Modeling fMRI Time Series Using a Non-Linear Method

By

Christopher Paul Meller

A Dissertation Submitted to the Faculty of

the Graduate School of Biomedical Sciences

of the Medical College of Wisconsin in

Partial Fulfillment of the Requirements

for the Degree of Master of Science

Milwaukee, Wisconsin

December 2004

Abstract

Magnitude-only data models have dominated fMRI analysis since the inception of the technology. However, when such analyzes are done half the data is disregarded which may or may not contain valuable information. This overlooked data is called the phase and is restricted to values between $-\pi$ and π. When phase time series are analyzed, ordinary least squares (OLS) regression has been the technique of choice. However, OLS models perform poorly when "wrap-around" or low SNR is present. We have explored alternatives to the OLS model which will account for the angular response of the phase while also allowing us the flexibility to develop similar hypothesis tests. We adopt a model by Fisher and Lee (1992) to our analysis of the generally discarded fMRI phase time series and show its improvement over the OLS model for both parameter estimation and data prediction for various conditions. We fit the model to both simulated data and acquired data from an actual fMRI experiment and found an improvement in parameter estimation along with modeling for the Fisher and Lee method in the simulated data while detailing potential benefits when used with experimental acquired data. Finally, we look at a map of statistics relating association of the observed voxel phase time courses with a reference function in our acquired data and show the possible detection of biological information in the generally discarded phase.

Acknowledgments

I wish to express my gratitude to my advisor, Dr. Daniel Rowe, for suggesting the research problem along with providing guidance and support in my studies.

I appreciated all the comments and suggestion by my committee members:
Dr. Raymond Hoffmann and Dr. Brent Logan

Thanks to all the students, faculty, and staff from the Division of Biostatistics for their constant motivation.

Most importantly, I would like to thank my wife, Laura, for her constant love, support, and the many sacrifices she made during my graduate school career.

Contents

List of Figures

Chapter 1

Introduction

Magnetic resonance imaging (MRI) is an invaluable tool used by clinicians and physicians alike to investigate and diagnose biological phenomena in both animals and humans. An MRI of the knee can reveal whether ligaments are intact or torn, while an MRI of the brain can detect the difference between gray and white matter. Different pulse sequences are able to be constructed with enough sensitivity to look at tumors, abnormalities in blood vessels, or bone damage [8]. Functional magnetic resonance imaging (fMRI) can be thought of as many volumetric MRIs over time. This allows the investigator to localize and monitor different mental processes in different parts of the brain. An example using fMRI is shown when looking at dyslexia in children [1]. Those children affected by the disorder had different brain activity patterns than those who did not suffer from dyslexia. However, it was shown that the neural activity in the affected children can be corrected to match those children who

read normally within a few weeks by using reading instruction . This is just one of the many examples which fMRI was used as a tool by investigators looking to better understand and treat individuals with conditions that affect millions a year.

Functional MRI is a collection of MRI slices over a period of time. Usually in experimental fMRI some type of stimuli is introduced, such as finger tapping or flashing lights, and then removed to measure changes in brain response. This is usually done in a number sets of some predetermined length. We will use eight sets lasting 16 seconds each for both "on" and "off" function in both simulated and real data.

Those areas of the brain that exhibit cognitive function to different types of stimuli is of interest to investigators. When neurons in a given location synapse for brain function they use up bound oxygen in the hemoglobin of oxygenated blood. The physiological response in the brain is to overcompensate with a surplus of oxygenated blood in that location. It is the change in the magnetic field in this location due to an increase of oxygenated blood that is measured and not neuronal firing. A statistically significant increase in oxygenated blood in a location or voxel associated with the presentation of a stimulus is called activation [2].

Currently, the most common method being used to detect activation in fMRI is a general linear model with magnitude-only data. However, this model disregards all the phase data and any information it may contain. We will look at modeling the phase data that is generally disregarded by using a model which is more flexible

than the ordinary least squares (OLS) regression model, the current standard for this type of analysis. Specifically, we will implement a model developed by Fisher and Lee (1992) on both simulated and actual experimental fMRI time series data then lay the framework for making inferences about the possible phase activations. In doing this we will have accounted for the angular nature of the response and provide an improvement in dealing with the "wrap-around" issue. In the past the only way that has been used to account for "wrap-around" in fMRI data was to artifically "unwrap" it and proceed with an analysis.

Chapter 2

Background

The MRI data we acquire generally after an inverse Fourier transform (IFT) has taken place is composed of two parts, a real ($y_{R\ell,m,n}$) and imaginary ($y_{I\ell,m,n}$) image parts, where (ℓ, m, n) are the location coordinates, slice, row, and column, of the voxel on the 3D volume image. We are neglecting a third coordinate, ℓ, because we will deal only with a single slice of an MRI scan over time. These parts can be broken down in the following way for a single time point,

$$y_{Rm,n} = \rho_{R_{m,n}} + \eta_{R_{m,n}} \tag{2.0.1}$$

$$y_{Im,n} = \rho_{I_{m,n}} + \eta_{I_{m,n}} \tag{2.0.2}$$

where $\rho_{R_{m,n}}$ and $\rho_{I_{m,n}}$ are the true signals while $\eta_{R_{m,n}}$ and $\eta_{I_{m,n}}$ are the additive random noise in each channel [17]. It is assumed that the noise has the following distribution, $(\eta_{R_{m,n}}, \eta_{I_{m,n}})' \sim N(0, \Sigma)$. Combining the channels into a complex number,

$y_{m,n}$, we get

$$y_{m,n} = \left[\rho_{R_{m,n}} + \eta_{R_{m,n}}\right] + i\left[\rho_{I_{m,n}} + \eta_{I_{m,n}}\right]. \tag{2.0.3}$$

However, the above notation is dealing only with a particular voxel at one time point, MRI. In functional MRI (fMRI) we observe the same voxel over a regularly sampled discrete time series. Therefore, we must introduce the necessary subscript, t, to describe these different time points,

$$y_{m,n,t} = \left[\rho_{R_{m,n,t}} + \eta_{R_{m,n,t}}\right] + i\left[\rho_{I_{m,n,t}} + \eta_{I_{m,n,t}}\right]. \tag{2.0.4}$$

FMRI can be thought of as a 4D lattice of complex numbers where time is the fourth dimension.

The most common methodology used currently in fMRI activation detection is the magnitude-only data model. This can only be explored after transforming from rectangular coordinates $y_{Rm,n}$ and $y_{I\,m,n}$ into polar coordinates leaving the investigator with both a magnitude and phase value for each voxel and time point. This is a non-unique transformation of the data. The magnitude-only model discards the phase. Using phase-only data, we will explore task related phase changes which will be called "phase activation."

Rowe and Logan (2004) introduced a complex nonlinear multiple regression model for magnitude activation that included a phase imperfection, θ, and a set of time dependant covariates, x_t' [17]. They define X as a $n \times (q+1)$ design matrix where n is the number of time points for each voxel and $q+1$ is the number of columns.

This design is composed of the individual rows x_t'. A simple design matrix has three columns which represent the intercept, a linear trend over time, and the reference function. The justification for a linear trend column in the design matrix is justified in a paper by Smith et al. (1999) in which he notices a low frequency drift over time in both cadavers and phantom fMRI scans [20]. Rowe and Logan's model has the following form

$$\begin{pmatrix} y_{Rt,m,n} \\ \\ y_{It,m,n} \end{pmatrix} = \begin{pmatrix} x_t'\beta\cos(\theta) \\ \\ x_t'\beta\sin(\theta) \end{pmatrix} + \begin{pmatrix} \eta_{Rt,m,n} \\ \\ \eta_{It,m,n} \end{pmatrix} \tag{2.0.5}$$

with $\theta \in (-\pi, \pi]$ and $\beta \in \mathbb{R}^{q+1}$ being fixed but unknown parameter coefficients estimated voxel by voxel. As the reader can clearly see they are using both parts of the complex data at once rather than discarding the phase and using the magnitude-only data model. This model specifies the phase imperfection to be constant over time in a given voxel, however the authors expanded this model by incorporating an unrestricted phase that is unique at each time point. The magnitude-only data model was shown by Rowe and Logan to be equivalent to their complex model when unrestricted phase is present [18].

2.1 Magnitude Model

The most common method used to compute activations is the magnitude-only data model, which will be denoted by,

$$r_t = \left\{ \left[x_t' \beta \cos(\theta) + \eta_{R_t} \right]^2 + \left[x_t' \beta \sin(\theta) + \eta_{I_t} \right]^2 \right\}^{\frac{1}{2}}. \tag{2.1.6}$$

To reduce the amount of needed subscripts and using the fact that voxels are treated individually we will focus on a single time series at time t,

$$r_t = \left\{ \left[x_t' \beta \cos(\theta) + \eta_{R_t} \right]^2 + \left[x_t' \beta \sin(\theta) + \eta_{I_t} \right]^2 \right\}^{\frac{1}{2}}. \tag{2.1.7}$$

The magnitude is typically modeled with a normal distribution but this is only asymptotically true for a large signal-to-noise ratio (SNR) defined by ρ/σ [17]. To find the marginal distribution of the magnitude, which is what we have when we throw out the phase, we must integrate out the phase from the following joint distribution,

$$f(r,\phi) \;\; = \;\; (2\pi\sigma^2)^{-1} r e^{-\frac{r^2+\rho^2-2r\rho(cos(\phi)cos(\theta)+sin(\phi)sin(\theta))}{2\sigma^2}}, r,\ \rho,\ \sigma^2 > 0 \ -\pi < \phi, \theta \leq \pi, \tag{2.1.8}$$

bringing us to the exact distribution which is a Ricean and derived in Appendix A.1. The Ricean has the following density function,

$$f(r|\rho,\sigma^2) = \frac{r}{\sigma^2} e^{-\frac{r^2+\rho^2}{2\sigma^2}} I_0\left(\frac{r\rho}{\sigma^2}\right),\ r > 0,\ \rho > 0,\ \sigma^2 > 0. \tag{2.1.9}$$

In the above equality $I_0(\cdot)$ is a zeroth order modified Bessel function of the first kind. The subscript t is momentarily suppressed for clarity. If the SNR is zero the Ricean

distribution simplifies to a Rayleigh distribution. Using the derivation by Rowe and Logan (2004) we will expand the squares in the above magnitude model and use the Taylor's expansion that states $\sqrt{1+u} \approx 1 + u/2$ for the first two terms.

$$\begin{aligned} r_t &= \left\{ [x_t'\beta]^2 + \left[\eta_{R_t}^2 + \eta_{I_t}^2\right] + 2\,[x_t'\beta]\,[\eta_{R_t}\cos\theta + \eta_{I_t}\sin\theta] \right\}^{\frac{1}{2}} && (2.1.10) \\ &= [x_t'\beta]\left[1 + \frac{2\,[\eta_{R_t}\cos\theta + \eta_{I_t}\sin\theta]}{[x_t'\beta]} + \frac{\left[\eta_{R_t}^2 + \eta_{I_t}^2\right]}{[x_t'\beta]^2}\right]^{\frac{1}{2}} && (2.1.11) \\ &\approx x_t'\beta + \epsilon_t && (2.1.12) \end{aligned}$$

where $\epsilon_t = \eta_{Rt}\cos\theta + \eta_{Rt}\sin\theta \sim N(0, \sigma^2)$ is temporally independent and identically distributed. If the data is not independent than the investigator should pre-whiten the data as described in the Appendix of Rowe and Logan (2004). Expressed in matrix notation we write the model as

$$r \;=\; X\;\beta \;+\; \epsilon \qquad (2.1.13)$$

where $\epsilon \sim N(0, \sigma^2 I_n)$. Here X is a design matrix of dimension $n \times p$, r and ϵ are $n \times 1$ vectors of magnitude-only observations and errors, and β is a $(q+1) \times 1$ vector. We see that the model reduces to a standard linear multiple regression model for large SNR.

Finding the maximum likelihood estimates and deriving likelihood ratio tests dealing with linear hypothesis of $C\beta = 0$ are straight forward and are derived in the Appendix A.3 [15, 17]. We see that the unrestricted maximum likelihood estimates are,

$$\hat{\beta} \;=\; (X'X)^{-1}\,X'r \qquad (2.1.14)$$

$$\hat{\sigma}^2 = \frac{(r - X\hat{\beta})'(r - X\hat{\beta})}{n} \tag{2.1.15}$$

with the restricted MLE's being,

$$\tilde{\beta} = \Psi\hat{\beta} \tag{2.1.16}$$

$$\tilde{\sigma}^2 = \frac{(r - X\tilde{\beta})'(r - X\tilde{\beta})}{n} \tag{2.1.17}$$

and $\Psi = I - (X'X)^{-1}C'(C(X'X)^{-1}C')^{-1}C$. Using the asymptotic χ^2_w distribution of -2 log of a likelihood ratio test we note that the test statistic is

$$-2\log\Lambda = n\log\left(\frac{\tilde{\sigma}^2}{\hat{\sigma}^2}\right). \tag{2.1.18}$$

The degrees of freedom, w, is equal to the full row rank of C [15].

