998), we let  $\pi_0 = P(H_0)$  be the prior probability assigned to  $H_0$  and define the prior probability over  $\Theta$  to be

$$\pi(\theta) = \begin{cases} \pi_0, & \theta = \theta_0 \\ (1 - \pi_0)g(\theta), & \theta \neq \theta_0 \end{cases},$$

where  $g(\theta)$  is a conditional prior probability over the alternative values. The Bayes factor for assessing a null hypothesis  $H_0$  versus an alternative  $H_1$  is defined as

$$B(x) = \frac{\frac{P(H_0|x)}{P(H_1|x)}}{\frac{P(H_0)}{P(H_1)}}. \quad (1)$$

Let  $g(\theta)$  to be a continuous density on  $\{\theta \in \Theta : \theta \neq \theta_0\}$  or  $\Theta$ . Thus, for sharp hypothesis testing, the Bayes factor in equation (1) is equivalent to

$$B(x) = \frac{f(x|\theta_0)}{m_g(x)}, \quad (2)$$

where

$$m_g(x) = \int f(x|\theta)g(\theta)d\theta.$$

The Bayes factor definition in (2) combines prior information with the sample information in the likelihood. In other words, it can be considered as a weighted likelihood ratio of  $H_0$  to  $H_1$ . The weight function for a Bayes factor is the prior density,  $g$ , for the unknown parameter,  $\theta$ . Thus, a Bayes factor offers a way of evaluating evidence against a null hypothesis. The interpretation of Bayes factors is similar to the usual likelihood ratio. For instance, if we observe a value of  $B = \frac{1}{10}$  means that  $H_1$  is supported ten times as much by the data as is  $H_0$ . In addition to a Bayes factor, the posterior probability of  $H_0$  is also a relevant measure that can be used for the same purpose. In current setup, it is easy to see that the posterior probability is

$$P(H_0|x) = \left[1 + \frac{(1 - \pi_0) m_g(x)}{\pi_0 f(x|\theta_0)}\right]^{-1} = \left[1 + \frac{(1 - \pi_0) 1}{\pi_0 B}\right]^{-1}. \quad (3)$$

Following Grenander (1981), we extend the finite dimensional Bayes factor in equation (2) to a more general setting where the parameter and the sample spaces are infinite dimensions. Let  $(\Omega, \mathcal{F})$  be a probability space,  $X$  be a random object (measurable mapping) from  $(\Omega, \mathcal{F})$  to  $(\mathcal{X}, \mathcal{F}_1)$ ,  $\Theta$  be an abstract parameter space, and  $\mathcal{F}_2$  be a  $\sigma$ -algebra of subsets of  $\Theta$ . Suppose  $G$  is a probability measure on  $(\Theta, \mathcal{F}_2)$  and for each  $\theta \in \Theta$ ,  $P_\theta$  is the probability measure induced by  $X$  on  $(\mathcal{X}, \mathcal{F}_1)$ . In addition, for each  $A \in \mathcal{F}_1$ , let  $P_\theta(A) \in \mathcal{F}_2$  be a measurable function of  $\theta$ . Let  $P_G(A) = \int_\Theta P_\theta(A)G(d\theta)$  be the marginal measure of  $X$  on  $(\mathcal{X}, \mathcal{F}_1)$ . For a given  $\theta_0 \in \Theta$ , consider the measures  $P_{\theta_0}$  and  $P_G$ . According to the Lebesgue decomposition theorem, there exists a set  $A_0 \in \mathcal{F}_1$  with  $P_G(A_0) = 0$  and a nonnegative function  $f$  integrable with respect to  $P_G$  such that for any  $A \in \mathcal{F}_1$  we have

$$P_{\theta_0}(A) = \int_A f(x)P_G(dx) + P_{\theta_0}(A \cap A_0).$$

The function  $f$  is the Radon-Nikodym (R-N) derivative of  $P_{\theta_0}$  with respect to  $P_G$  and denoted by

$$f(x) = \frac{dP_{\theta_0}}{dP_G}(x).$$

In this setting, for testing  $H_0 : \theta = \theta_0$  versus  $H_1 : \theta \neq \theta_0$ , we define the Bayes factor to be

$$B(x) = \begin{cases} \infty & x \in A_0 \\ f(x) & x \notin A_0 \end{cases}. \quad (4)$$

Two problems that arise with this definition are how to obtain  $A_0$  and how to calculate  $f$ . The set  $A_0$  has either  $P_{\theta_0}(A_0) = 1$  or  $P_{\theta_0}(A_0) = 0$ . If  $P_{\theta_0}(A_0) = 1$ , the two measures  $P_{\theta_0}$  and  $P_G$  are orthogonal and a perfect decision about  $H_0$  can be made. If  $P_{\theta_0}$  and  $P_G$  are equivalent then  $P_{\theta_0}(A_0) = 0$  and  $f \neq 0$  with  $P_G$  probability of 1. Grenander (1981) provides a simultaneous characterization of  $A_0$  and  $f$  for cases where the observation  $x$  can be represented as a countable number of variates. In Section 4, we applied this notion to a common signal detection problem where a Gaussian measure model is used for both the observation and the conditional prior measure over some Hilbert spaces. In Section 3 we provide the model and preliminary results for our approach.

