

University of Northern Colorado

## Scholarship & Creative Works @ Digital UNC

---

Dissertations

Student Research

---

8-2021

### **BAYESIAN APPROACH TO THE MIXTURE OF GAUSSIAN RANDOM FIELDS AND ITS APPLICATION TO AN FUNCTIONAL MAGNETIC RESONANCE IMAGING STUDY**

Mozhdeh Forghaniarani

Follow this and additional works at: <https://digscholarship.unco.edu/dissertations>

---

©2021

Mozhdeh Forghaniarani

ALL RIGHTS RESERVED

UNIVERSITY OF NORTHERN COLORADO

Greeley, Colorado

The Graduate School

BAYESIAN APPROACH TO THE MIXTURE OF GAUSSIAN  
RANDOM FIELDS AND ITS APPLICATION TO AN  
FUNCTIONAL MAGNETIC RESONANCE  
IMAGING STUDY

A Dissertation Submitted in Partial Fulfillment  
of the Requirements for the Degree of  
Doctor of Philosophy

Mozhdeh Forghaniarani

College of Education and Behavioral Sciences  
School of Educational Research Leadership and Technology  
Department of Applied Statistics and Research Methods

August, 2021

This dissertation by: Mozhdeh Forghaniarani

Entitled: *Bayesian Approach to the Mixture of Gaussian Random Fields and its Application to an Functional Magnetic Resonance Imaging Study*

has been approved as meeting the requirement for the Degree of Doctor of Philosophy in College of Education and Behavioral Sciences in Department of Applied Statistics and Research Methods

Accepted by the Doctoral Committee

---

Khalil Shafie, Ph.D., Research Advisor

---

William Merchant, Ph.D., Committee Member

---

Chia-Lin Tsai, Ph.D., Committee Member

---

Nathaniel Eldredge, Ph.D., Faculty Representative

Date of Dissertation Defense \_\_\_\_\_

Accepted by the Graduate School

---

Jeri-Anne Lyons, Ph.D.  
Dean of the Graduate School  
Associate Vice President for Research

## ABSTRACT

Forghaniarani, Mozhdeh. *Bayesian Approach to the Mixture of Gaussian Random Fields and its Application to an fMRI Study*. Published Doctor of Philosophy dissertation, University of Northern Colorado, 2021.

Due to the functional nature of fMRI data, random field theory is used as a remedy to the multiple comparisons problem in brain signal detection. Traditionally, a Gaussian random field model is fitted to the functional data using this approach. However, fMRI data are not homogeneous, and there exist multiple underlying classes in functional data, so traditional inferential methods may fail. Here, we proposed a new model for signal detection in fMRI data in which we addressed the heterogeneity in such data. The proposed model is a mixture of two Gaussian random fields. We developed a Bayesian approach for hypothesis testing by using the notion of Bayes factor in infinite-dimensional parameter spaces. For such spaces, the Bayes factor is defined based on the concept of the Radon-Nikodym derivative. In our model, the Bayes factor is interpreted as the inverse of the expected value of a likelihood ratio with respect to the prior density of the model parameters. Obtaining the Bayes factor in infinite-dimensional parameter spaces is not analytically tractable, and we needed to compute it through numerical methods. Our methodology is empirically justified by Monte Carlo simulations and illustrated by an analysis of the simulated dataset.

To Farhang, Farnoosh, Jyoti and my Mom

## ACKNOWLEDGEMENTS

I would like to express my sincere gratitude and appreciation to my great advisor Dr. Khalil Shafie for his excellent guidance, encouragement, caring, and support. I have learned a lot from him during my academic journey.

I am also grateful to the rest of my supervisory committee, Dr. William Merchant, Dr. Chia-Lin Tsai, and Dr. Nathaniel Eldredge for their helpful suggestions, flexibility, and inspiration.

My sincere appreciation also goes to Dr. Babak Shahbaba, Dr. Hernando Ombao, and Dr. Siamak Noorbaloochi who have provided me with their support and encouragement specifically during the journey of the dissertation. Also, I would like to thank Dr. Ali Sharifi for his invaluable help in carrying out the computational part of this work.

I would also like to express my thanks to all those who have helped, guided, or encouraged me in any way during my years of graduate study in the department of Applied Statistics and Research Methods. Among them, I should specifically thank Keyleigh Gurney, the department's secretary.

Also, I would like to thank Carol Steward for her time and help with revising my thesis.

I also wish to express my thankfulness to my son, Farhang, my sister, Farnoosh, my brother, Jyoti, my nephew, Shayan, and my lovely nieces Neela and Ella. I am grateful for having them and their support.

The last but not least person that I would like to thank is my mother, Toloo. I should present my thankfulness to my mother for her support. I am sure that her prayers for me empowered me during this journey.

## TABLE OF CONTENTS

CHAPTER		Page
I	INTRODUCTION	1
	Statistical Analysis of Functional Magnetic Resonance Imaging Data .	2
	Problem Statement . . . . .	4
	Purpose of Study . . . . .	5
	Research Questions . . . . .	6
	Definition of Terms . . . . .	7
II	REVIEW OF LITERATURE	9
	Introduction . . . . .	10
	General Concept . . . . .	12
	Gaussian Random Fields . . . . .	22
	Radon-Nikodym Derivative and Hypothesis Testing . . . . .	24
	Signal Detection in the Gaussian Random Fields . . . . .	33

	Finite Mixture Models and Identifiability . . . . .	40
III	METHODOLOGY	44
	Gaussian Random Field Mixture Model and its Identifiability . . . . .	46
	Radon-Nikodym Derivative in Gaussian Mixture Models . . . . .	47
	Likelihood Approach for Gaussian Mixture Models . . . . .	49
	Bayes Factor for Mixture of Gaussian Random Fields . . . . .	51
	Estimating $\pi$ through Gibbs Sampling . . . . .	53
	Simulation Study . . . . .	56
IV	RESULTS	61
	Result for the Bayes Factor . . . . .	61
	Result for Parameters Estimation . . . . .	68
V	CONCLUSIONS	72
	Findings and Discussion . . . . .	73
	Limitations and Suggestions for Future Research . . . . .	74

REFERENCES 76

APPENDIX A R CODE FOR MONTE CARLO SIMULATION 79

## LIST OF TABLES

1	Bayes Factor Grading . . . . .	30
2	Schemes of the Parameters where $\sigma_{01} = 0.2$ , $\sigma_{02} = 0.4$ , . . . . .	60
3	The Bayes Factor and its power where $\sigma_{01} = 0.2$ , $\sigma_{02} = 0.4$ , $n = 10$ . . . . .	64
4	The Bayes Factor and its power where $\sigma_{01} = 0.2$ , $\sigma_{02} = 0.4$ , $n = 20$ . . . . .	65
5	Five-Number Summary of Bayes Factor For Schemes 1-4 . . . . .	67
6	Five-Number Summary of Bayes Factors For Schemes 5-8 . . . . .	67
7	Five-Number Summary of Bayes Factors For Schemes 9-12 . . . . .	69
8	Parameter estimates where $\sigma_{01} = 0.2$ , $\sigma_{02} = 0.4$ , $n = 10$ . . . . .	71

## LIST OF FIGURES

1	A sample path of spherical Gaussian random field . . . . .	34
2	A sample path of a smoothed spherical Gaussian random field . . . . .	36
3	A mixture of two bi-variate Normal densities . . . . .	41
4	Boxplots of Bayes Factor with respect to Parameter Scheme . . . . .	63
5	Empirical Bayes Factor Distribution for Scheme 1-4 . . . . .	66
6	Empirical Bayes Factor Distribution for Scheme 5-8 . . . . .	68
7	Empirical Bayes Factor Distribution for Scheme 9-12 . . . . .	69
8	Empirical Bayes Factor Distribution for Scheme 13-16 . . . . .	70
9	Empirical Bayes Factor Distribution for Scheme 17-20 . . . . .	70
10	Empirical Bayes Factor Distribution for Scheme 21-24 . . . . .	71

## CHAPTER I

### INTRODUCTION

Modern and advanced technologies of data collections, in recent years have produced very detailed and informative images, many extremely complex. Some of these advanced technologies of collecting data are positron emission tomography (PET), magnetic resonance imaging (MRI), functional magnetic resonance imaging (fMRI), and satellite imaging methods. The imaging technique used in a medical setting to produce images of the inside of the human body is known as Magnetic resonance imaging (*MRI*). *MRI* is based on the principles of nuclear magnetic resonance, a spectroscopic technique used by scientists to obtain microscopic chemical and physical information about molecules. An *MRI* scanner consists of a large and very strong magnet in which the patient lies. A radio wave antenna is used to send signals to the body and then receive signals back. These returning signals are converted into images by a computer attached to the scanner. Images of almost any part of the body can be obtained using *MRI* technique, although *MRI* Scanners are more suitable for looking at the non-bony parts or soft tissues of the body like the brain and nerves. These tissues are seen much more clearly with *MRI* than with regular x-rays and CAT scans. A disadvantage of *MRI* is its higher cost compared to a regular x-ray or CAT scan. Also, CAT scans are frequently superior for looking at the bones to *MRI*.

While *MRI* concentrates on the structure of the brain, functional *MRI* (*fMRI*) is a technique introduced to obtain functional information from the central nervous system. This technique extends anatomical imaging of *MRI* to include localization of the active brain areas during perceptions, actions, visual and cognitive tasks. Activation of an area of the brain causes an increase in blood flow in that area which is greater than that needed to keep up with the oxygen demands of the tissues. It results in a net increase in intravascular oxyhemoglobin and a decrease in deoxyhemoglobin. Deoxyhemoglobin is paramagnetic, resulting in a decrease to signal intensity coming from the tissues. Less deoxyhemoglobin as a consequence of an increase in blood flow results in an overall increase in signal, (Pauling & Coryell, 1936). Sophisticated image processing techniques are used to brain images of these flow changes. The *fMRI* technique offers opportunities for the investigation of the human brain's functional organization. Such techniques deal with the area of the body (for example brain) in which we see an increase in blood flow, or "activation", due to stimulation conditions. The main statistical problem in signal detection is to specify the regions of the brain in which the signal (activation) exists and separate them from the rest of the brain where no activation can be detected (the noise).

### **Statistical Analysis of Functional Magnetic Resonance Imaging Data**

The problem of signal detection in *fMRI* data is statistically equivalent to a problem of hypothesis testing. Regarding the nature of *fMRI* data, the multiple comparisons problem is the consequence of numerous statistical tests being done in each image. The most common method to deal with the multiple comparison problem is the Bonferroni correction method, (Bender & Lange, 2001). However, applying the

Bonferroni method to fMRI data would not be appropriate because first, fMRI data includes a large number of small elements, pixels/voxels, should be tested. Also, some degree of spatial autocorrelation exists in the data. Therefore, to solve the multiple comparison problem in the fMRI study, several other methods were proposed which “random field theory” is one of them, Bennett et al. (2009).

When applying the random field theory, statistical inferences are mostly done in abstract infinite-dimensional spaces. Because of the complexity of “abstract inference”, researchers have been confronted with challenging theoretical problems. Grenander (1981) and Adler (1981) are the pioneers of developing theoretical aspects of “abstract inference”. Fundamental parts of random field theory were developed by Adler in 1981, and it was later applied by Worsley to the fMRI study to solve the multiple comparison problem in testing the signal. Worsley et al. (1992) and Worsley (1994) have shown that the images of the brain can be modeled as a Gaussian random field  $X(\mathbf{t})$ , where  $\mathbf{t} \in \mathbb{R}^N$  is a location vector in the brain  $C \subset \mathbb{R}^N$ ,  $N = 3$ . To test the signal, they chose  $X_{\max}$  the global maxima of the random field in  $C$  as the test statistic for detecting signals in the brain. To reduce the signal-to-noise ratio, it is common to spatially smooth the images with a filter before analysis. The Matched Filter Theorem of signal processing states that a signal added to white noise is best detected by smoothing with a filter whose shape matches that of the signal  $f(\mathbf{t})$ . Siegmund and Worsley (1995) considered the situations that, after smoothing by the Gaussian kernel, observations can be decomposed into a fixed signal plus a random noise, where the noise is modeled as a particular stationary Gaussian random field in  $N$ -dimensional Euclidean space. The signal is assumed that has the form of a known function centered at an unknown location and multiplied by an unknown

amplitude. There are many examples where the signal scale or width is assumed to be known after smoothing. However, Siegmund and Worsley (1995) considered the case where the width of the signal is unknown. Therefore, their proposed Gaussian random field model was in an  $N + 1$ -dimensional space,  $N$ -dimension for the location  $\mathbf{t}$ , and one dimension for the width of the smoothing kernel,  $\sigma$ . Such a smooth random field,  $X(\mathbf{t}, \sigma)$  is known as the “Gaussian scale-space random field”. They studied the classical testing problem to detect the signal, and their test statistic was the maximum of the “Gaussian scale-space random field”. They used “the expected Euler characteristic of the excursion set” of the random field and “the volume of tubes” to derive an approximate distribution of  $X_{\max}$  under the null hypothesis of no signal. Most of the methods introduced in the references are based on classical likelihood ratio methods introduced by Grenander (1981) using the Radon-Nikodym derivative. Shafie and Noorbaloochi (2001) extended the definition of Bayes factor by using Radon-Nikodym derivative to abstract spaces, and introduced the Bayesian testing for signal detection in noisy images for the cases that the observations can be decomposed into an unknown signal and a random noise. They assumed that both signal and noise are the elements of a Hilbert space.

### **Problem Statement**

In practical situations, we have often an idea about the area that the signal can occur. Especially for the Gaussian scale-space random field, we may have a piece of historical information about the amplitude, location, and scale of the signal. Therefore it would be very useful if we use this information in the Bayesian viewpoint for signal detection.

The main work of this dissertation is based on the study of the Bayesian testing for signal detection in Gaussian scale-space random field in the cases that shape of the signal is known, and the scale, location and amplitude parameters have some known prior distribution.

Recalling that the study of random fields is the study of random functions, in modeling fMRI data, we are modeling random functions indexed by a location vector  $t$  in the brain  $C \subset \mathbb{R}^3$ . When modeling brain activity, it is common to model functional data of the brain by using one Gaussian random field. However, assuming that the brain region being homogeneous is not practical. Therefore, when the functional data are not homogeneous, and there exist multiple classes of functional data, modeling images as one random field like a Gaussian random field may fail in signal detection. In the next section, we explain our model that can be a solution to this issue.

### **Purpose of Study**

The motivation of this dissertation is to detecting signals embedded in the brain functional imaging data, so-called “signal detection”. Considering that the brain region is heterogeneous, there exist multiple classes of functional data, when studying fMRI data. In this work, we propose a new model so-called “mixture of Gaussian random fields”, to incorporate both functional and heterogeneous properties of the data. In the current study, the images are modeled as a mixture of two Gaussian random fields, however, most results can be extended to the class of finite mixtures of Gaussian random fields with more than two components.

## Research Questions

In this study, the averaged image obtained from fMRI technology is modeled as a mixture of two real-valued Gaussian scale space random fields  $X_1$  and  $X_2$ . The model can be formulated as

$$X = zX_1 + (1 - z)X_2 \quad (1)$$

where,

$$p(z) = \begin{cases} \pi & z = 1 \\ 1 - \pi & z = 0 \end{cases}$$

The objectives of this dissertation is to propose a Bayesian test statistic to test the existence of signal in the cases that shape of signals are known, and the amplitude, location, and scale parameters in two random fields,  $\boldsymbol{\theta}_1 = (\xi_1, \mathbf{t}_0, \sigma_{01})$  and  $\boldsymbol{\theta}_2 = (\xi_2, \mathbf{s}_0, \sigma_{02})$  have some known prior distributions. Testing the existence of signal is statistically equivalent to the problem of testing the following hypotheses:

$$\begin{cases} H_0 : \xi_1 = 0 \ \& \ \xi_2 = 0 \\ H_1 : \xi_1 > 0 \ \& \ \xi_2 > 0 \end{cases}$$

In this dissertation, we demonstrate the methodology to investigate the following research questions:

Q1 How to develop a Bayesian approach for testing the signal in the mixture model  $X(\mathbf{t}, \sigma)$  for two-dimensional images?

Q2 How the performance of the Bayes factor can change with respect to different parameter schemes?

### Definition of Terms

**Amplitude.** The amplitude of a wave refers to the maximum amount of displacement of a particle on the “medium” from its rest position. In a sense, the amplitude is the distance from rest to crest.

**Euler characteristic (EC).** A topological measure of the statistical parametric map after thresholding.

**Excursion set.** A set of points where a random field exceeds a fixed threshold value.

**Finite dimensional distributions.** A collection of distribution functions for a random field.

**Functional magnetic resonance imaging (fMRI).** A neuroimaging technique used to study brain functions.

**Gaussian kernel.** A smoothing kernel used to smooth data.

**Gaussian random field.** A type of random field where all of the finite dimensional distributions are multivariate joint Gaussian (normal) distributions.

**Pixel.** A measure of unit in a two-dimensional image.

**Random fields.** A collection of random variables defined over a subset of  $N$ -dimensional Euclidean space.

**Scale space.** A range of smoothing widths is used to create an extra scale dimension to the data.

**Smoothing kernel.** A function used to filter images.

**Spatial smoothing.** A procedure that replaces the BOLD signal in a voxel with the average of BOLD signals from neighboring voxels.

**Voxel.** A measure of unit in a three-dimensional image.

**Width of the smoothing kernel.** A parameter used to determine the amount of spatial smoothing applied to the image. In current study it was controlled by  $\sigma$ .

## CHAPTER II

### REVIEW OF LITERATURE

As it is claimed earlier, the motivation of this dissertation is to detecting signals embedded in the brain functional imaging data, so called “signal detection”. Because in the current study, the images are modeled as a finite mixture of Gaussian random fields, in this chapter, we introduce the main approaches in signal detection which are based on the random field theory. Before we begin our own study of random fields in the next chapter, we should take time for a brief review of the theories related to our study on random fields. In this chapter, we will study general properties of random fields. Also, we introduce the class of Gaussian random fields and its properties. The focus of this chapter is on the Bayesian approach in the signal detection which proposed by Shafie and Noorbaloochi (2001), and extended by Rohani et al. (2006). Their approach and methodologies are foundation of this dissertation. Furthermore, in this chapter, we are going to have an introduction about mixture models in general, and the Gaussian mixture model as a specific case.

The format of this chapter is as follows. In section 1, we give an introduction about the main approaches based on the random field theory in signal detection. In section 2, we review some of the general concepts and definitions in the mathematical foundations of random fields. In section 3, Gaussian random fields as one of the most important random fields are introduced and some of their properties are briefly reviewed.

In section 4, we explain about the notion of Radon-Nikodym derivative and its interpretation in hypothesis testing. In section 5, we review the likelihood and Bayesian approaches for signal detection when the image is modeled as a Gaussian random field. In section 6, we introduce mixture models, and Gaussian mixture models. Also, we review the problem related to identifiability in mixture models.

### **Introduction**

Random field theory as a sophisticated mathematical work was developed by Adler in 1981, and later the theory was applied extensively by Keith Worsley in the study of brain functional imaging to rectify the multiple comparisons problem. Worsley et al. (1992), and Worsley (1994) have shown that the images of the brain can be modeled as a Gaussian random field  $X(\mathbf{t})$ , where  $\mathbf{t} \in \mathbb{R}^N$  is a location vector in the brain  $C \subset \mathbb{R}^N$ ,  $N = 3$ . Siegmund and Worsley (1995) assumed that images can be decomposed into a deterministic signal and a homogeneous  $N$ -dimensional Gaussian random field so called as random noise. They considered the situations that the signal has the form of a known function, i.e a Gaussian form, centered at an unknown location  $\mathbf{t}_0$ , and multiplied by an unknown amplitude  $\xi$ , with an unknown parameter  $\sigma_0$  as the width of signal. Usually before analysis, the images may be spatially smoothed to enhance the signal to noise ratio. By the Matched Filter Theorem in the signal processing, a signal which is added to a white noise is best detected by smoothing with a filter whose shape matches that of the signal,  $f(\mathbf{t})$ . Siegmund and Worsley 1995 considered a Gaussian kernel to smooth the image which was modeled as a Gaussian random field. In their model, the scale of Gaussian kernel,  $\sigma$  is unknown, and they considered it as an extra dimension for

the space of the random field. The smooth random field in their model,  $X(\mathbf{t}, \sigma)$  is known as a Gaussian scale space random field.

In this work, we employ Gaussian scale space random field models, and we consider the unknown  $\sigma$  as a new index-parameter for the random functions in a Hilbert space.

To detect the signal in Gaussian scale space random field, Siegmund and Worsley (1995) studied the classical testing. Their test statistic was the maximum of the Gaussian random field in a  $(N + 1)$ -dimensional “scale space”,  $N$ -dimensions for the location and one dimension for the width or scale of the smoothing kernel. They used the expected Euler characteristic of the excursion set of the random field to derive an approximate distribution of  $X_{\max}$  under the null hypothesis of no signal. Shafie and Abravesh (2016) showed that for Gaussian scale space,  $X_{\max}$  is equivalent to the likelihood ratio test statistic.

In practical situations, specially for Gaussian random fields, we may have a historical information about the amplitude, location and scale of the signal. This information can be used in the Bayesian viewpoint for signal detection. Another preference of Bayesian approach in signal detection is the ability to apply different smoothing functions to the sample path. Shafie and Noorbaloochi (2001) introduced the Bayesian testing for signal detection in noisy images. They extended the definition of Bayes factor to abstract spaces by using Radon-Nikodym derivative. They assumed that observations can be decomposed into an unknown signal plus a random noise but they considered the cases that both signal and noise are the elements of a Hilbert space. Rohani et al. (2006) studied the Bayesian testing for signal detection in Gaussian scale

space random field in the cases that shape of the signal is known, and the scale, location and amplitude parameters have some known prior distributions. They considered both cases of smooth and non-smooth scale space Gaussian random fields. They studied the generalization of Bayes factor for Gaussian scale space random fields by using Radon-Nikodym derivative. Their approach is the foundation of this study, therefore, in this chapter, we will explain about their methodology in Bayesian testing more in details.