This has been the usual way to model activation in fMRI but this method also removes half the data by not looking at the phase [17]. When we convert from real-imaginary rectangular coordinates to polar coordinates we generate a magnitude, r, and an angle, ϕ.

In the next sections we will look at building a model from the phase-only data and detect "phase" activation. In doing so we will gain an insight into the possibility of information being contained in the phase which is disregarded in the magnitude-only data model. Even if the phase-only data model doesn't model activation as well as the magnitude model it will at least give us another method to investigate activation or a biological response to the presentation of stimulus. The complex model by Rowe and Logan (2004) is yet another method developed which detects statistically

significant changes or activation in the magnitude but from the raw complex data [17]. Their model shows great promise and improved power over the industry standard, the magnitude-only data model. Allowing the investigator to look at the entire data and not artifically generated magnitude-only data is another positive attribute. As mentioned, we will be exploring possible biological information in the generally omitted phase-only data while disregarding the magnitude-only data.

Chapter 3

Theory

As previously stated, the phase data is often disregarded when trying to determine task related activation in fMRI. However, phase represents half the data in each voxel so intuitively we would like to develop a model analogous to the magnitude-only data and make inferences. We will see that as with the magnitude-only data model ordinary least squares (OLS) can also be used when modeling phase-only data when various assumptions about the data hold. Finally, we will account for the angular response property of the phase in which π and $-\pi$ are adjacent to each other by discussing methods discussed by Gould [6], Johnson and Wehrly [10], along with Fisher and Lee [5] to model the data.

3.1 Ordinary Least Squares

We begin to look more closely at the phase-only data model by first deriving its distribution, ϕ, which is outlined in Appendix A.1 and shown in Rowe and Logan (2004). Below in Equation 3.1.2 we have the joint distribution of the magnitude r and the phase ϕ along with the marginal distribution of ϕ by integrating out r. We are suppressing subscripts, t, on r, ρ, and ϕ to make our notation simpler however the reader should keep in mind they vary over time.

$$f(r,\phi) = (2\pi\sigma^2)^{-1} r e^{-\frac{r^2+\rho^2-2r\rho(cos(\phi)cos(\theta)+sin(\phi)sin(\theta))}{2\sigma^2}},\ r,\ \rho,\ \sigma^2 > 0\ -\pi < \phi, \theta \leq \pi, \tag{3.1.1}$$

$$f(\phi) = \frac{e^{-\frac{\rho^2}{2\sigma^2}}}{2\pi}\left[1 + \frac{\rho\ cos(\phi-\theta)}{2\pi\sigma} e^{\frac{\rho^2\ cos^2(\phi-\theta)}{2\sigma^2}} \Phi\left(\frac{\rho\ cos(\phi-\theta)}{\sigma}\right)\right], \tag{3.1.2}$$

where $\Phi(\cdot)$ is the cumulative distribution of a standard normal. Since the phase distribution is quite intimidating at first glance we will look at the asymptotic distribution for large SNR [17]. This distribution simplifies to a normal distribution with mean θ and variance $\tau^2 = \sigma^2/\rho^2$. If the SNR $= 0$ or $\rho = 0$, the distribution limits to the following uniform,

$$f(\phi) = \frac{1}{2\pi},\ -\pi < \phi \leq \pi. \tag{3.1.3}$$

The asymptotic normal distribution will be used when we build our OLS phase regression model. Doing this will make our large SNR assumption imperative if we are to make valid inferences without using numerical methods.

To be consistent with Rowe and Logan (2004) we will give the following argument in

support of the normal limiting distributions for the phase and explain the assumptions which must be present for it to hold thus making OLS a reasonable model. To deal with our time series we introduce the subscript t to distinguish our observed data at different times,

$$\phi_t = \tan^{-1}\left[\frac{\rho_t \sin(\theta_t) + \eta_{I_t}}{\rho_t \cos(\theta_t) + \eta_{R_t}}\right]. \tag{3.1.4}$$

Recall that the distribution assumptions are on η_{I_t} and η_{R_t} which are both distributed as $N(0, \sigma^2)$. Focusing on the argument of the arctan function, call it R_t.

$$\begin{aligned} R_t &= \frac{\rho_t \sin(\theta_t) + \eta_{I_t}}{\rho_t \cos(\theta_t) + \eta_{R_t}} \\ &= \frac{[\rho_t \sin(\theta_t) + \eta_{I_t}]\,\frac{1}{\rho_t \cos(\theta_t)}}{[\rho_t \cos(\theta_t) + \eta_{R_t}]\,\frac{1}{\rho_t \cos(\theta_t)}} \\ &= \frac{\tan(\theta_t) + N(0, \frac{\sigma^2}{\rho_t^2 \cos^2(\theta_t)})}{1 + N(0, \frac{\sigma^2}{\rho_t^2 \cos^2(\theta_t)})}. \end{aligned} \tag{3.1.5}$$

Examining the final equality we can approximate this further when θ_t is small. The ratio of non central normal random variables has been discussed in the statistical literature by Marsaglia [13]. When θ_t is small and the SNR is large, we observe $\cos^2(\theta_t)$ to be close to one and the variances of the random components are very small. We can approximate the denominator as unity and using the Taylor expansion of $\tan(\theta_t) \approx \theta_t$ for small θ_t and reduce R_t to be

$$\begin{aligned} R_t &\approx \frac{N\left(\theta_t, \frac{\sigma^2}{\rho_t^2}\right)}{N\left(1, \frac{\sigma^2}{\rho_t^2}\right)} \\ &\approx N\left(\theta_t, \frac{\sigma^2}{\rho_t^2}\right). \end{aligned} \tag{3.1.6}$$

Now looking at the arctan of the above distribution we see that the Taylor expansion

of $\tan^{-1}(\theta_t)$ is $\approx \theta_t - \frac{\theta_t^3}{3} + \frac{\theta_t^5}{5}$. Since we have the restriction that θ_t is small we can neglect all the terms raised to powers greater than one. This agrees with the marginal distribution of ϕ_t we derived earlier. So,

$$\begin{aligned} \phi_t &= \tan^{-1}\left(\frac{\rho\sin(\theta_t)+\eta_{I_t}}{\rho\cos(\theta_t)+\eta_{R_t}}\right) \\ &\approx N\left(\theta_t, \frac{\sigma^2}{\rho_t^2}\right) \end{aligned} \tag{3.1.7}$$

for larger SNR and small θ_t. However, this argument is only valid for very small θ_t. More generally the limiting distribution can be derived from the marginal phase distribution for large SNR and θ_t.

As mentioned, phase data is rarely analyzed in fMRI, and the times it is, OLS is the modeling methodology of choice [3]. Before an OLS model can be fit, the time series needs to be artifically "unwrapped." The process of artifically "unwrapping" the data and then fitting an OLS model to it will be described shortly. Menon (2002) looked at phase time series after unwrapping and noticed an approximate linear association with the magnitude data which he believed was attributed to large blood vessels [14]. The OLS model will be briefly summarized then an alternative model will be introduced along with its advantages and disadvantages.

The OLS phase-only data model can be written for an arbitrary voxel in the following form,

$$\phi_t = z_t'\gamma + \delta_t, \tag{3.1.8}$$

where ϕ_t are observed phase angle measurements at each time point t and z_t' is a

specific row of the design matrix Z in Equation 3.1.9, γ are the fixed but unknown phase regression coefficients, and δ_t is the measurement error. This measurement error is assumed to originate from the following normal distribution, $N(0, \tau_t^2)$ where $\tau_t^2 = \sigma^2/\rho_t^2$, and the design matrix for our example Z is constructed in the following way for three covariates,

$$Z = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 1 \\ \vdots & \vdots & \vdots \\ 1 & 17 & -1 \\ \vdots & \vdots & \vdots \\ 1 & 256 & -1 \end{pmatrix}. \tag{3.1.9}$$

The above design matrix includes a column of 1's to model the intercept, a column of counting numbers to account for a possible linear trend over time [20], and a third column with alternating sets of 16 1's and -1's to account for the a task related function. Sets of 16 are used in our matrix to coincide with both our simulated and acquired data which have stimulus lengths of 16 seconds. We define SNR by $\rho_t/\sigma = (\beta_0 + \beta_1 t + \beta_2 z_{2t})/\sigma$ but $\beta_1 t + \beta_2 z_{2t}$ is generally very small when compared to β_0 within the brain and zero outside the brain so we utilize the following approximation, $\rho_t/\sigma \approx \beta_0/\sigma$ and thus the variance τ^2 becomes constant over time. When we describe the phase-only data with the OLS model we can make inferences using standard linear contrasts tests, H_0: $D\gamma = 0$ vs H_1: $D\gamma \neq 0$. Our contrast matrix, D, is defined to be a $w \times (q+1)$ matrix. This testing procedure leads to the unrestricted, $(\hat{\gamma}, \hat{\tau}^2)$,

and restricted, $(\tilde{\gamma}, \tilde{\tau}^2)$, maximum likelihood estimates for γ and τ^2 having the familiar form,

$$\hat{\gamma} = (Z'Z)^{-1}Z'\phi \tag{3.1.10}$$

$$\hat{\tau}^2 = \frac{(\phi - Z\hat{\gamma})'(\phi - Z\hat{\gamma})}{n} \tag{3.1.11}$$

$$\tilde{\gamma} = \Psi\hat{\gamma} \tag{3.1.12}$$

$$\tilde{\tau}^2 = \frac{(\phi - Z\tilde{\gamma})'(\phi - Z\tilde{\gamma})}{n} \tag{3.1.13}$$

and $\Psi = I - (Z'Z)^{-1}D'[D(Z'Z)^{-1}D']^{-1}D$. Using the asymptotic χ^2_w distribution of -2 log of a likelihood ratio test we note that the test statistic is

$$-2\log\Lambda = n\log\left(\frac{\tilde{\tau}^2}{\hat{\tau}^2}\right). \tag{3.1.14}$$

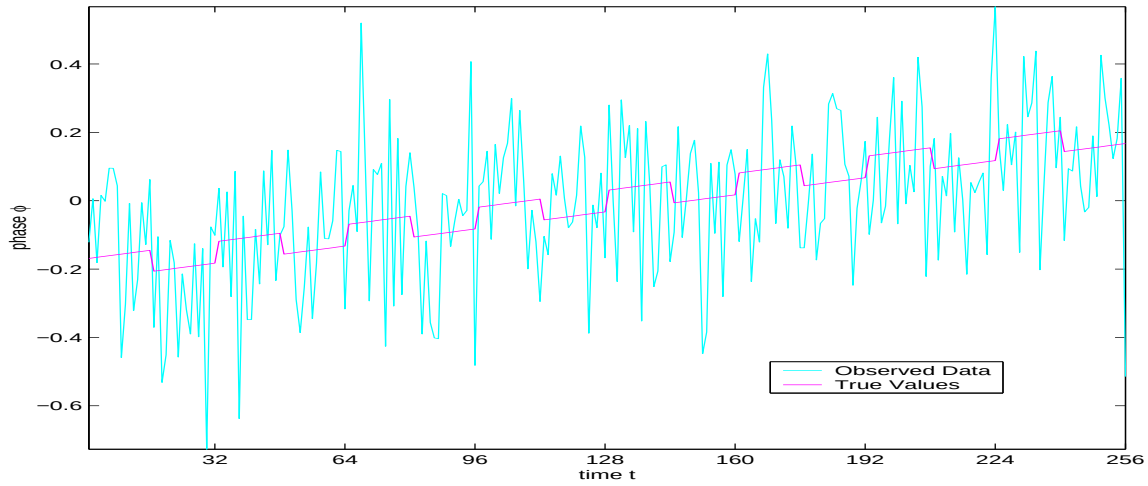
The degrees of freedom, w, is equal to the full row rank of D [15]. We also may test individual hypothesis by using the large sample normal statistics because the number of time points present.