### 3. Gaussian measure on a Hilbert space

We are interested in characterizing the Bayes factor when observations are Hilbert valued random objects. Let  $\mathcal{H}$  be a real separable Hilbert space with inner product operator  $\langle \cdot, \cdot \rangle$ ,  $\mathcal{B}$  be the class of all Borel subsets of  $\mathcal{H}$ , and  $\mathcal{H}^*$  be the dual space of  $\mathcal{H}$ . The characteristic function of a measure,  $m$ , on  $(\mathcal{H}, \mathcal{B})$  for  $h \in \mathcal{H}$  is

$$\varphi(h) = \int_{\mathcal{H}} e^{i\langle x, h \rangle} dm(x).$$

Assuming  $m$  to be a probability measure on  $\mathcal{H}$ , then  $\varphi$  is a continuous positive definitoninite functional on  $\mathcal{H}$  and enjoys almost all properties of characteristic functions on Euclidean spaces. Varadhan (1968), Kuo (1975), and Bogachev (1998) provided a detailed review for the measures on Hilbert space. We need the following definitions and theorem for developing the method proposed in Section 4.

**Definition 1.** If  $m$  is a measure on  $\mathcal{H}$ , then the mean function,  $\mu$ , is an element of  $\mathcal{H}$  definitioned by

$$\langle \mu, h \rangle = \int \langle x, h \rangle dm(x), \quad \text{for all } h \in H.$$

**Definition 2.** If  $m$  is a measure on  $\mathcal{H}$ , then the covariance operator is a bilinear function on  $\mathcal{H}$  definitioned by

$$\langle Sh_1, h_2 \rangle = \int \langle h_1, x \rangle \langle h_2, x \rangle dm(x).$$

The covariance is a symmetric positive definitoninite bilinear form and the operator  $S$  definitioning the bilinear form is a positive semidefinitioninite Hermitian operator.

A Gaussian probability measure  $P$  on  $\mathcal{H}$  is a measure such that its characteristic function has the form

$$\varphi(h) = \exp(i \langle \mu, h \rangle - \frac{1}{2} \langle Sh, h \rangle),$$

where  $\mu$  and  $S$  are the mean and covariance of  $P$ , respectively. Following, we present the Theorems and Lemmas on the equivalence of two Gaussian measures on a Hilbert space which are needed for further developments. Proofs of these statements can be found in Varadhan (1968).

**Theorem 1.** *Two Gaussian measures definitioned on a Hilbert space are either orthogonal or equivalent.*

**Lemma 1.** *Suppose  $P_1 = HN(\mu_1, S_1)$  and  $P_2 = HN(\mu_2, S_2)$ , where  $HN(\cdot, \cdot)$  denotes the normal distribution definitioned on Hilbert Space. Let  $P'_1 = HN(0, S_1)$  and  $P'_2 = HN(0, S_2)$ . If  $P_1$  is not orthogonal to  $P_2$ , then  $P'_1$  and  $P'_2$  are not orthogonal either.*

**Lemma 2.** *Suppose  $P_1 = HN(0, S_1)$  and  $P_2 = HN(0, S_2)$  are equivalent. Then there exists a Hermitian bounded invertible operator  $T$  such that  $T - I$  belongs to Hilbert- Schmidt class and*

$$S_2 = S_1^{1/2} T S_1^{1/2}.$$

**Theorem 2.** *Let  $P_1 = HN(0, S)$  and  $P_2 = HN(\mu, S)$ . Then  $P_1$  and  $P_2$  are either equivalent or orthogonal depending on whether or not  $\mu$  is in  $R(S^{1/2})$ , where  $R(S^{1/2})$  denotes the range of operator  $S^{1/2}$ .*

**Theorem 3.**  *$P_1 = HN(\mu_1, S_1)$  is either orthogonal to  $P_2 = HN(\mu_2, S_2)$  or is equivalent to it. For equivalence, the following assertions are necessary and sufficient*

$$HN(\mu_1, S_1) \text{ is equivalent to } HN(\mu_2, S_1),$$

and

$$HN(\mu_2, S_1) \text{ is equivalent to } HN(\mu_2, S_2).$$

Suppose  $P_1$  and  $P_2$  are two Gaussian measures on  $\mathcal{H}$  with means  $\mu_1$  and  $\mu_2$  and non-singular covariance operators  $S_1$  and  $S_2$ , respectively.

