2016

Novel Statistical Methodologies in Analysis of Position Emission Tomography Data: Applications in Segmentation, Normalization, and Trajectory Modeling

Daniel B. Shin
University of Pennsylvania, dbshin@mail.med.upenn.edu

Novel Statistical Methodologies in Analysis of Position Emission Tomography Data: Applications in Segmentation, Normalization, and Trajectory Modeling

Abstract

Position emission tomography (PET) is a powerful functional imaging modality with wide uses in fields such as oncology, cardiology, and neurology. Motivated by imaging datasets from a psoriasis clinical trial and a cohort of Alzheimer's disease (AD) patients, several interesting methodological challenges were identified in various steps of quantitative analysis of PET data. In Chapter 1, we consider a classification scenario of bivariate thresholding of a predictor using an upper and lower cutpoints, as motivated by an image segmentation problem of the skin. We introduce a generalization of ROC analysis and the concept of the parameter path in ROC space of a classifier. Using this framework, we define the optimal ROC (OROC) to identify and assess performance of optimal classifiers, and describe a novel nonparametric estimation of OROC which simultaneous estimates the parameter path of the optimal classifier. In simulations, we compare its performance to alternative methods of OROC estimation. In Chapter 2, we develop a novel method to normalize PET images as an essential preprocessing step for quantitative analysis. We propose a method based on application of functional data analysis to image intensity distribution functions, assuming that that individual image density functions are variations from a template density. By modeling the warping functions using a modified function-on-scalar regression, the variations in density functions due to nuisance parameters are estimated and subsequently removed for normalization. Application to our motivating data indicate persistence of residual variations in standardized image densities. In Chapter 3, we propose a nonlinear mixed effects framework to model amyloid-beta (Aβ), an important biomarker in AD. We incorporate the hypothesized functional form of Aβ trajectory by assuming a common trajectory model for all subjects with variations in the location parameter, and a mixture distribution for the random effects of the location parameter address our empirical findings that some subjects may not accumulate Aβ. Using a Bayesian hierarchical model, group differences are specified into the trajectory parameters. We show in simulation studies that the model closely estimates the true parameters under various scenarios, and accurately estimates group differences in the age of onset.

Degree Type

Dissertation

Degree Name

Doctor of Philosophy (PhD)

Graduate Group

Epidemiology & Biostatistics

First Advisor

Russell T. Shinohara

Second Advisor

Andrea B. Troxel

Keywords

Classification, Image normalization, Nonlinear models, ROC

Subject Categories

Biostatistics

NOVEL STATISTICAL METHODOLOGIES IN ANALYSIS OF POSITION EMISSION TOMOGRAPHY DATA: APPLICATIONS IN SEGMENTATION, NORMALIZATION, AND TRAJECTORY MODELING

Daniel B. Shin

A DISSERTATION

in

Epidemiology and Biostatistics

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2016

Supervisor of Dissertation

_____

Russell T. Shinohara

Assistant Professor of Biostatistics

Graduate Group Chairperson

_____

John H. Holmes, Professor of Medical Informatics in Epidemiology

Dissertation Committee

Andrea B. Troxel, Professor of Biostatistics

Haochang Shou, Assistant Professor of Biostatistics

Joel M. Gelfand, Professor of Dermatology, Professor of Epidemiology

Nehal N. Mehta, Chief, Section of Inflammation and Cardiometabolic Diseases, NHLBI

NOVEL STATISTICAL METHODOLOGIES IN ANALYSIS OF POSITION EMISSION TOMOGRAPHY DATA: APPLICATIONS IN SEGMENTATION, NORMALIZATION, AND TRAJECTORY MODELING

DEDICATION

To my parents, who have prayed many prayers for me to finish school, and to my loving wife Yu Xi, who endured my frequent bemoaning of the rigors of graduate school.

ACKNOWLEDGEMENT

My sincere gratitude to “Taki” Shinohara for taking a chance with me as his first graduate student. His endless patience and unwavering support allowed me to push through many obstacles throughout my graduate program.

I am forever indebted to Joel Gelfand for his decade-long mentorship and helping me find my calling in biostatistics. Without him, I would still be contemplating my career choices.

Many thanks to my dissertation committee members, Andrea Troxel, Haochang Shou, and Nehal Mehta, for their generous guidance and feedback.

My gratitude extends to Farrah Mateen, Jaroslow Harezlak, Abass Alavi, Paul Yushkevich, Sharon Xie, Ciprian Crainiceanu, RajaNandini Muralidharan, and many others who have shared their expertise and research that helped shape my dissertation research.

Special thanks to Marissa Fox and Cathy Vallejo, who have been my cheerleaders in Blockley.

ABSTRACT

NOVEL STATISTICAL METHODOLOGIES IN ANALYSIS OF POSITION EMISSION TOMOGRAPHY DATA: APPLICATIONS IN SEGMENTATION, NORMALIZATION, AND TRAJECTORY MODELING

Daniel B. Shin

Russell T. Shinohara

Position emission tomography (PET) is a powerful functional imaging modality with wide uses in fields such as oncology, cardiology, and neurology. Motivated by imaging datasets from a psoriasis clinical trial and a cohort of Alzheimer's disease (AD) patients, several interesting methodological challenges were identified in various steps of quantitative analysis of PET data. In Chapter 1, we consider a classification scenario of bivariate thresholding of a predictor using an upper and lower cutpoints, as motivated by an image segmentation problem of the skin. We introduce a generalization of ROC analysis and the concept of the parameter path in ROC space of a classifier. Using this framework, we define the optimal ROC (OROC) to identify and assess performance of optimal classifiers, and describe a novel nonparametric estimation of OROC which simultaneous estimates the parameter path of the optimal classifier. In simulations, we compare its performance to alternative methods of OROC estimation. In Chapter 2, we develop a novel method to normalize PET images as an essential preprocessing step for quantitative analysis. We propose a method based on application of functional data analysis to image intensity distribution functions, assuming that that individual image density functions are variations from a template density. By modeling the warping functions using a modified function-on-scalar regression, the variations in density functions due to nuisance parameters are estimated and subsequently removed for normalization. Application to our motivating data indicate persistence of residual variations in standardized image densities. In Chapter 3, we propose a nonlinear mixed effects framework to model amyloid-beta (Aβ), an important biomarker in AD. We incorporate the hypothesized functional form of Aβ trajectory by assuming a common trajectory model for all subjects with variations in the location parameter, and a mixture distribution for the random effects of the location parameter address our empirical findings that some subjects may not accumulate Aβ. Using a Bayesian hierarchical model, group differences are specified into the trajectory parameters. We show in simulation studies that the model closely

estimates the true parameters under various scenarios, and accurately estimates group differences in the age of onset.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

CHAPTER 1

INTRODUCTION

The field of biomedical imaging has transformed in recent decades from simple applications, such as visualization of broken bones using projection radiographs, to vastly more complex applications, such as quantification of brain activity using functional magnetic resonance imaging. Advances in imaging technology has made it possible to observe real-time *in vivo* disease processes, e.g. the level of impaired bone mineralization or the amyloid plaque burden in the brains of Alzheimer's disease (AD) patients, that until recently could only be seen post mortem or using invasive methods such as biopsies. The transition from qualitative to quantitative applications in biomedical imaging, along with increasing understanding of the relationship between various disease biomarkers and clinical outcomes, has highlighted the importance of robust statistical methods.