### General Concept

In this section, we will have a brief excursion through the mathematical foundations of probability and random fields which will be used in this thesis.

**Definition II.1.** A *vector* or *linear space* is a set  $L$  along with two operations defined on its elements, addition and scalar multiplication, under which  $L$  is algebraically closed. That is, for any  $x, y \in L$ , and  $\alpha \in R$ , the sum  $x + y$  and the scalar product  $\alpha x$  are also elements of  $L$ . Furthermore, the operations conform to distributive laws, i.e.

$$\alpha(x + y) = \alpha x + \alpha y.$$

These are the familiar properties of N-dimensional Euclidean space,  $R^N$ . In dealing with random fields, however, we encounter *function spaces* which is vector spaces whose elements are functions.

**Definition II.2.** (Aubin, 1977) A norm on a linear space  $L$  is a non-negative real-valued function satisfying

i.  $\|x\| > 0$  for  $x \neq 0 \quad \forall x \in L$

ii.  $\|x\| = 0 \iff x = 0$

$$\text{iii. } \|\alpha x\| = |\alpha| \|x\| \quad \forall \alpha \in R, \quad \forall x \in L$$

$$\text{iv. } \|x + y\| \leq \|x\| + \|y\|, \quad \forall x, y \in L.$$

**Definition II.3.** A linear space  $L$  equipped with a norm  $\|\cdot\|$ , is called a *normed linear space*, and denoted by  $(L, \|\cdot\|)$ .

In general there are many possible norms that can be defined on a given vector space. For  $R^n$ , the common norm is the Euclidean norm

$$\|x\|_2 \equiv \left[ \sum_{i=1}^n x_i^2 \right]^{1/2}$$

Given a norm, we can measure the *distance* between two vectors by defining a function known as a *metric*,  $d(x, y) = \|x - y\|$ .

A norm also carries with itself the abstract notion of distance, and so a *normed linear space* is also a *metric space*. In the context of function spaces, the only norm we will need to consider is the so called *p-norm*, defined by

$$\|f\|_p \equiv \left[ \int |f|^p d\mu \right]^{1/p} \quad (1)$$

**Definition II.4.** A linear space is said to be *complete* if and only if each Cauchy sequence (that is, each sequence such that  $\|x_n - x_m\| \rightarrow 0$  as  $n, m \rightarrow \infty$ ) converges to a point in the space. In effect, completeness guarantees that there is always an  $x$  in the space that the sequence converges to it.

**Definition II.5.** An *inner product* on a (real) linear space  $H$  is a real-valued function defined on  $H \times H$ , denoted  $\langle \cdot | \cdot \rangle$ , with the following properties:

$$\text{i. } \langle x, x \rangle \geq 0 \quad \forall x \in H$$

$$\text{ii. } \langle x, x \rangle = 0 \iff x =$$

$$\text{iii. } \langle x, y \rangle = \langle y, x \rangle \quad \forall x, y \in H$$

$$\text{iv. } \langle \alpha x + \beta y, z \rangle = \alpha \langle x, z \rangle + \beta \langle y, z \rangle. \quad \forall \alpha, \beta \in \mathbb{R}, \quad \forall x, y, z \in H$$

**Definition II.6.** A linear space with an inner product is called an *inner product space*.

So a space with an inner product is automatically a normed space because when an inner product is given we can always define a norm by

$$\|x\| \equiv \sqrt{\langle x, x \rangle} \tag{2}$$

**Definition II.7.** A complete normed linear space  $(L, \|\cdot\|)$  whose norm comes from an inner product is called a *Hilbert space*.

Euclidian N-space is a Hilbert space with its inner product being to the standard dot product.

**Definition II.8.** We say that two vectors  $x$  and  $y$  in a Hilbert space are *orthogonal*, and write  $x \perp y$ , if

$$\langle x, y \rangle = 0.$$

Similarly, a vector  $x$  is said to be orthogonal to a set  $S \subset L$ , denoted  $x \perp S$ , if  $x \perp y$  for all  $y \in S$ .

**Definition II.9.** A set  $B$  in Hilbert space  $H$  is called an *orthogonal system* or *orthonormal* if:

- i.  $\|x\| = 1, \quad \forall x \in B,$
- ii.  $x \perp y, \quad \forall x, y \in B \text{ such that } x \neq y.$

In addition,  $B$  is said to be complete if the only  $x \in H$  such that  $x \perp B$  is  $x = 0$ .

**Theorem II.1.** (Ash, 1972) Let  $B = \{x_\alpha, \alpha \in I\}$  be an orthonormal subset of the Hilbert space  $H$ . The following conditions are equivalent:

- i.  $B$  is an orthonormal basis.
- ii.  $B$  is a complete orthonormal set.
- iii.  $B$  spans  $H$ , i.e. it is the smallest closed subspace of  $H$  containing all elements of  $B$ .
- iv. For all  $x \in H, x = \sum_\alpha \langle x, x_\alpha \rangle x_\alpha.$
- v. For all  $x, y \in H, \langle x, y \rangle = \sum_\alpha \langle x, x_\alpha \rangle \langle x_\alpha, y \rangle.$
- vi. For all  $x \in H, \|x\|^2 = \sum_\alpha |\langle x, x_\alpha \rangle|^2.$

One of the most important properties of Hilbert space is that its elements can be approximated by projecting them onto some convenient subspace, typically one of finite dimension.

**Theorem II.2.** (*Orthogonal projection Theorem*) Let  $M$  be a closed linear subspace of an Hilbert space  $H$ . Then there exists a unique mapping  $\pi^M$  of  $H$  to  $M$  such that for every  $x \in H$ ,

$$x - \pi^M(x) \perp y, \quad y \in M.$$

**Theorem II.3.** If  $H$  is a separable Hilbert space, then every complete orthogonal systems is countable. If  $\{e_n\}$  is such a system, we have for all  $x \in H$

$$x = \sum_{n=0}^{\infty} \langle x, e_n \rangle e_n$$

and

$$\|x\|^2 = \sum_{n=0}^{\infty} \langle x, e_n \rangle^2.$$

Here, we want to introduce measure spaces, therefore, we review the required definitions.

**Definition II.10.** (Williams, 1991) Let  $\mathcal{X}$  be a space.  $\mathcal{A}$ , a collection of subsets of  $\mathcal{X}$  is an *algebra* if:

- i.  $\mathcal{X} \in \mathcal{A}$
- ii.  $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$
- iii.  $A, B \in \mathcal{A} \Rightarrow A \cup B \in \mathcal{A}$

**Definition II.11.** (Williams, 1991)  $\mathcal{A}$  is a  $\sigma$ -*algebra* if it is an algebra and for  $A_n \in \mathcal{A}, n \in \mathbb{N}$ , we have  $\cup A_n \in \mathcal{A}$ .

**Definition II.12.** A space  $\mathcal{X}$  and a  $\sigma$ -algebra  $\mathcal{A}$  on  $\mathcal{X}$  is called a *measurable space*  $(\mathcal{X}, \mathcal{A})$ .

**Definition II.13.** (Williams, 1991) Let  $(\mathcal{X}, \mathcal{A})$  be a measurable space. A map  $\mu : \mathcal{A} \rightarrow [0, \infty]$  is a *measure* if it is countably additive, meaning if  $A_i \cap A_j = \phi$  for  $\{A_n : n \in \mathbb{N}\} \subset \mathcal{A}$ , then

$$\mu(\cup_n A_n) = \sum_n \mu(A_n)$$

A measure  $\mu$  assigns positive numbers to sets  $A : \mu(A) \in \mathbb{R}$ . For example where  $A$  is a subset of Euclidean space,  $\mu(A)$  can be length, area, or volume. Or if  $A$  is an event,  $\mu(A)$  can be the probability of the event.

**Definition II.14.** (Williams, 1991) The triple  $(\mathcal{X}, \mathcal{A}, \mu)$  is called a *measure space*.

**Definition II.15.** (Williams, 1991) A measure space  $(\Omega, \mathcal{A}, \mathcal{P})$  is a probability space if  $\mathcal{P}(\Omega) = 1$ . In this case,  $\mathcal{P}$  is called a *probability measure*.

**Definition II.16.** A *random variable*  $X(\omega)$  is a real-valued measurable function in a probability space.

If  $(\Omega, \mathcal{F}, \mu)$  is a measure space, and  $f$  is a real valued Borel measurable function on this space, such that for given  $p > 0$ ,  $\int_{\Omega} |f|^p d\mu < \infty$ . Then the collection of all such functions, denoted by  $L^p(\Omega, \mathcal{F}, \mu)$ , is a function space (Ash (1972)).

Given a probability space  $(\Omega, \mathcal{F}, \mu)$ , the measure space  $\mathbb{L}^2(\Omega, \mathcal{F}, \mu)$  is a Hilbert subspace of measurable functions on  $(\Omega, \mathcal{F}, \mu)$  with respect to the inner product below:

$$\langle f, g \rangle \equiv \int_{\Omega} fg d\mu \tag{3}$$

Now suppose  $(\Omega, \mathcal{F}, P)$  is a probability space,  $T$  is an  $N$ -dimensional Euclidian space and  $(T, \mathcal{B}, \lambda_N^*)$  is the Lebesgue measure space. So the space of squared integrable real functions, denoted by  $L^2(T, \mathcal{B}, \lambda_N^*)$  is a Hilbert space. Therefore any measurable random field with finite second moment, can be considered as a Hilbert valued random object from  $(\Omega, \mathcal{F}, P)$  to  $L^2(T, \mathcal{B}, \lambda_N^*)$ .

Another definition that we need is the definition of a Gaussian measure. To introduce that, we need to review some more definitions as follow.

**Definition II.17.** Let  $m$  be a measure on  $(\mathcal{H}, \mathcal{B})$ . Its *characteristic functional*  $\phi(y)$  is defined for any  $y \in \mathcal{H}$  by

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

If  $m$  is a probability measure on  $\mathcal{H}$  then  $\phi$  is a positive definite continuous function on  $\mathcal{H}$  and  $|\phi(y)| \leq 1$ .

**Definition II.18.** If  $m$  is a measure on  $\mathcal{H}$  then its mean  $\mu$  is an element of  $\mathcal{H}$  defined by

$$\langle \mu, y \rangle = \int_{\mathcal{H}} \langle x, y \rangle dm(x)$$

for all  $y \in \mathcal{H}$ .

If  $\int_{\mathcal{H}} \|x\| dm < \infty$ , then  $\mu$  exists and

$$\|\mu\| \leq \int_{\mathcal{H}} \|x\| dm(x).$$

**Definition II.19.** If  $m$  is a measure on  $\mathcal{H}$  then its *covariance operator* is a bilinear functional on  $\mathcal{H}$ , defined by

$$\langle Sy_1, y_2 \rangle = \int_{\mathcal{H}} \langle y_1, x \rangle \cdot \langle y_2, x \rangle dm(x)$$

We should note that the covariance is a symmetric positive definite bilinear functional.

**Definition II.20.** A *Gaussian measure*  $m$  on  $\mathcal{H}$  is one such that for every vector  $y \in \mathcal{H}$  the distribution of  $\langle x, y \rangle$  is a one-dimensional Gaussian distribution. Equivalently  $m$  is a *Gaussian measure* iff its characteristic functional has the form

$$\phi(y) = \exp \left[ i \langle \mu, y \rangle - \frac{\langle Sy, y \rangle}{2} \right]$$

where  $\mu$  is its mean and  $S$  is its covariance.

In this work, the Gaussian measure  $m$  with mean  $\mu$  and covariance  $S$  on a Hilbert space is denoted by  $HN(\mu, S)$ .

As we will see in the next chapters, in practical situations we need the eigenvalues and eigenfunctions of the covariance function. Thus we introduce an approach to compute these values. Now let  $T$  be a compact interval in  $\mathcal{R}^N$  and suppose that on  $T \times T$  we have a continuous, real-valued, non-negative definite (covariance) function  $R(\mathbf{s}, \mathbf{t})$ . Consider the integral equation

$$\int_T R(\mathbf{s}, \mathbf{t}) \phi(\mathbf{t}) d\mathbf{t} = \lambda \phi(\mathbf{s}), \quad \text{for } \mathbf{s} \in T \quad (4)$$

**Definition II.21.** A nonzero number  $\lambda$  for which there exists a function  $\phi$  satisfying (4) such that  $\int_T |\phi(\mathbf{t})|^2 d\mathbf{t} < \infty$  is called *eigenvalue* of the integral equation. The corresponding function  $\phi$  is called an eigenfunction.

In general the integral equation (4) will yield an infinite number of eigenvalues  $\lambda_1, \lambda_2, \dots$  with corresponding eigenfunctions  $\phi_1, \phi_2, \dots$ . One can assume that the sequence of eigenvalues is non-increasing and the eigenfunctions are orthonormal, i.e.

$$\int_T \phi_i(\mathbf{t})\phi_j(\mathbf{t}) d\mathbf{t} = \begin{cases} 1 & \text{for } i = j \\ 0 & \text{for } i \neq j \end{cases}. \quad (5)$$

So these eigenfunctions form an orthonormal basis for  $L^2(T, \mathcal{B}, \lambda_N^*)$ .

### General Concepts in Random Fields

The study of random fields is the same as the study of random functions. A *random field* is in fact a random function which is defined over some Euclidean spaces. As an example of a random field, consider an ocean surface that is parameterized by a hypothetical zero plane passing through the surface of ocean by the point  $(t_1, t_2)$ . Let the function  $X(t_1, t_2, t_3)$  denote the height of ocean above the point  $(t_1, t_2)$  at time  $t_3$ . Assuming  $X$  being random in some sense is reasonable, and so we can model  $X$  as a three-dimensional random field. If we hold time  $t_3$  fixed, we obtain a two-dimensional field; (See Adler, 2010).

Adler (2010) gave two different approaches to defining a random field. We chose the following definition that satisfies this study; this definition of a random field is more natural in a modeling context.

**Definition II.22.** Let  $(\Omega, \mathcal{F}, P)$  be a probability measure. A collection of random variables  $\{X(\mathbf{t}), \mathbf{t} \in \mathbb{R}^N\}$  equipped with a collection of distribution functions  $F_{\mathbf{t}_1, \dots, \mathbf{t}_n}$  on  $\mathcal{B}^n$ ,  $n = 1, 2, \dots$ , and  $\mathbf{t}_j$ 's  $\in \mathbb{R}^N$  is called a *random field* if for an arbitrary finite set of  $\mathbf{t}_1, \dots, \mathbf{t}_n$ , and a collection of measures  $F_{\mathbf{t}_1, \dots, \mathbf{t}_n}$  on  $\mathcal{B}^n$

$$F_{\mathbf{t}_1, \dots, \mathbf{t}_n}(B) = P[(X(\mathbf{t}_1), \dots, X(\mathbf{t}_n)) \in B], \quad \forall B \in \mathcal{B}^n \quad (6)$$

We should recall that  $\mathcal{B}$  denotes the Borel  $\sigma$ -algebra generated by the half-open intervals in  $\mathcal{R}$ ,  $I = (a_1, b_1]$ .

The collection of all such measures or equivalently the corresponding distribution functions is known as the family of finite dimensional (**fi-di**) distributions of the random field  $X$ , and in general it is the fi-di distributions that we work with in the study of a random field.

For a random field  $X(\mathbf{t})$ , the mean and covariance functions can be defined respectively as,

$$m(\mathbf{t}) = E[X(\mathbf{t})]$$

$$R(\mathbf{s}, \mathbf{t}) = E[(X(\mathbf{s}) - m(\mathbf{s}))(X(\mathbf{t}) - m(\mathbf{t}))]$$

In the random variables case, two variables  $X$  and  $Y$  are equivalent if  $P\{X = Y\} = 1$ . This implies that, for all purposes, these variables are indistinguishable. Two random fields  $X(t)$  and  $Y(t)$  are said to be equivalent fields if

$$P\{X(t) = Y(t)\} = 1 \quad \text{for every } t \in \mathcal{R}^N$$

Two equivalent processes generate equivalent measures but are not necessarily indistinguishable in every sense. But when we are dealing with separable fields only, equivalent fields are indistinguishable.

As in the stochastic processes, in the study of random fields there is a powerful property known as homogeneity or stationarity.

**Definition II.23.** (Adler, 2010) A real-valued random field  $X(\mathbf{t})$  is said to be *strictly homogeneous* or *stationary* if, for each arbitrary  $k$ , and any real number  $x_1, \dots, x_k$  and any  $(k+1)$  points  $\tau, \mathbf{t}_1, \dots, \mathbf{t}_k$  in  $\mathbb{R}^N$  the following condition on its fi-di distributions is satisfied:

$$F_{\mathbf{t}_1, \dots, \mathbf{t}_k}(x_1, \dots, x_k) = F_{\mathbf{t}_1 + \tau, \dots, \mathbf{t}_k + \tau}(x_1, \dots, x_k).$$

**Definition II.24.** (Adler, 2010) A real-valued random field  $X(\mathbf{t})$  is said to be *weakly homogeneous* if for any points  $\mathbf{t}, \mathbf{s}$  in  $\mathbb{R}^N$ , the covariance function exists and the following two conditions are satisfied:

$$E[X(\mathbf{t})] = \text{constant},$$

$$R(\mathbf{s}, \mathbf{t}) = \text{Cov}[X(\mathbf{s}), X(\mathbf{t})] = R(\mathbf{s} - \mathbf{t})$$

**Definition II.25.** (Adler, 2010) A weakly homogeneous random field is called an *isotropic* if the covariance function  $R(\mathbf{s}, \mathbf{t})$  depends only on the distance between  $\mathbf{t}$  and  $\mathbf{s}$ , ( $\|\mathbf{t} - \mathbf{s}\|$ ).

### Gaussian Random Fields

Same as the case of Gaussian random variables, Gaussian random fields have some interesting properties that make them as the most important random fields. A

random field whose all fi-di distributions are multivariate Gaussian distributions is known as Gaussian random fields. A Gaussian random field  $X(\mathbf{t})$  is completely determined once we specify its mean and Covariance functions. It can be easily seen from the form of multivariate normal density that if a real-valued Gaussian field has a constant mean and a covariance function that is dependent only on  $\mathbf{s} - \mathbf{t}$ , then the field is homogeneous.

Now suppose  $(\Omega, \mathcal{F}, P)$  is a probability space,  $T$  an  $N$ -dimensional Euclidian space and  $(T, \mathcal{B}, \lambda_N)$  is the Lebesgue measure space. Also it is known that the space of squared integrable real functions with respect to  $\lambda_N$  on  $T$  is a Hilbert space, which is denoted by  $L^2(T, \mathcal{B}, \lambda_N)$ . Therefore any measurable random field, with finite second moment, can be considered as a Hilbert valued random object from  $(\Omega, \mathcal{F}, P)$  to  $L^2(T, \mathcal{B}, \lambda_N)$ . Now, let  $X(\mathbf{t})$  be a Gaussian random field on  $(\Omega, \mathcal{F}, P)$  with mean  $m(\mathbf{t})$  and covariance function  $R(\mathbf{s}, \mathbf{t})$ , then the probability measure induced by  $X$  on  $L^2(T, \mathcal{B}, \lambda_N)$  is a Gaussian measure on this Hilbert space, whose mean is  $m(\mathbf{t})$  and the covariance operator is obtained as

$$S(f(\mathbf{t})) = \int_T R(\mathbf{s}, \mathbf{t})f(\mathbf{s})d\lambda_N(\mathbf{s}) \quad (7)$$

This Gaussian measure on the Hilbert space  $L^2(T, \mathcal{B}, \lambda_N)$  is a Hilbert normal denoted by  $HN(\mu, S)$ . For more details, (See Rohani, 2003). Several important properties of Gaussian random fields which we will apply in this study are,

- (i) Every linear combination of Gaussian random fields is a Gaussian random field.

Also, derivatives of any differentiable Gaussian random field are also Gaussian random fields.

(ii) Let  $dW(\mathbf{t})$  be a Gaussian white noise, and  $k(\mathbf{t}, \mathbf{s})$  be a deterministic function, then

$$X(\mathbf{t}) = \int_{\mathbb{R}^N} k(\mathbf{t}, \mathbf{h}) dW(\mathbf{h}) \quad (8)$$

will be a Gaussian random field. The function  $k(\mathbf{t}, \mathbf{h})$  is called the kernel function.

This kind of random field has many applications in image processing (Worsley et al., 1992; Worsley, 1994; Siegmund & Worsley, 1995). Note that by an appropriate selection of the kernel function, one can smooth a Gaussian white noise.

(iii) Let  $dW(\mathbf{t})$  be a Gaussian white noise, and

$$X(\mathbf{t}) = \int_{\mathbb{R}^N} k(\mathbf{t}, \mathbf{h}) dW(\mathbf{h})$$

The mean and covariance function of  $X(\mathbf{t})$  are

$$m(\mathbf{t}) = 0 \quad , \quad R(\mathbf{t}, \mathbf{s}) = \int_{\mathbb{R}^N} k(\mathbf{t}, \mathbf{h}) k(\mathbf{s}, \mathbf{h}) d\mathbf{h}$$

### **Radon-Nikodym Derivative and Hypothesis Testing**

Radon-Nikodym derivative is one of the fundamental concepts of the methodology in this thesis. We begin this section by reviewing Radon-Nikodym theorem, and its generalization. Then, we discuss the relationship of the classical and Bayesian test statistic with Radon-Nikodym(R-N) derivative.

**Theorem II.4.** (Radon-Nikodym Theorem) Let  $(\mathcal{X}, \mathcal{A}, \mu)$  be a  $\sigma$ -finite measure space, and suppose  $\nu \ll \mu$ . Then there exists  $\mu$ -measurable function  $f$ , such that

$$\nu(A) = \int_A f d\mu \quad \forall A \in \mathcal{A}$$

The function  $f = \frac{d\nu}{d\mu}$  is the Radon-Nikodym derivative of  $\nu$  w.r.t.  $\mu$ .

We should note that  $\nu \ll \mu \Leftrightarrow \nu$  has a density w.r.t.  $\mu$ .

The following corollary is a generalization of Radon-Nikodym Theorem.