The angular response of the phase-only OLS model can cause modeling problems at the wrap around junction if not accounted for properly thus resulting in poor parameter estimation and incorrect inferences. When a combination of the three following conditions are valid we can describe the phase data with little concern using our standard OLS regression model. These conditions are 1) a large SNR, 2) our baseline angle, γ_0, is not near either the upper boundary of π or the lower boundary of $-\pi$, and 3) the linear trend is small enough that the data does not rise or lower beyond π and $-\pi$. The large SNR assumption makes the probability that

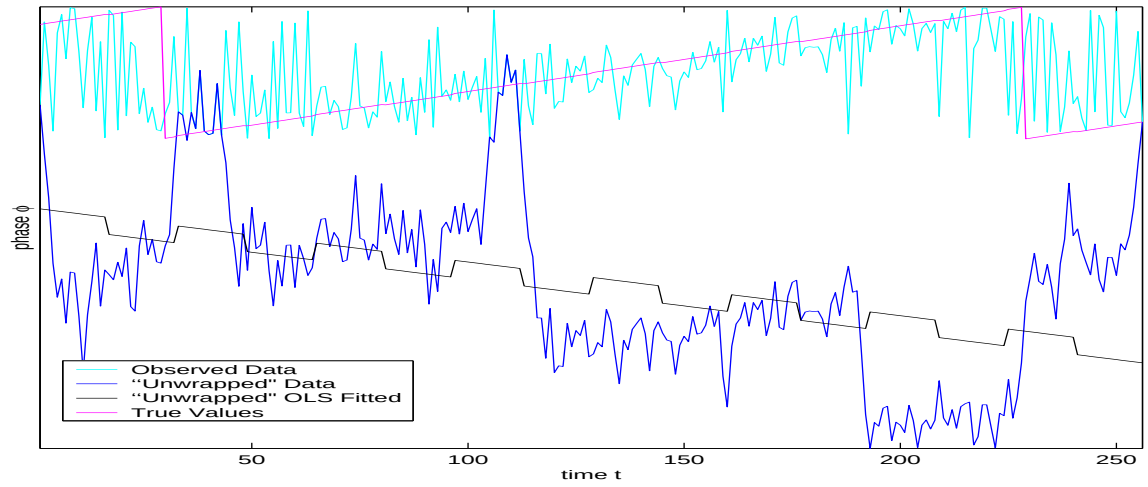
successive measurements have a large difference between them to be very small. A simulated time series following the aforementioned conditions is plotted in Figure 3.1(a).

However, in real acquired fMRI data the mean phase in voxels varies greatly and these stringent conditions are in general not met. Often within an fMRI data set, phase angles are observed close to $\pm\pi$. An example of phase data which "wraps" around the phase boundary is seen in Figure 3.1(b) where the cyan time series is the original data and the blue time series is the "unwrapped" data. As previously mentioned, the method generally used to deal with this issue is "unwrapping." Unwrapping is the process of beginning with the first observation, proceeding through the time series and "flagging" an observation in which the next point in the time series have an absolute difference greater than or equal to a predefined value, generally π, then shifting the rest of the time series by multiples of $\pm 2\pi$. The process is repeated to the end of the time series. However, when high SNR becomes suspicious the model fails and the investigator needs another model to work with.

Often times simply "unwrapping" the data and fitting an OLS regression line to the data is sufficient as in Figure 3.1(a). This situation can be defined as "control" data in which no "wrap-around" is present and the SNR is large. However, as we see in Figure 3.1(b) this isn't always true. We see the fitted OLS regression line (black) for the "unwrapped" data (blue) doesn't model the data correctly and often this will lead to incorrect inferences. The associated reference function jumps are very



(a) "Control" fMRI Phase Data



(b) Original, "Unwrapped" Phase Data, and OLS Fitted Line

Figure 3.1: Part a: Shows data where no "unwrapping" is needed. Part b: Shows an example of an improper OLS fitted line (Black) to "unwrapped" phase time series (Blue) along with the original data (Cyan) and true values (Magenta).

small when looking at the true values (magenta) causing the line to appear nearly straight. The fitted regression line has a negative slope while the true value of the slope coefficient is positive. This implies the need for another model that can account for the angular response. Simply "unwrapping" the data does not guarantee OLS to be a correct method to deal with the "wrap-around" issue, especially when the SNR is low. Instead, we explore various angular models and look at fitting the data one. Implementing an angular model procedure to deal with the angular nature of the response will allow us to relax the large SNR requirement.

3.2 Circular Phase Model

3.2.1 Distribution

Since the unwrapped OLS model fails for various conditions including a low SNR we will explore distributions and models that deal with data similar to our phase measurements. The literature dealing with angular regressions include models by Gould [6], Johnson and Wehrly [10], along with Fisher and Lee [5]. We will expand on these and choose a preferred model for our simulated and real fMRI phase data.

First, we would like to introduce the most common distribution used with angular random variables and which also is the corner stone of our angular regression techniques. This distribution goes by two names, the circular normal and the Von Mises.

We will use these two interchangeably. The conditional distribution of the phase given the magnitude is actually a circular normal distribution [9] and is derived in Appendix A.2. This fact alone should give us confidence in using models based on this distribution. The circular normal has the following probability density function (PDF),

$$f(\phi) = \frac{e^{\kappa \cos(\phi - \theta)}}{2\pi I_0(\kappa)}, \quad -\pi < \phi, \theta \leq \pi, 0 < \phi, \theta < \infty. \tag{3.2.15}$$

This distribution has mean θ and concentration κ.

3.2.2 Models

As describe in Jammalamadaka and SenGupta (2001), Gould (1969) introduced a model which estimated the overall mean direction of each data point using the following form,

$$\theta_t = \gamma_0 + \sum_{j=1}^{m} \gamma_j z_{jt} (\text{mod } 2\pi) \tag{3.2.16}$$

with $j = 1, \ldots, m$, where m is the number of covariates, and $t = 1 \ldots n$, n being the number of time points in a single voxel. The modulus, denoted by mod, is defined to be the remainder of each $\gamma_j z_{jt}$ quantity when dividing by the argument, 2π. The log likelihood for this angular regression with a single regressor,γ, i.e. $m = 1$, case is

$$\log L = -n \log 2\pi \ + \ \kappa \sum_{t=1}^{n} \cos(\phi_t - \gamma_0 - z_{1t}\gamma_1) - n \log I_0(\kappa). \tag{3.2.17}$$

As discussed by Johnson and Wehrly, this model has some advantages [10]. These advantages include the ability to test $\gamma_1 = 0$ and the ability to estimate the unknown

parameters using an iterative process which Gould laid out in his paper. Yet, this model has certain limitations that arise when considering it for our fMRI phase time series analysis. This model assumes the angle continuously "wraps" around resulting in what is referred to as the "barber's pole" form. However, in fMRI data the phase angle rarely wraps around more than once unless the SNR is extremely low as would by seen outside the brain. Johnson and Wehrly also discussed the fact that the MLEs are non-unique in this model because the likelihood function has infinitely many equally high peaks.

Johnson and Wehrly (1978) also introduced an alternative model to be used for angular data with linear predictors. Their model assumed that the predictor variable, z_{1t}, is random and has a known distribution. This allows them to derive the joint distribution of ϕ_t and z_{1t},

$$f(\phi_t, z_{1t}) \;=\; 2\pi g\left[2\pi(F_1(\phi_t) - F_2(z_{1t}))\right] f_1(\phi_t) f_2(z_{1t}),\ 0 \le \phi_t < 2\pi,\ -\infty < z_{1t} < \infty$$

where $F(\phi_t), F(z_{1t}), f(\phi_t), f(z_{1t})$ are marginal cumulative density functions (CDFs) and marginal PDFs for ϕ_t and z_{1t} respectively. The function $g(\cdot)$ is an a priori defined density on a unit circle. This density $g(\cdot)$ is often chosen to be the Von Mises distribution because it is the most well known. The authors accordingly chose $g(\cdot)$ to be the Von Mises distribution for their example, making the conditional distribution of ϕ_t given z_{1t} to be the following,

$$f(\phi|Z) \;=\; \left[2\pi I_0(\kappa)\right]^{-n} e^{\kappa \sum_{t=1}^{n} \cos(\phi_t - \gamma_0 - 2\pi F(z_{1t}))}.$$

where $\phi = (\phi_1, \ldots, \phi_n)'$. The authors also show that the maximum likelihood estimates for the mean angle, γ_0, and the concentration, κ, are well defined for this model. From this distribution Johnson and Wehrly show the maximum likelihood estimates of γ_0 and κ to be $\hat{\gamma_0} = \bar{z}_0$ and $\hat{\kappa}$ from $\bar{R} = A(\hat{\kappa})$, where

$$\bar{R}\cos(\bar{z_0}) = \frac{1}{n}\sum_{t=1}^{n}\cos(\phi_t - 2\pi F(z_t)) \tag{3.2.18}$$

and

$$\bar{R}\sin(\bar{z}_0) = \frac{1}{n}\sum_{t=1}^{n}\sin(\phi_t - 2\pi F(z_t)), \tag{3.2.19}$$

$$A(\bar{R}) = I_1(\hat{\kappa})/I_0(\hat{\kappa}) \tag{3.2.20}$$

and $I_m(\cdot)$ is the modified Bessel function of the first kind and order m [10].

This model does differ from Gould's angular model because it restricts ϕ_t from "wrapping" around more than once. In other words, the model by Johnson and Wehrly doesn't allow more than one complete rotation of the phase on the "barber pole" as z_{1t} varies over time. Often times this is a reasonable assumption as in fMRI phase data.

However, this model does have drawbacks of its own including the need to know the distribution of z_{1t}. We often do not know the PDF or CDF of our predictor variables and tend to view them as fixed numbers that don't originate from a known distribution. Also, this model only allows for a single predictor z_{1t}. We may like to choose a design matrix Z as described in the previous section thus allowing us to include both a possible linear trend and reference function into our model.

Fisher and Lee (1992) generalize the Johnson and Wehrly model to allow for multiple predictor variables and relaxed the need for distributional assumptions on the design matrix [5]. Their model assumes the angular observations, $\phi_1, \ldots, \phi_n$, are independent and follow a Von Mises distribution with a constant concentration parameter, κ, that is the same across the series. In other words, $\phi_1, \ldots, \phi_n$ each originate from a Von Mises distribution with mean θ_t and concentration κ. Since they keep the concentration parameter fixed over time they model using only mean directions. They mention a model that fixes the mean and models the concentration but this is not appropriate for functional "phase" activation. The fixed concentration model is given by

$$\theta_t \quad = \quad \gamma_0 + g(u_t'\gamma) \tag{3.2.21}$$

where the design matrix U, with the t^{th} row $u_t' = (u_{1t}, \ldots, u_{(q+1)t})$, is comprised of all columns except the first in the design matrix Z, as defined in the previous section. Of course U can have any combination of columns the investigator chooses but we will focus on our choice. There is no need to include the baseline column in U because the intercept is already estimated within the model. The "link" function $g(\cdot)$ has the purpose of mapping the result to the range of $-\pi$ to π. As stated by the authors this link function corrects the non-identifiability problem of MLEs that is present in the Gould model [5]. One possible link function that we will use which is given by Fisher and Lee is,

$$g(\cdot) \quad = \quad 2\tan^{-1}\left(\text{sgn}(\cdot)\,|\cdot|^{\xi}\right) \tag{3.2.22}$$

where sgn$(\cdot)$ is the operator that returns the sign of its argument and the transformation parameter, ξ, can be estimated from the data similar to the Box-Cox transformation [4, 5].

Another form used by the authors to model the directional mean was

$$\theta_t \;=\; \gamma_0 \pm 2\pi g(z_i). \tag{3.2.23}$$

Here z are lying on the bounded region $[0,1]^k$ after scaling and k being the number of columns in our design matrix. This allows the link function, $g(\cdot)$, to be within some flexible parametric family of k-dimensional distribution. The authors gave an example of having the link function equal the incomplete beta distribution, $I_x(\alpha,\beta)$,for the single covariate case. This distribution has the following PDF,

$$I_x(\alpha,\beta) \;=\; \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\int_{t=0}^{x} t^{\alpha-1}(1-t)^{\beta-1}dt. \tag{3.2.24}$$

We won't be exploring this model of the directional mean because we want to allow our design matrix to be as robust as needed while defining the link function to be as simple as possible.