**Theorem 4.** *Suppose  $P_1 = HN(\mu_1, S)$  and  $P_2 = HN(\mu_2, S)$  are equivalent. Then, the Radon-Nikodym derivative of  $P_1$  with respect to  $P_2$ ,  $f(x)$ , is*

$$f(x) = \exp \left( \sum_{i=1}^{\infty} \frac{(\mu_{1i} - \mu_{2i})}{\lambda_i} \left[ Z_i(x) - \frac{\mu_{1i} + \mu_{2i}}{2} \right] \right),$$

where, for  $i = 1, 2, \dots, \infty$ ,  $Z_i(x) = \langle x, e_i \rangle$ ,  $\mu_{ki} = \langle e_i, \mu_k \rangle$ , and  $\lambda_i$ s are eigenfunctions and their corresponding eigenvalues of  $S$ .

**Theorem 5.** *Suppose  $P_1 = HN(\mu, S_1)$  and  $P_2 = HN(\mu, S_2)$  are equivalent. Then, the Radon-Nikodym derivative of  $P_1$  with respect to  $P_2$  denoted by  $f(x)$  is*

$$f(x) = \exp \left( -\frac{1}{2} \sum_{i=1}^{\infty} \left[ (Z_i(x) - \mu_i)^2 \left(1 - \frac{1}{\lambda_i}\right) - \log \lambda_i \right] \right), \quad (5)$$

where, for  $i = 1, 2, \dots, \infty$ ,  $\mu_i$  is the mean value of  $Z_i(x)$  given  $P_1$ . In equation (5),  $Z_i(x) = \sum_{j=1}^{\infty} a_{ij} u_j(x)$  where, for  $j = 1, 2, \dots, \infty$ ,  $u_j(x) = \langle e_j, x \rangle / \sqrt{s_j}$  with  $e_j$  and  $s_j$  denoting the eigenfunctions and their corresponding eigenvalues of  $S_1$ , respectively. Furthermore,  $\lambda_j$  and  $g_j$  are the eigenfunctions and eigenvalues of  $T$ , respectively, where  $T$  is such that  $S_2 = S_1^{1/2} T S_1^{1/2}$ , and  $g_j = \sum_{k=1}^{\infty} a_{jk} e_k$ .

## 4. Signal detection

Let  $(\mathcal{T}, \mathcal{F}, m)$  be a measurable space and  $L_2(\mathcal{T}, m)$  be a Hilbert space of squared integrable real functions on  $\mathcal{T}$ . Let  $\mathcal{X} = \Theta = L_2(\mathcal{T}, m)$  and  $\mathcal{F}_1$  and  $\mathcal{F}_2$  be the Borel sigma field of  $L_2(\mathcal{T}, m)$ . For each  $\theta \in \Theta$ , let  $P_\theta$  be a Gaussian measure with mean  $\theta$  and covariance operator,  $\rho$ . In addition, let  $G$ , the prior measure on  $\Theta$ , to be a Gaussian measure with known mean  $\mu \in \Theta$  and covariance operator,  $\tau$ . Using the characteristic function approach, it can be easily proved that the marginal measure of  $X$ ,  $P_G$ , is also a Gaussian measure with the mean  $\mu$  and covariance operator  $\rho + \tau$ .

For Bayesian hypothesis testing of  $H_0 : \theta = 0$  versus  $H_1 : \theta \neq 0$ , we should know if  $P_0$  and  $P_G$  are equivalent or orthogonal. If they are orthogonal, then a perfect decision can be made between  $H_0$  and  $H_1$ . Otherwise, we need to derive the R-N derivative of  $P_0$  with respect to  $P_G$ . The following result provides a sufficient condition.