This dissertation focuses on positron emission tomography (PET) and statistical methodologies inspired by problems encountered during the analysis of PET data. The motivation comes from the Vascular Inflammation in Psoriasis (VIP) trial. The VIP trial is a multicenter, randomized controlled trial to assess the effects of adalimumab, a biological systemic therapy, on systemic vascular inflammation in subjects diagnosed with moderate to severe psoriasis, as compared to narrow-band ultraviolet B phototherapy or placebo. The primary outcome measure is the change in total vascular inflammation in aortic segments as assessed using ^{18}F-fludeoxyglucose (FDG) PET/computer tomography (CT) between baseline and end of study. The timing of the start of the dissertation research coincided with the start of the VIP trial, and the opportunity naturally arose to observe and participate in the analytic planning of the trial. This opportunity eventually evolved into the first two chapters of this dissertation.

Chapter 2 is inspired by a novel application of FDG-PET/CT to quantify the level of skin inflammation. With the goal to quantify and correlate PET signals with clinical psoriasis severity measures, skin segmentations need to be superimposed on the FDG-PET images that contain measures of metabolic activity. A method for image segmentation, or the partitioning of an image into segments of interest, is devised for the skin using the CT image, where a bivariate threshold (i.e. upper and lower cutpoints are used) is applied to a binary tissue mask image transformed using a Gaus-

sian filter. This chapter addresses the performance assessment of bivariate thresholding such as the interval classification used in the segmentation method. However, after the methodology was developed, it was noticed that proper co-registration, or alignment, of PET image to the CT segmentations became impossible due to significant movements of the extremities during image acquisition in the VIP trial. A dataset from serum sodium levels in patients hospitalized for fulminant bacterial meningitis fit the classification scenario was substituted as the motivating example.

While parallels can be drawn from the classic binary classification scenario and the receiver operating characteristic (ROC) analysis, these classic approaches are inadequate and restrictive for classification beyond simple univariate binary thresholding. Simple classification scenarios involving thresholding a predictor using a cutpoint offer straightforward assessments of classifier performance using the ROC curve, but in complex scenarios, such as bivariate thresholding, the ROC curve is not specified. Chapter 2 presents a generalization of the ROC analysis, introduces the concept of the parameter path in the ROC space of a classification scenario, and defines the optimal ROC (OROC) to identify and assess performance of optimal classifiers. Interval classification, where a predictor is thresholded using two cutpoints, is used to illustrate the OROC estimation and show that nonparametric estimation (NPE) procedure simultaneously produces the parameter path of the optimal classifier. Alternative semiparametric and parametric methods for OROC estimation are presented: the generalized additive model (GAM), and the maximum likelihood estimation (MLE) based on a profile likelihood. The performance of the NPE, GAM, and MLE in Monte Carlo simulations and application to the serum sodium dataset are presented.

Chapter 3 covers image normalization, a major topic in quantitative PET. Image normalization is widely viewed as an essential preprocessing step for quantitative analysis. While great advancements have been made in normalization for magnetic resonance imaging (MRI), quantitative analysis of positron emission tomography (PET) primarily involves the use of standardized uptake values (SUV) which aim to account for major sources of nuisance variation. However, these units are highly susceptible to variations in imaging protocol and physiology. A normalization method based on the application of functional data analysis to image intensity distribution functions is proposed in this chapter, with the assumption that individual density functions are variations from a template density function. Warping functions, which transform the template density into individual densities, capture these variations. By modeling the warping functions using a modified function-on-scalar

regression, the variations in density functions due to nuisance parameters are estimated and subsequently removed for normalization. This chapter outlines image intensity density normalization and includes the application to the VIP trial dataset. Additionally, the normalized densities show correlations with cardiovascular biomarkers that are not present in the original densities.

Chapter 4 presents another interesting PET data problem in the realm of neuroimaging. The problem presents itself in the form of a plot (Figure 4.1) of longitudinal measurements of amyloid-beta (Aβ) in the brain, using florbetapir-PET. There are striking features that show 1) differences by genotypes, 2) an outline of a common trajectory function, and 3) possible clustering of subjects who many have significant delay in Aβ accumulation. Aβ is an important biomarker in AD, and a general hypothesis exists regarding the functional form of its longitudinal trajectory. Surprisingly, even with various hypotheses on the shape of Aβ trajectories, no studies have integrated this knowledge in modeling Aβ trajectories, with most analyses instead relying on basic linear mixed effects models. This chapter describes a new approach using nonlinear mixed effects framework to model Aβ trajectories as measured by florbetapir-PET in a cohort of patients from the Alzheimer's Disease Neuroimaging Initiative (ADNI). The hypothesized functional form of Aβ trajectory is incorporated by assuming 1) a common trajectory function for all subjects, with variations in the location parameter, and 2) a mixture distribution for the random effects of the location parameter to address an empirical findings that some subjects may not accumulate Aβ. Using a Bayesian hierarchical model, group differences are specified into the trajectory parameters. Monte Carlo simulation results are presented to show the performance of model estimates under various scenarios. Application to the ADNI data show an estimated difference of 21 years in the onset of Aβ between average and high risk group based on APOE genotype.

CHAPTER 2

ESTIMATION OF THE OPTIMAL ROC IN COMPLEX CLASSIFICATION SETTINGS

2.1. Introduction

Assessment of classifier performance is well established for simple biomedical scenarios. The most common methods quantify the tradeoff between correct and incorrect disease classifications to determine the performance of a biomarker, and these methods are prevalent in binary classification settings. While these methods may be sufficient for classification that involves thresholding a continuous score, they are not be able to accommodate complex classifications such as multivariate thresholding. In the setting of multiple thresholds, the traditional single threshold method cannot be used to assess the performance of the classifier. Alternatively, the performance measure of the classifier must accurately reflect the classification rule.