**Corollary II.1.** Let  $P_0$  and  $P_1$  be probability measures on  $(\mathcal{X}, \mathcal{F}_1)$ . There exists a set  $A_0 \in \mathcal{F}_1$  with  $P_0(A_0) = 0$ , and a non-negative  $P_0$ -measurable function  $f$ , such that for every  $A \in \mathcal{F}_1$

$$P_1(A) = \int_A f(x) dP_0(x) + P_1(A \cap A_0) \quad (9)$$

The function  $f$  is called the Radon-Nikodym (R-N) derivative of  $P_1$  with respect to  $P_0$ , and denoted by

$$f(x) = \frac{dP_1}{dP_0}(x) \quad (10)$$

and so in general the Radon-Nikodym derivative can be defined as

$$\left\{ \begin{array}{ll} +\infty & \text{for } x \in A_0 \\ f(x) & \text{for } x \notin A_0 \end{array} \right. .$$

If the set  $A_0$  has  $P_1(A_0) = 1$ , two measures  $P_0$  and  $P_1$  is said to be perpendicular to each other, and in this case  $A_0$  plays the role of a critical region and will allow us, without any errors, to discriminate between  $H_0$  and  $H_1$ . If  $P_1$  and  $P_0$  are equivalent then  $P_1(A_0) = 0$  and the second part of r.h.s. in (9) is zero, and  $f \neq 0$  with  $P_0$ -probability 1. Intermediate cases occur when  $P_0$  and  $P_1$  neither perpendicular nor equivalent, but surprisingly seldom for cases of practical interest. Now, first, we discuss the role of R-N derivative in testing hypothesis using the likelihood approach, and then we introduce the Bayesian testing procedure by applying R-N derivative.

### **Likelihood Approach**

In this part, we briefly discuss the likelihood approach for testing hypothesis in a random field. At first the simple null hypothesis ( $H_0$ ) versus the simple alternative hypothesis ( $H_1$ ) is considered.

#### ***Simple Alternative Hypothesis***

Suppose  $\mathcal{X}$  is a sample space corresponding to random object  $X$ ,  $\mathcal{F}_1$  a  $\sigma$ -field of subsets of  $\mathcal{X}$ , and  $\Theta = \{\theta_0, \theta_1\}$  is the parameter space. In the standard Neyman-Pearson approach for testing  $H_0$  against  $H_1$ , we deal with the subset  $W$  of the sample space (the critical region), and the decision rule that if the observed sample  $x \in W$ , we reject  $H_0$ ; otherwise we accept it. Of course, and because of the existence of errors of type  $I$ ,  $\alpha$ , and type  $II$ ,  $\beta$ , we are interested in decision rules that balance these errors. The famous Neyman-Pearson lemma tells us how to minimize the probability of second error for fixed error of type  $I$ . The Neyman-Pearson lemma is based on the likelihood ratio. The Radon-Nikodym derivative plays the role of likelihood ratio in an abstract space. So the

basic problem is to calculate the Radon-Nikodym derivative and analyze its properties.

By the corollary II.1, a Radon-Nikodym derivative can be interpreted as of the likelihood ratio.

The following lemma gives the critical region for testing simple hypothesis

$H_0 : \theta = \theta_0$ , against  $H_1 : \theta = \theta_1$ .

**Lemma II.1.** (Neyman-Pearson lemma) Let  $f(x)$  be Radon-Nikodym derivative of  $P_1$  with respect to  $P_0$ . Also let for a given  $c \geq 0$  the critical region  $W$  be of the form

$$W = \{x \mid f(x) \geq c\} \subset \mathcal{X} \quad (11)$$

Then no other critical region at the same level of significance has greater power, i.e. if  $W'$  is such that  $P_0(W') \leq P_0(W)$ , then  $P_1(W') \leq P_1(W)$ .

This lemma introduces the methodology of finding the best critical region  $W$  at the given level of significance.

### ***Composite Alternative Hypothesis***

Now we consider testing the simple null hypothesis  $H_0 : \theta = \theta_0$ , against the composite alternative hypothesis  $H_1 : \theta \neq \theta_0$ . Suppose  $\mathcal{X}$  is the sample space, and  $\Theta$  is the parameter space. Again we suppose for every  $\theta \in \Theta$ ,  $P_\theta$  is a probability measure on  $(\mathcal{X}, \mathcal{F}_1)$ . Therefore following generalized likelihood ratio test statistics, we can define a test statistic as:

$$f(x) = \sup_{\theta^* \in \Theta} f_{\theta^*}(x) = \sup_{\theta^* \in \Theta} \frac{dP_{\theta^*}}{dP_{\theta_0}}(x) \quad (12)$$

where  $\frac{dP_{\theta^*}}{dP_{\theta_0}}(x)$  is the Radon-Nikodym derivative of  $P_{\theta^*}$  with respect to  $P_{\theta_0}$ . However as in the classical case,  $f(x)$  itself has not necessarily the Radon-Nikodym derivative interpretation. Therefore the Neyman-Pearson lemma for abstract space in testing the simple null hypothesis versus the simple alternative is not extended to the simple against composite alternative hypothesis testing. But we can define the critical region  $W$  same as the simple hypothesis case, of the form

$$W = \{x \mid \sup_{\theta^* \in \Theta} f_{\theta^*}(x) \geq c\}.$$

### **Bayesian Approach**

Here, we review the Bayesian testing of simple hypothesis versus a simple or composite alternative. For this, suppose again  $\mathcal{X}$  is the sample space corresponding to a random object  $X$ ,  $\mathcal{F}_1$  is a  $\sigma$ -field of subsets of  $\mathcal{X}$ .

#### ***Simple Alternative Hypothesis***

Let  $\Theta = \{\theta_0, \theta_1\}$ , and  $(\Theta, \mathcal{F}_2)$  be a parameter space. For testing simple hypothesis  $H_0 : \theta = 0$ , against  $H_1 : \theta = 1$ , in a Bayesian approach, Radon-Nikodym derivative  $f(x)$  in (3) can be applied as a test statistic. Here,  $\pi_0$  and  $\pi_1$  are denoted as the two prior probabilities of  $\theta = 0$  and  $\theta = 1$ , and  $P_0$  and  $P_1$  as a probability measures on  $(\mathcal{X}, \mathcal{F}_1)$ , respectively, under the null and alternative hypotheses. Rohani et al. (2006) by the following theorem extended the classical construction of the best Bayesian test.

**Theorem II.5.** If  $P_1$  is absolutely continuous with respect to  $P_0$  and if the prior probabilities of the two hypotheses  $H_0$  and  $H_1$  are  $\pi_0$  and  $\pi_1$ , respectively, the best test is

given by the critical region

$$W = \{x | f(x) > \frac{\pi_0}{\pi_1}\}$$

### ***Composite Alternative Hypothesis***

Now we consider testing the simple null hypothesis  $H_0 : \theta = \theta_0$  against the composite hypothesis  $H_1 : \theta \neq \theta_0$ . Suppose  $X$  is an observable quantity with density  $f(x | \theta)$ , where  $\theta$ , the parameter of interest, and  $X$  both are elements of the Euclidian space  $\mathbb{R}^N$ .

There are two Bayesian criteria for testing  $H_0$ , the Bayes factor and the Bayesian posterior probability of  $H_0$ , Berger (2013).

### ***Bayes factor***

The Bayes factor for assessing a null hypothesis  $H_0$  against an alternative  $H_1$  generally is defined as

$$B(x) = \frac{P(H_0 | x)}{P(H_1 | x)} \div \frac{P(H_0)}{P(H_1)}$$

Let  $G(\theta)$  be a continuous prior distribution of  $\theta$  over its parameter space,  $\Theta$ . It is easy to see that for the above sharp null testing problem, the Bayes factor is equivalent to

$$\mathbf{B} = \frac{f(x|\theta_0)}{m_G(x)} \tag{13}$$

where

$$m_G(x) = \int f(x|\theta)dG(\theta) \tag{14}$$

The Bayes factor incorporates prior information with the sample information in the likelihood. It can be considered as a weighted likelihood ratio of  $H_0$  to  $H_1$ , where the weight function for a Bayesian is the prior distribution  $G(\theta)$ . So it can be used to evaluate evidence against a null hypothesis, and its interpretation is the same as that of the usual likelihood ratio. For example,  $B = 0.2$  means that the observations support for  $H_1$  is five times of that of  $H_0$ , Rohani et al. (2006). In other words, if  $B < 1$ , the null hypothesis  $H_0$  is not supported. Similarly, if  $B > 1$ ,  $H_0$  will be supported. And if  $B = 1$ ,  $H_0$  and  $H_1$  are not preferred over each other.

From Jeffreys (1998), the common grading of the Bayes factor is given in Table 1.

**Table 1**

*Bayes Factor Grading*

Grade	Bayes Factor ( $B$ )	Result
0	$B > 1$	$H_0$ supported
1	$10^{-1/2} < B < 1$	Weak evidence against $H_0$ ,
2	$10^{-1} < B < 10^{-1/2}$	Evidence against $H_0$ substantial
3	$10^{-3/2} < B < 10^{-1}$	Evidence against $H_0$ strong
4	$10^{-2} < B < 10^{-3/2}$	Evidence against $H_0$ very strong
5	$0 < B < 10^{-2}$	$H_0$ Evidence against $H_0$ decisive

*Posterior probability*

If a Bayesian specifies the prior probability  $\pi_0$  on  $H_0$  in addition to  $G$ , then the posterior probability of  $H_0$  is

$$\begin{aligned}
 P(H_0|x) &= \left[ 1 + \frac{(1 - \pi_0)}{\pi_0} \cdot \frac{m_G(x)}{f(x|\theta_0)} \right]^{-1} \\
 &= \left[ 1 + \frac{(1 - \pi_0)}{\pi_0} \cdot \frac{1}{B} \right]^{-1}
 \end{aligned} \tag{15}$$

For the Bayes factor criteria, the results in (12) and (13) are for finite-dimensional sample spaces. By extending these results to the abstract sample spaces, Shafie and Noorbaloochi (2001) showed that  $f(x | \theta)$  is the Radon-Nikodym derivative. Now we review their measure theoretic approach that they applied to generalize the notion of Bayes factor for infinite dimensional sample spaces.

Let  $P_G(A) = \int_{\Theta} P_{\theta}(A)dG(\theta)$  be the marginal probability measure of  $X$  on  $(\mathcal{X}, \mathcal{F}_1)$ . Then for a given  $\theta_0 \in \Theta$ ,  $P_{\theta_0}(A)$  and  $P_G(A)$  are both probability measures on  $(\mathcal{X}, \mathcal{F}_1)$ . The following lemma which is based on Grenander (1981) essentially ensures that the most powerful non-randomized tests are based on the R-N derivative of  $P_1$  with respect to  $P_0$ .

**Lemma II.2.** (Rohani, 2003) Let for any  $\theta_0 \in \Theta$ ,  $P_{\theta_0}$  be a probability measure on  $(\mathcal{X}, \mathcal{F}_1)$ . Then there exists a set  $A_0 \in \mathcal{F}_1$  with  $P_G(A_0) = 0$ , and a nonnegative  $P_G$ -measurable function  $f$ , such that for every  $A \in \mathcal{F}_1$

$$P_{\theta_0}(A) = \int_A f(x)dP_G(x) + P_{\theta_0}(A \cap A_0) \quad (16)$$

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

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

Regarding the above lemma, for testing  $H_0 : \theta = \theta_0$  against  $H_1 : \theta \neq \theta_0$ , Shafie and Noorbaloochi (2001) defined the Bayes factor as

$$\mathbf{B}(x) = \begin{cases} \infty & x \in A_0 \\ f(x) & x \notin A_0 \end{cases} \quad (17)$$

They showed that the two extreme cases in corollary II.1 can also occur in this case. If the set  $A_0$  has  $P_{\theta_0}(A_0) = 1$ , then two measures  $P_G$  and  $P_{\theta_0}$  is said to be perpendicular to each other, and in this case  $A_0$  as a critical region will allow us, without any error, to discriminate between  $H_0$  and  $H_1$ . If  $P_{\theta_0}$  and  $P_G$  are equivalent then the second part of the right hand side in (16) is zero and  $P_{\theta_0}(A_0) = 0$  and  $f \neq 0$  with  $P_G$ -probability 1.

**Lemma II.3.** (Rohani, 2003) Let  $(\Theta, \mathcal{F}_2, G)$  be a probability space and for any  $\theta \in \Theta$ ,  $P_\theta$  be a probability measure on  $(\mathcal{X}, \mathcal{F}_1)$ . In addition, suppose that  $P_G$  is the marginal probability measure of  $X$  on  $(\mathcal{X}, \mathcal{F}_1)$ . If all of  $P_\theta$ 's ( $\theta \in \Theta$ ) are equivalent, then  $P_G$  is equivalent to all of them.

**Theorem II.6.** (Rohani, 2003) Under the conditions of lemma (II.3), there exists  $A_0 \in \mathcal{F}_1$  with  $P_G(A_0) = 0$ , for which the Bayes factor for testing  $H_0 : \theta = \theta_0$  versus  $H_1 : \theta \neq \theta_0$  is given by

$$\mathbf{B}(x) = \begin{cases} \infty & x \in A_0 \\ \left[ \int \frac{dP_\theta}{dP_{\theta_0}}(x) dG(\theta) \right]^{-1} & x \notin A_0 \end{cases} \quad (18)$$

Note that  $f(x) = \frac{dP_{\theta_0}}{dP_G}(x)$  is R-D derivative of  $P_{\theta_0}$  with respect to  $P_G$ , and

$$\frac{dP_{\theta_0}}{dP_G}(x) = \left[ \frac{dP_G}{dP_{\theta_0}}(x) \right]^{-1} = \left[ \int_{\Theta} \frac{dP_{\theta}}{dP_{\theta_0}}(x) dG(\theta) \right]^{-1}.$$

$f(x | \theta) = \frac{dP_{\theta}}{dP_{\theta_0}}(x)$ , the density of  $P_{\theta}$  with respect to  $P_{\theta_0}$  can be considered as the likelihood function.

### Signal Detection in the Gaussian Random Fields

In this section, we review the likelihood and Bayesian approaches for signal detection in the Gaussian scale space random field. Suppose that a signal is present and we observe the field satisfying

$$dZ(\mathbf{t}) = \xi \sigma_0^{-N/2} f[\sigma_0^{-1}(\mathbf{t} - \mathbf{t}_0)] d\mathbf{t} + dW(\mathbf{t}) \quad (19)$$

where  $f$ , the shape function is a positive, and smooth function satisfying  $\int f(\mathbf{t})^2 d\mathbf{t} = 1$ .

Here, our choice of shape function is the Gaussian form of

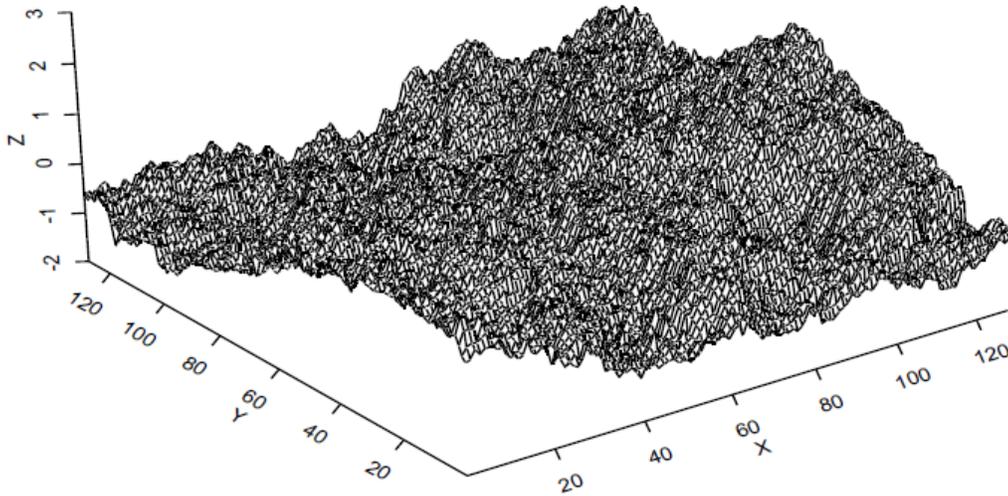
$$f(\mathbf{t}) = \pi^{-N/4} \exp(-\|\mathbf{t}\|^2/2) \quad (20)$$

Also,  $\mathbf{t}_0 \in \mathbb{R}^N$  is the unknown location of the signal, and the unknown parameters  $\xi \geq 0$ , and  $\sigma_0 > 0$  are respectively represent the amplitude and scale of the signal, and  $dW$  is assumed to be a Gaussian white noise.

A random field satisfying (19), assuming  $f$  is Gaussian, is a non-smooth Gaussian random field. Figure (1) shows a sample path of a non-smooth Gaussian random field.

**Figure 1**

*A sample path of spherical Gaussian random field*



For smoothing the random field in (19), two scenarios could be considered in this study. First, the width of smoothing kernel  $\sigma$  being fixed, and second, the scale space method where the kernel width  $\sigma$  is not fixed. Here, we have chosen the second scenario. After smoothing the random field in (19) with the kernel of  $\sigma^{-N/2}k[\sigma^{-1}(\mathbf{h} - \mathbf{t})]$ , where  $k$  is chosen to equal  $f$ , and  $\sigma$  is unknown, we have the Gaussian scale space random field,

$$X(\mathbf{t}, \sigma) = \sigma^{-N/2} \int f[\sigma^{-1}(\mathbf{h} - \mathbf{t})] dZ(\mathbf{h}) \quad (21)$$

which has the mean and covariance functions as follow (see Siegmund & Worsley, 1995),

$$\mu(\mathbf{t}, \sigma; \xi, \mathbf{t}_0, \sigma_0) = \xi(\sigma_0\sigma)^{-N/2} \int f[\sigma_0^{-1}(\mathbf{h} - \mathbf{t}_0)] f[\sigma^{-1}(\mathbf{h} - \mathbf{t})] d\mathbf{h} \quad (22)$$

$$R((\mathbf{t}_1, \sigma_1), (\mathbf{t}_2, \sigma_2)) = \sigma_1\sigma_2^{-N/2} \int f[\sigma_1^{-1}(\mathbf{h} - \mathbf{t}_1)] f[\sigma_2^{-1}(\mathbf{h} - \mathbf{t}_2)] d\mathbf{h} \quad (23)$$

Figure 2 is a realization of a Gaussian scale space random field.

Recalling the assumption of  $k = f$ , we can write the expanded form of (21) as follows

$$\begin{aligned} X(\mathbf{t}, \sigma) &= (\sigma_0\sigma)^{-N/2}\xi \int f[\sigma_0^{-1}(\mathbf{h} - \mathbf{t}_0)] f[\sigma^{-1}(\mathbf{h} - \mathbf{t})] d\mathbf{h} \\ &\quad + \sigma^{-N/2} \int f[\sigma^{-1}(\mathbf{h} - \mathbf{t})] dW(\mathbf{h}) \end{aligned} \quad (24)$$

If we denote the first and second term in (24) respectively by  $\mu$  and  $W^*$ , we have

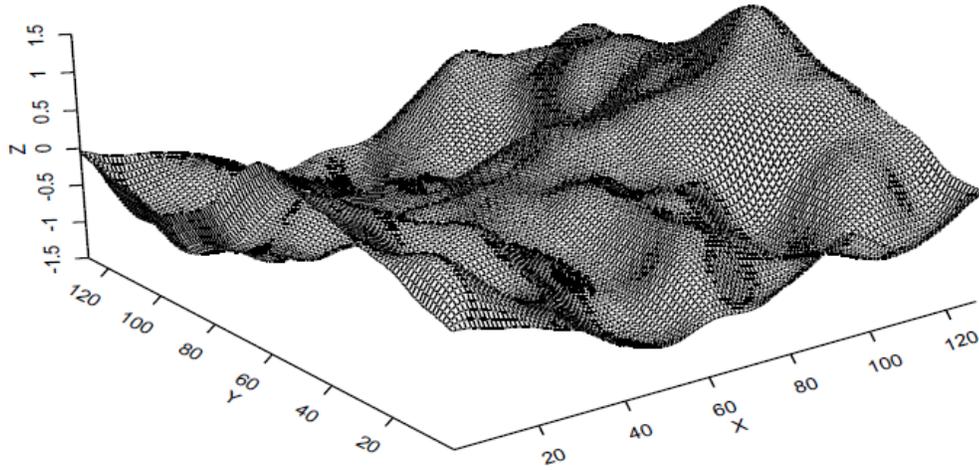
$$X(\mathbf{t}, \sigma) = \mu(\mathbf{t}, \sigma; \xi, \mathbf{t}_0, \sigma_0) + W^*(\mathbf{t}, \sigma) \quad (25)$$

where  $\mu(\mathbf{t}, \sigma; \xi, \mathbf{t}_0, \sigma_0)$  is given in (22), and  $W^*(\mathbf{t}, \sigma)$  is an N-dimensional Gaussian random field with zero mean, unit variance and covariance function given in (23).

In the next theorem proved by Parzen (1961), the generalized likelihood ratio test statistic for the hypothesis  $H_0 : \xi = 0$ , against  $H_1 : \xi > 0$  is obtained. The theorem shows that the generalized likelihood ratio test for smoothed Gaussian random field in (21) is the global maxima of the sample path, as stated in (Siegmund & Worsley, 1995).

**Figure 2**

*A sample path of a smoothed spherical Gaussian random field*



**Theorem II.7.** Let  $X(\mathbf{t}, \sigma)$  be a Gaussian random field in the form of (24) with  $k = f$ , where  $dW$  is a white noise. For testing the null hypothesis  $H_0 : \xi = 0$  against the composite alternative  $H_1 : \xi > 0$ , the generalized likelihood ratio test statistic obtained from (12) is equivalent to

$$X_{\max} = \sup_{\mathbf{t}} X(\mathbf{t}).$$

**Corollary II.2.** Suppose  $X(\mathbf{t}, \sigma)$  is a Gaussian random field of the form (24), with kernel  $k = f$ , then the mean of this random field has the form of  $\xi R((\mathbf{t}, \sigma), (\mathbf{t}_0, \sigma_0))$ , and the covariance function  $R((\mathbf{t}_1, \sigma_1), (\mathbf{t}_2, \sigma_2))$ , is as in (23).