**Theorem 6.** *If  $P_0 = HN(0, \rho)$  and  $G = HN(\mu, \tau)$  are equivalent then  $P_0$  and  $P_G$  are orthogonal.*

*Proof.* Suppose  $P_0$  and  $P_G$  are equivalent. Let  $G' = HN(0, \tau)$ , then according to Lemma 1 and Theorem 1  $P_0$  and  $P_{G'}$  are equivalent. From Lemma 2, there exists an invertible bounded Hermitian operator  $T$  such that  $(T - I)$  is of Hilbert-Schmidt class and

$$\rho + \tau = \rho^{1/2} T \rho^{1/2},$$

or

$$\tau = \rho^{1/2} (T - I) \rho^{1/2}. \quad (6)$$

On the other hand, since  $P_0$  and  $G'$  are equivalent, there exists an invertible bounded Hermitian operator  $D$  such that  $(D - I)$  is of Hilbert-Schmidt class and

$$\tau = \rho^{1/2} D \rho^{1/2} \quad (7)$$

Comparing (6) and (7) shows both operators  $T - I$  and  $T - 2I$  are of Hilbert-Schmidt class which is a contradiction. The following example shows that the orthogonality of  $P_0$  and  $G$  does not yield the equivalence of  $P_0$  and  $P_G$ .  $\square$

**Example.** Let  $P_0 = HN(0, \rho)$  and  $G = HN(\mu, \rho)$  where  $\mu \notin R(\rho^{1/2})$ . In this case, according to Theorem 2,  $P_0$  and  $G$  are orthogonal. If  $P_0$  and  $P_G = HN(\mu, 2\rho)$  are equivalent, then by Theorem 3, so are  $P_0 = HN(0, \rho)$  and  $G = HN(\mu, \rho)$ , which is a contradiction.

This example can be generalized for any  $P_0 = HN(0, \rho)$  and  $G = HN(\mu, \tau)$ , where  $\mu \notin R(\rho^{1/2})$ . In the following theorem, we consider a special case where  $P_0$  and  $P_G$  are equivalent and the R-N derivative has a simple form of the exponential of a diagonal form. A general form of the theorem is proved in Varberg (1967) and Chung and Rajput (1981). Here, we give a direct proof for an special case.

**Theorem 7.** Let  $\mathcal{H} = L_2(\mathcal{T}, m)$ , where  $\mathcal{T} \subset \mathbb{R}^n$  is a compact set and  $m$  is the Lebesgue measure. In addition, let  $P_0 = HN(0, \rho)$  and  $G = HN(0, \tau)$ , where  $\rho$  is continuous on  $\mathcal{T} \times \mathcal{T}$ . Furthermore,  $\tau(\mathbf{t}, \mathbf{s}) = \sum_{i=1}^{\infty} \frac{\lambda_i^2}{1-\lambda\lambda_i} \phi_i(\mathbf{t})\phi_i(\mathbf{s})$ , with  $\{\lambda_i\}$  and  $\{\phi_i\}$  are eigenvalues and the corresponding eigenfunctions of the operator induced by  $\rho$ , respectively, and  $\lambda > 0$  is such that for  $i = 1, 2, \dots$ ,  $1 - \lambda\lambda_i > 0$ . Then,  $P_0$  and  $P_G$  are equivalent and

$$\frac{dP_0}{dP_G} = D^{-1/2}(\lambda) \exp\left(-\frac{\lambda}{2} \int_{\mathcal{T}} X^2(\mathbf{t}) dm(\mathbf{t})\right), \tag{8}$$

where  $D(\lambda) = \prod_{i=1}^{\infty} (1 - \lambda\lambda_i)$ .

*Proof.* Since  $\sum_{i=1}^{\infty} \lambda_i < \infty$ ,  $1/(1 - \lambda\lambda_i)$  is bounded and therefore the series in the definition of  $\tau$  is absolutely convergent. In addition,  $\rho$  is continuous, so using the Mercer's theorem, we have

$$\rho(\mathbf{t}, \mathbf{s}) = \sum_{i=1}^{\infty} \lambda_i \phi_i(\mathbf{t})\phi_i(\mathbf{s}).$$

Thus

$$\begin{aligned} r(\mathbf{t}, \mathbf{s}) = \rho(\mathbf{t}, \mathbf{s}) + \tau(\mathbf{t}, \mathbf{s}) &= \sum_{i=1}^{\infty} \left[ \frac{\lambda\lambda_i^2}{1 - \lambda\lambda_i} + \lambda_i \right] \phi_i(\mathbf{t})\phi_i(\mathbf{s}) \\ &= \sum_{i=1}^{\infty} \frac{\lambda_i}{1 - \lambda\lambda_i} \phi_i(\mathbf{t})\phi_i(\mathbf{s}). \end{aligned}$$

Let  $\gamma_{ij} = (1 - \lambda\lambda_i)\delta_{ij}$  where  $\delta$  is the Kroneker delta. Then, the operator  $\Gamma$  definitioned by  $[\gamma_{ij}]_{i,j=1}^{\infty}$  is invertible, bounded, and since

$$\sum_{i,j} (\gamma_{ij} - \delta_{ij})^2 = \sum \lambda^2 \lambda_i^2 < \infty,$$

the  $(\Gamma - I)$  is a Hilbert-Schmidt operator. It can be easily proved that  $\rho = r^{1/2}\Gamma r^{1/2}$ . Hence, by Lemma 2,  $P_0$  and  $P_G$  are equivalent.