Fisher and Lee give us equations to obtain parameter estimates and make inferences using the mean model $\theta_t = \gamma_0 + g(u_t'\gamma)$ with the link function defined as $g(\cdot) = 2\tan^{-1}(\cdot)$. Of course the link function can be chosen differently but we will illustrate fitting the Fisher and Lee model with this particular choice. First, we define the natural log likelihood, denoted log, of the Von Mises distribution,

$$\log L = -n\log 2\pi \;\; - n\log I_0(\kappa) \;\; + \kappa\sum_{t=1}^{n}\cos(\phi_t - \gamma_0 - g(u_t'\gamma)). \tag{3.2.25}$$

Second, the authors define the following

$$\begin{aligned} v_t &= \sin(\phi_t - \gamma_0 - g(u_t'\gamma)), \\ v &= (v_1, \ldots, v_n)', \\ U &= (u_1, \ldots, u_n)' \\ G &= \mathrm{diag}(g'(u_1'\gamma), \ldots, g'(u_n'\gamma)) \\ S &= \frac{1}{n}\sum_{t=1}^{n} \sin(\phi_t - g(u_t'\gamma)) \\ C &= \frac{1}{n}\sum_{t=1}^{n} \cos(\phi_t - g(u_t'\gamma)) \\ R &= (S^2 + C^2)^{1/2}. \end{aligned}$$

In the above, variables have the following dimensions: v is a $n \times 1$ vector, G is a $n \times n$ matrix; while S, C, and R are scalars. The function, $g'(\cdot)$ is defined to be the derivative of the link function and not the transpose. Next, the MLEs are found by solving the following equations,

$$U'Gv = 0 \tag{3.2.26}$$

$$R\sin(\hat{\gamma}_0) = S, \tag{3.2.27}$$

$$R\cos(\hat{\gamma}_0) = C, \tag{3.2.28}$$

$$A(\hat{\kappa}) = R \tag{3.2.29}$$

where $A(\kappa) = I_1(\kappa)/I_0(\kappa)$. Fisher and Lee describe an iterative procedure for finding a solution to these equations. They also noted that centering the individual columns of U around their means will optimize the numeric calculations. We begin with an

initial value for $\hat{\gamma}$ and calculate values for S, C, and R from the above equations. An updated value of $\hat{\gamma}$, denoted by $\hat{\gamma}^*$ is then found by solving the following equation for $\hat{\gamma}^*$,

$$(U'G^2U)(\hat{\gamma}^* - \hat{\gamma}) \;=\; U'G^2y \tag{3.2.30}$$

where $y = (y_1, \ldots, y_n)'$ and $y_t = \frac{v_t}{[A(\hat{\kappa})g'(u_t'\hat{\gamma})]}$. We also find an updated estimate for γ_0 at each iteration by solving $\gamma_0 = \tan^{-1}(S/C)$ and κ from $A(\kappa)$. The updated estimate of $\hat{\gamma}^*$ is recursively placed back into the formulas until a convergence criteria is achieved such as a set number of iterations or until the iterative values differ by less than some pre-defined amount.

Fisher and Lee also give us a solution to find the large sample asymptotic variance of the estimated coefficient vector,

$$var(\hat{\gamma}) \;=\; \frac{1}{\hat{\kappa}A(\hat{\kappa})}\left\{\left(U'G^2U\right)^{-1} + \frac{\left(U'G^2U\right)^{-1}U'gg'U\left(U'G^2U\right)^{-1}}{\left(n - g'U\left(U'G^2U\right)^{-1}U'g\right)}\right\},$$

which will allow us to draw inferences on our γ's where g is a vector whose elements are the diagonal elements of G. They also describe the asymptotic variance for $\hat{\gamma}_0$ and $\hat{\kappa}$ to be equal to $1/(nA'(\hat{k}))$ and $[2(n-q)\hat{\kappa}A(\hat{\kappa})]^{\frac{1}{2}}$, respectively where $A'(\kappa)$ is the derivative of the ratio of Bessel functions with respect to $\hat{\kappa}$. The variance of the asymptotic normal limiting distribution of the Von Mises is $1/\kappa$ [9].

This now provides us with estimates for out model parameters including the phase regression coefficients, γ, and their variances. We can then use a large sample normal

approximation to test the hypothesis, H_0: $\gamma_m = 0$ verses H_1: $\gamma_m \neq 0$ using the constructed test statistics, $\hat{\gamma}_m / \sqrt{var(\hat{\gamma}_m)}$. Alternatively one could set up linear contrast hypothesis tests and obtain the $-2\log$ of the ratio of the unrestricted likelihood over the restricted and use the asymptotic χ^2_w distribution to make inferences.

Chapter 4

Simulations

We generated data to simulate activation in a voxel which is similar to that observed from a bilateral finger tapping fMRI block design experiment as described in the next Chapter [17]. The simulated time series consisted of $n = 256$ points where the true values for the data are known before random noise according to a pre-specified distribution is added.

Simulated fMRI data is constructed according to a general non-linear model as described by Rowe and Logan (2005) which for the magnitude consists of an intercept β_0; a time trend coefficient β_1; and a coefficient β_2 for a reference function, related to a block experimental design [18]. We also include regression coefficients on our phase change which consists of γ_0, γ_1, and γ_2. These specified γ's are also determinant of an intercept, trend, and a reference function. This allows the complex-valued data to

have the following form,

$$y_t = [(\beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t})\cos(\theta_t) + \eta_{Rt}] + i\,[(\beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t})\sin(\theta_t) + \eta_{It}] \quad (4.0.1)$$

where

$$\theta_t = (\gamma_0 + \gamma_1 u_{1t} + \gamma_3 u_{2t}) \quad (4.0.2)$$

$t = 1, \ldots, n$, and $(\eta_{Rt}, \eta_{It}) \sim N(0, I_2\sigma^2)$. After creating the data we obtained the respective phase time series by taking the four quadrant arctangent of the imaginary component over the real component which we have previously shown to have the complicated distribution given in Equation 3.1.2 which is normal for large SNR.

For the current simulations we looked at two basic cases for the fMRI phase time series and fit both the standard OLS model along with the Fisher and Lee angular model to the data. We looked at the two fits graphically and examined the parameter estimates along with the variances of the parameter estimates for regression coefficients γ with each model. The two cases include a "control" case where there is no "wrap-around" present and a "test" case where the issue of "wrap-around" is present, both at equal SNR and magnitude parameters. We use the first time series to compare the two models when OLS could be used with little concern and verify the Fisher and Lee model is working properly. The latter phase time series will demonstrate to the reader that the OLS model poorly fits the data when the stringent conditions are not met. We will denote estimates under the unconstrained alternative hypothesis for the OLS model to be $\hat{\gamma}_{OLS}$ and those from the Fisher and Lee model with $\hat{\gamma}_{FL}$

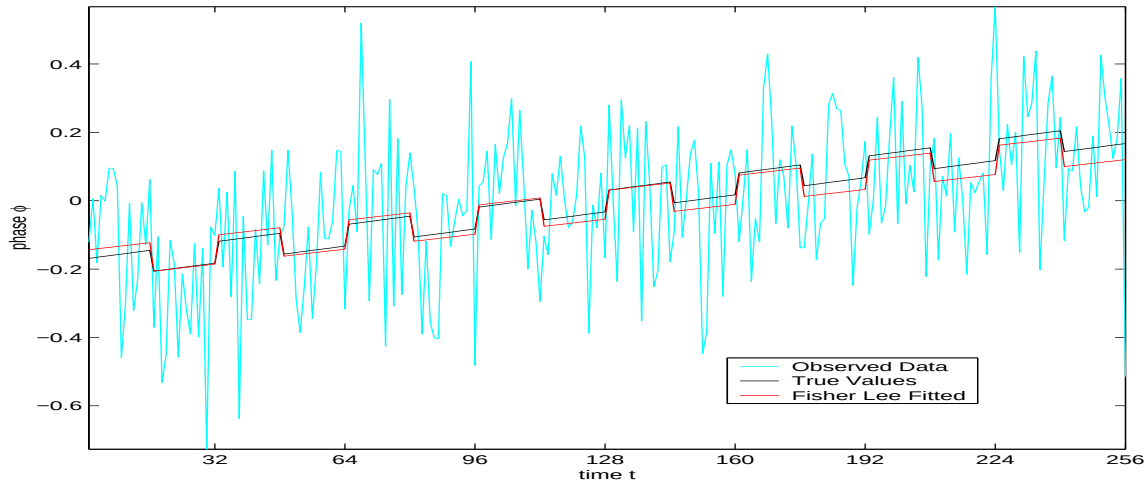
The first time series has an SNR=10 with the true values being $(\beta_0, \beta_1, \beta_2) = (0.25, 0.008, 0.05)$, $(\gamma_0, \gamma_1, \gamma_2) = (0, 0.2, 8)$, and $\sigma^2 = 0.05$. We obtain estimates for the OLS model to be $\hat{\gamma}_{0_{OLS}} = -0.011$, $\hat{\gamma}_{1_{OLS}} = 0.175$, and $\hat{\gamma}_{2_{OLS}} = 10.805$. The estimates for the Fisher and Lee angular regression are $\hat{\gamma}_{0_{FL}} = -0.011$, $\hat{\gamma}_{1_{FL}} = 0.088$, and $\hat{\gamma}_{2_{FL}} = 5.396$. The respective variances are $var(\hat{\gamma}_{0_{OLS}}) = 0.0001$, $var(\hat{\gamma}_{1_{OLS}}) = 0.0004$, and $var(\hat{\gamma}_{2_{OLS}}) = 9.699$ for the OLS model while they are $var(\hat{\gamma}_{0_{FL}}) = 0.000072$, $var(\hat{\gamma}_{1_{FL}}) = 0.0001$, and $var(\hat{\gamma}_{2_{FL}}) = 2.411$ for the angular model. The OLS model estimate $\hat{\sigma}^2$ was equal to 0.037 while the Fisher and Lee model estimate of $1/\hat{\kappa}$ equaled 0.0363. In Figure 4.1(a) the true phase signal is shown with the black line, the observed time series (with noise) is plotted with cyan and the Fisher and Lee fitted model is plotted as red. Since that the OLS and Fisher and Lee models gives us identical curves when no "wrap-around" is present we omitted the overlapping plot of the OLS curve. The decisions based on test the statistics for H_0: $\gamma_2 = 0$ vs H_1: $\gamma_2 \neq 0$ also agreed when using an asymptotic 5% two sided critical value of 1.96. The test statistic values were 3.469 for the OLS model and 3.476 for the Fisher and Lee model.

The second time series also has an SNR=10 with identical β and σ^2 values as the previous simulation but now with true values of $(\gamma_0, \gamma_1, \gamma_2) = (-1, 3.5, 40)$. We obtained the estimates for the OLS model to be $\hat{\gamma}_{0_{OLS}} = 0.351$, $\hat{\gamma}_{1_{OLS}} = 0.323$, and $\hat{\gamma}_{2_{OLS}} = 17.127$. The estimates for the Fisher and Lee angular regression are $\hat{\gamma}_{0_{FL}} = -1.012$, $\hat{\gamma}_{1_{FL}} = 2.951$, and $\hat{\gamma}_{2_{FL}} = 36.414$. The respective variances are

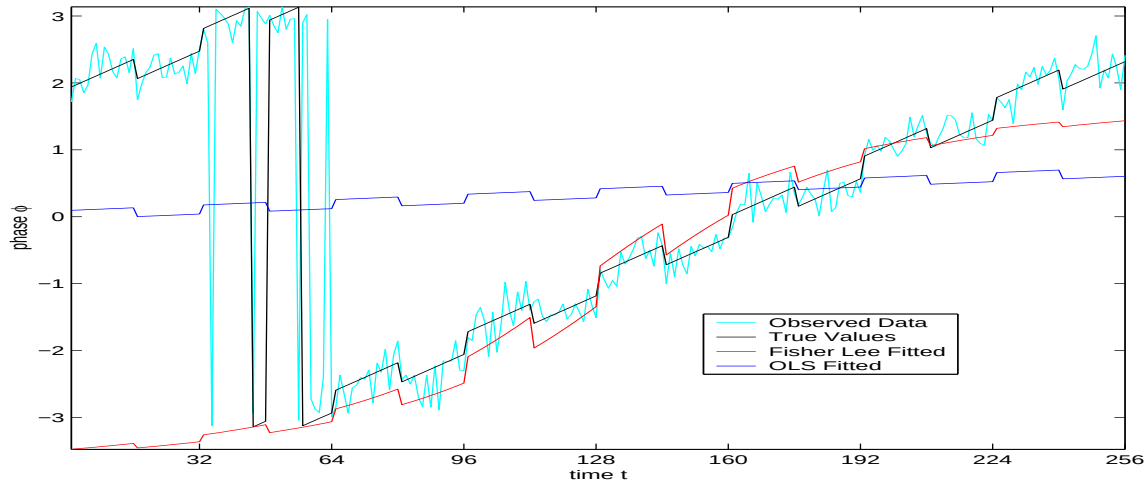
$var(\hat{\gamma}_{0_{OLS}}) = 0.013$, $var(\hat{\gamma}_{1_{OLS}}) = 0.04$, and $var(\hat{\gamma}_{2_{OLS}}) = 874.597$ for the OLS model while they are $var(\hat{\gamma}_{0_{FL}}) = .00042$, $var(\hat{\gamma}_{1_{FL}}) = 0.011$, and $var(\hat{\gamma}_{2_{FL}}) = 53.471$ for the Fisher and Lee angular model. The OLS model estimate $\hat{\sigma}^2$ was equal to 0.2997 while the Fisher and Lee model estimate of $1/\hat{\kappa}$ equaled 0.1905. The decisions based on the test statistics for H_0: $\gamma_2 = 0$ vs H_1: $\gamma_2 \neq 0$ do not agree when using an asymptotic two sided 5 % critical value of 1.96. The test statistic values were 0.579 for the OLS model and 4.9792 for the Fisher and Lee model.