The receiver operating characteristic (ROC) curve is the standard method of assessing classifier performance, owing largely to the simplicity of its graphical and statistical interpretation (Metz, 1978; Pepe, 2003). The ROC curve is defined in the binary setting, where two distinct classes are assumed to have different (i.e. separated) distributions of a common measure such that partial separation of distributions can be achieved using a single cutoff. In this particular setting, the ROC curve has been extensively characterized and its properties are well understood.

In the framework of the simple thresholding of a dependent variable that allows the tradeoff between sensitivity and specificity to be used, extensions have been developed for more complex decision rules, with the emphasis primarily on ternary classification such as trichotomous classification (Dreiseitl, Ohno-Machado, and Binder, 2000; Mossman, 1999) and transitional (ordinal) classification (Alonzo and Nakas, 2007; Nakas and Alonzo, 2007). These classifier assessment methods rely on a seemingly natural extension of the ROC curve, the ROC surface, whose volumetric interpretation is similar to the area under the ROC curve (AUC) (Dreiseitl, Ohno-Machado, and Binder, 2000; Mossman, 1999; Nakas and Yiannoutsos, 2004; Nakas and Alonzo, 2007). However, traditional ROC analysis suffers the major limitation of an assumption of classification based on univariate thresholds, i.e. it assumes risk increases or decreases monotonically with the biomarker. Consider a scenario where the classification is based on a more complex decision rule,

such as biomedical image segmentation (i.e. classification of features of an image as anatomical or disease classes) using a multimodal histogram where a feature of interest has values only within a specific interval (Chen, 2008). A histogram of the image intensity values may assume a multimodal distribution, and a segmentation algorithm may partition the histogram using thresholds based on features such as valleys in the histogram. Another scenario to consider is predicting survival of patients based on biomarkers at admission to a hospital in patients with fulminant bacterial meningitis (Muralidharan, Mateen, and Rabinstein, 2014), where serum sodium level within a specific range may predict survival better than using single threshold. In such cases, traditional ROC analysis, which can only assess the sensitivity and specificity tradeoff for one threshold at a time, are no longer useful.

We propose a simple generalization of ROC analysis that does not impose a binary cutoff restriction and can easily accommodate various classification scenarios. In Section 2.2, we introduce optimal ROC framework and apply this concept using a simple case that involves classification bounded by two thresholds, or interval classification. We demonstrate nonparametric, semiparametric, and parametric estimation procedures for the generalized ROC analysis, and show that the nonparametric estimation procedure has the benefit of simultaneously estimating the optimal thresholds for a given sensitivity or specificity. In Section 2.4, we compare the performance of the various proposed ROC methods in simulation studies. In Section 2.5, we apply these methods to the study of serum sodium levels for predicting survival in bacterial meningitis patients.

2.2. Description

2.2.1. Classical terminology

We adopt the traditional terminology from the simple binary classification scenario. The true class membership $Y = \{1, 0\}$ is predicted using a continuous predictor X. For a given threshold parameter θ, an observation with $X > \theta$ is classified as $Y = 1$. Consequently, the true positive rate (TPR), or sensitivity, is defined as the proportion of correctly classified positives (denoted as Y):

$$TPR(\theta) = P(X > \theta | Y = 1) = P(X_Y > \theta). \tag{2.1}$$

The false positive rate (FPR), or 1-specificity, is defined as the proportion of incorrectly classified negatives (denoted as $\bar{Y}$):

$$FPR(\theta) = P(X > \theta | Y = 0) = P(X_{\bar{Y}} > \theta). \tag{2.2}$$

While the plot of TPR as a function of $FPR(\theta)$ traditionally represents the ROC curve (Pepe, 2003), we use a more general representation of the ROC as a vector-valued function,

$$ROC(\theta) = \{FPR(\theta), TPR(\theta)\}, \text{for } \theta \in \Theta. \tag{2.3}$$

In traditional binary classification, $\Theta = \mathbb{R}$, and the plot is a monotonic curve. Specifically, the ROC curve is a function of the data generating distribution, the classifier f, and its parameter θ. In simple classification, the classifier is based on a single parameter θ and it can be summarized as an indicator function

$$f(\theta, X) = I(X > \theta). \tag{2.4}$$

The performance of the classifier is summarized using the AUC by integrating TPR over FPR.

2.2.2. Parameter path of the ROC curve and the optimal ROC

The tradeoff between TPR and FPR is graphically summarized by the ROC curve. In the simple classification scenario, TPR as a function of $FPR(\theta)$ is sufficient to describe the performance of the classifier over Θ as an ROC curve. In the more general case, $\theta \in \Theta$ must be properly structured to represent a classifier in ROC space.

For a classifier f, we define a continuous path $\boldsymbol{\phi}$,

$$\boldsymbol{\phi} : (0, 1) \mapsto \boldsymbol{\Theta}, \tag{2.5}$$

where $\lim_{s \to 0} \boldsymbol{\phi}(s)$ corresponds to (0,0) in the ROC plot, and $\lim_{s \to 1} \boldsymbol{\phi}(s)$ corresponds to (1,1). In the simple classification case, $\lim_{s \to 0} \boldsymbol{\phi}(s) = -\infty$ and $\lim_{s \to 1} \boldsymbol{\phi}(s) = \infty$. Conceptually, $\boldsymbol{\phi}$ can be represented as a sequence of θ that plots the ROC curve from (0,0) to (1,1). We denote the functional space of all $\boldsymbol{\phi}$ by $\boldsymbol{\Phi}$, and in simple classification $\mathrm{Im}(\boldsymbol{\phi}) \equiv \mathbb{R}, \ \forall \boldsymbol{\phi} \in \boldsymbol{\Phi}$.

In a given classification schema, $\mathbf{\Phi}$ produces set of all possible ROC curves. Since it is convention to attribute one ROC curve to a classifier, from now we refer to ϕ as a classifier. Revisiting the ROC, it can now be defined as a function of the classifier and data:

$$ROC(F_{Y,\bar{Y}}, \boldsymbol{\phi}, s) = \{FPR(\phi(s)), TPR(\phi(s))\},\ s \in [0,1]. \tag{2.6}$$

Furthermore, we generalize the AUC as the integration of TPR over ϕ transformed to FPR:

$$AUC = \int_{\phi} TPR(\theta)\, dFPR(\theta) \cdot d\theta. \tag{2.7a}$$

It can be reparameterized as an integral over FPR:

$$AUC = \int_{0}^{1} TPR(\boldsymbol{\phi}(s))\, dFPR(\boldsymbol{\phi}(s)). \tag{2.7b}$$