In the next theorem, Parzen (1961) obtained the generalized likelihood ratio test statistic for the hypothesis  $H_0 : \xi = 0$ , against  $H_1 : \xi > 0$ . This theorem shows that the

generalized likelihood ratio test for smooth Gaussian random field (24) is the global maxima of the sample path, (see Siegmund & Worsley, 1995).

**Theorem II.8.** Let  $X(\mathbf{t}, \sigma)$  be a Gaussian random field in the form of (24) with  $k = f$ , where  $dW(\mathbf{t})$  is a white noise. For testing the null hypothesis  $H_0 : \xi = 0$  vs. the composite alternative  $H_1 : \xi > 0$ , the generalized likelihood ratio test statistic obtained from (12) is equivalent to

$$X_{\max} = \sup_{\mathbf{t}, \sigma} X(\mathbf{t}, \sigma).$$

Recalling that  $HN(\mu, S)$  denotes a Gaussian measure on a Hilbert space where  $\mu$  and  $S$  are, respectively, its mean and covariance operators, based on the above theorem, Rohani (2003) found the log of likelihood ratio by obtaining the Radon-Nikodym derivative of  $P_1^* = HN(\mu^*, S)$  with respect to  $P_0 = HN(0, S)$ . Moreover, he used it to make the Bayesian statistical testing which we will review it later here.

$$\log[f_{\mu^*}(x)] = \log\left[\frac{dP_1^*}{dP_0}(x)\right] = \xi x(\mathbf{t}^*, \sigma^*) - \xi^2/2 \quad (26)$$

A good approximation of the  $P$ -value associated with the generalized likelihood ratio test statistic  $X_{\max}$ , for smooth (or non-smooth) scale space random fields, is the use of expected Euler characteristic of the excursion set, (see Siegmund & Worsley, 1995).

**Definition II.26.** The *excursion set* of  $N$ -dimensional random field  $X(\mathbf{t}, \sigma)$  above the level  $x$  in a subset  $T$  of  $\mathbb{R}^N$  is defined as

$$A_x(X, T) = \{\mathbf{t} \in T : X(\mathbf{t}, \sigma) \geq x\}, \quad T \subset \mathbb{R}^N,$$

and denoted by  $A_x$ .

As the threshold level  $x$  increases, from a theorem presented by Adler (2010), for very high thresholds near the global supremum of  $X(\mathbf{t}, \sigma)$ , the Euler characteristic is 1 if  $X_{\max} > x$ , and zero otherwise. Thus the expected Euler characteristic of the excursion set approximates the  $P$ -value of  $X_{\max}$ , (Hasofer, 1978),

$$P\{X_{\max} \geq x\} \approx E[\chi(A_x)]$$

For the two dimensional case, the Euler characteristic counts the number of connected components of a set, minus the number of “holes”. The advantage of this approximation is that it can be found exactly. Siegmund and Worsley (1995) have used the expected Euler characteristic in scale space Gaussian random fields.

Now we review the Bayesian testing which Shafie et al. (2003) proposed for signal detection in noisy image that its shape is unknown and considered as element of a Hilbert space. For the case of a smooth Gaussian random field, Rohani et al. (2006) found the Bayes factor of simple hypothesis  $H_0 : \xi = 0$  against  $H_1 : \xi > 0$ .

**Corollary II.3.** Let  $X(\mathbf{t}, \sigma)$  be the Gaussian random field

$$X(\mathbf{t}, \sigma) = \mu(\mathbf{t}, \sigma ; \xi, \mathbf{t}_0, \sigma_0) + W^*(\mathbf{t}, \sigma),$$

where

$$\mu = (\sigma_0 \sigma)^{-N/2} \xi \int f[\sigma_0^{-1}(\mathbf{h} - \mathbf{t}_0)] k[\sigma^{-1}(\mathbf{h} - \mathbf{t})] d\mathbf{h},$$

$$W^* = \sigma^{-N/2} \int k [\sigma^{-1}(\mathbf{h} - \mathbf{t})] dW(\mathbf{h}),$$

and  $dW(\mathbf{t})$  is a white noise and width of the signal,  $\sigma$  is unknown. Assume  $\rho$  is the covariance function of  $W(\mathbf{t})$ , and  $R(\rho^{1/2})$  is the range of  $\rho^{1/2}$ . For testing  $H_0 : \xi = 0$  against  $H_1 : \xi > 0$ , if  $\mu(\mathbf{t}, \sigma ; \xi, \mathbf{t}_0, \sigma_0) \in R(\rho^{1/2})$ , then  $P_0$  and  $P_G$  are equivalent and the Bayes factor is

$$\frac{dP_0}{dP_G} = \left[ \int_{(\xi, \mathbf{t}, \sigma)} \exp[\xi x(\mathbf{t}, \sigma) - \xi^2/2] dG(\xi, \mathbf{t}) \right]^{-1} \quad (27)$$

We should note that the above Bayes factor is the inverse of Radon-Nikodym derivative of  $P_G$  w.r.t  $P_0$ .

### Two Dimensional Case

In this part, we review the application of results in the previous part to a 2-dimensional smooth Gaussian scale space random field which is done by Rohani (2003).

Let  $X(\mathbf{t}, \sigma)$  be a Gaussian scale space random field with the form (24), where  $T = [0, 1] \times [0, 1]$  whose mean and variance respectively are zero and one, and covariance function comes from (23). Given a number of smooth path functions and prior distribution function on  $\theta = (\xi, (t_{01}, t_{02}), \sigma_0)$ , Rohani (2003) calculated the Radon-Nikodym derivative in (27) for 2-dimensional case as follows.

$$\mathbf{B}(x) = \left[ \int_{(t_{01}, t_{02}, \sigma_0)} \frac{\Phi(x)}{\phi(x)} dG_2(t_{01}) dG_3(t_{02}) dG_4(\sigma_0) \right]^{-1} \quad (28)$$

where  $\phi(\cdot)$  and  $\Phi(\cdot)$  denote the density and distribution functions of the standard Gaussian, and  $x = x(t_{01}, t_{02}, \sigma_0)$ . In the above, priors on the parameters of interest  $(\xi, t_{01}, t_{02}, \sigma_0)$  are assumed mutually independent and noted as  $G_1(\xi), G_2(t_{01}), G_3(t_{02}), G_4(\sigma_0)$ , where  $\xi$  has an improper prior distribution on  $[0, \infty)$ .

In the next chapter, we generalize the above Bayes factor for the Gaussian random field mixture models.

### Finite Mixture Models and Identifiability

Finite mixture models are very popular statistical modeling techniques for different random phenomena. The best use of these models is when multiple populations contribute to the observed outcome, and there exists an unobserved heterogeneity. In other words, the population from which we are sampling is heterogeneous, and there are multiple sub-populations.

The simplest finite mixture models are a class of models that combine a finite number of probability distribution functions, so called components, to better model the data.

A finite mixture density of  $Y$  with  $C$  components is given by

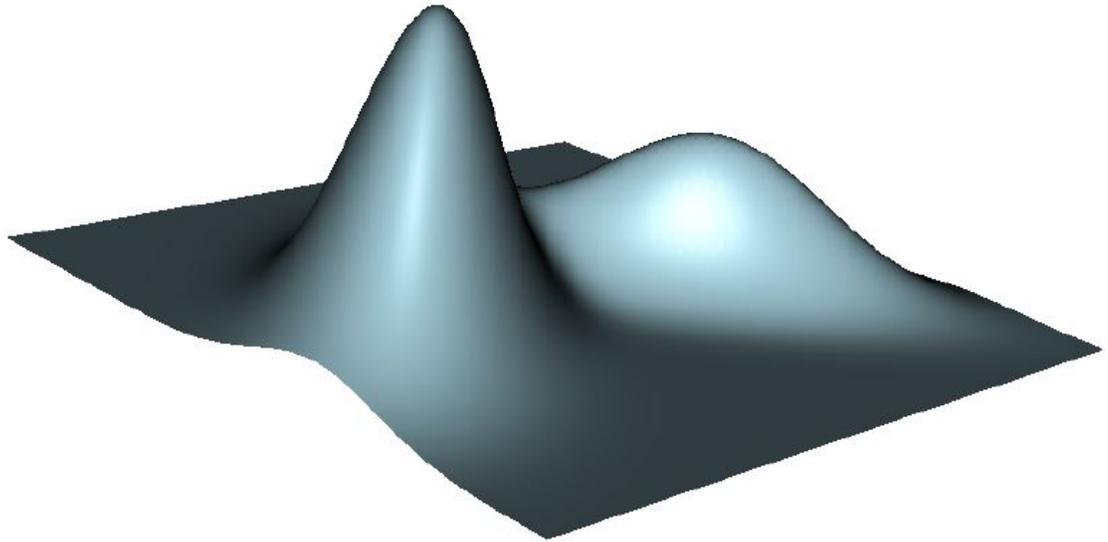
$$Y \sim \sum_{i=1}^C \pi_i f(\cdot | \theta_i) \quad (29)$$

For the weight components  $\pi_i$ s, it holds

$$\sum_{i=1}^C \pi_i = 1 \quad \text{and} \quad \pi_i > 0, \quad \forall i = 1, \dots, C$$

**Figure 3**

*A mixture of two bi-variate Normal densities*



In 29, it assumes that data  $Y$  are drawn from a density modeled as a convex combination of components each specified by  $f$ .

When a mixture distribution is consisting of  $C$  Gaussian components  $X_i \sim N(\mu_i, \Sigma_i)$ , and mixture weights be  $\pi_i$  for  $i = 1, \dots, C$ , we have a Gaussian mixture model,

$$Y \sim \sum_{i=1}^C \pi_i \phi_{X_i}(x)$$

In other words, the density of a Gaussian mixture is a convex combination of Gaussian densities. Figure (3), courtesy of Green (2019), shows a mixture of two bi-variate Normal densities.

Any finite mixture model is generally being represented by its parameter vector  $\Theta$ , which consists of the component weights,  $\pi_i$ s, and component parameters,  $\theta_i$ s of specific

distributions in the mixture. Due to a mapping from parameter space to the model space, identifiability is very important to obtain model parameter estimates. In fact, in the finite mixture models, estimating and testing parameters of distributions can be meaningfully discussed if the family of mixing distributions is identifiable. For a model to be identifiable, the mapping from the parameter space into the model space should be one-to-one. So that the result of mapping would be a unique parameter vector, (Forghaniarani & Shafie, 2020).

Let  $\Theta$  denote the space of parameters for  $C$  component mixtures, a mixture model in (29) is identifiable if the following condition is fulfilled:

$$\forall i, j \in \{1, \dots, C\} : i \neq j \Rightarrow \theta_i \neq \theta_j$$

“In a pioneer work, Teicher (1961) studied the identifiability of finite mixtures and showed that the class of all mixtures of a one-parameter additively-closed family of distributions is identifiable. Since then, identifiability has been proved in many special cases. Teicher (1963) proposed a sufficient condition for the identifiability of a finite mixture and applied it to the Normal and Gamma families”, (Forghaniarani & Shafie, 2020).

Now consider a mixture model that its components are Gaussian distribution of functions, i.e. Gaussian processes or Gaussian random fields. This type of Mixtures is an interesting and useful alternative to mixture of high-dimensional Normals. Huang et al. (2014) studied identifiability of a mixture of Gaussian processes as follows,

$$Y(t) \sim \sum_{c=1}^C \pi_c N \{ \mu_c(t), G_c(s, t) \} \quad (30)$$

where  $\mathcal{C}$  be a latent class variable with a discrete distribution  $P(\mathcal{C} = c) = \pi_c$  for  $c = 1, 2, \dots, C$ . Moreover,  $N\{\mu_c(t), G_c(s, t)\}$  denotes a Gaussian process  $\{X(t) : t \in \mathbb{T}\}$  with mean  $\mu_c(t)$ , and covariance function  $\text{Cov}\{X(s), X(t)\} = G_c(s, t)$ . And  $\mathbb{T}$  is a closed and bounded time interval  $[0, T]$ . As a condition of identifiability in (30), they proved the following theorem.

**Theorem II.9.** Suppose  $G_c(s, t)$  is a positive definite and bivariate smooth function of  $s$  and  $t$  and  $\mu_c(t)$  is a smooth function of  $t$  for any  $c = 1, \dots, C$ . Let  $\mathbf{S} = \{t \in \mathbb{T} : (\mu_i(t), G_i(t, t)) = (\mu_j(t), G_j(t, t)) \text{ for some } 1 \leq i \neq j \leq C\}$ . If the complement of  $\mathbf{S}$  is not empty, then the proposed mixture of Gaussian processes in (30) is identifiable.

We will use the above theorem to conclude the identifiability of the mixture of Gaussian random fields provided in the next chapter.

## CHAPTER III

### METHODOLOGY

In this study, the averaged image obtained from fMRI technology is modeled as a mixture of two real-valued Gaussian scale-space random fields  $X_1(\mathbf{t}, \sigma_1)$  and  $X_2(\mathbf{t}, \sigma_2)$ ,  $\mathbf{t} \in T \subset \mathbb{R}^N$ , with their mean functions  $\mu_1(\mathbf{t}, \sigma_1)$  and  $\mu_2(\mathbf{t}, \sigma_2)$  and covariance functions  $R_1((\mathbf{t}_1, \sigma_{11}), (\mathbf{t}_2, \sigma_{12}))$  and  $R_2((\mathbf{t}_1, \sigma_{21}), (\mathbf{t}_2, \sigma_{22}))$ . The model can be formulated as

$$X(\mathbf{t}, \sigma_1, \sigma_2) = zX_1(\mathbf{t}, \sigma_1) + (1 - z)X_2(\mathbf{t}, \sigma_2) \quad (1)$$

where,

$$p(z) = \begin{cases} \pi & z = 1 \\ 1 - \pi & z = 0 \end{cases}$$

The objective of this dissertation is to propose a Bayesian procedure to test the existence of signal in the cases that shape of signals are known, and the amplitude, location, and scale parameters in two random fields,  $\boldsymbol{\theta}_1 = (\xi_1, \mathbf{t}_0, \sigma_{01})$  and  $\boldsymbol{\theta}_2 = (\xi_2, \mathbf{s}_0, \sigma_{02})$  have some known prior distributions. Testing the existence of signal is statistically equivalent to the problem of testing the hypotheses such that

$$\left\{ \begin{array}{l} H_0 : \xi_1 = 0 \ \& \ \xi_2 = 0 \\ H_1 : \xi_1 > 0 \ \& \ \xi_2 > 0 \end{array} \right.$$

Outlines of this chapter demonstrate the methodology to investigate the following research questions:

- Q1 How to develop a Bayesian approach for testing the signal in the mixture model  $X(\mathbf{t}, \sigma)$  for two-dimensional images?
- Q2 How the performance of the Bayes factor can change with respect to different parameter schemes?

In this chapter, we assume that our images are realizations of a mixture of two homogeneous Gaussian scale space random fields. We suppose that each component of the mixture has a signal with the form of a known function, here a Gaussian form, centered at an unknown location  $\mathbf{t}_0$ , and multiplied by an unknown amplitude  $\xi$ . We also assume that the parameter of scale or width of signal,  $\sigma_0$  is not known.

Our objective is to detect the signal when images are modeled as (1). Therefore, in section 1, we will give more details about the model, and discuss its identifiability. In section 2, we obtain the R-N derivative for the mixture model (1). In section 3, for model (1), we discuss the likelihood ratio test statistic for testing the signal when the model is a mixture of two Gaussian random fields. In section 4, we propose the Bayesian procedure by employing the Radon-Nikodym derivative in the signal detection problem and apply our approach to the two-dimensional case. In section 5, as side work, we present the Gibbs sampling method that through it, we estimate  $\pi$  for a two-dimensional case. In

section 6, we provide the simulation scheme for the numerical study on the model (1) in a two-dimensional case.

### Gaussian Random Field Mixture Model and its Identifiability

Assume that we have two real-valued Gaussian scale space random fields  $X_1(\mathbf{t}, \sigma_1)$  and  $X_2(\mathbf{t}, \sigma_2)$ , each satisfying

$$X(\mathbf{t}, \sigma) = \mu(\mathbf{t}, \sigma; \xi, \mathbf{t}_0, \sigma_0) + W^*(\mathbf{t}, \sigma) \quad (2)$$

Wherefore, their mean functions would be

$$\mu(\mathbf{t}, \sigma; \xi, \mathbf{t}_0, \sigma_0) = \xi(\sigma_0\sigma)^{-N/2} \int f[\sigma_0^{-1}(\mathbf{h} - \mathbf{t}_0)] f[\sigma^{-1}(\mathbf{h} - \mathbf{t})] d\mathbf{h} \quad (3)$$

and their covariance functions would be

$$R((\mathbf{t}_1, \sigma_1), (\mathbf{t}_2, \sigma_2)) = \sigma_1\sigma_2^{-N/2} \int f[\sigma_1^{-1}(\mathbf{h} - \mathbf{t}_1)] f[\sigma_1^{-1}(\mathbf{h} - \mathbf{t}_2)] d\mathbf{h} \quad (4)$$

Let  $\boldsymbol{\theta}_1 = (\xi_1, \mathbf{t}_0, \sigma_{01})$  and  $\boldsymbol{\theta}_2 = (\xi_2, \mathbf{s}_0, \sigma_{02})$  be the corresponding parameter vectors to  $X_1$  and  $X_2$  where  $\xi_1 \geq 0$ ,  $\mathbf{t}_0 \in T \subset \mathbb{R}^N$  and  $\sigma_{01} > 0$  respectively represent the amplitude, location and scale of the signal in  $X_1$ , and  $\xi_2 \geq 0$ ,  $\mathbf{s}_0 \in T$  and  $\sigma_{02} > 0$  respectively represent the amplitude, location and scale of the signal in  $X_2$ . Therefore, the

Gaussian mixture model (1) can be written as

$$\begin{aligned} X(\mathbf{t}, \sigma_1, \sigma_2) &= z\mu_1(\mathbf{t}, \sigma_1; \xi_1, \mathbf{t}_0, \sigma_{01}) + (1 - z)\mu_2(\mathbf{t}, \sigma_2; \xi_2, \mathbf{s}_0, \sigma_{02}) \\ &+ zW_1^*(\mathbf{t}, \sigma_1) + (1 - z)W_2^*(\mathbf{t}, \sigma_2) \end{aligned} \quad (5)$$

where,

$$p(z) = \begin{cases} \pi & z = 1 \\ 1 - \pi & z = 0 \end{cases}$$

Here,  $W_j^*(\mathbf{t}, \sigma_j)$ ,  $j = 1, 2$ , is a smooth N-dimensional Gaussian random field with zero mean, unit variance and the covariance function satisfying equation (4).

For hypothesis testing to be valid, the identifiability of the mixture models should be studied. By the theorem (II.9) proved by Huang et al. (2014), we can conclude that model (1) is identifiable.

### **Radon-Nikodym Derivative in Gaussian Mixture Models**

Assume  $X_1$  and  $X_2$  are Gaussian random fields in  $\mathcal{X} = L^2(T, B, \lambda)$ , so  $P_1$  and  $P_2$  are Gaussian measures on  $\mathcal{X} = L^2$  induced by  $X_1$  and  $X_2$ . From Chapter II, we know that if  $P$  is a probability measure on  $(\mathcal{X}, \mathcal{F}_1)$ , then there exist a set  $A_0 = \{x \mid x \in \mathcal{X}\}$  in  $\mathcal{F}_1$  with  $P_0(A_0) = 0$ , and a non-negative  $P_0$ -measurable function  $f$  such that  $f(x) = \frac{dP}{dP_0}(x)$  be Radon-Nikodym derivative of  $P$  with respect to  $P_0$ .