To obtain the R-N derivative, note that  $\Gamma$  and  $r$  have the same eigenfunctions  $\phi_i$  and their corresponding eigenvalues are  $\{1 - \lambda\lambda_i\}$  and  $\{\frac{\lambda_i}{1-\lambda\lambda_i}\}$ , respectively. Therefore, correspond to  $a_{ij}$ ,  $g_i$ ,  $s_i$ ,  $e_i$ , and  $\lambda_i$  in Theorem 5 are  $1, \phi_i, \lambda_i/(1 - \lambda\lambda_i), \phi_i$ , and  $\{1 - \lambda\lambda_i\}$ , respectively. Hence,

$$u_i(X) = Z_i(X) = \sqrt{\frac{1 - \lambda\lambda_i}{\lambda_i}} \int_{\mathcal{T}} X(\mathbf{t})\phi_i(\mathbf{t}) dm(\mathbf{t}),$$

and

$$\sum_{i=1}^{\infty} Z_i^2 [1 - 1/(1 - \lambda\lambda_i)] = \lambda \sum_{i=1}^{\infty} \int_{\mathcal{T}} X(\mathbf{t})\phi_i(\mathbf{t}) dm(\mathbf{t}) = \lambda \int_{\mathcal{T}} X^2(\mathbf{t}) dm(\mathbf{t}).$$

If we substitute all these in (5), we get the result. □

Using Theorem 4, we have the following for the case  $\mu \neq 0$ .

**Theorem 8.** Under the conditions of Theorem 7, if  $G = HN(\mu, \tau)$  with  $\mu \in R(\rho^{1/2})$ , we have

$$\frac{dP_0}{dP_G} = D^{-1/2}(\lambda) \exp \left[ - \left( \frac{\lambda}{2} \int_{\mathcal{T}} X^2(\mathbf{t}) dm(\mathbf{t}) + \sum_{i=1}^{\infty} \frac{\mu_i(1 - \lambda\lambda_i)}{\lambda_i} (Z_i + \frac{\mu_i}{2}) \right) \right], \quad (9)$$

where  $\mu_i = \int_{\mathcal{T}} \mu(\mathbf{t}) \phi_i(\mathbf{t}) dm(\mathbf{t})$ , and  $Z_i = \int_{\mathcal{T}} X(\mathbf{t}) \phi_i(\mathbf{t}) dm(\mathbf{t})$ .

## 5. Application to fMRI data

In this section, we apply the proposed methodology on a fMRI scan of the human brain. Note that the main purpose of this data analysis is to show that proposed approach can be used for fMRI data analysis.

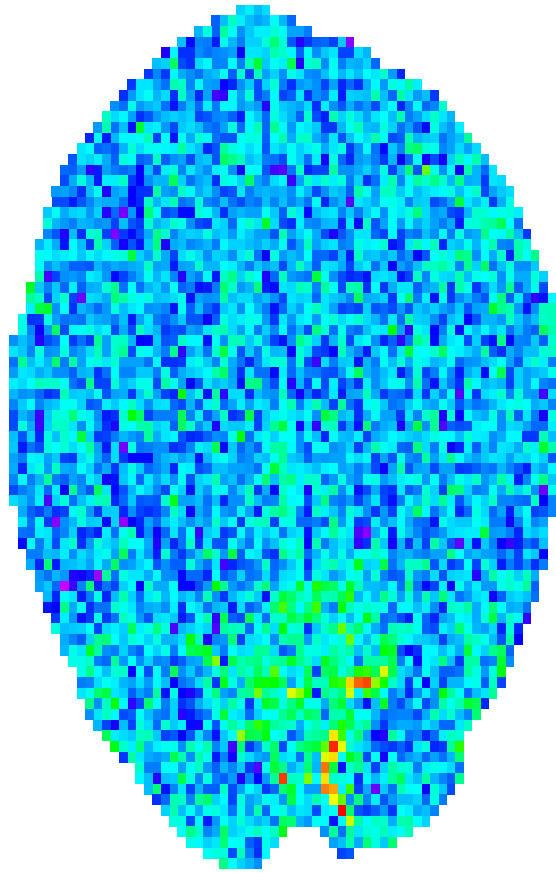


Figure 1: The cosine component of the fMRI data.