In the case where FPR is an invertible function whose domain is $\boldsymbol{\phi}(s),\ s \in [0,1]$, substituting $u = FPR(\boldsymbol{\phi}(s))$, (2.7b) becomes more familiar:

$$AUC = \int_{0}^{1} TPR(FPR^{-1}(u))\, du. \tag{2.7c}$$

The concept of the optimal ROC is straightforward. For a given classification scenario in $\mathbf{\Phi}$, we are interested in a classifier ϕ that achieves the best classification. Namely, we want ϕ that achieves the optimal ROC, with optimality defined by a feature (generally the AUC) of the ROC curve:

$$\arg\max_{\boldsymbol{\phi}} AUC(\boldsymbol{\phi}) := \{\boldsymbol{\phi} \mid \forall \boldsymbol{\tau} \in \mathbf{\Phi} : AUC(\boldsymbol{\tau}) \leq AUC(\boldsymbol{\phi})\}. \tag{2.8}$$

In simple classification, $\{\mathrm{Im}(\boldsymbol{\phi})\} \equiv \mathbb{R}$ implicitly defines a unique $\boldsymbol{\phi}$ that spans from $-\infty$ to ∞, therefore it is the optimal ROC classifier.

2.2.3. Interval classification

The concept of optimal ROC may seem trivial in simple classification, but its importance is evident in classification that uses more than a single parameter. To illustrate the optimal ROC, we describe a more complex classification schema with two real parameters, $\boldsymbol{\theta} = \{\theta_1, \theta_2\}$, and with the

classification function defined as

$$f(\boldsymbol{\theta}, X) = I(\theta_1 < X < \theta_2). \tag{2.9}$$

We refer to this as interval classification, and the decision rule is also straightforward. In this case, the TPR is:

$$TPR(\boldsymbol{\theta}) = P(X_Y > \theta_1 \cap X_Y < \theta_2). \tag{2.10}$$

Similarly, FPR is defined as

$$FPR(\boldsymbol{\theta}) = P(X_{\bar{Y}} > \theta_1 \cap X_{\bar{Y}} < \theta_2). \tag{2.11}$$

A contour plot of $FPR(\boldsymbol{\theta})$ and $TPR(\boldsymbol{\theta})$ (Fig 2.1) best illustrates $\boldsymbol{\Theta}$, now a subset of $\mathbb{R}^2$, and $\boldsymbol{\Phi}$ of this schema. As before, define a path represented as an arc along the surface with end points parameterized as 0 and 1, respectively representing (FPR, TPR) of (0,0) and (1,1). Let $\boldsymbol{\phi}$ be a continuous function that maps the arc to the parameters that constitute an ROC curve, such that $\lim_{s\to 0} \boldsymbol{\phi}(s) = \{\theta, \theta\}$ for some $\theta \in \mathbb{R}$ and $\lim_{s\to 1} \boldsymbol{\phi}(s) = \{-\infty, \infty\}$. In this scenario, $\mathrm{Im}(\boldsymbol{\phi})$ is a set of $\boldsymbol{\theta} \in \mathbb{R}^2$ that defines an ROC curve. It is evident that a unique $\boldsymbol{\phi}$ does not exist; rather, many possible classifiers exist. The corresponding ROC curves of the four $\boldsymbol{\phi}$s highlighted in (Fig 2.2) show a wide classifier performance range, but $\boldsymbol{\phi}_1$ has the largest AUC.

The AUC in interval classification has the following form:

$$\begin{aligned} AUC &= \int_{\boldsymbol{\phi}} TPR(\boldsymbol{\theta})\, dFPR(\boldsymbol{\theta})\, d\boldsymbol{\theta} \\ &= \int_{\boldsymbol{\phi}} P(X_Y \geq \theta_1 \cap X_Y \leq \theta_2) \cdot \\ &\qquad d\{P(X_{\bar{Y}} \geq \theta_1 \cap X_{\bar{Y}} \leq \theta_2)\} d\boldsymbol{\theta}. \end{aligned}$$

The conventional interpretation of the AUC is the probability of correctly ordering diseased and non-diseased in simple classification (Pepe, 2003). This interpretation is more difficult in complex classification settings; however, for the simple case where F_Y and $F_{\bar{Y}}$ are symmetric about 0 and

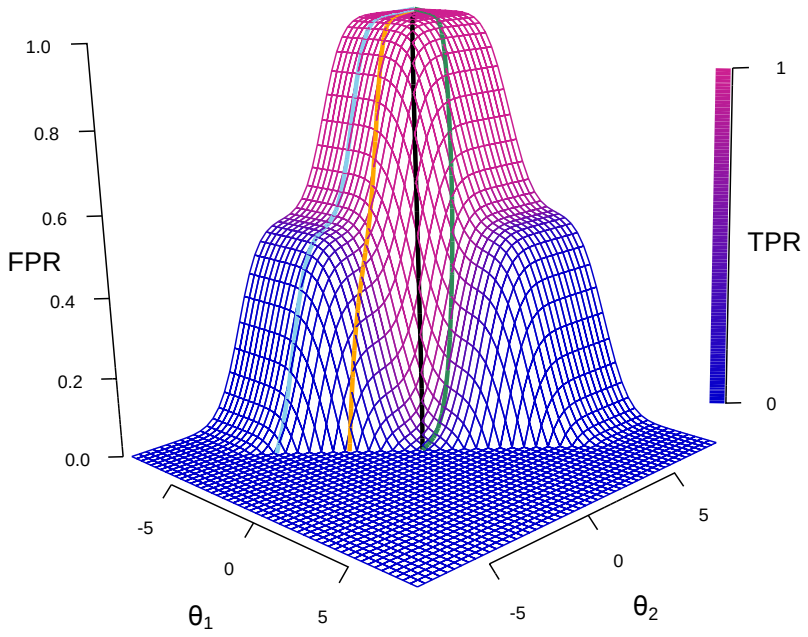

Figure 2.1: Plot of $FPR(\boldsymbol{\theta})$ and $TPR(\boldsymbol{\theta})$. $TPR(\boldsymbol{\theta})$ is the color overlay of $FPR(\boldsymbol{\theta})$, represented as the surface plot. The colored curves along the surface represent sample ϕ's. The black curve represents the ϕ of the optimal ROC. The corresponding ROC curves are shown in Figure 2.2. The plot represents underlying data densities in Figure 2.3a ($\mu = 2.5$).

$\theta_1 = -\theta_2$ for all points along ϕ, it can be easily seen that

$$AUC = P(|X_Y| < |X_{\bar{Y}}|), \tag{2.12}$$

which is similar to the probabilistic interpretation of AUC in the classical case.

2.3. OROC estimation

We present three methods for estimating the optimal ROC for our interval classification scenario, but these methods are generalizable to many binary classification scenarios.