Also, we know if  $P \ll P_0$ , we can consider  $f$  as the density of  $P$  w.r.t  $P_0$ . By this, if  $P_1 \ll P_0$  and  $P_2 \ll P_0$ , we know that there exist  $f_1(x) = \frac{dP_1}{dP_0}(x)$  and  $f_2(x) = \frac{dP_2}{dP_0}(x)$  which  $f_1$  is the density of  $P_1$  w.r.t  $P_0$ , and  $f_2$  is the density of  $P_2$  w.r.t  $P_0$ . Now, consider

the mixture model (1). Because  $X$  is also a random field in  $\mathcal{X} \equiv L^2$ , we assume  $P^*$  be the measure induced by  $X$  on  $\mathcal{X}$ . Based on the definition of mixtures, model (1) can be written in the following form

$$f(x) = \pi f_1(x_1) + (1 - \pi) f_2(x_2)$$

or more precisely,

$$\frac{dP^*}{dP_0}(x) = \pi \frac{dP_1}{dP_0}(x_1) + (1 - \pi) \frac{dP_2}{dP_0}(x_2) \quad (6)$$

Here, we want to find  $f_1(x)$ ,  $f_2(x)$  to find  $f(x)$ . By definition of Gaussian probability measure in Hilbert space, we have  $P_1 = HN(\mu_1, S)$ , and  $P_2 = HN(\mu_2, S)$ . We also assume  $P_0 = HN(0, S)$ , and use the result from Chapter II in (26). So we have the Radon-Nikodym derivative of  $P_j$  w.r.t  $P_0$  which is the likelihood function as following,

$$f_j(x; \theta_j) = \frac{dP_j}{dP_0}(x) = \exp \left\{ \xi_j x_j(t_0, \sigma_0) - \xi_j^2 / 2 \right\} \quad \text{for } j = 1, 2 \quad (7)$$

For this study, we should have more than one realization of the field. Assuming to have  $n$  sample paths from mixture of two Gaussian random fields, we can obtain R-N derivative. Assume  $x_1, \dots, x_n$  are observed from model (1), so from (7), the likelihood function for  $x_i$ ,  $i = 1, \dots, n$  can be written as

$$f(x_i | \pi, \theta_1, \theta_2) = \pi f_1(x_i | \theta_1) + (1 - \pi) f_2(x_i | \theta_2) \quad (8)$$

or

$$f(x_i, z_i | \pi, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = [\pi f_1(x_i | \boldsymbol{\theta}_1)]^{z_i} [(1 - \pi) f_2(x_i | \boldsymbol{\theta}_2)]^{(1-z_i)} \quad (9)$$

From (9),

$$\begin{aligned} f(\mathbf{x}, \mathbf{z} | \pi, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) &= \prod_{i=1}^n [\pi f_1(x_i | \boldsymbol{\theta}_1)]^{z_i} [(1 - \pi) f_2(x_i | \boldsymbol{\theta}_2)]^{(1-z_i)} \\ &= \prod_{i=1}^n [\pi \exp\{\xi_1 x_i(\mathbf{t}_0, \sigma_{01}) - \xi_1^2/2\}]^{z_i} \\ &\quad \times [(1 - \pi) \exp\{\xi_2 x_i(\mathbf{s}_0, \sigma_{02}) - \xi_2^2/2\}]^{(1-z_i)} \end{aligned}$$

In section 5, the following likelihood function can be used to estimate  $\pi$  and other parameters.

$$\begin{aligned} f(\mathbf{x}, \mathbf{z} | \pi, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) &= \prod_{i=1}^n \exp\left\{z_i [\xi_1 x_i(\mathbf{t}_0, \sigma_{01}) - \xi_1^2/2]\right\} \\ &\quad \times \exp\left\{(1 - z_i) [\xi_2 x_i(\mathbf{s}_0, \sigma_{02}) - \xi_2^2/2]\right\} \pi^{z_i} (1 - \pi)^{(1-z_i)} \end{aligned} \quad (10)$$

### Likelihood Approach for Gaussian Mixture Models

In Chapter II, we learned that for testing the simple  $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$  against  $H_a = \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$ , the generalized likelihood ratio test statistic is

$$f(x) = \sup_{\boldsymbol{\theta}^* \in \Theta} f_{\boldsymbol{\theta}^*}(x) = \sup_{\boldsymbol{\theta}^* \in \Theta} \frac{dP_{\boldsymbol{\theta}^*}}{dP_{\boldsymbol{\theta}}}(x)$$

Also, regarding Theorem II.7 proved by Parzen (1961), when  $X(\mathbf{t}, \sigma)$  is a Gaussian random field in the form of

$$X(\mathbf{t}, \sigma) = \mu(\mathbf{t}, \sigma; \xi, \mathbf{t}_0, \sigma_0) + W(\mathbf{t}, \sigma),$$

and  $k = f$ , the generalized likelihood ratio test statistic for testing  $H_0 : \xi = 0$  vs.

$H_a : \xi > 0$  is

$$X_{\max} = \sup_{\mathbf{t}, \sigma} X(\mathbf{t}, \sigma)$$

Rohani (2003) showed that for the Gaussian scale space random field  $X(\mathbf{t}, \sigma)$ ,

$$X_{\max} = \sup_{\mathbf{t}_0, \xi, \sigma_0} [\xi x(\mathbf{t}_0, \sigma_0) - \xi^2/2]$$

Moreover,  $P(X_{\max} \geq x_0) \approx E[\chi(A_{x_0})]$ , where  $\chi(A_{x_0})$  is the Euler characteristic of the excursion set of  $X(\mathbf{t}, \sigma)$ ,  $A_{x_0} = \{(\mathbf{t}, \sigma) \in (T \times [\sigma_1^*, \sigma_2^*]), X(\mathbf{t}, \sigma) \geq x_0\}$ . For the mixture model (1), the generalized likelihood ratio test statistic is

$$X_{\max} = zX_{1\max} + (1 - z)X_{2\max} \tag{11}$$

with

$$P(X_{\max} \geq x_0) = \pi \cdot P(X_{1\max} \geq x_0) + (1 - \pi) \cdot P(X_{2\max} \geq x_0),$$

therefore, we have

$$P(X_{\max} \geq x_0) \approx \pi E[\chi(A_1)] + (1 - \pi)E[\chi(A_2)] \quad (12)$$

where  $A_j = \{(\mathbf{t}, \sigma) \in (T \times [\sigma_1^*, \sigma_2^*]), X_j(\mathbf{t}, \sigma) \geq x_0\}$  for  $j = 1, 2$ .

For a future study, this result can be applied to two dimensional case where  $T \subset \mathbb{R}^2$ . In that case, to approximate  $P(X_{\max} \geq x_0)$ , the expected Euler characteristic that Siegmund & Worsley (1995) obtained for a Gaussian random field can be used.

### Bayes Factor for Mixture of Gaussian Random Fields

In Chapter II, we learned that for testing  $H_0 : \xi = 0$ , against  $H_a = \xi > 0$ , the Bayes factor is as follows,

$$B(x) = \begin{cases} \infty & x \in A_0 \\ \left[ \int \frac{dP_{\boldsymbol{\theta}}}{dP_{\boldsymbol{\theta}_0}}(x) dG(\boldsymbol{\theta}) \right]^{-1} & x \notin A_0 \end{cases}$$

We can use the above definition of Bayes factor for testing  $H_0 : \boldsymbol{\xi} = \mathbf{0}$ , against  $H_a = \boldsymbol{\xi} > \mathbf{0}$ , where  $\boldsymbol{\theta} = (\pi, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$  is the vector of parameters in the model (1), and  $\boldsymbol{\xi} = (\xi_1, \xi_2)$ . Note that  $\boldsymbol{\theta}_1 = (\xi_1, \mathbf{t}_0, \sigma_{01})$ , and  $\boldsymbol{\theta}_2 = (\xi_2, \mathbf{s}_0, \sigma_{02})$  are corresponding parameter vectors of Gaussian random fields,  $X_1(\mathbf{t}, \sigma_1)$  and  $X_2(\mathbf{t}, \sigma_2)$ , respectively, and  $G(\boldsymbol{\theta})$  is a prior distribution of  $\boldsymbol{\theta}$  over its parameter space  $\Theta$ , and  $P_{\boldsymbol{\theta}}$  is a probability measure on  $(\mathcal{X}, \mathcal{F}_1)$ . Moreover, for any  $A \in \mathcal{F}_1$ ,  $A = \{x \mid x \in \mathcal{X} \equiv L^2\}$ ,  $P_G(A) = \int_{\Theta} P_{\boldsymbol{\theta}}(A) dG(\boldsymbol{\theta})$  is the marginal probability measure of  $X$  on  $(\mathcal{X}, \mathcal{F}_1)$ , and  $P_G(A_0) = 0$ .

Finding the above Bayes factor, analytically is not tractable, and we need to compute it through numerical methods. By writing the Bayes factor as

$$B(x) = \left[ \int \frac{dP_{\boldsymbol{\theta}}}{dP_{\boldsymbol{\theta}_0}}(x) dG(\boldsymbol{\theta}) \right]^{-1} = \left[ \int \frac{dP_{\boldsymbol{\theta}}}{dP_{\boldsymbol{\theta}_0}}(x) g(\boldsymbol{\theta}) d(\boldsymbol{\theta}) \right]^{-1},$$

it can be interpreted as the inverse of expected value of likelihood ratio,  $\frac{dP_{\boldsymbol{\theta}}}{dP_{\boldsymbol{\theta}_0}}(x)$  w.r.t the prior density,  $g(\boldsymbol{\theta})$ . Regarding this perspective, we are going to find the Bayes factor for model (1).

To apply the above definition of the Bayes factor, we need to assume priors on the model parameters. We assume  $\pi$  and  $\boldsymbol{\theta}_1$ , and  $\boldsymbol{\theta}_2$  are independent in prior and their prior distribution is  $(G_{\pi} \cdot G_{\boldsymbol{\theta}_1} \cdot G_{\boldsymbol{\theta}_2})$ , therefore,  $dG(\boldsymbol{\theta}) = dG_{\pi} \cdot dG_{\boldsymbol{\theta}_1} \cdot dG_{\boldsymbol{\theta}_2}$ . Based on (6) and (8),  $B(x)$  is formulated as,

$$\frac{1}{B(x)} = \int_{[0,1] \times \Theta_1 \times \Theta_2} \prod_{i=1}^n \left[ \pi f_1(x_i | \boldsymbol{\theta}_1) + (1 - \pi) f_2(x_i | \boldsymbol{\theta}_2) \right] dG_{\pi} dG_{\boldsymbol{\theta}_1} dG_{\boldsymbol{\theta}_2} \quad (13)$$

where,  $f_1(x_i | \boldsymbol{\theta}_1) = \exp\{\xi_1 x_i(\mathbf{t}_0, \sigma_{01}) - \xi_1^2/2\}$ , and

$f_2(x_i | \boldsymbol{\theta}_2) = \exp\{\xi_2 x_i(\mathbf{s}_0, \sigma_{02}) - \xi_2^2/2\}$ .

By (13) and through samples from the prior density, obtaining  $B(x)$  is numerically possible. For this study, we borrowed our choice of priors on  $\boldsymbol{\theta}_1 = (\xi_1, \mathbf{t}_0, \sigma_{01})$  and  $\boldsymbol{\theta}_2 = (\xi_2, \mathbf{s}_0, \sigma_{02})$  from Rohani (2003), and Rohani et al. (2006).

Assuming  $\pi$ ,  $\xi_1$ ,  $\xi_2$ ,  $\mathbf{t}_0$ ,  $\mathbf{s}_0$ ,  $\sigma_{01}$ ,  $\sigma_{02}$  are independent in prior, the priors are as follows:

- $\pi | \boldsymbol{\alpha} \sim \text{Beta}(\alpha_1, \alpha_2)$ ,  $\alpha_1 = 2$ ,  $\alpha_2 = 2$

- $\sigma_{01}, \sigma_{02} \stackrel{\text{iid}}{\sim} \text{Inv-Gamma}(\beta_1, \beta_2)$ ,  $\beta_1 = 4$ ,  $\beta_2 = 0.5$
- $\mathbf{t}_0, \mathbf{s}_0 \stackrel{\text{iid}}{\sim} TN(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{a}, \mathbf{b})$ , where  $\boldsymbol{\mu} = (0.5, 0.5)^T$  and covariance matrix,  $\boldsymbol{\Sigma} = \begin{pmatrix} 1 & .5 \\ .5 & 1 \end{pmatrix}$ , and the lower and upper truncation points  $\mathbf{a} = (0, 0)^T$ ,  $\mathbf{b} = (1, 1)^T$ .
- $\xi_1, \xi_2 \stackrel{\text{iid}}{\sim} \text{Unif}(0, 5)$

In Chapter IV, to compute the Bayes factor, we will apply “prior.sampler” function in the two-dimensional case where  $T \subset \mathbb{R}^2$ . The “prior.sampler” function written in R by the author of this work is provided in Appendix.

We should note that in the case we only have one realization of the field, the following formula can be used to obtain the Bayes factor.

$$\frac{1}{B(y)} = E(\pi) \frac{1}{B_1(x_1)} + [1 - E(\pi)] \frac{1}{B_2(x_2)} \quad (14)$$

In section 6, we provide the simulation scheme for the numerical study on the two-dimensional case.

Although to obtain the Bayes factor by employing (13), we do not need to estimate  $\pi$  or other model parameters, as side work and initiation for the future study, we run Gibbs sampling to estimate  $\pi$ ,  $\xi_1$ ,  $\xi_2$ ,  $\mathbf{t}_0$ ,  $\mathbf{s}_0$ ,  $\sigma_{01}$ ,  $\sigma_{02}$ . The details are displayed in the next section.

### **Estimating $\pi$ through Gibbs Sampling**

To estimate  $\pi$  and the other parameters in the model (1), we can use Gibbs sampling for mixture models. Gibbs sampler is an MCMC method that originated by Geman (1984), and is a useful approach to draw the sample from the joint posterior

distribution when its form is complicated and difficult to handle. In Gibbs sampler, instead of directly sampling from joint posterior distribution, we sample from conditional posterior distributions. By the theorem of Markov Chain, for a sufficiently large sample, the result can be viewed as a random sample from posterior  $p(\cdot | data)$ .

By obtained samples from the posterior  $P(\boldsymbol{\theta} | Data)$ , where  $\boldsymbol{\theta} = (\pi, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$  and  $\boldsymbol{\theta}_1 = (\xi_1, \mathbf{t}_0, \sigma_{01})$  and  $\boldsymbol{\theta}_2 = (\xi_2, \mathbf{s}_0, \sigma_{02})$ , we find marginal posterior distribution  $P(\pi | \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, data)$ , and the mean of this distribution would be an estimate for  $\pi$ . To find  $P(\pi | \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, data)$ , besides the likelihood function, we need to specify prior distributions. Our choice of priors on  $\boldsymbol{\theta}_1 = (\xi_1, \mathbf{t}_0, \sigma_{01})$  and  $\boldsymbol{\theta}_2 = (\xi_2, \mathbf{s}_0, \sigma_{02})$  are the same as presented priors in section 4. We have also used function (10) as the likelihood function.

The following are the steps in the Gibbs sampling algorithm.

### Gibbs Sampling Algorithm

- Randomly generate  $(\pi^{(1)}, \mathbf{z}^{(1)}, \xi_1^{(1)}, \xi_2^{(1)}, \mathbf{t}_0^{(1)}, \mathbf{s}_0^{(1)}, \sigma_{01}^{(1)}, \sigma_{02}^{(1)})$
- Where  $T$  is the number of iterations, for  $t = 1, \dots, T$ , do the following:
  1. Draw  $\pi^{(t+1)} \sim p(\pi | \mathbf{z}^{(t)}, \xi_1^{(t)}, \xi_2^{(t)}, \mathbf{t}_0^{(t)}, \mathbf{s}_0^{(t)}, \sigma_{01}^{(t)}, \sigma_{02}^{(t)}, data)$
  2. Draw  $z^{(t+1)} \sim P(z_n | \mathbf{z}_{-n}^{(t)}, \xi_1^{(t)}, \xi_2^{(t)}, \mathbf{t}_0^{(t)}, \mathbf{s}_0^{(t)}, \sigma_{01}^{(t)}, \sigma_{02}^{(t)}, \pi^{(t+1)}, data)$
  3. Draw  $(\xi_1, \xi_2)^{(t+1)} \sim P(\boldsymbol{\xi} | \mathbf{t}_0^{(t)}, \mathbf{s}_0^{(t)}, \sigma_{01}^{(t)}, \sigma_{02}^{(t)}, \pi^{(t+1)}, \mathbf{z}^{(t+1)}, data)$
  4. Draw  $\mathbf{t}_0^{(t+1)} \sim P(\mathbf{t}_0 | \mathbf{s}_0^{(t)}, \sigma_{01}^{(t)}, \sigma_{02}^{(t)}, \pi^{(t+1)}, \mathbf{z}^{(t+1)}, \boldsymbol{\xi}^{(t+1)}, data)$
  5. Draw  $\mathbf{s}_0^{(t+1)} \sim P(\mathbf{s}_0 | \sigma_{01}^{(t)}, \sigma_{02}^{(t)}, \pi^{(t+1)}, \mathbf{z}^{(t+1)}, \boldsymbol{\xi}^{(t+1)}, \mathbf{t}_0^{(t+1)}, data)$
  6. Draw  $(\sigma_{01}, \sigma_{02})^{(t+1)} \sim P(\boldsymbol{\sigma}_0 | \pi^{(t+1)}, \mathbf{z}^{(t+1)}, \boldsymbol{\xi}^{(t+1)}, \mathbf{t}_0^{(t+1)}, \mathbf{s}_0^{(t+1)}, data)$

- By the theorem of Markov Chain, for a sufficiently large  $t$ ,

$(\pi^{(t)}, \mathbf{z}^{(t)}, \xi_1^{(t)}, \xi_2^{(t)}, \mathbf{t}_0^{(t)}, \mathbf{s}_0^{(t)}, \sigma_0^{(t)})$  can be viewed as a random sample of posterior  $p(\cdot \mid data)$ .

In this part, we want to obtain the forms of different conditional posterior distributions.

$$1. P(\pi \mid \mathbf{z}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, data) \propto \prod_{i=1}^n (\pi)^{z_i} (1 - \pi)^{(1-z_i)} \cdot (\pi)^{\alpha_1-1} (1 - \pi)^{\alpha_2-1}$$

$$P(\pi \mid \mathbf{z}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, data) \equiv \text{Beta} \left( \alpha_1 + \sum_{i=1}^n z_i, \alpha_2 + n - \sum_{i=1}^n z_i \right)$$

$$2. P(z_i \mid \mathbf{z}_{(-i)}, \pi, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, data) \propto \exp \left\{ z_i [\xi_1 x_i(\mathbf{t}_0, \sigma_{01}) - \xi_1^2/2] \right\} \\ \times \exp \left\{ (1 - z_i) [\xi_2 x_i(\mathbf{s}_0, \sigma_{02}) - \xi_2^2/2] \right\} \\ \times (\pi)^{z_i} (1 - \pi)^{z_i-1}$$

$$P(z_i \mid \mathbf{z}_{(-i)}, \pi, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, data) \propto (\pi f_1)^{z_i} [(1 - \pi) f_2]^{(1-z_i)}$$

$$P(z_i \mid \mathbf{z}_{(-i)}, \pi, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, data) \equiv \text{Bern} \left( \frac{\pi f_1}{\pi f_1 + (1 - \pi) f_2} \right)$$

$$3. P(\xi_1 \mid \mathbf{t}_0, \sigma_{01}, \boldsymbol{\theta}_2, \pi, \mathbf{z}, data) \propto \prod_{i=1}^n e^{\{z_i [\xi_1 x_i(\mathbf{t}_0, \sigma_{01}) - \xi_1^2/2]\}} \cdot P(\xi_1) \cdot I_{(\xi_1 > 0)}$$

$$P(\xi_1 \mid \mathbf{t}_0, \sigma_{01}, \boldsymbol{\theta}_2, \pi, \mathbf{z}, data) \equiv TN \left( \frac{\sum_{i=1}^n z_i x_i(\mathbf{t}_0, \sigma_{01})}{\sum_{i=1}^n z_i}, \frac{1}{\sum_{i=1}^n z_i} \right)$$

$$P(\xi_2 \mid \mathbf{s}_0, \sigma_{02}, \boldsymbol{\theta}_1, \pi, \mathbf{z}, data) \propto \prod_{i=1}^n e^{\{z_i [\xi_2 x_i(\mathbf{s}_0, \sigma_{02}) - \xi_2^2/2]\}} \cdot P(\xi_2) I_{(\xi_2 > 0)}$$

$$P(\xi_2 \mid \mathbf{s}_0, \sigma_{02}, \boldsymbol{\theta}_1, \pi, \mathbf{z}, data) \equiv TN \left( \frac{\sum_{i=1}^n (1 - z_i) x_i(\mathbf{s}_0, \sigma_{02})}{\sum_{i=1}^n (1 - z_i)}, \frac{1}{\sum_{i=1}^n (1 - z_i)} \right)$$

The above truncated normal distributions are truncated at  $\xi_j \geq 0$  for  $j = 1, 2$ .

4. Because there is not any closed forms for the following conditional posterior, we will use Metropolis-Hasting method to generate  $\mathbf{t}_0$ ,  $\mathbf{s}_0$ ,  $\sigma_{01}$ ,  $\sigma_{02}$ , from following distributions.

$$P(\mathbf{t}_0 \mid \xi_1, \sigma_{01}, \boldsymbol{\theta}_2, \pi, \mathbf{z}, data),$$

$$P(\mathbf{s}_0 \mid \xi_2, \sigma_{02}, \boldsymbol{\theta}_1, \pi, \mathbf{z}, data),$$

$$P(\sigma_{01} \mid \xi_1, \mathbf{t}_0, \boldsymbol{\theta}_2, \pi, \mathbf{z}, data),$$

$$P(\sigma_{02} \mid \xi_2, \mathbf{s}_0, \boldsymbol{\theta}_1, \pi, \mathbf{z}, data).$$

The choice of proposal distribution used in Metropolis-Hasting sampler is based on the choice of priors for  $\mathbf{t}_0$ ,  $\mathbf{s}_0$ ,  $\sigma_{01}$ ,  $\sigma_{02}$ . Thus, the chosen proposal distribution is:

$$Q(t^*, \sigma_0^*) = \frac{\beta_2^{\beta_1}}{\Gamma(\beta_1)} \sigma_0^{-(\beta_1+1)} e^{-\beta_2/\sigma_0} \frac{\exp\left\{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right\}}{\int_{\mathbf{a}}^{\mathbf{b}} \exp\left\{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right\} d\mathbf{x}}$$

where,  $\beta_1 = 4$ ,  $\beta_2 = 0.5$ ,  $\mathbf{a} = (0, 0)^T$ ,  $\mathbf{b} = (1, 1)^T$ ,  $\boldsymbol{\mu} = (0.5, 0.5)^T$ , and  $\boldsymbol{\Sigma} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ .

Results of the above Gibbs and Metropolis-Hastings samplers are given in Chapter IV.

### Simulation Study

In this section, we apply the theoretical results of the previous sections on a data set obtained by simulation for *fMRI* images of the human brain. We use the results to

provide the simulation scheme for the numerical study on the model (1) in a two-dimensional case. Our simulation is based on the Monte Carlo method, and we will employ R packages to run the simulations. The image resolution we plan to have is  $64 \times 64$ , which is commonly used in practice. We need to generate  $n$  images assuming each is for one subject. We will consider two values of 10 and 20 for  $n$  to generate various series of images for each analysis under different parameter schemes. As before, the null hypothesis is  $\xi = 0$ , or equivalently there is no signal, and the alternative hypothesis is  $\xi > 0$  for at least one of the two random fields. To do the hypothesis testing, we apply the Bayes factor approach for the scale-space model. For spatial smoothing of images, a Gaussian-shaped kernel is used before any analysis. The Gaussian-shaped kernel is centered to ensure that the entire image is evenly smooth.

Here, we assume that our images are realizations of a mixture of two homogeneous Gaussian random fields. We suppose that each component of the mixture has a signal with the form of a known function, here a Gaussian form, centered at an unknown location  $t_0$ , and multiplied by an unknown amplitude  $\xi$ . We also assume that the parameter of scale or width of signal,  $\sigma_0$  is not known.