By looking at Figure 4.1(b) we can see that reference function or third column of U has a strong correlation with the phase time series values and we would agree with the Fisher and Lee decision that it is statistically significant. In Figure 4.1(b) the true noiseless phase signal is shown with the black line, the simulated observed time series (with noise) is plotted with cyan, the OLS fitted model is signified with blue, and the Fisher and Lee fitted model is plotted as red. We see that the OLS and Fisher and Lee models no longer give us consistent fits. The OLS model doesn't fit the data well and is situated between most of the observations. The estimated values for the γ's were much closer to the true values for the Fisher and Lee model compared with the OLS model. Also, the variances for the Fisher and Lee estimates are much smaller especially for $\hat{\gamma}_{2_{FL}}$ which is the coefficient of primary interest. Further extending our exploration of the γ_2 estimates for both models we generated 10,000 random data sets with the same parameter values and compared the estimates for γ_2 using both models. The sample mean for all 10,000 estimates using the Fisher and Lee model

along with the OLS model were 32.41 and 9.64, respectively. The sample variance for these 10,000 estimates of γ_2 were 13.28 and 209.22 for the Fisher and Lee along with the OLS model, respectively. This suggests that the Fisher and Lee angular regression model is an improved alternative to OLS when modeling fMRI phase time series because in has the potential for more accurate and precise estimates when dealing with a reference function coefficient.



(a) OLS Fitted Line



(b) Both OLS and Fisher and Lee Fitted Line

Figure 4.1: Part a: True signal (Black), observed signal (Cyan), and Fisher and Lee fitted line (Red). Part b: True signal (Black), observed signal (Cyan), OLS fitted line (Blue), and Fisher and Lee fitted line (Red)

Chapter 5

Real Data Example

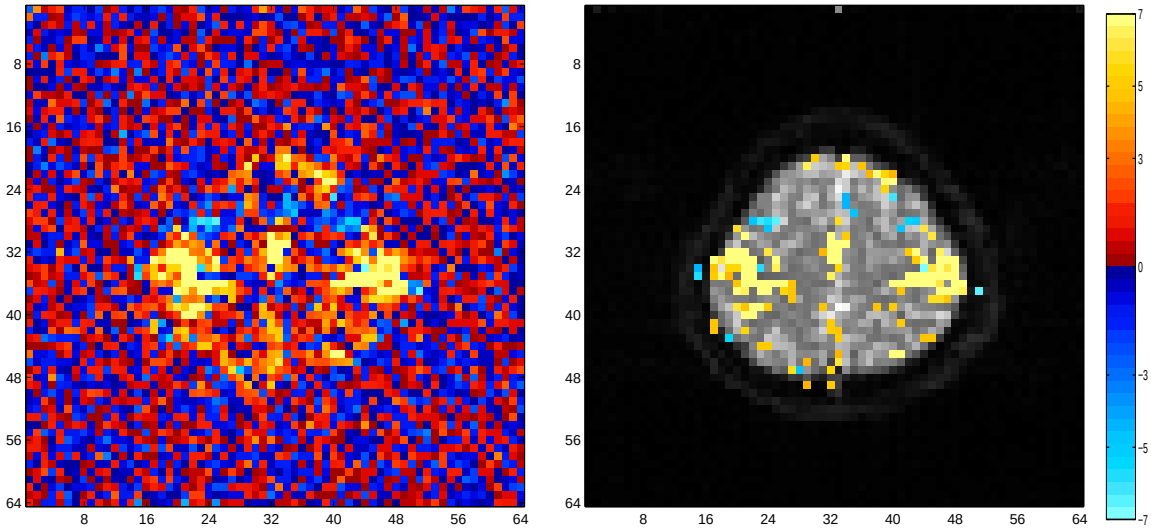
We now will compare the two models using actual data acquired during a sequential bilateral finger tapping fMRI block design experiment. There are a total of 4096 voxel time series comprised of n=256 observations per voxel. These 256 observations consists of alternating sets of 16 observations on and off. An axial slice through the motor cortex is examined here. Scanning used a 1.5T GE Signa scanner, where 5 axial slices of 64×64 where acquired. Voxels were 3.125 × 3.125mm in plane and 5mm thick with TR=1000ms, TE=47ms, and filtering was applied to remove low frequency and respiration noise. Before examining the phase-only data, we looked at the magnitude-only data. The activation regions for the magnitude-only OLS model can be seen in Figure 5.1.

If the phase-only data is to contain information regarding possible biological phenomena in the brain we would like to see "phase" activations to be in similar locations

as the "magnitude" activations. The association between magnitude-only and phase-only time series observed by Menon for areas with large blood vessels would suggest "phase" activations can be found in similar places as "magnitude" activations given such blood vessels are present [14]. Any similarities between the statistics for the magnitude-only and phase-only models would strengthen the idea that valuable information is discarded when magnitude-only data is analyzed solely.

We fit each time series with both an OLS regression and a Fisher and Lee angular regression then compared the results. Due to the erratic nature of phase data outside of the brain we observed an increased number of statistically significant "phase" activations in this area with the Fisher and Lee model. These are observed with the Fisher and Lee model and not with the OLS model because the Fisher and Lee model is able to fit a more accurate model because of the lower SNR restriction. These "phase" activations outside the brain are obviously not due to any biological phenomena but solely the result of a more robust model fit.

An issue which arises when implementing the Fisher and Lee model with our given link function is convergence of the γ estimates for a small number of voxel time series. When a large concentration of the points in a given series are near $\pm\pi/2$, the limits of our arctangent function, the Fisher and Lee model does not always to converge. However, this issue is easily corrected by subtracting the angular mean from the entire series for those particular voxels before fitting with the Fisher and Lee method. The investigator is then able to obtain estimates and make inferences using this improved



(a) (b)

Figure 5.1: Part a shows the unthresholded OLS test statistics for H_0: $\beta_2 = 0$ using the magnitude-only data. Part b shows the thresholded statistics

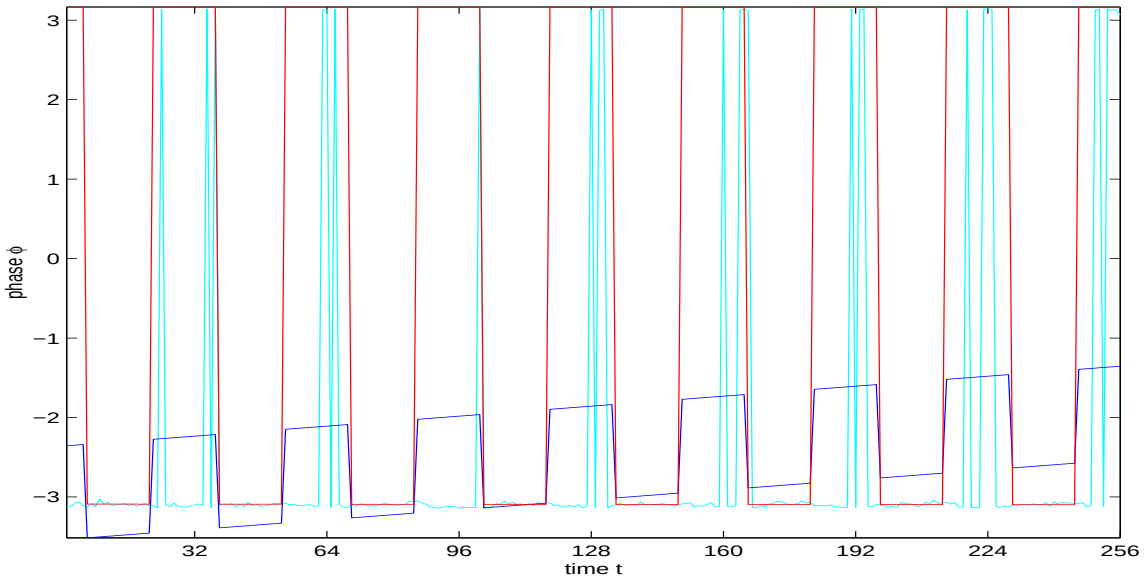
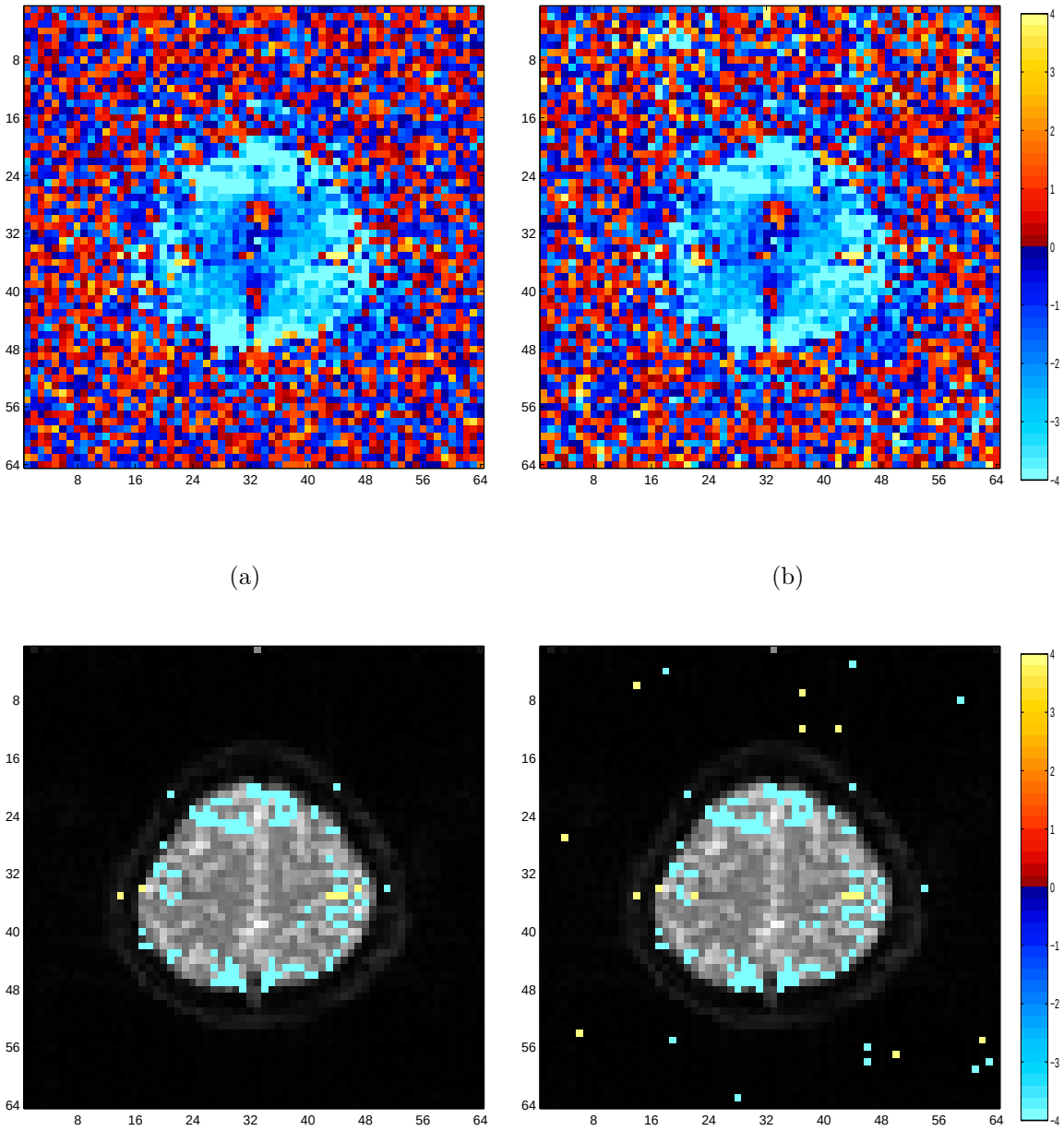


Figure 5.2: Actual acquired fMRI phase time series (Cyan), OLS fitted values (Blue), along with Fisher and Lee fitted values (Red).

model. An example which shows the robustness of the Fisher and Lee model can be seen in Figure 5.2 where the observed phase time series from an actual experimental voxel is shown in cyan while the fitted series for the OLS along with the Fisher and Lee models are shown with blue and red, respectively. It is easily seen that the Fisher and Lee angular regression accurately models the large "spikes" in the data during the finger tapping task and stays within the boundaries of $-\pi$ and π. Yet, the OLS fit doesn't accurately model the correct intensity of the "spikes" and actually has fitted values below the range of $-\pi$. It should be noted that this figure is truly a three dimensional cylinder instead of a two dimensional plot where the "spikes" are observed jittering around the $\pm\pi$ boundary. In other words, the plot takes the long way around the cylinder to connect points.