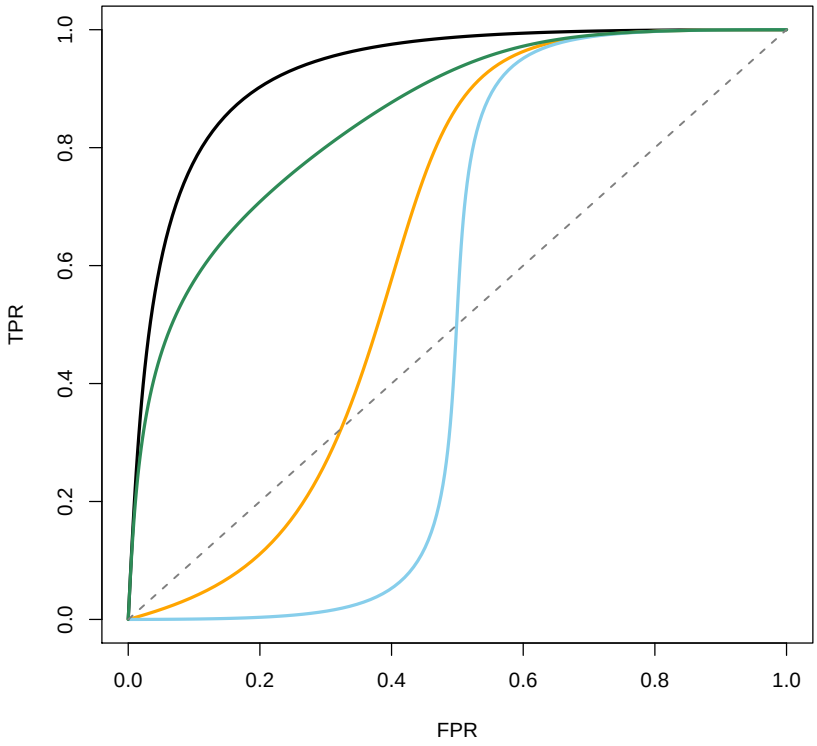

Figure 2.2: ROC curves corresponding to ϕ's in Fig 2.1.

2.3.1. Nonparametric estimation

Ideally, we want to estimate the ϕ that satisfies the definition of optimality (2.8) for a given classification scenario. From a random sample $(X_1, Y_1), \ldots, (X_n, Y_n)$, we may simultaneously estimate the ϕ and optimal ROC curve nonparametrically. Formally, for each possible value of $FPR = s$, we choose the parameters $\boldsymbol{\theta}$ that have the corresponding FPR and maximum TPR:

$$\widehat{\phi}_{Opt}(s) = \underset{\boldsymbol{\theta}:FPR(\boldsymbol{\theta})=s}{\arg\max}\ TPR(\boldsymbol{\theta}) := \{\boldsymbol{\theta} \mid \forall \boldsymbol{\pi} : TPR(\boldsymbol{\pi}) \leq TPR(\boldsymbol{\theta})\}. \tag{2.13}$$

Monotonicity is guaranteed since $FPR(\theta_1, \theta_2) \leq FPR(\theta_1 - \delta_1, \theta_2 + \delta_2)$ and $TPR(\theta_1, \theta_2) \leq TPR(\theta_1 - \delta_1, \theta_2 + \delta_2)$ for any $\delta_1, \delta_2 > 0$. $\widehat{\phi}_{Opt}(u)$ is the estimator for ϕ in (2.8); maximization of TPR over all values of FPR results in maximizing the AUC. Alternatively, for a given value of $TPR = s$, $\boldsymbol{\theta}$ could also be chosen to have the minimum FPR:

$$\underset{\boldsymbol{\theta}:TPR(\boldsymbol{\theta})=s}{\arg\min}\ FPR(\boldsymbol{\theta}) := \{\boldsymbol{\theta} \mid \forall \boldsymbol{\pi} : FPR(\boldsymbol{\pi}) \geq FPR(\boldsymbol{\theta})\}. \tag{2.14}$$

The two methods are equivalent and yield the same optimal ROC curve, since TPR for $\hat{\boldsymbol{\theta}}$ obtained for given value of FPR returns $\hat{\boldsymbol{\theta}}$ by definition, and vice versa.

The nonparametric method simultaneously estimates the classifier ϕ and the optimal ROC curve. In the interval classification scenario, ϕ is represented by a continuous path along the ROC surface in Figure 2.1, but ϕ may include any set of discontinuous or disjointed paths that give rise to the optimal ROC. From this point, we refer to the nonparametric optimal ROC classifier estimation as *NPE*. Appendix A is the R code for nonparametric estimation of the optimal ROC AUC.

2.3.2. Semiparametric estimation

An alternative method for estimating the optimal ROC for the interval classification is a generalized additive model (GAM) using quadratic penalized regression splines with smoothing parameters selected by REML (Hall, Hyndman, and Fan, 2004; Hastie and Tibshirani, 1986) for the classification of class membership. The model is simply

$$\text{logit}\{\Pr(Y = 1|X)\} = f(X) \tag{2.15}$$

where the function f is a smooth function of X, modeled using thin plate regression splines (Wood, 2003). For interval classification, the general shape of f is a smooth curve with a single hump centered between the intervals and low values outside of the interval (e.g. a quadratic curve).

The performance assessment of ϕ is straightforward: the predicted value from (2.15) can be thresholded as in the classical scenario, and a simple area under the ROC curve can be employed to estimate the optimal ROC. Additionally, $f^{-1}\{\text{logit}(\hat{p})\} \in \{\text{Im}(\hat{\phi})\}$ so ϕ can be estimated indirectly, and we refer to this semiparametrically estimated classifier as the *GAM classifier*. We implement this using the R (R Core Team, 2014) package mcgv (Wood, 2000, 2004, 2011).

2.3.3. Parametric estimation

Another alternative method for optimal interval classification is specifying a parametric model. For our interval classification scenario, we can incorporate the complex decision rule directly into a generalized linear model and maximize the model likelihood $\mathcal{L}$ under

$$\text{logit}\{\Pr(Y = 1|S)\} = \alpha + \beta I(\theta_1 < X < \theta_2). \tag{2.16}$$

Our parameter of interest is $\boldsymbol{\theta} = \{\theta_1, \theta_2\}$. We estimate $\{\boldsymbol{\theta}, \alpha, \beta\}$ using a profile likelihood $\mathcal{L}_{\boldsymbol{\theta}}(\hat{\alpha}_{\boldsymbol{\theta}}, \hat{\beta}_{\boldsymbol{\theta}})$. For each $\boldsymbol{\theta}$ using a grid search, we find $\hat{\alpha}$ and $\hat{\beta}$. Subsequently, we evaluate

$$\hat{\boldsymbol{\theta}}_{Opt} = \arg\max_{\boldsymbol{\theta}} \mathcal{L}_{\theta}(\hat{\alpha}_{\boldsymbol{\theta}}, \hat{\beta}_{\boldsymbol{\theta}}), \tag{2.17}$$

which we refer to as the *MLE classifier*. We assess the performance of the MLE classifier using ROC curve generated from the logistic model or direct classification.