Under the alternative hypothesis, the simulated images contain not just the noise but also the signals. Therefore, images with signals under various schemes are generated from model (1). The signals are manipulated through the following parameters:

- (a) weight in the mixture ( $\pi$ ), with two levels:  $\pi = 0.3, 0.5$ . By this, we consider two different scenarios for model (1).

- (b) amplitude ( $\xi$ ), where there are three levels:  $\xi = 0, 1, 2.5$ . With two random fields in the model, the amplitudes of two signals,  $\xi_1$  and  $\xi_2$ , could be different resulting in six combinations of  $\xi_1$  and  $\xi_2$  as displayed in Table 2
- (c) distance, where there are two levels: first, the signals close to each other (Near), and second, the signals far from each other (Far). The operational definition of these two levels is illustrated in the following subsection.
- (d) scale ( $\sigma_0$ ), where there are two levels: 0.2, 0.4. With two random fields in the model, we would have 3 combinations for  $\sigma_{01}$  and  $\sigma_{02}$ . However, to avoid having too many scenarios, we have chosen one combination of ( $\sigma_{01} = 0.2, \sigma_{02} = 0.4$ ).

### **Choices of the Levels of the Parameters**

Because there is not any previous simulation study on this model, as the first investigation of this topic, the levels of the above model parameters (weight, amplitude, distance, and scale) are generally chosen based on the previous studies with signals from one Gaussian random field. Coordinates of two signals,  $(t_{01}, t_{02})$  and  $(s_{01}, s_{02})$ , with two levels are:

Near: (0.5,0.5) and (0.7,0.7)

Far: (0.1,0.6) and (0.6,0.1)

In previous research by (Shafie et al., 2003; Siegmund & Worsley, 1995), for the values of amplitude and scale in the simulation,  $\xi = 6$  and  $\sigma_0 = 1$  are used. However, these values are not applicable here because with all the simulated images being unit

squares in the current study, a signal with  $\xi = 6$  and  $\sigma_0 = 1$  is too big to be contained nicely within a unit square. Thus, more reasonable values for amplitude and scale are chosen instead.

### **Simulation Scheme**

In summary, regarding two values for  $n$  and all combinations of parameters, a total of 48 schemes are considered for the simulation study. The simulation scheme is displayed in Table 2. Through the simulation, first, we estimate  $\pi$  in the model (1) through Gibbs and Metropolis-Hasting samplers. Furthermore, we obtain the Bayes factor using Monte Carlo simulation. After computing the Bayes factors under different schemes, we will use the common grading of the Bayes factor given in Table 1 to reject or accept the existence of the signal. Moreover, the power of the test will be calculated for each scheme.

**Table 2**

*Schemes of the Parameters where  $\sigma_{01} = 0.2$ ,  $\sigma_{02} = 0.4$ ,*

No.	$\pi$	$\xi_1$	$\xi_2$	Distance
1	0.3	0	0	Near
2		0	1	
3		0	2.5	
4		1	1	
5		1	2.5	
6		2.5	2.5	
7	0.3	0	0	Far
8		0	1	
9		0	2.5	
10		1	1	
11		1	2.5	
12		2.5	2.5	
13	0.5	0	0	Near
14		0	1	
15		0	2.5	
16		1	1	
17		1	2.5	
18		2.5	2.5	
19	0.5	0	0	Far
20		0	1	
21		0	2.5	
22		1	1	
23		1	2.5	
24		2.5	2.5	

## CHAPTER IV

### RESULTS

In this chapter, we give the results of the simulation study for all parameter schemes. Bayes factors are obtained through Monte Carlo simulation from the prior distribution for a mixture of two scale-space Gaussian random fields, and results are displayed in Tables 3 and 4. The number of iterations considered for each simulation was 1000. Moreover, the number of simulations for each scheme was chosen to be 1000 to obtain the power of the test. Even though, estimating parameters in the model was not required for testing, we have estimated all parameters  $\pi$ ,  $\xi_1$ ,  $\xi_2$ ,  $t_0$ ,  $s_0$ ,  $\sigma_{01}$ ,  $\sigma_{02}$  in the model (1). Estimation results obtained through Gibbs and Metropolis-Hasting samplers are given in Table 8 for several schemes. The Appendix includes the R code used for the simulation study.

#### **Result for the Bayes Factor**

Tables 3 and 4 show the Bayes factors corresponding to different parameter schemes under the selected prior distributions of  $(\pi, \theta_1, \theta_2)$  in Chapter *III*. These results are obtained through the Monte Carlo method and by applying the formula in (13).

We should recall that in the hypothesis testing in which the Bayes factor is being used as a criterion, a large value of the Bayes factor is evidence in support of  $H_0$ .

In this simulation study, values of all obtained Bayes factors are between  $1.2 \times 10^{-87}$ , and 0.995. And, the maximum observed Bayes factors in all simulated

images including images with no signals are 0.995. Also, to uncover the behavior of the Bayes factor in our model, empirical distributions of Bayes factor under different parameter schemes are displayed in *Figures 5 – 10*. With regard to the findings, the grading system in Table 1 cannot be employed to make a decision about testing  $H_0 : \xi = 0$ . Essentially, Jeffreys' grading system implicitly assumes values for the Bayes factor in the finite dimensions, so this grading system is invalid to be applied in infinite-dimensional cases. Our justification for observing small values of the Bayes factor in the model (1) is the nature of the likelihood function in (8). Consequently, we need to choose a proper grading system for the Bayes factor in such a study.

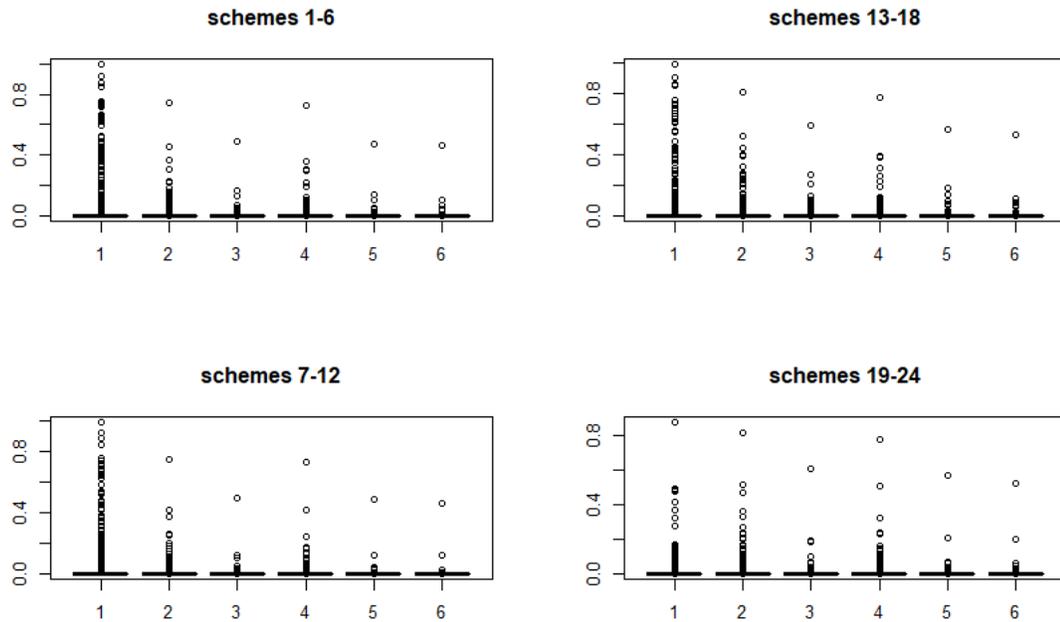
In Figure (4), different box-plots show that the larger values of Bayes factor are mostly related to schemes in which  $\xi_1$ , and  $\xi_2$  are 0. For more details about the range of Bayes factors, tables 5, 6, and 7 are provided in Appendix for several parameter schemes.

Regarding the provided information about the values of the Bayes factor, we chose a threshold for decision-making. The choice of threshold was based on the empirical distributions of the Bayes factor for the schemes in which images are realizations of a mixture model with components being Gaussian random fields with no signals. For observed Bayes factors in schemes 1, 7, 13, and 19, different percentiles are presented in the following table. We have chosen the 75<sup>th</sup> percentile as a threshold to distinguish the large values of the Bayes factor in support of the  $H_0$ . Although we made a justification, this choice is whatsoever arbitrary.

65%	75%	85%	95%
$2.27 \times 10^{-5}$	$3.68 \times 10^{-4}$	$7.58 \times 10^{-3}$	$1.33 \times 10^{-1}$

**Figure 4**

*Boxplots of Bayes Factor with respect to Parameter Scheme*



Therefore,

$$\begin{cases} \text{if } B > 3.68 \times 10^{-4} & \text{accept the } H_0 \\ \text{if } B \leq 3.68 \times 10^{-4} & \text{reject the } H_0 \end{cases}$$

Values of observed Bayes factors, for different parameter schemes and their corresponding power of test are displayed in Tables 3 and 4. The result of the decision-making is based on the chosen threshold,  $3.68 \times 10^{-4}$ . The threshold was used to calculate the power. As illustrated in Chapter III, there are 24 parameter schemes manipulating different factors of  $n$  simulated images, including  $\pi$ , amplitude, and distance. The discussion about these results are provided in Chapter V.

**Table 3**

The Bayes Factor and its power where  $\sigma_{01} = 0.2$ ,  $\sigma_{02} = 0.4$ ,  $n = 10$

scheme	$\pi$	$\xi_1$	$\xi_2$	Distance	Bayes Factor	Decision	Power%
1	0.3	0	0	Near	$2.66 \times 10^{-2}$	Accept $H_0$	29
2		0	1		$2.29 \times 10^{-6}$	Reject $H_0$	88
3		0	2.5		$4.48 \times 10^{-15}$	Reject $H_0$	96
4		1	1		$5.54 \times 10^{-7}$	Reject $H_0$	90
5		1	2.5		$7.99 \times 10^{-15}$	Reject $H_0$	97
6		2.5	2.5		$5.2 \times 10^{-15}$	Reject $H_0$	97
7	0.3	0	0	Far	$2.86 \times 10^{-2}$	Accept $H_0$	30
8		0	1		$4.61 \times 10^{-6}$	Reject $H_0$	90
9		0	2.5		$1.56 \times 10^{-13}$	Reject $H_0$	97
10		1	1		$3.66 \times 10^{-6}$	Reject $H_0$	91
11		1	2.5		$1.94 \times 10^{-13}$	Reject $H_0$	97
12		2.5	2.5		$5.84 \times 10^{-15}$	Reject $H_0$	98
13	0.5	0	0	Near	$2.79 \times 10^{-2}$	Accept $H_0$	30
14		0	1		$6.44 \times 10^{-6}$	Reject $H_0$	86
15		0	2.5		$1.94 \times 10^{-12}$	Reject $H_0$	95
16		1	1		$1.43 \times 10^{-6}$	Reject $H_0$	90
17		1	2.5		$1.81 \times 10^{-13}$	Reject $H_0$	96
18		2.5	2.5		$2.17 \times 10^{-14}$	Reject $H_0$	97
19	0.5	0	0	Far	$2.73 \times 10^{-9}$	Reject $H_0$	12
20		0	1		$2.47 \times 10^{-5}$	Reject $H_0$	88
21		0	2.5		$1.63 \times 10^{-11}$	Reject $H_0$	96
22		1	1		$3.52 \times 10^{-5}$	Reject $H_0$	90
23		1	2.5		$3.43 \times 10^{-12}$	Reject $H_0$	97
24		2.5	2.5		$1.44 \times 10^{-11}$	Reject $H_0$	97

**Table 4**

*The Bayes Factor and its power where  $\sigma_{01} = 0.2$ ,  $\sigma_{02} = 0.4$ ,  $n = 20$*

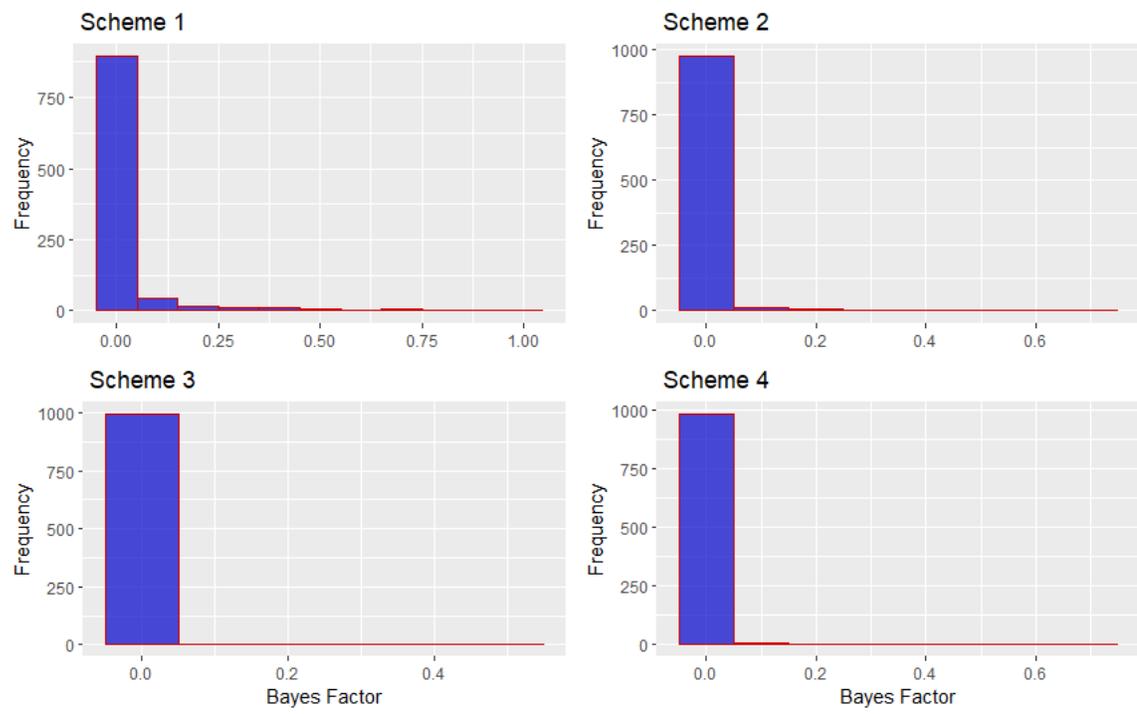
scheme	$\pi$	$\xi_1$	$\xi_2$	Distance	Bayes Factor	Decision	Power %
1	0.3	0	0	Near	$1.6 \times 10^{-4}$	Reject $H_0$	15
2		0	1		$5.8 \times 10^{-12}$	Reject $H_0$	96
3		0	2.5		$2.9 \times 10^{-28}$	Reject $H_0$	99
4		1	1		$3.8 \times 10^{-13}$	Reject $H_0$	97
5		1	2.5		$2 \times 10^{-28}$	Reject $H_0$	99
6		2.5	2.5		$9.4 \times 10^{-29}$	Reject $H_0$	99
7	0.3	0	0	Far	$1.3 \times 10^{-3}$	Accept $H_0$	15
8		0	1		$2.8 \times 10^{-11}$	Reject $H_0$	96
9		0	2.5		$6.0 \times 10^{-25}$	Reject $H_0$	99
10		1	1		$4.2 \times 10^{-12}$	Reject $H_0$	97
11		1	2.5		$6.3 \times 10^{-25}$	Reject $H_0$	99
12		2.5	2.5		$1.6 \times 10^{-23}$	Reject $H_0$	99
13	0.5	0	0	Near	$1.4 \times 10^{-11}$	Reject $H_0$	3
14		0	1		$2.6 \times 10^{-10}$	Reject $H_0$	94
15		0	2.5		$2.7 \times 10^{-22}$	Reject $H_0$	98
16		1	1		$8.7 \times 10^{-13}$	Reject $H_0$	97
17		1	2.5		$3.7 \times 10^{-25}$	Reject $H_0$	99
18		2.5	2.5		$8.8 \times 10^{-28}$	Reject $H_0$	99
19	0.5	0	0	Far	$1.2 \times 10^{-3}$	Accept $H_0$	16
20		0	1		$1.3 \times 10^{-28}$	Reject $H_0$	99
21		0	2.5		$6.5 \times 10^{-22}$	Reject $H_0$	99
22		1	1		$9.3 \times 10^{-10}$	Reject $H_0$	97
23		1	2.5		$1.5 \times 10^{-21}$	Reject $H_0$	99
24		2.5	2.5		$6.5 \times 10^{-22}$	Reject $H_0$	99

**More Figures and Tables**

Following figures and tables include more details about distributions of Bayes factors under different parameter schemes. These information are used to choose the threshold for decision making.

**Figure 5**

*Empirical Bayes Factor Distribution for Scheme 1-4*



**Table 5***Five-Number Summary of Bayes Factor For Schemes 1-4*

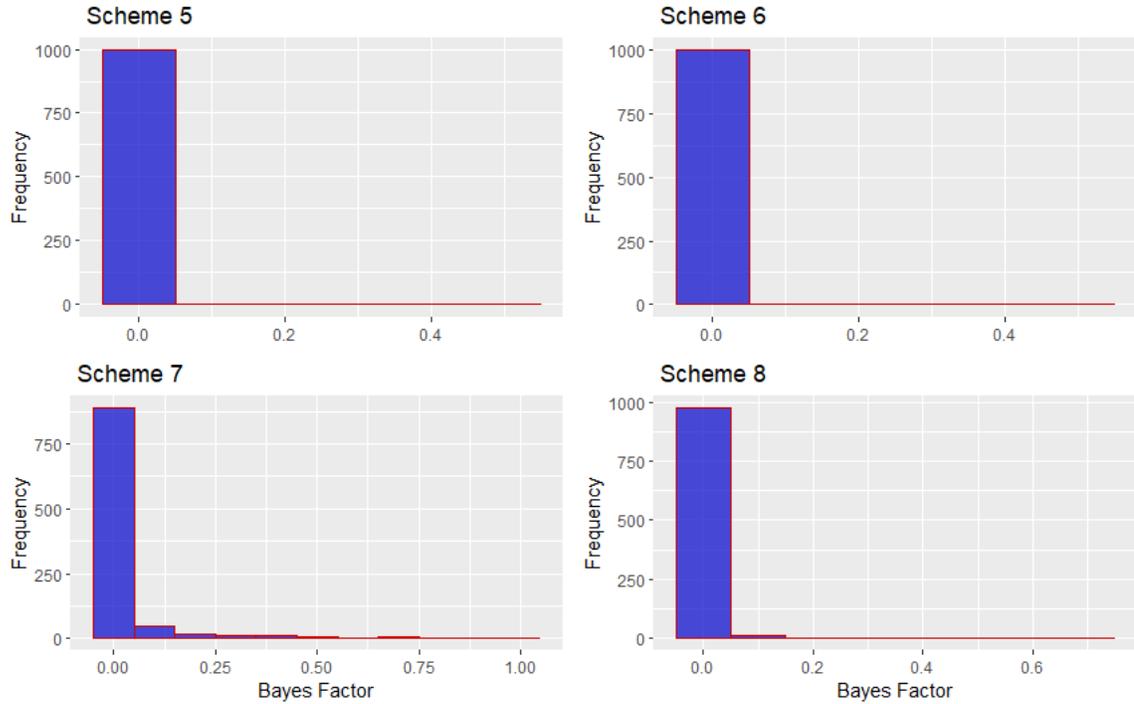
Scheme 1	Scheme 2	Scheme 3	Scheme 4
Min. :0.0000000	Min. :0.0000000	Min. :0.0000000	Min. :0.0000000
1st Qu.:0.0000000	1st Qu.:0.0000000	1st Qu.:0.0000000	1st Qu.:0.0000000
Median :0.0000034	Median :0.0000000	Median :0.0000000	Median :0.0000000
Mean :0.0318689	Mean :0.0055949	Mean :0.001194	Mean :0.0038750
3rd Qu.:0.0011446	3rd Qu.:0.0000015	3rd Qu.:0.0000000	3rd Qu.:0.0000002
Max. :0.9948877	Max. :0.7464239	Max. :0.487341	Max. :0.7284476

**Table 6***Five-Number Summary of Bayes Factors For Schemes 5-8*

Scheme 5	Scheme 6	Scheme 7	Scheme 8
Min. :0.0000000	Min. :0.0000000	Min. :0.0000000	Min. :0.0000000
1st Qu.:0.0000000	1st Qu.:0.0000000	1st Qu.:0.0000000	1st Qu.:0.0000000
Median :0.0000000	Median :0.0000000	Median :0.0000033	Median :0.0000000
Mean :0.001039	Mean :0.0008439	Mean :0.0320692	Mean :0.0046506
3rd Qu.:0.0000000	3rd Qu.:0.0000000	3rd Qu.:0.0013094	3rd Qu.:0.0000003
Max. :0.473745	Max. :0.4638732	Max. :0.9860340	Max. :0.7489678

**Figure 6**

*Empirical Bayes Factor Distribution for Scheme 5-8*



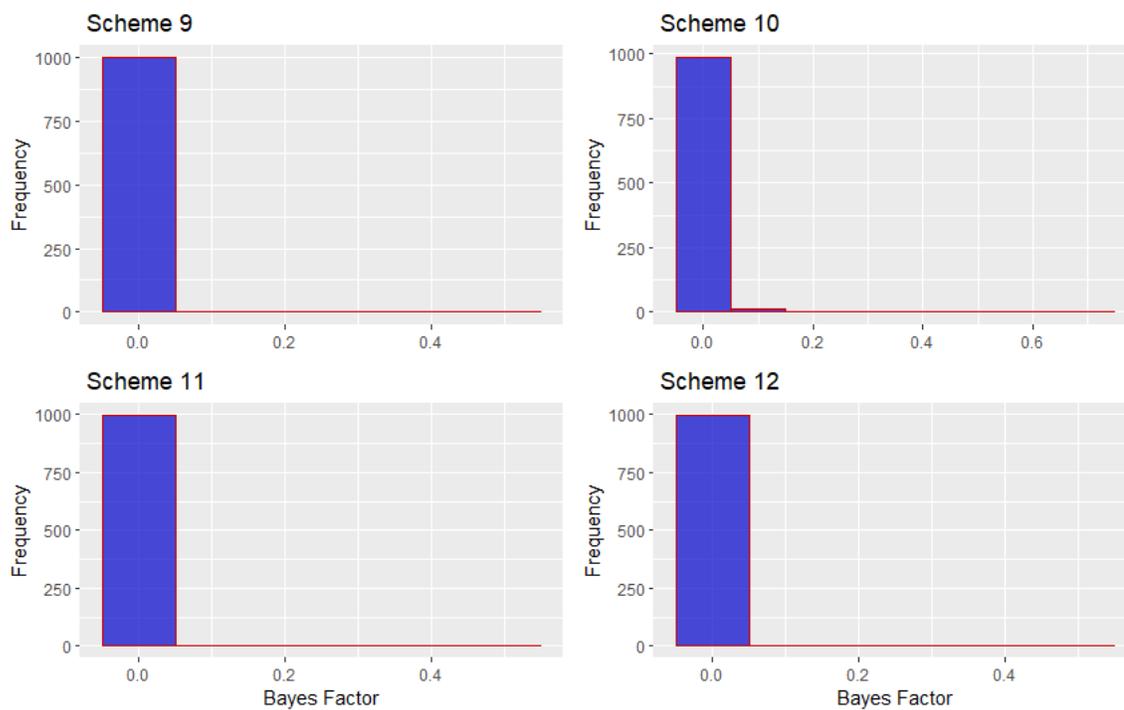
### **Result for Parameters Estimation**

Table 8 shows the estimated parameters corresponding several different parameter schemes. To estimate parameters, we obtained samples from the posterior

$P(\pi, \xi_1, \xi_2, \mathbf{t}_0, \mathbf{s}_0, \sigma_{01}, \sigma_{02} \mid Data)$  through Gibbs and Metropolis-Hasting samplers.