Since we are interested in possible associations between the phase-only data and a pre-defined reference function we constructed large sample asymptotic z-values to make inference about γ_2. We monitored these statistics for all 4096 voxels in both the OLS and the Fisher and Lee models. In Figures 5.3(a) and 5.3(b) we show the unthresholded statistics map for both OLS and the Fisher and Lee models. We see the statistics within the brain appear to agree almost identically because the time series don't exhibit large variation similar to the region outside the brain. It should be noted that some of the phase time series on the border of the brain had angular means near the $\pm\pi/2$ value and needed to be centered to obtain convergence. After applying a threshold that was Bonferroni corrected for multiple comparisons as described in Logan and Rowe (2004) with 5% family wise error (FWE) rate we present the activation maps in Figures 5.3(c) and 5.3(d) [12]. Again we see similar conclusions to the OLS model with the exception of additional "phase" activations outside the brain. Again, we attribute this to the robustness of the Fisher and Lee model when applied to very noisy data. Since the Fisher and Lee model more accurately characterizes troublesome data we would expect more statistical significance to be found for the reference function coefficient, γ_2. Troublesome voxels can be seen in the brain as well as seen by Figure 5.2 which is the "yellow" voxel from within the right side of the region. The focal positive "magnitude" activations agree with those of the "phase" activations which would imply valuable information in the phase-only data.



(a) (b)

(c) (d)

Figure 5.3: Parts a and b Shows the unthresholded OLS and Fisher and Lee test statistics for H_0: $\gamma_2 = 0$ using the phase-only data. Parts c and d Shows the thresholded test statistics for both OLS and Fisher and Lee models using the same hypothesis.

Chapter 6

Conclusion

Modeling fMRI phase time series with OLS regression results in some troublesome phenomena which include poor fit, incorrect parameter estimation, and inaccurate test statistics. Most of these problem arise from the issue of "wrap-around" in the time series. We discussed three linear-circular regression techniques and concluded that the model proposed by Fisher and Lee (1992) was the best choice for the problem at hand. It allowed us to define a design matrix which could account for several regressors including a linear trend and a reference function from which to make and test hypothesis using a large sample asymptotic z statistic. We were able to implement a numeric algorithm given by Fisher and Lee to obtain our coefficient estimates along with the variances of the coefficient estimate and thus test hypotheses.

Simulations were performed for both a "control" case with no "wrap-around" and a "test" case with "wrap-around" present. In the case of no "wrap-around" we found

that the fitted line for the Fisher and Lee model was nearly identical to that of the OLS model. The test statistics also were very similar and the conclusion for each agreed. This would imply the models are equivalent under these conditions. In the generated time series where "wrap-around" was present we noted that the estimates for the Fisher and Lee model were much closer to the true values compared with the OLS model. Also, the test statistics for the Fisher and Lee model differed from those for the OLS model and led to different conclusions with respect to γ_2 for the case with "wrap-around." We also showed during our second set of simulations that the γ_2 estimates were more accurate and precise when we compared the Fisher and Lee model to the OLS model for the 10,000 simulated phase-only data sets created with the same parameters. This should give the reader confidence in implementing this model when the data contains "wrap-around."

The actual phase time series we presented solidified our findings that the Fisher and Lee model is an excellent choice for fitting with fMRI phase-only data. We showed a specific voxel example where OLS was unable to match the modeling accuracy of the Fisher and Lee model and actually contained fitted values not consistent with the fMRI phase angular property. Fisher and Lee's model is less suseptible to low SNR problems and if applied to other real data examples with more noise we would expect it to perform at as well as OLS. This newly implemented angular model does detect temporal correlations between the phase time course and a reference function in many of the same places the OLS model does. We suggest using the Fisher and

Lee angular model as both a comparative tool along with a exploratory method for fMRI phase data. It may be more beneficial for the investigator to implement this new angular method on a voxel-wise basis where the fitted values can be compared to both the observed data and the OLS fit allowing the appropriate modeling decision to be made. Finally, as we have shown with simulated data the Fisher and Lee model has less assumptions that need to be met and often times describes the data better compared with OLS regression.

Appendix A

Appendix

A.1 Joint and Marginal Distributions

We acquire complex valued data during magnetic resonance imaging (MRI) [17]. For notational purpose we will denote y_R to be real component and y_I to be the imaginary component in a voxel at particular time. We also assume each component is independently normally distributed such that $y_R = a + \eta_R$ and $y_I = b + \eta_I$ where η_R and $\eta_I \sim N(0, \sigma^2)$. Defining a and b to be $\rho\cos(\theta)$ and $\rho\sin(\theta)$, respectively.

We need to write out the joint distribution of y_R and y_I,

$$f(y_R, y_I) = (2\pi\sigma^2)^{-1} e^{-\frac{[y_R - \rho\cos(\theta)]^2 - [y_I - \rho\sin(\theta)]^2}{2\sigma^2}} \tag{A.1.1}$$

and define our magnitude, $r = \sqrt{y_R^2 + y_I^2}$, and our phase, $\phi = \tan^{-1}\left(\frac{y_I}{y_R}\right)$. The following Figure A.1 is a 3D histogram from random samples of the joint distribution

of the magnitude and the phase for an SNR of 0.5, $\sigma^2 = 0.05, \rho = 0.5$.

We define $\tan^{-1}(\cdot)$ as the four quadrant version. We now observe $y_R = r\cos(\phi)$ and $y_I = r\sin(\phi)$ and the Jacobian of the transformation to be r.

Looking at the numerator of the exponent we substitute for y_R and y_I then expand the squares,

$$\begin{aligned}
\left[y_R - \rho\cos(\theta)\right]^2 &= \left[r\cos(\phi) - \rho\cos(\theta)\right]^2 \\
&= r^2\cos^2(\phi) - 2r\rho\cos(\phi)\cos(\theta) + \rho^2\cos^2(\theta) \\
\left[y_I - \rho\sin(\theta)\right]^2 &= \left[r\sin(\phi) - \rho\sin(\theta)\right]^2 \\
&= r^2\sin^2(\phi) - 2r\rho\sin(\phi)\sin(\theta) + \rho^2\sin^2(\theta)
\end{aligned}$$

then adding them together,

$$\begin{aligned}
&= r^2\left[\cos^2(\phi) + \sin^2(\phi)\right] - 2r\rho\left[\cos(\phi)\cos(\theta) + \sin(\phi)\sin(\theta)\right] + \rho^2\left[\cos^2(\theta) + \sin^2(\theta)\right] \\
&= r^2 + \rho^2 - 2r\rho\left[\cos(\phi)\cos(\theta) + \sin(\phi)\sin(\theta)\right]. \qquad \text{(A.1.2)}
\end{aligned}$$

Proceeding with the transformation,

$$\begin{aligned}
f(r,\phi) &= (2\pi\sigma^2)^{-1} r e^{-\frac{r^2+\rho^2-2r\rho(\cos(\phi)\cos(\theta)+\sin(\phi)\sin(\theta))}{2\sigma^2}} \\
&= (2\pi\sigma^2)^{-1} r e^{\frac{-r^2-\rho^2}{2\sigma^2}} e^{\frac{-2r\rho(\cos(\phi)\cos(\theta)+\sin(\phi)\sin(\theta))}{2\sigma^2}}. \qquad \text{(A.1.3)}
\end{aligned}$$

Now that we have the joint distribution of r and ϕ we will obtain their marginal distributions. First, we will look at the marginal distribution of the magnitude, r.

$$f(r) = \frac{r}{\sigma^2} e^{\frac{-r^2-\rho^2}{2\sigma^2}} \int_{\phi=0}^{2\pi} \frac{1}{2\pi} e^{\frac{-2r\rho(\cos(\phi)\cos(\theta)+\sin(\phi)\sin(\theta))}{2\sigma^2}} d\phi \qquad \text{(A.1.4)}$$

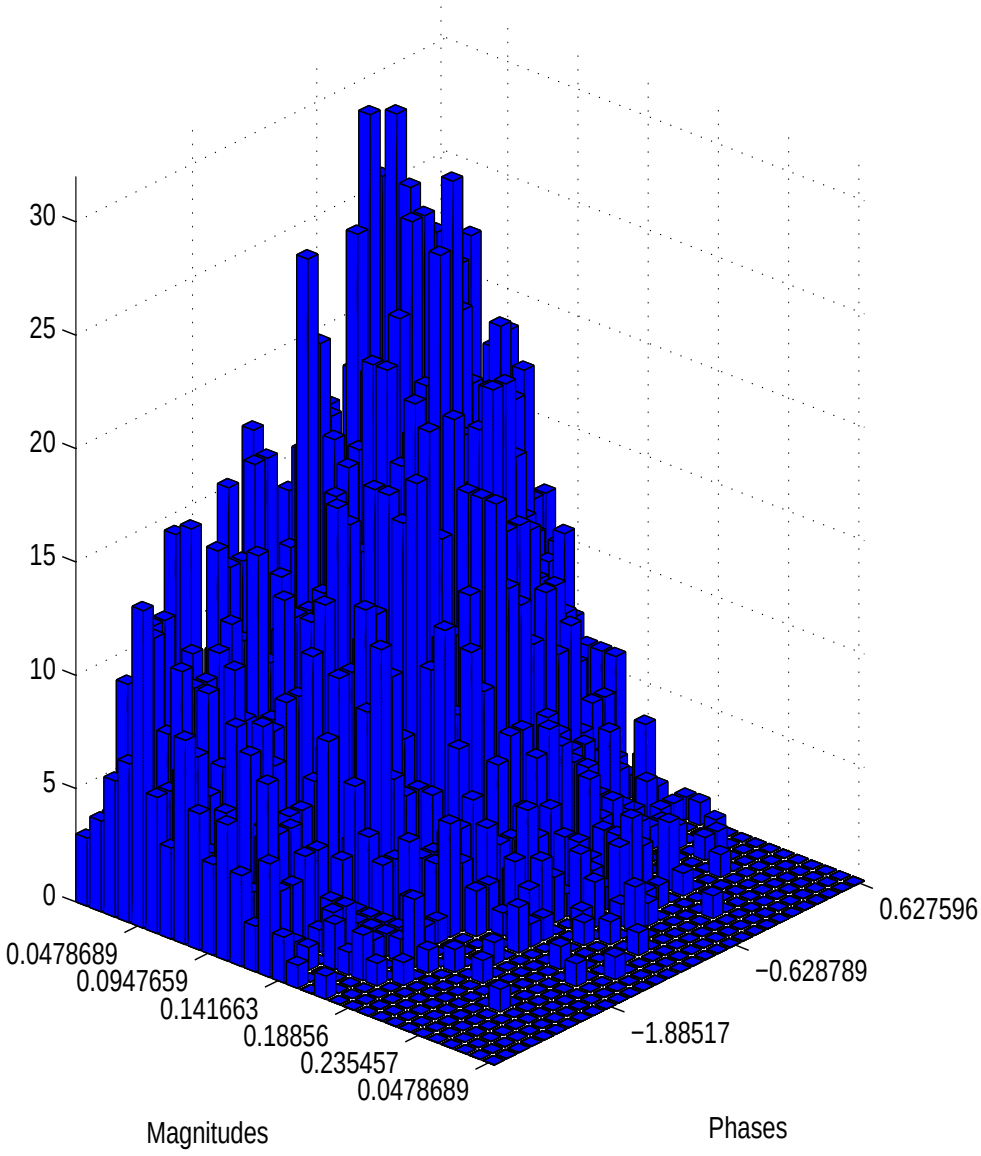


Figure A.1: 3D histogram of joint distributions

defining $\alpha = \phi - \theta$ we get,

$$f(r) = \frac{r}{\sigma^2} e^{-\frac{r^2+\rho^2}{2\sigma^2}} \int_{-\theta}^{2\pi-\theta} e^{-\frac{r\rho\cos(\alpha)}{\sigma^2}} d\alpha. \tag{A.1.5}$$

Next, we integrate over our new limits and recognize the integral as a zeroth order modified Bessel function of the first kind. This resulting PDF is known as the Ricean distribution which is,

$$\begin{aligned} f(r) &= \frac{r}{\sigma^2} e^{-\frac{r^2+\rho^2}{2\sigma^2}} \int_{0}^{2\pi} e^{-\frac{r\rho\cos(\gamma)}{\sigma^2}} d\gamma & \text{(A.1.6)} \\ &= \frac{r}{\sigma^2} e^{-\frac{r^2+\rho^2}{2\sigma^2}} I_0\left(\frac{r\rho}{\sigma^2}\right). & \text{(A.1.7)} \end{aligned}$$

It is known that the limiting distribution is normal with mean ρ and variance σ^2 for large SNR or as $\rho \to \infty$. This can be seen in Figure A.2 where the exact distribution is plotted for the following SNRs (0, 0.2, 0.5, 1, 2.5, 5, 10).