2.4. Simulation

To compare the proposed estimation techniques for the optimal ROC, we simulate from data-generating distributions consistent with the interval threshold setting, where X_Y is sandwiched by $X_{\bar{Y}}$. We first consider symmetric distributions for simplicity using bimixture distributions, and we then consider estimating ϕ using asymmetric distributions. The data are independently generated for training and validation.

2.4.1. Simulation - symmetric distributions

Three different scenarios are considered for our simulations, with variations in distributional overlap between X_Y and $X_{\bar{Y}}$. In the world simulation, we use a normal distribution and a mixture distribution of normals (Figure 2.3a). For each observations $i = 1, ..., N$, we randomly sample $X_{Yi} \sim \mathcal{N}(0, 1)$, and for each observations $i = N + 1, ..., 2N$, we randomly sample $X_{\bar{Y}i} \sim \frac{1}{2}\mathcal{N}(-\mu, 1) + \frac{1}{2}\mathcal{N}(\mu, 1)$. In the second simulation, the distribution of the score for cases is modified to a uniform distribution (Figure 2.3b). For each observations $i = 1, ..., N$, we randomly sample $X_{Yi} \sim Unif(-1.5, 1.5)$, and for each observations $i = N + 1, ..., 2N$, we randomly sample $X_{\bar{Y}i} \sim \frac{1}{2}\mathcal{N}(-\mu, 1) + \frac{1}{2}\mathcal{N}(\mu, 1)$. The third simulation involves a biuniform distribution flanking a uniform distribution (Figure 2.3c); the distributional overlap is no longer transitional. For each observations $i = 1, ..., N$, we randomly sample $X_{Yi} \sim Unif(-1.5, 1.5)$, and for each observations $i = N + 1, ..., 2N$, we randomly sample $X_{\bar{Y}i} \sim \frac{1}{2}Unif(-\mu - 1.5, -\mu + 1.5) + \frac{1}{2}Unif(\mu - 1.5, \mu + 1.5)$.

2.4.2. Simulation results

We simulate $B = 1000$ datasets for each parameter combination and obtained AUC performance measures for the training and validation sets. Figure 2.4 summarizes the mean AUCs. In all simulations, the AUC increases with greater separation between X_Y and $X_{\bar{Y}}$ (i.e. increasing μ).

While there is a tendency for the NPE to overfit the data when the separation is small (i.e. greater overlap between X_Y and $X_{\bar{Y}}$), more so than the GAM and the MLE, the NPE performs better than GAM and MLE in small sample sizes and smaller separations in distributions. With sharp distributions (i.e. uniform, simulations 2 and 3 in Fig 2.4), the NPE once again performs better than GAM and MLE in smaller sample sizes and separations. In general, the optimal ROCs from NPE and GAM methods are more similar than those of the MLE method, and this may be attributed to the estimation of a single point in the MLE ϕ .

2.4.3. ϕ estimation

Figure 2.5 illustrates the NPE of ϕ. We simulated data using asymmetric distributions with the following parameters: $X_{Yi} \sim \mathcal{N}(1, 1)$, $X_{\bar{Y}i} \sim \frac{1}{2}\mathcal{N}(-0.5, 1) + \frac{1}{2}\mathcal{N}(2, 1)$, and $N = 500$ each for X_Y and $X_{\bar{Y}}$. The estimation procedure according to (2.13) is used.

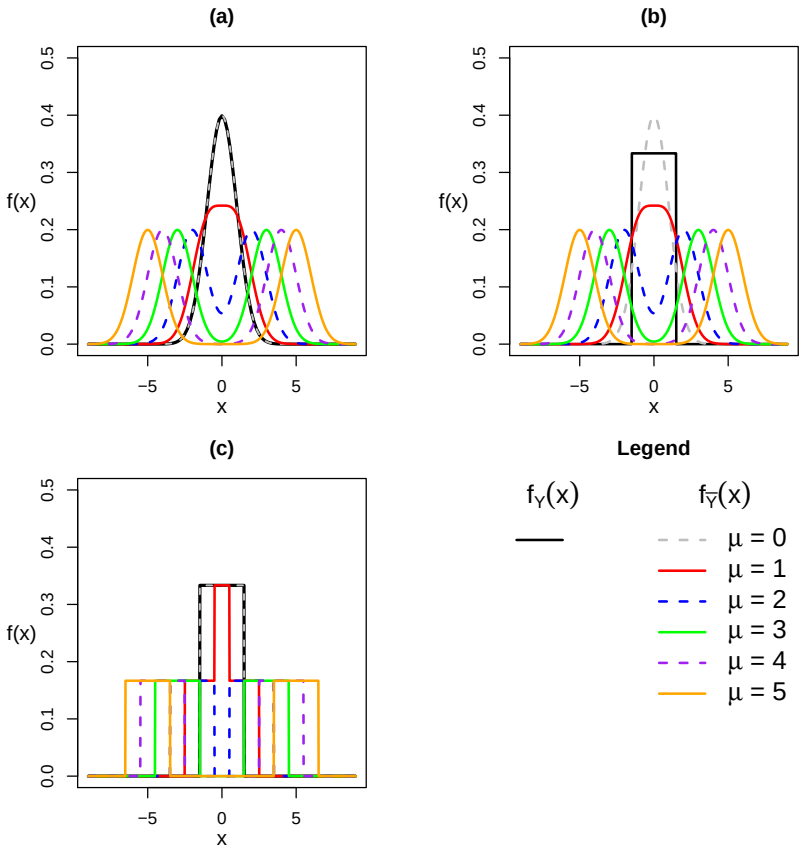

Figure 2.3: Simulation densities for X_Y and $X_{\bar{Y}}$. (a) mixture of normals for $f_{\bar{Y}}(x)$ and normal for $f_Y(x)$; (b) mixture of normals for $f_{\bar{Y}}(x)$ and uniform for $f_Y(x)$uniform; (c) mixture of uniforms for $f_{\bar{Y}}(x)$ and uniform for $f_Y(x)$.