We have employed the likelihood function (10), and prior distributions in Chapter *III*.

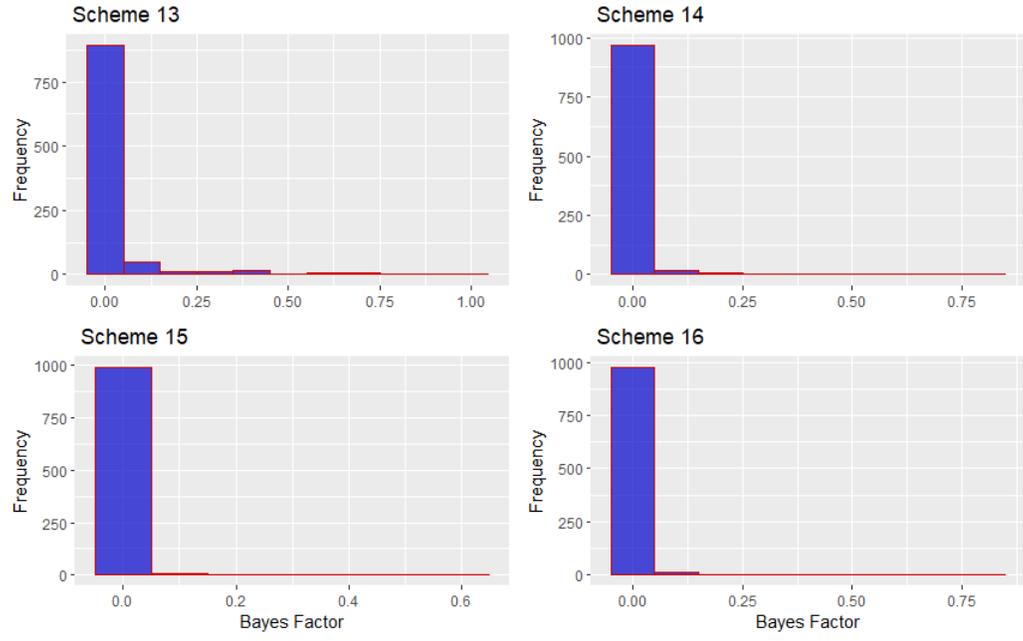
The discussion on results in this table is provided in Chapter *V*.

**Figure 7***Empirical Bayes Factor Distribution for Scheme 9-12***Table 7***Five-Number Summary of Bayes Factors For Schemes 9-12*

Scheme 9	Scheme 10	Scheme 11	Scheme 12
Min. :0.0000000	Min. :0.0000000	Min. :0.0000000	Min. :0.0000000
1st Qu.:0.0000000	1st Qu.:0.0000000	1st Qu.:0.0000000	1st Qu.:0.0000000
Median :0.0000000	Median :0.0000000	Median :0.0000000	Median :0.0000000
Mean :0.0009265	Mean :0.0033190	Mean :0.0007734	Mean :0.0006803
3rd Qu.:0.0000000	3rd Qu.:0.0000001	3rd Qu.:0.0000000	3rd Qu.:0.0000000
Max. :0.4952471	Max. :0.7272340	Max. :0.4811155	Max. :0.4609896

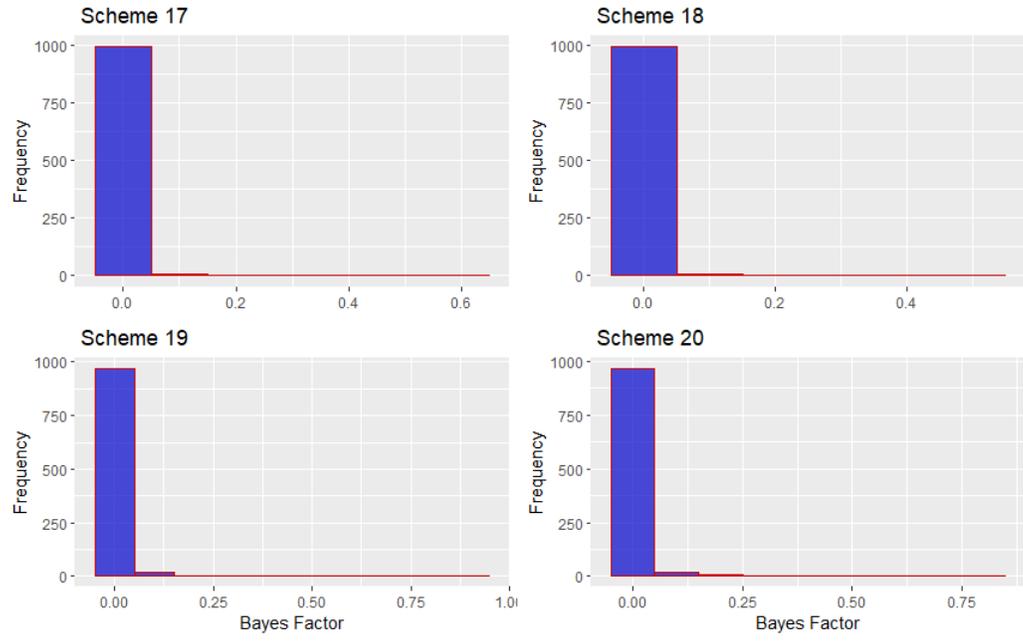
**Figure 8**

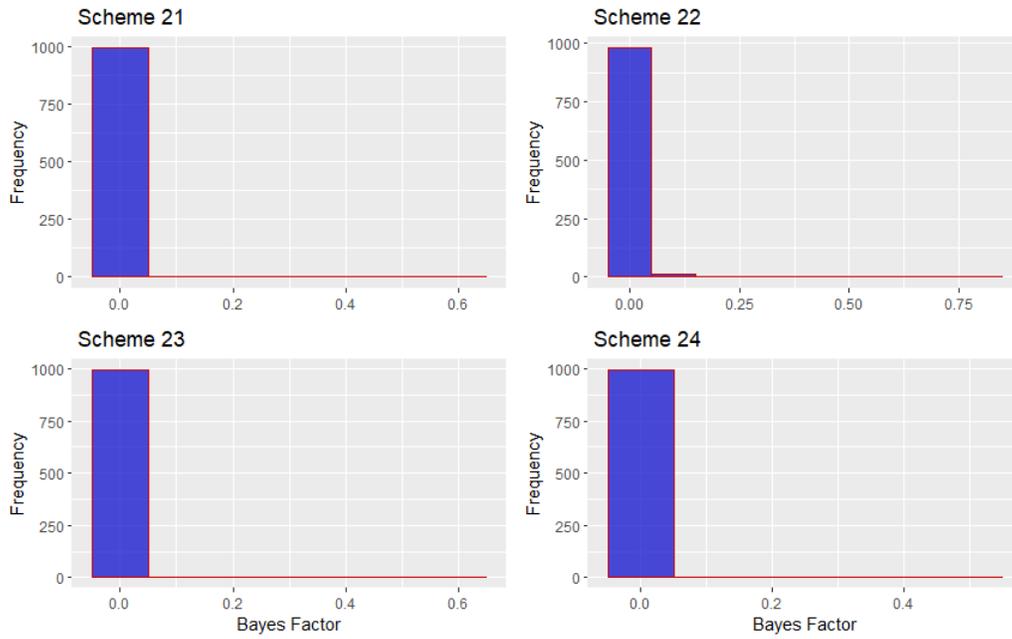
*Empirical Bayes Factor Distribution for Scheme 13-16*



**Figure 9**

*Empirical Bayes Factor Distribution for Scheme 17-20*



**Figure 10***Empirical Bayes Factor Distribution for Scheme 21-24***Table 8***Parameter estimates where  $\sigma_{01} = 0.2$ ,  $\sigma_{02} = 0.4$ ,  $n = 10$* 

$\pi$	$\hat{\pi}$	$\xi_1, \xi_2$	$\hat{\xi}_1, \hat{\xi}_2$	Distance	$\hat{t}_{01}, \hat{t}_{02}$	$\hat{s}_{01}, \hat{s}_{02}$	$\hat{\sigma}_{01}, \hat{\sigma}_{02}$
0.3	0.75	(0, 1)	(1.9, 0.6)	Near	(0.56, 0.64)	(0.62, 0.69)	(0.42, 0.38)
0.3	0.74	(2.5, 2.5)	(2.8, 1.8)	Near	(0.52, 0.49)	(0.49, 0.48)	(0.44, 0.44)
0.3	0.75	(2.5, 2.5)	(2.9, 1.9)	Far	(0.12, 0.59)	(0.12, 0.59)	(0.46, 0.45)
0.5	0.75	(0, 2.5)	(2.9, 1.9)	Near	(0.66, 0.64)	(0.67, 0.66)	(0.43, 0.42)
0.5	0.78	(2.5, 2.5)	(2.9, 1.8)	Far	(0.74, 0.17)	(0.59, 0.16)	(0.45, 0.43)

## CHAPTER V

### CONCLUSIONS

The purpose of the current study was to propose a Bayesian approach to the Gaussian mixture model (1) to detect a signal in fMRI data. The problem of detecting signals is statistically equivalent to hypothesis testing. To test the existence of signal, we developed the Bayesian testing approach which Shafie et al. (2003) proposed for signal detection using the notion of Bayes factor,  $B(x)$ . For infinite-dimensional parameter space, they defined the Bayes factor based on the concept of the Radon-Nikodym derivative. We developed their Bayes factor definition to obtain a Bayesian criteria for testing the proposed model (1). Finding the Bayes factor in an infinite-dimensional case is not analytically tractable, and we needed to compute it through numerical methods. In this study, under 48 schemes of the model parameters, two-dimensional images were simulated. To obtain  $B(x)$  for each set of images, the formula in (13) was applied.

Assuming parameters  $\pi$ ,  $\xi_1$ ,  $\xi_2$ ,  $\mathbf{t}_0$ ,  $\mathbf{s}_0$ ,  $\sigma_{01}$ ,  $\sigma_{02}$  are independent in prior, our choice of priors were as follows:

- $\pi \mid \alpha \sim Beta(\alpha_1, \alpha_2)$ ,  $\alpha_1 = 2$ ,  $\alpha_2 = 2$
- $\sigma_{01}, \sigma_{02} \stackrel{iid}{\sim} Inv\text{-Gamma}(\beta_1, \beta_2)$ ,  $\beta_1 = 4$ ,  $\beta_2 = 0.5$
- $\mathbf{t}_0, \mathbf{s}_0 \stackrel{iid}{\sim} TN(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{a}, \mathbf{b})$
- $\xi_1, \xi_2 \stackrel{iid}{\sim} Unif(0, 5)$

We used the bivariate truncated Normal as prior on  $\mathbf{t}_0, \mathbf{s}_0$  with  $\boldsymbol{\mu} = (0.5, 0.5)^T$  and covariance matrix  $\boldsymbol{\Sigma} = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$ , where the lower and upper truncation points were  $\mathbf{a} = (0, 0)^T, \mathbf{b} = (1, 1)^T$ .

To approximate the integrals in (13), after taking 1000 samples of  $\boldsymbol{\theta} = (\pi, \xi_1, \xi_2, \mathbf{t}_0, \mathbf{s}_0, \sigma_{01}, \sigma_{02})$  from prior distributions, Monte Carlo method was employed. In addition, via 1000 simulations for each parameter scheme, the power of testing  $H_0 : \boldsymbol{\xi} = \mathbf{0}$ , against  $H_a = \boldsymbol{\xi} > \mathbf{0}$  was obtained. The result of computing Bayes factors and test-powers were displayed in Chapter *IV* for all 48 parameter schemes (Tables 3 and 4).

### Findings and Discussion

For this work, we had to answer two main questions. The first was how to develop a Bayesian procedure for testing the signal in the model (1) for two-dimensional images. The second was how the performance of the Bayes factor can change with respect to different parameter schemes. In Chapter *III*, we developed a Bayesian approach to model (1), and in Chapter *IV*, we illustrated our approach by analysis of the simulated data. To answer the first question, we found Bayes factors for all schemes and by them, we decided to reject or accept  $H_0$ . To answer the second question, we carried out a simulation study to obtain the powers of the test corresponding to each scheme.

Here, we should notice that our choice of threshold for decision-making about testing  $H_0$ , where the Bayes factor used as a criterion is somehow arbitrary. Regarding information in Tables 3 and 4, it is obvious that this choice of threshold resulted in very low power of the test for schemes with no signals, and large powers for ones with signals.

Here we can discuss that the choice of power of the test can be considered as an important factor for choosing a threshold to make a decision on accepting/rejecting the  $H_0$ .

Furthermore, Tables 3 and 4 show that the proposed Bayesian approach for testing the signal is successful for all the schemes in which at least one of the components in the model (1) is a Gaussian random field with a non-zero mean function.

We illustrated in Chapter *III* that there are 24 parameter schemes for each scenario of  $n$  manipulating different factors of simulated images, including  $\pi$ , amplitude, and distance. Information in Tables 3 and 4 show the effect of  $\pi$ , amplitude, and distance on the power of the Bayes factor. It is seen that the higher powers are mostly for the schemes in which  $\pi = 0.3$ , and the centers of signals in two random fields are far from each other. Given that, the effect of amplitudes on the power is of course clear for all schemes.

### **Limitations and Suggestions for Future Research**

We found that the Bayesian approach of signal detection within noisy images when the image is modeled as a mixture of two Gaussian scale-space random fields is a suitable procedure. However, we had some limitations of the above analysis that must be noticed.

To evaluate our Bayesian approach to the proposed model (1), we carried out a simulation study by generating simulated data. But, the results of this study should be developed for the application on real fMRI data. For instance, in the cases that the shape of signals are not known, smoothing with the kernel of  $\sigma^{-N/2}k[\sigma^{-1}(\mathbf{h} - \mathbf{t})]$  is not

justifiable, and so we must apply the non-smooth random fields such as the one satisfying (19).

As it illustrated in Chapter *IV*, the grading system in Table 1 cannot be employed to make a decision about testing  $H_0 : \xi = \mathbf{0}$ . Therefore, we need to construct a proper grading system for the Bayes factor in this type of study. Even though we somehow used the information from empirical distributions of the Bayes factor to find the threshold for testing, our choice was arbitrary. As it was seen in Tables 3 and 4, the study of the power of the Bayes factor can help researchers to construct a proper grading system for these types of study.

The Bayes factor proposed in (13) was only applied to a mixture model with two Gaussian scale-space random fields. However, due to its abstract generality, the proposed Bayes factor can be applied to the mixture models with more components.

Although the estimation of parameters in the model (1) was not the purpose of this study, the Gibbs and Metropolis samplers were employed to obtain the empirical posterior distribution  $P(\pi, \xi_1, \xi_2, \mathbf{t}_0, \mathbf{s}_0, \sigma_{01}, \sigma_{02} \mid Data)$ . By sampling from the posterior distribution, we estimated  $\pi, \xi_1, \xi_2, \mathbf{t}_0, \mathbf{s}_0, \sigma_{01}, \sigma_{02}$  for some of schemes and results are displayed in Table(8). As it is seen, the estimation for parameters  $\pi, \xi_1, \xi_2$  was not successful. However, results of estimation for parameters  $\mathbf{t}_0, \mathbf{s}_0, \sigma_{01}, \sigma_{02}$  seem pretty decent. This part of our work can be a field for the future studies. Of course applying machine learning methods might increase the level of success in estimating parameters of model (1).

## REFERENCES

- Adler, R. J. (1981). *The geometry of random fields* (1st ed.). New York: Wiley.
- Adler, R. J. (2010). *The geometry of random fields*. SIAM.
- Ash, R. (1972). *Probability and real analysis*. John Wiley & Sons, New York.
- Aubin, J. (1977). *Abstract and applied analysis*. Wiley-Interscience, New York.
- Bender, R., & Lange, S. (2001). Adjusting for multiple testing: When and how? *Journal of Clinical Epidemiology*, *54*, 343–349.
- Bennett, C. M., Wolford, G. L., & Miller, M. B. (2009). The principled control of false positives in neuroimaging. *Social Cognitive and Affective Neuroscience*, *4*(4), 417–422. doi: 10.1093/scan/nsp053
- Berger, J. O. (2013). *Statistical decision theory and bayesian analysis*. Springer Science & Business Media.
- Forghaniarani, M., & Shafie, K. (2020). *Identifiability of finite mixtures of truncated poisson regression models*. Under Review.
- Geman, S., & Geman, D. (1984). Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on pattern analysis and machine intelligence*(6), 721–741.
- Green, P. J. (2019). Introduction to finite mixtures. In *Handbook of mixture analysis* (pp. 3–20). Chapman and Hall/CRC.
- Grenander, U. (1981). *Abstract inference*. Wiley-New York.

- Hasofer, A. (1978). Upcrossings of random fields. *Advances in Applied Probability*, 10, 14–21.
- Huang, M., Li, R., Wang, H., & Yao, W. (2014). Estimating mixture of gaussian processes by kernel smoothing. *Journal of Business & Economic Statistics*, 32(2), 259–270.
- Jeffreys, H. (1998). *The theory of probability*. OUP Oxford.
- Parzen, E. (1961). An approach to time series analysis. *The Annals of Mathematical Statistics*, 32(4), 951–989.
- Pauling, L., & Coryell, C. D. (1936). The magnetic properties and structure of hemoglobin, oxyhemoglobin and carbonmonoxyhemoglobin. *Proceedings of the National Academy of Sciences*, 22(4), 210–216.
- Rohani, M. F. (2003). *Bayesian approach to the Gaussian scale space random fields*. Unpublished doctoral dissertation. Shahid Beheshti University, Tehran.
- Rohani, M. F., Shafie, K., & Noorbaloochi, S. (2006). A Bayesian signal detection procedure for scale-space random fields. *The Canadian Journal of Statistics*, 34(2), 311–325.
- Shafie, K., & Abravesh, A. (2016). RKHS approach for signal detection in rotation and scale space random fields. *Journal of Statistical Theory and Applications*, 15(3), 214–220.
- Shafie, K., & Noorbaloochi, S. (2001). *Signal detection in noisy images: A Bayesian approach*. (Tech. Rep.). Shahid Beheshti University.
- Shafie, K., Sigal, B., Siegmund, D., & Worsley, K. J. (2003). Rotation space random fields with an application to fmri data. *The Annals of Statistics*, 31(6), 1732–1771.

- Siegmund, & Worsley. (1995). Testing for a signal with unknown location and scale in a stationary gaussian random field. *The Annals of Statistics*, 608–639.
- Teicher, H. (1961). Identifiability of mixtures. *The annals of Mathematical statistics*, 32(1), 244–248.
- Teicher, H. (1963). Identifiability of finite mixtures. *The annals of Mathematical statistics*, 1265–1269.
- Williams, D. (1991). *Probability with martingales*. Cambridge university press.
- Worsley, K. J. (1994). Local maxima and the expected euler characteristic of excursion sets of  $\chi^2$ , f and t fields. *Advances in Applied Probability*, 13–42.
- Worsley, K. J., Evans, A. C., 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.