The limiting distribution for zero SNR or $\rho = 0$ is Rayleigh which has the following density function,

$$f(r) = \frac{r}{\sigma^2} e^{-\frac{r^2}{2\sigma^2}}. \tag{A.1.8}$$

Nearly all fMRI analysis are done with magnitude-only data, however it is important to explore the phase distribution because it contains the other half of the data. Models can be developed using the phase-only but first we need to derive its distribution. Starting with the same joint distribution we must integrate out the magnitude, r.

$$\begin{aligned} f(\phi) &= \frac{1}{2\pi\sigma^2} e^{-\frac{\rho}{2\sigma^2}} \int_{0}^{\infty} e^{-\frac{1}{2\sigma^2}(r^2 - 2\rho r\cos(\phi-\theta))} dr & \text{(A.1.9)} \\ &= \frac{1}{2\pi\sigma^2} e^{-\frac{\rho}{2\sigma^2}} e^{\frac{1}{2\sigma^2}(\rho\ \cos(\phi-\theta))^2} \int_{0}^{\infty} r e^{-\frac{1}{2\sigma^2}(r-\rho\ \cos(\phi-\theta))^2} dr. & \text{(A.1.10)} \end{aligned}$$

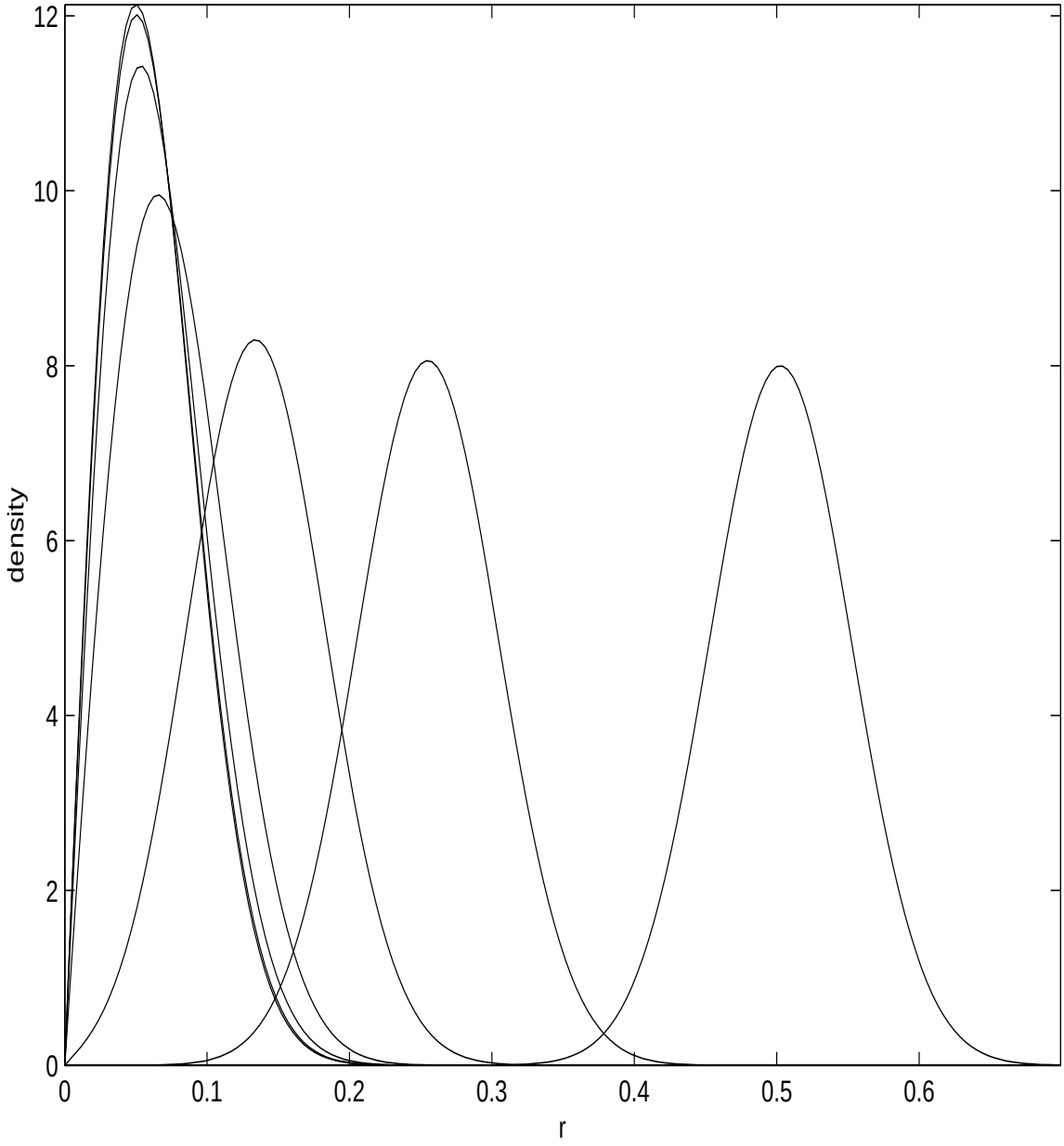


Figure A.2: Ricean distribution for a variety of different SNR values left to right [0, 0.2, 0.5, 1, 2.5, 5, 10].

Defining $z^2 = \frac{(r-\rho\ \cos(\phi-\theta))^2}{\sigma^2}$ we get,

$$\begin{aligned} f(\phi) &= \frac{1}{2\pi\sigma^2} e^{-\frac{\rho}{2\sigma^2}} e^{\frac{1}{2\sigma^2}(\rho\ \cos(\phi-\theta))^2} \int_{-\infty}^{\frac{\rho\ \cos(\phi-\theta)}{\sigma}} (z\sigma + \rho\ \cos(\phi-\theta))\, e^{-\frac{z^2}{2}} dz\ \sigma \\ &= \frac{1}{2\pi\sigma^2} e^{-\frac{\rho}{2\sigma^2}} e^{\frac{1}{2\sigma^2}(\rho\ \cos(\phi-\theta))^2} \\ &\times \left(\int_{-\infty}^{\frac{\rho\ \cos(\phi-\theta)}{\sigma}} z\ \sigma^2 e^{-\frac{z^2}{2}} dz + \int_{-\infty}^{\frac{\rho\ \cos(\phi-\theta)}{\sigma}} \sigma\rho\ \cos(\phi-\theta) e^{-\frac{z^2}{2}} dz \right) \\ &= \frac{1}{2\pi} e^{-\frac{\rho^2}{2\sigma^2}(1-\cos(\phi-\theta)^2)} \\ &\times \left(-e^{-\frac{\rho^2 \cos(\phi-\theta)^2}{2\sigma^2}} + \frac{\rho}{\sigma} \cos(\phi-\theta) \Phi\left(\frac{\rho\ \cos(\phi-\theta)}{\sigma} \right) \right) \\ &= \frac{e^{-\frac{\rho^2}{2\sigma^2}}}{2\pi} \left(1 + \frac{\rho\ \cos(\phi-\theta)}{2\pi\sigma} e^{\frac{\rho^2\ \cos^2(\phi-\theta)}{2\sigma^2}} \Phi\left(\frac{\rho\ \cos(\phi-\theta)}{\sigma} \right) \right) \end{aligned} \tag{A.1.11}$$

where $\Phi(\cdot)$ is the cumulative density function for the standard normal distribution. It has been shown that this distribution has a limiting normal density when SNR is large or $\rho \to \infty$ with mean θ and variance $\left(\frac{\sigma}{\rho}\right)^2$[17]. The following Figure A.3 shows the exact phase distribution for the following SNR values (0, 0.2, 0.5, 1, 2.5, 5), notice the distribution limiting to a normal with $\theta = \frac{\pi}{6}$.

A.2 Conditional Distribution

We will show and verify that the conditional distribution of the phase is a general Von Mises (Circular Normal) distribution. To derive the conditional distribution we need both the joint distribution, $f(\phi, r)$ and the marginal, $f(r)$ which were derived by Rowe and Logan (2004).

$$f(\phi, r) = (2\pi\sigma^2)^{-1} r e^{-\frac{r^2+\rho^2}{2\sigma^2}} e^{-\frac{r\rho\cos(\phi-\theta)}{\sigma^2}} \tag{A.2.12}$$

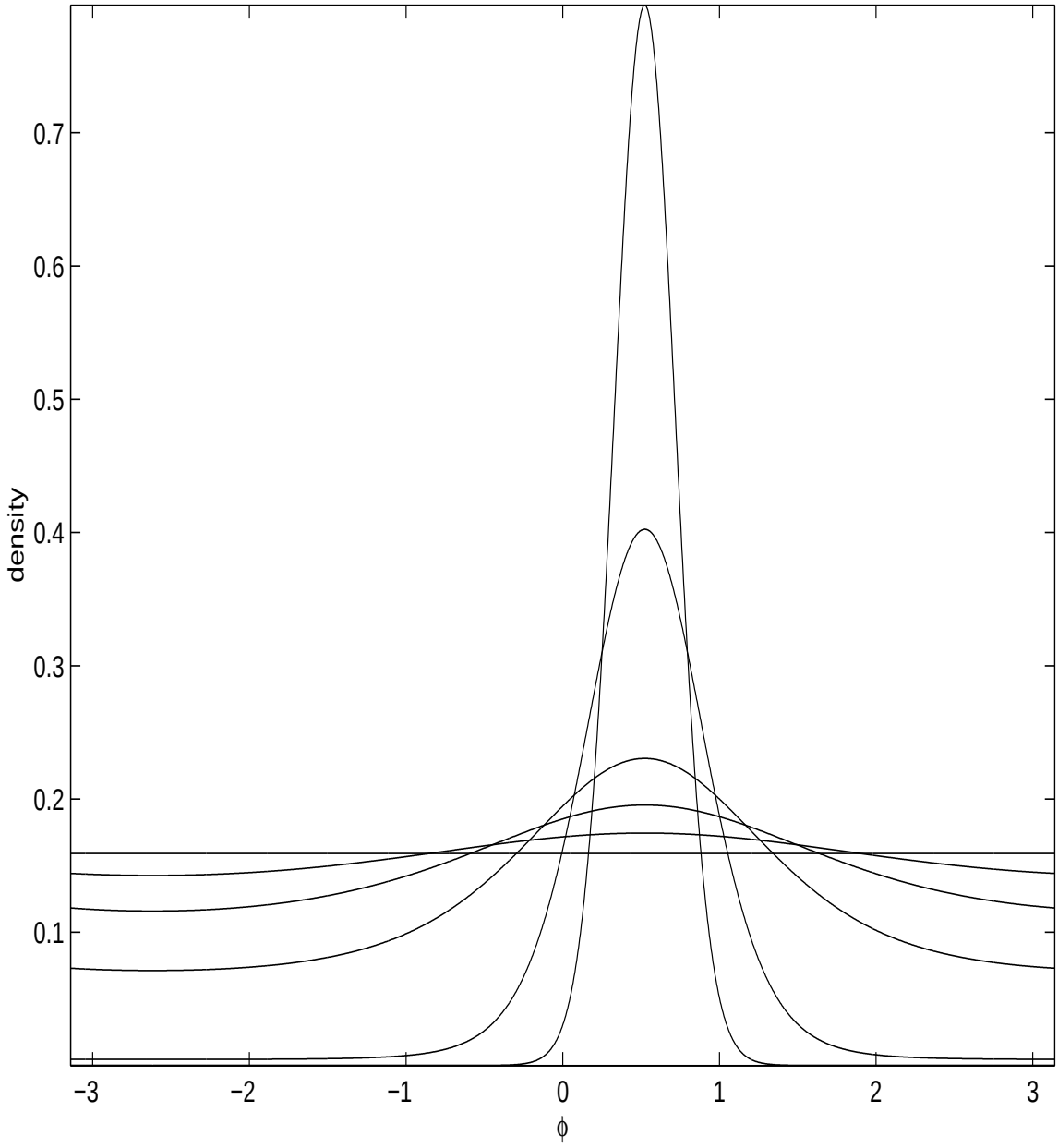


Figure A.3: Phase distribution for a variety of different SNR values from lowest peak to highest [0, 0.2, 0.5, 1, 2.5, 5].