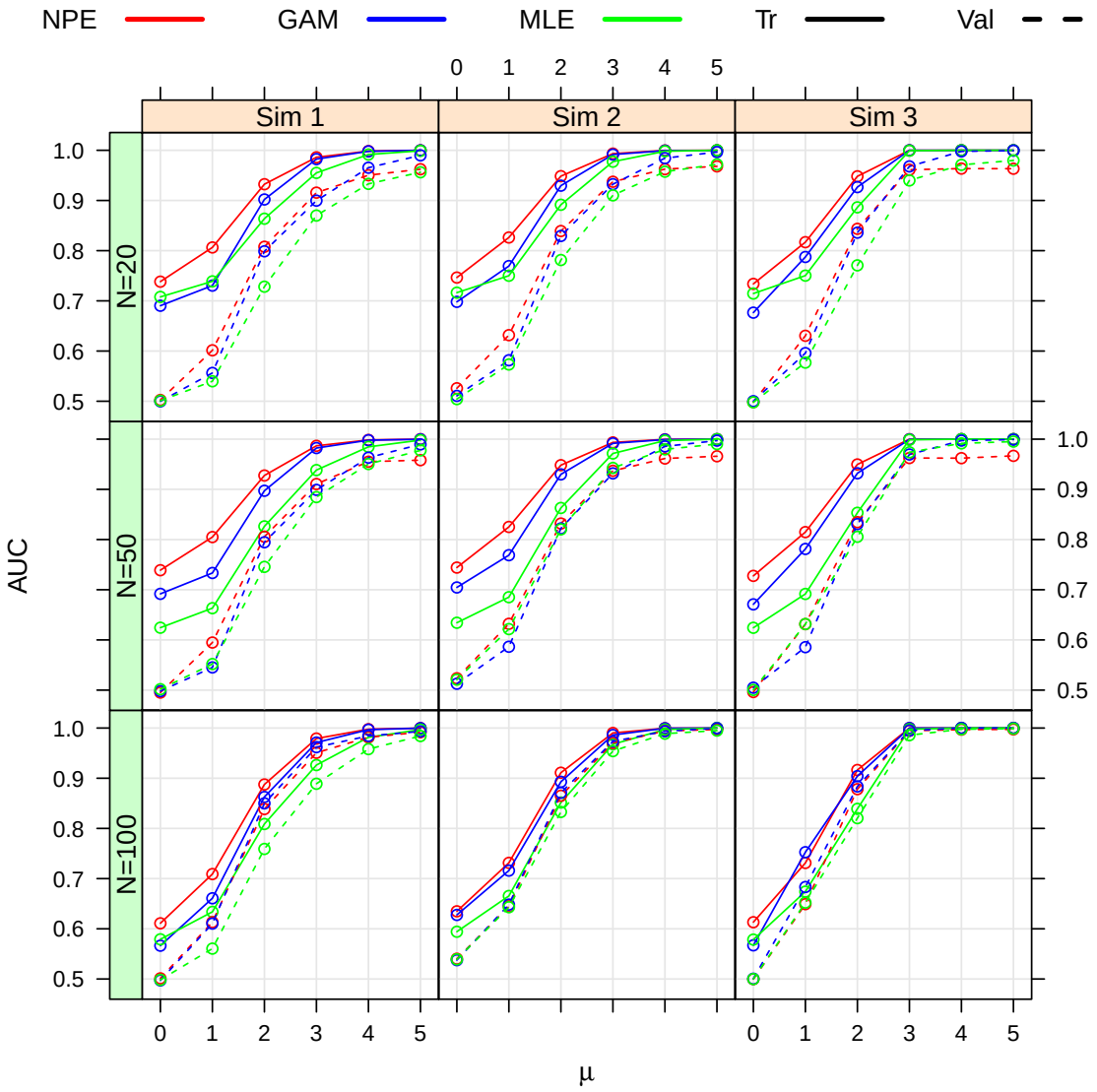

Figure 2.4: Mean AUCs (%) from 1000 simulations from independent training (Tr) and validation (Val) sets in simulations.

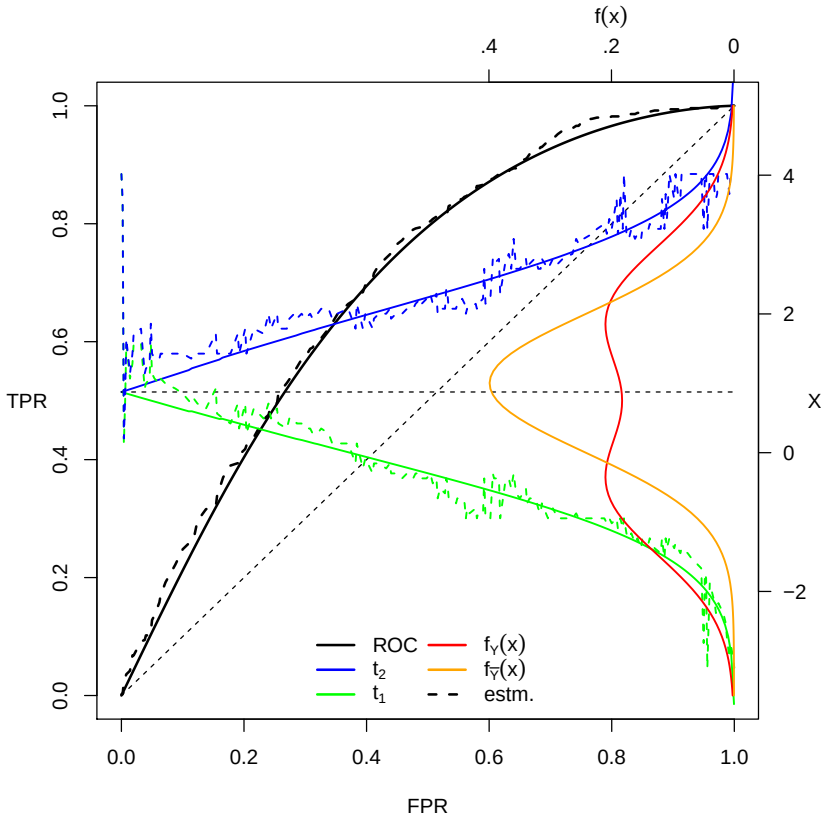

Figure 2.5: ROC curve and ϕ. Solid and dashed lines represent the truth (the data generating distributions shown in red and orange) and NPE of the ROC curve and ϕ.

The NPE ROC curve is remarkably close to the true optimal ROC curve. There is expected noise in $\widehat{\phi}$, especially at the beginning of the ROC curve, but the overall path is captured. The intervals between θ_1 and θ_2 in the true and estimated ϕ are consistent over the values of FPR.