## APPENDIX A

## R CODE FOR MONTE CARLO SIMULATION

```
#####  
#### Through this code, we want to find the Bayes factor in mixture  
#### of 2 Gaussian random fields by applying Monte carlo sampler.  
#### The dimension is d=2.  
  
par.scheme<-as.matrix(read.table(file="par.scheme"))  
SampleTheta<-as.matrix(read.table(file="sampletheta"))  
  
##### Required packages #####  
  
if (!requireNamespace("BiocManager", quietly = TRUE))  
install.packages("BiocManager")  
BiocManager::install()  
  
library(ExtDist)  
library(invgamma)  
library(truncnorm)  
library(truncdist)  
library(mvtnorm)  
library(tmvtnorm)  
library(coda)  
library(reshape)  
library(iterpc)  
library(doParallel)
```

```

library(parallel)
library(foreach)
library(EBImage)

##### Introducing Parameters #####
# xi_1 & xi_2 are amplitudes of two GRFs
# sigma0_1 & sigma0_2 are scales for two GRFs
# t0 & s0 are vectors of locations for centers of two GRFs
# Pi is the weight parameter in the mixture

#####
##### This function generates a realization of a non-smooth ###
### Gaussian RF for a specific vector of 'theta'
signal.generate <- function(N,xi,t0,sigma0){

noise <- matrix(rnorm(N*N),N,N) #white noise

pix_x <- (row(noise)-1)/(N-1) #Range [0,1]
pix_y <- (col(noise)-1)/(N-1)

mu <- (xi/(sigma0*sqrt(pi)))
  *exp((-0.5*((pix_x-t0[1])^2+(pix_y-t0[2])^2)/sigma0^2))
#mean function of Z(t)

RF <- mu + noise #RF is Z(t) that is a non-smooth GRF
return(RF)
}

```

```
#####
### generating an image from mixture of RF1 & RF2
datagen<-function(theta, N){

RF1 <- signal.generate(N,theta[2],c(theta[4], theta[5]), theta[8])
  #first non-smooth RF in the mixture model
RF2 <- signal.generate(N,theta[3],c(theta[6], theta[6]), theta[9])
  #second non-smooth RF in the mixture model
z <- rbinom(1,1,theta[1])
Image=z*RF1+(1-z)*RF2
return(Image)
}

#####
loglik<-function(sampletheta,data){
N<-dim(data)[1]

l1 <-log(sampletheta[1])+ sampletheta[2]*gblur(data,sampletheta[8])
[round(sampletheta[4]*(N-1))+1,round(sampletheta[5]*(N-1))+1]-
sampletheta[2]^2/2

l2 <-log((1-sampletheta[1]))+ sampletheta[3]*gblur(data,
sampletheta[9])[round(sampletheta[6]*(N-1))+1,
round(sampletheta[7]*(N-1))+1]-
sampletheta[3]^2/2

return(log(exp(l1)+exp(l2)))
}

```

```
#####

loglik.sim<-function(theta,sampletheta,N){
loglik( sampletheta,datagen(theta, N))
}

#####

cl <- makeCluster(8)
registerDoParallel(cl)

nsample<-1000
sampletheta<-SampleTheta[1:nsample,]
nsim<-1000
n<-10
N<-64

theta<-par.scheme[i,2:10]          #i in 1:48

f<-function(nsim){
foreach(isample=1:nsample, .combine=rbind) %do%
replicate(n,loglik.sim(theta,sampletheta[isample,],N))
}

LL<-lapply(1:nsim,f)
```

```
##### Simulation data analysis

Bayes.Factor<-function(k) {
  out<-exp(readRDS(filenamees[k]))
  1/apply(apply(out, c(1,2), prod), 1, mean)
}

filenamees<-paste0(rep("out", 48), as.character(1:48), rep(".RDS", 48))
BFL<-lapply(1:48, Bayes.Factor)
BF<-matrix(as.numeric(unlist(BFL)), ncol=48)
filenamees<-paste0(rep("out", 48), as.character(1:48), rep(".RDS", 48))
BFL<-lapply(1:48, Bayes.Factor)
BF<-matrix(as.numeric(unlist(BFL)), ncol=48)

#####
#####
#### Through this code, we want to estimate parameters in mixture of
#### two Gaussian random fields by applying Gibbs sampler and M-H.
#### The dimension is d=2, for N=64
####

##### Required packages #####

library(ExtDist)
library(invgamma)
```

```

library(truncnorm)
library(truncdist)
library(mvtnorm)
library(tmvtnorm)
library(coda)
library(reshape)
library(iterpc)
library(doParallel)
library(foreach)
cl <- makeCluster(8)
registerDoParallel(cl)

##### BEGINNING of FUNCTIONS-BLOCK #####

### This function generates a realization of a non-smooth ###
### Gaussian RF for a specific vector of 'theta'
signal.generate <- function(N,xi,t0,sigma0){

noise <- matrix(rnorm(N*N),N,N) #white noise

pix_x <- (row(noise)-1)/(N-1) #Range [0,1]
pix_y <- (col(noise)-1)/(N-1)

mu <- (xi/(sigma0*sqrt(pi)))
      *exp((-0.5*((pix_x-t0[1])^2+(pix_y-t0[2])^2)/sigma0^2))
#mean function of Z(t)

```

```

RF <- mu + noise      #RF is Z(t) that is a non-smooth GRF
return(RF)
}

### This function smooths the non-smooth GRF by ###
### a Gaussian kernel with scale of sigma
smooth2 <- function(N,Image,ftxyfilter,t0,l) {
  (Re(fft(fft(Image[, , l]) * ftxyfilter, inverse = T)))
    [round(t0[1]*(N-1))+1,round(t0[2]*(N-1))+1] /N^2
}

### This function after smoothing data, gives the value of ###
### GRF in the center of signal, x(t0_1,t0_2,sigma0)
signalCenter.value <- function(Image,t0,sig){
  #"Image" is an observed field(non-smooth)
  N<-dim(Image)[1]
  n<- dim(Image)[3]
  x <- 1:N
  xfilter <- (pi)^(-0.5) * exp(-0.5 *
    ((1/sig)*(x - mean(x)))^2)          #smoothing kernel
  xfilter <- sqrt(xfilter/sum(xfilter))
  xyfilter <- outer(xfilter, xfilter)
  ftxyfilter <- Mod(fft(xyfilter))    #fft:Fast Discrete Fourier Transform
  x_value<-rep(0,n)
  foreach(l = 1:n, .combine=c) %do% smooth2(N=N,Image=Image,
    ftxyfilter=ftxyfilter,t0,t0,l=1)
}

```

```

### This function calculates the probability for posterior ###
### of z[j] which is z of each image
prob_Bern <- function(p,x_1,xi_1,x_2,xi_2){
xi_min<- min(xi_1,xi_2)
f_1 <- (xi_1*x_1)-((1/2) *(xi_min^2 - xi_1^2))
f_2 <- (xi_2*x_2)-((1/2) *(xi_min^2 - xi_2^2))
p_z <- 1/(1 + p/(1-p)*exp(f_2-f_1))
z <- rbinom(1,1,p_z)      #generating Z from conditional posterior
Ber <- c(z,p_z)          #results of this function
return(Ber)
}

```

```

### This function returns the un-normalized log-likelihood ###
### function of ethal=(t0,sigma0_1) ll is log-likelihood function ###
###[Dissertation, (8)]

```

```

loglik1<-function(t0,sigma0,p,xi_1,data,z){
ll <-sum(z*(log(p)
          + xi_1*signalCenter.value(data,t0,sigma0)- xi_1^2/2))
return(ll)
}

```

```

### This function returns the un-normalized log-likelihood ###
### function of etha2=(s0,sigma0_2) ll is log-likelihood function
###[Dissertation, (8)]

loglik2<-function(s0,sigma0,p,xi_2,data,z){
ll <- sum((1-z)*(log(1-p)
          + xi_2*signalCenter.value(data,s0,sigma0)- xi_2^2/2) )
return(ll)
}

### This function returns the log-prior function of          ###
### etha=(t0,sigma0) lprior is log-prior on etha
### beta_1 is Shape parameter of Inv-Gamma as a prior on sigma0
### beta_2 is Scale parameter of Inv-Gamma as a prior on sigma0
### mu is MEAN VECTORS of PRIOR on t0 or s0
### Rho is CORRELATION MARTIX of PRIOR on t0 or s0
log.prior <- function(t0,sigma0,beta_1, beta_2,mu, Rho){

lprior <- dtrunc(sigma0, beta_1, beta_2,spec="invgamma"
          ,a=0.1,b=0.5,log = TRUE)+
dtmvnorm(t0, mean=mu, Rho,lower=c(0,0),
          upper=c(1,1),log=TRUE)

return(lprior)
}

```

```

## Introducing proposal dist. for Metropolis-Hasting Sampler ##
# We consider Inv-Gamma dist. as the proposal dist. for sigma0.
# beta_1 & beta_2 are Shape and Scale parameters of Inv-Gamma.
# We consider truncated bi-variate normal dist. as the proposal
# for t01, t02, so by this we assume a dependency for t01 and t02.
# mean=c(0.5, 0.5) is INITIAL MEAN VECTOR for t=(t01,t02)
# Rho=matrix(c(1,0.5,0.5,1),2,2) is INITIAL CORRELATION MATRIX
# for t=(t01,t02)
# c(0,0) gives lower boundaries for truncated bivariate normal
# c(1,1) gives upper boundaries for truncated bivariate normal

## This function returns the log-proposal function of ##
## etha=(t0,sigma0)

log.proposal <- function(t0,sigma0,beta_1, beta_2,mu, Rho){

lproposal <- dtrunc(sigma0, beta_1, beta_2,spec="invgamma",a=0.1
, b=0.5, log = TRUE)+
dtmvnorm(t0, mean=mu,Rho,lower=c(0,0), upper=c(1,1),log=TRUE)

return(lproposal)
}

```

```

### This function calculates log(r) in M-H step ###
### x is a vector of x(t0_1,t0_2,sigma0) in n images
### xStar is a vector of x(t0_star1,t0_star2,sigma0_star) in n images
### ll is [loglik1(t0_star,sig0_Star,Pi,xi_1=xi,Image,z)-
###          loglik1(t0,sig0,Pi,xi_1=xi,Image,z)]
### or [loglik2(s0_star,sig0_Star,Pi,xi_2=xi,Image,z)-
###          loglik2(s0,sig0,Pi,xi_2=xi,Image,z)]

calculate.r <- function(xi,t0,t0_star,sigma0,sig0_Star,beta_1
, beta_2,mu,Rho,ll){
# beta_1 & beta_2      #Hyper-parameters of INV-Gamma dist.
# mu                  #INITIAL MEAN VECTOR for t0=(t01,t02)
# Rho                 #INITIAL COVARIANCE MATRIX for t0=(t01,t02)

mean1=c(round(t0_star[1], digits = 2),round(t0_star[2], digits=2))
mean2=c(round(t0[1], digits = 2),round(t0[2], digits=2))

lr <- ll + log.prior(t0_star,sig0_Star,beta_1,beta_2,mu,Rho)-
      log.prior(t0,sigma0,beta_1,beta_2,mu,Rho)+
      log.proposal(t0,sigma0,beta_1,beta_2=sig0_Star,mu=mean1,Rho)-
      log.proposal(t0_star,sig0_Star,beta_1,beta_2=sigma0,mu=mean2,Rho)

r <- exp(lr)
return(r)

}

```

```

M.H.func <- function(I_MH,S,T0_1,T0_2,P,xi_1,xi_2,data,z,
                    beta_1,beta_2){

# beta_1 & beta_2      #Hyper-parameters of INV-Gamma dist.
# mu                  #INITIAL MEAN VECTOR for t0=(t01,t02)
# Rho                 #INITIAL COVARIANCE MATRIX for t0=(t01,t02)

# beta_1= 4; beta_2=0.5; mu=c(0.5,0.5);
# Rho=matrix(c(1,0.5,0.5,1),2,2);

#each row is for sigma0_1 & sigma0_2 in each iteration
Sig0 = matrix(0,I_MH,2)

Sig0[1,] <- S

T_0 = array(c(0),dim=c(I_MH,2,2))
T_0[1,,1] <- T0_1
T_0[1,,2] <- T0_2

sig0_star <- c()

for(h in 1:(I_MH-1)){
lambda <- Sig0[h,]          #Scale parameter of Inv-Gamma
#sig0_star[1] <- rtrunc(n = 1, beta_1,lambda[1],
                      spec="invgamma", a=.1, b=0.5)

#INV-Gamma, PROPOSAL for sigma0

```

```

sig0_star<-foreach(k=1:2,.combine=c) %do%
rtrunc(n = 1, beta_1,lambda[k], spec="invgamma", a=.1, b=0.5)
T0_star <- rtmvnorm(2, mean=c(0.5, 0.5),
  # PROPOSAL for t=(t01,t02) and s=(s01,s02)
  sigma=matrix(c(1, 0.5, 0.5, 1), 2, 2),
  lower=c(0, 0), upper=c(1, 1),
  algorithm="rejection")

xi_T=c(xi_1,xi_2)
ll_T<-foreach(k=1:2,.combine=c) %do%
(loglik1(T0_star[k,],sig0_star[k],P,xi_T[k],data,z)-
loglik1(T_0[h,,k],Sig0[h,k],P,xi_T[k],data,z))

r <- foreach(k=1:2,.combine=c) %do%
  (calculate.r(xi_T[k],T_0[h,,k],T0_star[k,],Sig0[h,k],
  sig0_star[k],beta_1,beta_2,mu=c(0.5,0.5),Rho=matrix(c(1,0.5,0.5,1),
  2,2),ll_T[k]))

for(k in 1:2){
if(1 <= r[k] || r[k]=="NaN" ){
Sig0[h+1,k] <- sig0_star[k]
T_0[h+1,,k] <- T0_star[k,]
}else{
b_v <- rbinom(1,1,r[k])

```

```

if(b_v == 1){
  Sig0[h+1,k] <- sig0_star[k]
  T_0[h+1,,k] <- T0_star[k,]
}else{
  Sig0[h+1,k] <- Sig0[h,k]
  T_0[h+1,,k] <- T_0[h,,k]
}
}
}
}
}
output <- cbind(Sig0,T_0[,,1],T_0[,,2])
return(output)
}

###                Gibbs Sampling                ###
### This function Generates the vector of "Theta"    ###

### Introducing priors on parameters ###
# Beta dist. for Pi
# Improper prior for xi_1 & xi_2 in [0,10]; 10 is an arbitrary
# extreme-large amplitude
# Inverse-Gamma for sigma0_1 & sigma0_2
# truncated bi-variate normal dist. for t0 & s0
# Bernoulli for latent variable z

Gibbs.func <- function(I,data){

```

```

##### First step in Gibbs: Generating initial values #####
# "I" is number of iterations, given at the beginning of simulation

alpha_1=2; alpha_2=2;          #Hyper-parameters of Beta dist.
beta_1= 4; beta_2=0.5;        #Hyper-parameters of INV-Gamma dist.

#each row is for sigma0_1 & sigma0_2 in each iteration
Sigma0 = matrix(0,I,2);

Sigma0[1,] <- rtrunc( 2,shape=beta_1,rate=beta_2,
                    spec="invgamma", a=.1, b=0.5 )

T0 = array(c(0),dim=c(I,2,2))
      #array of origins of signals from 2 GRFs
T0[1,,]=rtmvnorm(2, mean=c(0.5, 0.5),
                sigma=matrix(c(1, 0.5, 0.5, 1), 2, 2),
                lower=c(0, 0), upper=c(1, 1),algorithm="rejection")

#each row is for xi_1 & xi_2 in each iteration
xi = matrix(0,I,2);
xi[1,] <- runif(2,0,3)  #instead of improper, I used Uniform

Pi = c()                #each component is pi in an iteration
Pi[1] <- rbeta(1,alpha_1,alpha_2)

#each element in a row is a prob. of z=1 for one of images
P_z = matrix(0,I,n)

```

```

Z = matrix(0,I,n);

##each row is a vector of latent variables for n
# images in each iteration
Z[1,] <- rbinom(n,1,Pi[1])
f <- array(c(0),c(I,2,n))

m = c();
##each component is sum of latent variables for n
# images in each iteration
m[1] <- sum(Z[1,])

X <- array(c(0),c(I,n,2))
X[1,,] <- foreach(k=1:2,.combine=cbind) %do%
  signalCenter.value(data,T0[1,,k],Sigma0[1,k])
Lik = c() #vector of Likelihood functions of I iterations
Lik[1] <- exp(loglik1(T0[1,,1],Sigma0[1,1],Pi[1],xi_1=xi[1,1],
data,Z[1,])+loglik2(T0[1,,2],Sigma0[1,2],Pi[1],xi_2=xi[1,2],data,
Z[1,]) )

##### 2nd Step in Gibbs: Conditional posteriors #####

one_v = c(rep(1,n)) #related to generating xi_1 & xi_2
X_star = matrix(0,2,n) #related to calculating x_star in M-H step

### Beginning of the loop
for(i in 2:I){

```

```

print(i)
Pi[i] <- rbeta(1, alpha_1+m[i-1], alpha_2+n-m[i-1])
#generating Pi from conditional posterior in each iter.

X[i,,]<-foreach(k=1:2,.combine=cbind) %do%
signalCenter.value(data,T0[i-1,,k],Sigma0[i-1,k])

for(j in 1:n){
Ber <- prob_Bern(Pi[i],X[i,j,1],xi[i-1,1],X[i,j,2],xi[i-1,2])

Z[i,j] <- Ber[1]
#each row of Z is a vector of estimated latent vars of
#images in dataset for an iter.
P_z[i,j] <- Ber[2]
f[i,,j] <- Ber[3:4]
}
m[i] <- sum(Z[i,])

### generating Xi_1 & xi_2 from conditional posteriors
if(m[i] == 0){
xi[i,1] <- runif(1,0,3)
mu2 <- (1/n)*sum(X[i,,2])
sd2 <- sqrt(1/n)
xi[i,2] <- rtruncnorm(1, a=0, b=3, mean = mu2, sd = sd2)
}else if(m[i] == n){
xi[i,2] <- runif(1,0,3)
mu1 <- (1/n)*sum(X[i,,1])

```

```

sd1 <- sqrt(1/n)
xi[i,1] <- rtruncnorm(1, a=0, b=3, mean = mu1, sd = sd1)
}else{
mu1 <- (t(Z[i,])%*%X[i,,1])/m[i]
sd1 <- sqrt(1/m[i])
mu2 <- (t(one_v-Z[i,])%*%(X[i,,2]))/(n-m[i])
sd2 <- sqrt(1/(n-m[i]))
xi[i,1] <- rtruncnorm(1, a=0, b=3, mean = mu1, sd = sd1)
xi[i,2] <- rtruncnorm(1, a=0, b=3, mean = mu2, sd = sd2)
}

##### Metropolis-Hasting Sampler #####
#### Generating sigma0 & t0 from conditional posteriors
#### using Metropolis-Hasting

### Introducing proposal dist. for Metropolis-Hasting Sampler ###
# We consider exponential dist. as the candidate for sigma0
# We consider truncated bivariate normal dist. as the candidate for
# t01, t02, so by this we assume a dependency for t01 and t02.
# mean=c(0.5, 0.5) is INITIAL MEAN VECTOR for t=(t01,t02)
# sigma=matrix(c(1,0.5,0.5,1),2,2) is initial COV. for t=(t01,t02)

I_MH = 500          #Number of iterations in M-H step
                    #for each iteration of Gibbs

beta_1 = 4; beta_2 = 0.5    #Hyper-parameters of INV-Gamma dist.

```

```

output_MH <- M.H.func(I_MH,S=Sigma0[i-1,],
T0_1=T0[i-1,,1],T0_2=T0[i-1,,2],P=Pi[i],xi_1=xi[i,1],
xi_2=xi[i,2],data,z=Z[i,],beta_1,beta_2)

Sigma0[i,] <- output_MH[I_MH,1:2]
T0[i,,1] <- output_MH[I_MH,3:4]
T0[i,,2] <- output_MH[I_MH,5:6]
Lik[i] <- exp(loglik1(T0[i,,1],Sigma0[i,1],Pi[i],
xi_1=xi[i,1],data,Z[i,])+loglik2(T0[i,,2],Sigma0[i,2],
Pi[i],xi_2=xi[i,2],data,Z[i,]))

output.Gb <- cbind(Pi,xi,T0[,,1],T0[,,2],Sigma0,Lik)
return(output.Gb)
}

##### END of FUNCTIONS-BLOCK #####
##### Introducing Parameters #####
# xi_1 & xi_2 are amplitudes of two GRFs
# sigma0_1 & sigma0_2 are scales for two GRFs
# t0 & s0 are vectors of locations for centers of two GRFs
# Pi is the weight parameter in the mixture

```

```
##### Parameter Setting (scheme: 1 to 24) #####

spl_n <- rep(c(10,20),each=24)                #sample size
xi <- c(0,1,2.5) # Levels of amplitudes
amp <- matrix(c(0,0,0,1,0,2.5,1,1,1,2.5,2.5,2.5),6,2,byrow=T)
                #Combinations of xi's

loc <- matrix(c(.5,.5,.7,.7,.1,.6,.6,.1),2,4, byrow=T)
                #locations:t0 and s0

sigma01 <-rep(0.2,each=48)
sigma02 <-rep(0.4,each=48)

expand.grid.df(data.frame(xi=amp),data.frame(t=loc),
data.frame(sigma01),data.frame(sigma02))

Pi_true <- rep(c(0.3,0.5),each=12)           #weight parameter

par.scheme <- as.matrix(cbind(spl_n,Pi_true,rbind(comb,comb)))
                #parameter scheme

colnames(par.scheme) <- c("n","Pi_true","xi_1","xi_2","t0_1",
                "t0_2","s0_1","s0_2","sigma0_1","sigma0_2")
```

```
##### BEGINNING of SIMULATION #####

I=1000;          #Number of iterations
d <- 10          #Number of columns of "output_Gibbs"

out <- matrix(NA,nrow=I,ncol=d)

ii= 6
theta = par.scheme[ii,]
print(theta)

##### Generating n Data from the mixture #####
n=theta[1];      #Number of images; can be for n subjects or 1 subject
Pi_true=theta[2]          #true weight parameter
xi_1=theta[3];  xi_2=theta[4];          #true amplitudes
t0_true=theta[5:6];  s0_true=theta[7:8]          #true locations
sigma0_1=theta[9];  sigma0_2=theta[10];          #true scales

N=64

z <- c()          #true vector of z for n images
data <- array(0,c(N,N,n))          #includes n images

### generating n images from mixture of RF1 & RF2
for(l in 1:n){
RF1 <- signal.generate(N,xi_1,t0_true,sigma0_1)
  #first non-smooth RF in the mixture model
```

```

RF2 <- signal.generate(N,xi_2,s0_true,sigma0_2)
  #second non-smooth RF in the mixture model
z[1] <- rbinom(1,1,Pi_true)
Image=z[1]*RF1+(1-z[1])*RF2
data[, ,1] <- Image
}

```

```

#####      Gibbs Sampling      #####
### Introducing priors on parameters   ###
# Beta dist. for p
# Improper prior for xi_1 & xi_2 in [0,10];
# 10 is an arbitrary extreme-large amplitude
# Inverse-Gamma for sigma0_1 & sigma0_2
# truncated bi-variate normal dist. for t0 & s0
# Bernoulli for latent variable z
start_time <- Sys.time()
output_Gibbs <- Gibbs.func(I,data)
end_time <- Sys.time()
end_time - start_time

```