One of the common object in using fMRI is to locate the regions of the brain that respond to a simple visual stimulus (Kwong, Belliveau, Chesler, Goldberg, Weisskoff, Poncelet, Kennedy, Hoppel, Cohen, and Turner 1992). In an experiment at the Montreal Neurological Institute, a subject was given a simple visual stimulus, flashing red dots, presented through light-tight goggles (Ouyang, Pike, and Evans 1994). The stimulus was switched off for 4 scans, then on for 4 scans. This procedure was repeated 5 times which resulted in total of 40 scans. The time interval between pair of scans was 6 seconds and the stimulus period was  $\mathcal{T} = 48$  seconds. Hence, the data set consists of a time series of 40 2- $D$  images, each  $128 \times 128$  pixels of size 2 mm.

This data set has been analyzed by several researchers using the  $\chi^2$  scale space method (Worsley 2001), rotation space random field method (Shafie, Sigal, Siegmund, and Worsley 2003), and Bayesian method (Rohani *et al.* 2006). To apply the random field methodology, these researchers fitted a sine wave at each fixed pixel and extracted two images. These images were referred to as sine and cosine components of the data. The cosine component is shown in Figure 1. Here, we apply the proposed Bayes factor method to this image to find out if there is any sign of activation caused by the stimulus.

The Matched Filter Theorem justifies smoothing the image before any analysis with a filter of the form of signal. The smoothing assumption is essential for methods proposed by Worsley (2001) and Shafie *et al.* (2003). In order to get a smooth random field, the image is smoothed before analyzing data with a Gaussian filter. However, in practice the shape of the signal is not known. Therefore, smoothing with a kernel is not quite justifiable. Note that in our approach, we do not need this smoothing procedure.

We will consider the following two models:

**Model I.** Let

$$X(\mathbf{t}) = \theta(\mathbf{t}) + Z(\mathbf{t}), \quad \mathbf{t} \in \mathcal{T},$$

where  $\mathcal{T} = [0, 1] \times [0, 1]$ ,  $\theta \in L_2(\mathcal{T}, m)$ , and  $Z$  is Gaussian white noise in  $\mathcal{T}$ . Under these assumptions,  $P_0 = HN(0, \rho)$  with

$$\rho((t_1, t_2), (s_1, s_2)) = \begin{cases} 1 & t_1 = s_1, t_2 = s_2 \\ 0 & \text{otherwise} \end{cases}$$

This model is the same as the one used by Siegmund and Worsley (1995). The main difference is that here, the image is not pre-smoothed for the analysis. The model does not fit the implicit assumptions in section 4 as white noise cannot be realized in the Hilbert space. However, the theory can be extended to cover this case. For further information, see Balakrishnan (2012) and Adler and Taylor (2009).

Suppose  $\theta$  follows a  $HN(\mu, \tau)$  where  $\int \mu(\mathbf{t})^2 d\mathbf{t} < \infty$ , then  $P_0 = HN(0, I)$  and  $P_G = HN(\mu, \tau + I)$ . It is clear that  $P_0$  and  $P_G$  are equivalent and we have

$$f(x) = \exp \left[ -\frac{1}{2} \left( \sum_{i=1}^{\infty} (Z_i - \mu_i)^2 \frac{t_i}{t_i + 1} + \frac{2\mu_i}{t_i + 1} (Z_i + \frac{\mu_i}{2}) - \log(t_i + 1) \right) \right],$$

where for  $i = 1, 2, \dots, \infty$ ,  $Z_i = \langle x, e_i \rangle$ ,  $\mu_i = \langle \mu, e_i \rangle$ , and  $e_i$ s are the eigenvalues and corresponding eigenfunctions of  $\tau$ , respectively. For the one dimensional case,  $f(x)$  is similar to the criterion used for the detection of random Gaussian signal in Gaussian noise in signal processing literature. For further information, see Poor (2013). For a specific model, consider a Brownian sheet prior with mean 0 and the following covariance function

$$\tau((t_1, t_2), (s_1, s_2)) = \min(t_1, s_1) \min(t_2, s_2). \quad (10)$$

Since the eigenvalues and corresponding eigenfunctions of the Brownian motion on  $[0, 1]$  are  $\{\frac{1}{(i+1/2)^2\pi^2}\}$  and  $\{\sqrt{2} \sin((i+1/2)\pi t)\}$ , respectively (see Gikhman and Skorokhod (1969) page 189), the eigenvalues and eigenfunctions of  $\tau$  are  $\lambda_{ij} = \frac{1}{(i+1/2)^2(j+1/2)^2\pi^4}$  and  $\phi_{ij}(t_1, t_2) = 2 \sin((i+1/2)\pi t_1) \sin((j+1/2)\pi t_2)$ , respectively. For the observed image we obtain approximately 1.0045 as the Bayes factor.