We set $\sigma^2 = \frac{1}{\kappa}$ to obtain a more standard notation and get,

$$f(\phi, r) = \left(\frac{2\pi}{\kappa}\right)^{-1} re^{-\frac{(r^2+\rho^2)\kappa}{2}} e^{r\rho\kappa\cos(\phi-\theta)}. \quad \text{(A.2.13)}$$

By definition of conditional distributions we need to divide by the marginal distribution of r

$$\begin{aligned} f(\phi|r) &= \frac{\frac{1}{2\pi}e^{-\frac{\kappa}{2}(r^2-2\rho r\cos(\phi-\theta)+\rho^2)}}{re^{-\frac{\kappa}{2}(r^2+\rho^2)}I_0(r\rho\kappa)} \\ &= \frac{e^{\frac{2\kappa^*}{2}\cos(\phi-\theta)}}{2\pi I_0(\kappa^*)}. \end{aligned} \quad \text{(A.2.14)}$$

where $\kappa^* = r\rho\kappa$. This derived conditional distribution known distribution of a Von Mises (Circular Normal) with mean θ and concentration of $r\rho\kappa$. [9]

A.3 Linear Model and Hypothesis

We can characterize both the magnitude and phase as a linear model and construct generalized linear hypothesis tests. The models can be written as

$$\phi = U\gamma + \delta, \quad \text{(A.3.15)}$$

where $\delta \sim N(0, I_n\sigma^2)$. The likelihood function of (γ, σ^2) is given by

$$L(\gamma, \sigma^2) = (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{(\phi-U\gamma)'(\phi-U\gamma)}{2\sigma^2}}. \quad \text{(A.3.16)}$$

The maximum likelihood estimate (MLE) for γ can be found as follows,

$$\log L = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{(\phi-U\gamma)'(\phi-U\gamma)}{2\sigma^2} \quad \text{(A.3.17)}$$

$$\frac{\partial \log L}{\partial \gamma}|_{\gamma=\hat{\gamma}} = -\frac{1}{2\sigma^2}\left(-2U\phi + 2U'U\hat{\gamma}\right) \tag{A.3.18}$$

$$= 0 \tag{A.3.19}$$

which implies

$$\hat{\gamma} = (U'U)^{-1}U'\phi. \tag{A.3.20}$$

The MLE for σ^2 is obtained by the following,

$$\log L = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{(\phi - U\gamma)'(\phi - U\gamma)}{2\sigma^2} \tag{A.3.21}$$

$$\frac{\partial \log L}{\partial \sigma^2}|_{\sigma^2=\hat{\sigma}^2,\gamma=\hat{\gamma}} = -\frac{n}{2\sigma^2} + \frac{(\phi - U\gamma)'(\phi - U\gamma)}{2(\sigma^2)^2} = 0 \tag{A.3.22}$$

$$\hat{\sigma}^2 = \frac{(\phi - U\hat{\gamma})'(\phi - U\hat{\gamma})}{n}. \tag{A.3.23}$$

We are interested in constructing linear hypothesis dealing with the γ parameters so we use a likelihood ratio contrast test with the following null hypothesis, $D\gamma = b$. To account for this parameter restriction we will need to use a Lagrangian multiplier. The estimates which maximize the unrestricted likelihood the are the usual MLE's. We need to minimize the following function to obtain estimates which maximize the restricted likelihood,

$$S(\gamma, \lambda) = (\phi - U\gamma)'(\phi - U\gamma) + 2\lambda'(D\gamma - b). \tag{A.3.24}$$

We now proceed with the respective derivatives,

$$\frac{\partial S(\gamma, \lambda)}{\partial \gamma}|_{\gamma=\tilde{\gamma},\lambda=\tilde{\lambda},\sigma^2=\tilde{\sigma}^2} = -2U'\phi + 2U'U\tilde{\gamma} + 2D\tilde{\lambda}$$

$$\tilde{\gamma} = \hat{\gamma} - (U'U)^{-1}D\tilde{\lambda} \tag{A.3.25}$$

$$\frac{\partial S(\gamma,\lambda)}{\partial \lambda}|_{\gamma=\tilde{\gamma},\lambda=\tilde{\lambda},\sigma^2=\tilde{\sigma^2}} = 2D'\hat{\gamma} - 2b. \tag{A.3.26}$$

The implication from the last line is as follows,

$$b = D\tilde{\gamma} \tag{A.3.27}$$

$$= D'(\hat{\gamma} - (U'U)^{-1}D\tilde{\lambda}') \tag{A.3.28}$$

$$b - D'\hat{\gamma} = -D'(U'U)^{-1}D\tilde{\lambda} \tag{A.3.29}$$

$$\tilde{\lambda}' = (D'(U'U)^{-1}D)^{-1}(D'\hat{\gamma} - b)). \tag{A.3.30}$$

By setting $b = 0$ we get

$$\tilde{\lambda}' = (D'(U'U)^{-1}D)^{-1}D'\hat{\gamma}. \tag{A.3.31}$$

Substituting this into our estimate of γ we obtain,

$$\tilde{\gamma} = \hat{\gamma} - (U'U)^{-1}D(D'(U'U)^{-1}D)^{-1}D'\hat{\gamma} \tag{A.3.32}$$

$$= \left(I - (U'U)^{-1}D(D'(U'U)^{-1}D)^{-1}D'\right)\hat{\gamma} \tag{A.3.33}$$

$$= \Psi\hat{\gamma}, \tag{A.3.34}$$

where $\Psi = I - (U'U)^{-1}D'(D(U'U)^{-1}D')^{-1}D$. The only remaining estimate we need for the restricted likelihood is $\tilde{\sigma}^2$ stated below,

$$\tilde{\sigma}^2 = \frac{(\phi - U\tilde{\gamma})'(\phi - U\tilde{\gamma})}{n}. \tag{A.3.35}$$

To complete our likelihood ratio test we must take the ratio of the restricted likelihood divided by the unrestricted likelihood as follows,

$$\Lambda = \frac{\tilde{\sigma}^{2^{-\frac{n}{2}}} e^{-\frac{(\phi - U\tilde{\gamma})'(\phi - U\tilde{\gamma})/2}{(\phi - U\tilde{\gamma})'(\phi - U\tilde{\gamma})/n}}}{\hat{\sigma}^{2^{-\frac{n}{2}}} e^{-\frac{(\phi - U\hat{\gamma})'(\phi - U\hat{\gamma})/2}{(\phi - U\hat{\gamma})'(\phi - U\hat{\gamma})/n}}} \tag{A.3.36}$$

$$= \frac{\left(\tilde{\sigma}^2\right)^{-\frac{n}{2}}}{\left(\hat{\sigma}^2\right)^{-\frac{n}{2}}} \tag{A.3.37}$$

$$-2\ \log(\Lambda) = n \log\left(\frac{\tilde{\sigma}^2}{\hat{\sigma}^2}\right). \tag{A.3.38}$$

where $-2\log(\Lambda)$ has an asymptotic χ^2_w distribution, where w is the full row rank of D.

Bibliography

[1] E.H. Aylward, T.L. Richards, V.W. Berninger, W.E. Naggy, K.M. Field, and et al. Instructional Treatment Associated with Changes in Brain Activation in Children with Dyslexia. *Neurology*, 61:212–219, July 2003.

[2] P.A. Bandettini, A. Jesmanowicz, E.C. Wong, and J.S. Hyde. Processing Strategies for Time-Course Data Sets in Functional MRI of the Human Brain. *Magnetic Resonance in Medicine*, 30:161–173, 1993.

[3] J. Bodurka, J. Jesmanowicz, A. Hyde, J.S. Xu, H. Estkowski, and S.-J Li. Current-Induced Magnetic Resonance Phase Imaging. *Journal of Magnetic Resonance*, 137:265–271, 1999.

[4] G.E.P. Box and D.R. Cox. An Analysis of Transformations. *Journal of the Royal Statistical Society*, 26:296–311, 1964.

[5] N.I. Fisher and A.J. Lee. Regression Models for an Angular Response. *Biometrics*, 48:665–677, September 1992.

[6] A.L. Gould. A Regression Technique for Angular Variates. *Biometrics*, 25:683–700, 1969.

[7] E.M. Haacke, R.W. Brown, M.R. Thompson, and R. Venkatesan. *Magnetic Resonance Imaging, Physical Principles and Sequence Design*. Wiley-Liss, 1999.

[8] S.A. Huettel, A.W. Song, and G. McCarthy. *Functional Magnetic Resonance Imaging*. Sinauer, 2004.

[9] S.R. Jammalamadaka and A. SenGupta. *Topics in Circular Statistics*. World Scientific, 2001.

[10] R.A. Johnson and T.E. Wehrly. Some Angular-Linear Distributions and Related Regression Models. *Journal of the American Statistical Association*, 73(353):602–606, September 1978.

[11] P.J. Laycock. Optimal Design: Regression Models for Directions. *Biometrika*, 62:305–311, 1975.

[12] B.R. Logan and D.B. Rowe. An Evaluation of Thresholding Techniques in fMRI Analysis. *NeuroImage*, 22:95–108, 2004.

[13] G. Marsaglia. Ratios of Normal Variables and Ratios of Sums of Uniform Variables. *Journal of the American Statistical Association*, 60:193–204, 1965.

[14] R.S. Menon. Postacquisition Suppression of Large-Vessel BOLD Signals in High-Resolution fMRI. *Magnetic Resonance in Medicine*, 47:1–9, 2002.

[15] N. Ravishanker and D.K. Dey. *A First Course in Linear Model Theory*. Chapman & Hall/CRC, 2002.

[16] D.B. Rowe. *Multivariate Bayesian Statistics*. Chapman & Hall/CRC, 2003.

[17] D.B. Rowe and B.R. Logan. A Complex Way to Compute fMRI Activation. *NeuroImage*, 23(3):1078–1092, 2004.

[18] D.B. Rowe and B.R. Logan. Complex fMRI Analysis with Unrestricted Phase is Equivalent to a Magnitude-Only Model. *NeuroImage*, 24(2):xxx–xxx, 2005.

[19] J. Sijbers and A.J. den Dekker. Maximum Likelihood Estimation of Signal Amplitude and Noise Variance from MR Data. *Magnetic Resonance in Medicine*, 51:586–594, 2004.

[20] A.M. Smith, B.K. Lewis, U.E. Ruttimann, F.Q. Ye, T.M. Sinnwell, and et al. Investigation of Low Frequency Drift in fMRI Signal. *NeuroImage*, 9:526–533, 1999.

[21] T.E. Wehrly and R.A. Johnson. Bivariate Models for Dependence of Angular Observations and a Related Markoc Process. *Biometrika*, 66:255–256, 1979.

[22] L. Ying, J.X. Ji, D.C. Munson Jr., Z.P. Liang, R. Koetter, and B.J. Frey. A Robust and Efficient Method to Unwrap MR Phase Images. In *Proceedings of 11th Annual Meeting of International Society for Magnetic Resonance in Medicine*, page 782, July 2003.

□

This Thesis was prepared under the direction of the chairperson of Christopher Meller's supervisory committee and has been approved by all members of that committee. It was submitted to the Dean of the Graduate School of Biomedical Sciences, Medical College of Wisconsin and was approved as a partial fulfillment of the requirements for the degree of Master of Science.

Committee in Charge:

_____, Chairperson

_____

_____

_____

Dean of the Graduate School of Biomedical Sciences