Since the assumption of white noise may not be very realistic, we consider the following model.

**Model II.** Let

$$X(\mathbf{t}) = \theta(\mathbf{t}) + B(\mathbf{t}), \quad \mathbf{t} \in T,$$

where  $T = [0, 1] \times [0, 1]$ ,  $\theta \in L_2(T, m)$  and  $B$  is the standard Brownian sheet in  $T$ . Under these assumptions,  $P_0 = HN(0, \rho)$  where  $\rho(t_1, t_2)$  is the same as  $\tau$  in equation (10). To eliminate trivial cases and for simplicity, we assume  $\theta$  has a  $HN(0, \tau)$  prior where

$$\tau(\mathbf{t}, \mathbf{s}) = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \frac{\lambda \lambda_{ij}^2}{1 - \lambda \lambda_{ij}} \phi_{ij}(\mathbf{t}) \phi_{ij}(\mathbf{s}),$$



and  $0 < \lambda < \frac{81}{16}$  is a hyperparameter of the prior. Using Theorem 7, the discussion of model I, and the fact that for the observed image  $\int_T X(\mathbf{t})dm(\mathbf{t}) = 0.39$ , the Bayes factor for testing  $H_0 : \theta = 0$  versus  $H_1 : \theta \neq 0$  is

$$f_\lambda = D^{-1/2}(\lambda) \exp\left(-0.39 \frac{\lambda}{2}\right),$$

where  $D(\lambda) = \prod_{ij}(1 - \lambda\lambda_{ij})$ . Approximating the  $\log(1 - x)$  with  $-x$  and since  $\sum_{ij} \lambda_{ij} = 1/4$ , we have

$$f_\lambda \approx \exp(-0.07\lambda).$$

This means that for the values of  $0 < \lambda < 81/16$ , having  $HN(0, \tau)$  as a prior for  $\theta$ , we fail to reject the null hypothesis of no signal.

## 6. Conclusion

Research has greatly improved our understanding of the human brain. One of the common tools in brain image analysis is fMRI. However, relatively few studies have looked at this imaging technique as giving an observation as an infinite dimensional Hilbert random variable. In this paper, we extended the definition of the Bayes factor to a more general setting where the variables of interest can be members of an infinite dimensional Hilbert space. In this paper, our focus is to provide inferential tools for analyzing images obtained through fMRI techniques, but the approach might be used for analyzing other kinds of images or other kinds of data that can be expressed in the form of an image.

For the real data analysis, the Bayes factors for Models I and II find a small sign of a signal in Figure 1, which appears to show a signal. The Bayes factors are 1.0045 for Model I, and between 1 and about 0.70 for Model II. This would seem to suggest that this approach has low power, at least for the cosine component of this dataset. There are three possible explanations for why this might have happened.

- i The power might be greater if the time series for each voxel was reduced to a 2-D image using something other than the cosine component.
- ii In practice, to increase the signal-to-noise ratio, images are first smoothed using a smooth kernel. The models we have considered here, models I and II, are missing this feature.
- iii The original fMRI dataset consists of a time series of images. In our models, the time series of the images was reduced to a 2-D image. If we consider the time series of the images, as a time series of Hilbert valued variables, we might have better detected the signal.

The last two explanations are the topic of our current research.

## Acknowledgments

The authors would like to thank the editor and the referee for their careful reading and many constructive comments that led to substantial improvements in the article.

## References

- Adler RJ, Taylor JE (2009). *Random Fields and Geometry*. Springer Science & Business Media.

- Balakrishnan AV (2012). *Applied Functional Analysis*, volume 3. Springer Science & Business Media.
- Bogachev VI (1998). *Gaussian Measures*. 62. American Mathematical Society.
- Chung DM, Rajput BS (1981). “Equivalent Gaussian Measure whose RN Derivative Is the Exponential of a Diagonal Form.” *Journal of Mathematical Analysis and Applications*, **81**(1), 219–233.
- Flandin G, Penny WD (2007). “Bayesian fMRI Data Analysis with Sparse Spatial Basis Function Priors.” *NeuroImage*, **34**(3), 1108–1125.
- Friston KJ, Frith CD, Liddle PF, Frackowiak RSJ (1991). “Comparing Functional (PET) Images: The Assessment of Significant Change.” *Journal of Cerebral Blood Flow & Metabolism*, **11**(4), 690–699.
- Friston KJ, Holmes AP, Worsley KJ, Poline JP, Frith CD, Frackowiak RSJ (1994). “Statistical Parametric Maps in Functional Imaging: A General Linear Approach.” *Human Brain Mapping*, **2**(4), 189–210.
- Gikhman II, Skorokhod AV (1969). *Introduction to the Theory of Random Processes*. WB Saunders Co.
- Grenander U (1981). *Abstract Inference. 1981*. Wiley, New York.
- Jeffreys H (1998). *The Theory of Probability*. OUP Oxford.
- Kim S, Smyth P, Stern H (2006). “A Nonparametric Bayesian Approach to Detecting Spatial Activation Patterns in fMRI Data.” In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 217–224. Springer.
- Kuo HH (1975). “Gaussian Measures in Banach Spaces.” In *Gaussian Measures in Banach Spaces*, pp. 1–109. Springer.
- Kwong KK, Belliveau JW, Chesler DA, Goldberg IE, Weisskoff RM, Poncelet BP, Kennedy DN, Hoppel BE, Cohen MS, Turner R (1992). “Dynamic Magnetic Resonance Imaging of Human Brain Activity during Primary Sensory Stimulation.” *Proceedings of the National Academy of Sciences*, **89**(12), 5675–5679.
- Lazar N (2008). *The Statistical Analysis of Functional MRI Data*. Springer Science & Business Media.
- Lee PM (1997). *Bayesian Statistics*. Arnold Publication.
- Makni S, Idier J, Vincent T, Thirion B, Dehaene-Lambertz G, Ciuciu P (2008). “A Fully Bayesian Approach to the Parcel-based Detection-estimation of Brain Activity in fMRI.” *Neuroimage*, **41**(3), 941–969.
- Ouyang X, Pike GB, Evans AC (1994). “fMRI of Human Visual Cortex Using Temporal Correlation and Spatial Coherence Analysis.” In *13th Annual Symposium of the Society of Magnetic Resonance in Medicine*.
- Penny W, Friston K (2003). “Mixtures of General Linear Models for Functional Neuroimaging.” *IEEE Transactions on Medical Imaging*, **22**(4), 504–514.
- Poor HV (2013). *An Introduction to Signal Detection and Estimation*. Springer Science & Business Media.
- Rohani MF, Shafie K, Noorbaloochi S (2006). “A Bayesian Signal Detection Procedure for Scale-space Random Fields.” *Canadian Journal of Statistics*, **34**(2), 311–325.

- Shafie K, Sigal B, Siegmund D, Worsley KJ (2003). “Rotation Space Random Fields with an Application to fMRI Data.” *The Annals of Statistics*, **31**(6), 1732–1771.
- Siegmund DO, Worsley KJ (1995). “Testing for a Signal with Unknown Location and Scale in a Stationary Gaussian Random Field.” *The Annals of Statistics*, **23**(2), 608–639.
- Varadhan SRS (1968). “Stochastic Processes. Courant Institute of Math.” *Sciences, New York*.
- Varberg DE (1967). “Equivalent Gaussian Measures with a Particularly Simple Radon-Nikodym Derivative.” *The Annals of Mathematical Statistics*, **38**(4), 1027–1030.
- Worsley KJ (2001). “Testing for Signals with Unknown Location and Scale in a  $\chi^2$  Random Field, with an Application to fMRI.” *Advances in Applied Probability*, **33**(4), 773–793.
- Worsley KJ, Evans AC, Marrett S, Neelin P (1992). “A Three-dimensional Statistical Analysis for CBF Activation Studies in Human Brain.” *Journal of Cerebral Blood Flow & Metabolism*, **12**(6), 900–918.

**Affiliation:**

Khalil Shafie

Department of Applied Statistics and Research Methods  
University of Northern Colorado, Colorado, USA.

E-mail: [khalil.shafie@unco.edu](mailto:khalil.shafie@unco.edu)

Mohammad Reza Faridrohani

Department of Statistics, Faculty of Mathematical Sciences  
Shahid Beheshti University, Eveen, Tehran, Iran.

E-mail: [M\\_faridrohani@sbu.ac.ir](mailto:M_faridrohani@sbu.ac.ir)

Siamak Noorbaloochi

HSR & D Center for Care Delivery and Outcomes Research, VA Health Care System,  
Department of Medicine, University of Minnesota,  
Minneapolis, Minnesota, USA.

E-mail: [siamak@stat.umn.edu](mailto:siamak@stat.umn.edu)

Hossein Moradi Rekabdarkolae

Department of Mathematics and Statistics,  
South Dakota State University,  
Brookings, South Dakota, USA.

E-mail: [hossein.moradirekabdarkolae@sdstate.edu](mailto:hossein.moradirekabdarkolae@sdstate.edu)