2.5. Serum sodium data

The motivating dataset is from a retrospective study that sought to identify neurological factors associated with poor outcome in adult patients with fulminant bacterial meningitis (Muralidharan, Mateen, and Rabinstein, 2014). Serum sodium at admission was obtained in 39 patients hospitalized for fulminant bacterial meningitis at the Mayo Clinic in Rochester, Minnesota. The primary end-point of the study was in-hospital mortality.

Patients who have serum sodium levels outside of the ideal range of 136-145 mmol/liter may have

higher risk of poor outcomes (i.e. death) (Kratz et al., 2004), whereas serum sodium levels inside the ideal range would carry no additional risk of death. With the outcome defined as survival, we estimate the OROC using the NPE method and compare it with the GAM and MLE methods. Table 2.1 contains the complete data analyses along with bootstrapped 95% confidence intervals ($n = 1000$).

Although the GAM has the highest AUC at 0.83, it was not significantly different from the NPE and MLE methods (AUCs of 0.76 and 0.69, respectively). It is possible that due to a small sample size, outliers may force the flexible splines of the GAM to estimate a functional form that is cubic instead of quadratic in nature, thereby losing the fidelity to the classification decision rule. We also assess the performance of the different classifiers using subsampling at random at 30% and 50% of the observed sample size which shows with similar cross-validated results.

	NPE		GAM		MLE	
Sampling	Tr	Val	Tr	Val	Tr	Val
Complete data	75.5 (64.3-90.9)	-	82.9 (69.1-99.2)	-	68.6 (63.3-88.0)	-
30%	85.3	61.0	90.2	63.6	83.3	65.1
50%	80.7	66.7	83.5	70.3	76.3	65.8

Table 2.1: Serum sodium analysis: Complete data and mean AUCs (%) from 1000 resamples

2.6. Discussion

We generalize ROC-based performance assessment in complex classification settings by defining a classifier ϕ as a path in the ROC space. The example of interval thresholding illustrates the inherent limitations of applying the traditional binary decision rules associated with ROC curves, and the OROC offers an easy and intuitive framework to assess performance of interval classifiers.

Unlike the MLE classifier, where only a single classification table can be obtained, the NPE and GAM classifiers have flexibility in range of classification statistics. The NPE classifier, however, has the advantage of ease of interpretation; the directly estimated ϕ plotted over the ROC curve such in Figure 2.5 provides an intuitive look-up table for the classifier, and simple comparisons of new measurements to the optimal thresholds is straightforward. This simplicity is desirable in clinical diagnostic setting. In our example, $\widehat{\phi}$, the estimated path through Φ, is noisy, but a smooth function

may improve estimation of ϕ.

While GAM has the added benefit of natural accommodation of external covariates, which can be useful in applications such as tissue segmentation described above, existing methods for covariate-adjusted ROC analysis (Janes and Pepe, 2009) may be applicable to the NPE. Furthermore, the flexibility of the GAM may make it susceptible to influential outliers. This may result in loss of fidelity to an *a priori* classification rule, whereas the NPE and MLE methods are bound by the classification rule. Regarding comparison of NPE classifiers, existing tests using bootstrap estimates of standard errors of the AUCs can be easily performed (Hanley and McNeil, 1982) without being computationally expensive. Further work is necessary to investigate the asymptotic distribution of the NPE AUC with and without distributional assumptions.

In addition to improvements in classification performance in data scenarios described in this chapter, the simplicity of the performance measure of complex classifiers should make the application of NPE and GAM more compelling. The present work uses the bivariate classification setting example, but this methodology may be extended to multiple input scores $X_1, X_2, X_3, ...$ and threshold parameters $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3, ...$ Another interval thresholding scenario, albeit theoretical, may include multimodal thresholding, e.g. $I(t_{l_1} < S < t_{u_1} \cup t_{l_2} < S < t_{u_2} \cup ...)$, where clusters of intervals for class membership exist.

CHAPTER 3

INTENSITY NORMALIZATION OF PET IMAGES VIA DENSITY WARP REGRESSION

3.1. Introduction

Biomedical imaging modalities such as magnetic resonance imaging (MRI), computer tomography (CT), and positron emission tomography (PET) are established cornerstones of qualitative diagnostics that allow visualization of structures and physiological function in healthy and diseased subjects. Statistics have gradually transformed the use medical imaging from a qualitative to a quantitative tool, allowing greater discrimination and more comprehensive descriptions of disease status and prognosis. However, quantitative analysis of biomedical images is challenging due to the many sources of unwanted variation that confound the signal. These sources range from machine calibration and scan parameters to the patient's metabolic rate affected by ambient conditions (Coxson, 2013). Without comparability of measurements, formal statistical inference suffers from diminished power and potentially strong biases.

Image intensity normalization is generally acknowledged as a key preprocessing step in the analytical pipeline. Various methods exist in the literature, including histogram matching (Nyul and Udupa, 1999) and intensity normalization with respect to particular regions of interest (ROI) (Shinohara et al., 2014). In the latter example where particular anatomical structures are assumed to have similar physical consistency, such as the normal appearing white matter or cerebellar gray matter in the brain, simple z-score statistical normalization successfully removes a significant amount of nuisance variability due to parameters such as scanner and platform. More often than not, these benefits are conferred to imaging modalities with high resolutions (e.g. MRI) and well characterized tissue properties (e.g. density). In PET imaging, image resolution is bounded by an inherent uncertainty of the radionuclide tracer location, image reconstruction is dependent on near-perfect alignment of a reference image, and tissue-specific intensity is dependent on pharmacokinetics and physiological state of the body. The standardized uptake value (SUV), a relative measure of tracer uptake, compensates for the largest sources of signal variation, which include the amount of injected tracer, patient body weight, and radioactive decay. Summary measures of SUV and tissue-to-background ratio, which is calculated by dividing each SUV image by the subject's mean

SUV in reference tissues, are mainstays in quantitative PET. Increasingly, however, it is becoming clear that existing normalization methods for PET may be inadequate to handle these sources of variability (Huet et al., 2015; Keyes, 1995).

In this chapter, we propose a new statistical normalization strategy based on the application of functional data analysis (FDA) to image intensity distribution functions. FDA has been previously proposed for analyzing densities by treating empirical density curves as functional data (FD) objects (Alois Kneip, 2001) and conducting unsupervised analyses to investigate unwanted variation. One such method directly uses the empirical probability density functions (PDF) as FD objects (Delicado, 2011). Unfortunately, these normalizations suffer from undesirable properties, including the violation of regularity conditions of the Hilbert space-based methods when applied to density functions (Petersen and Mller, 2016). Outside of image analysis, supervised methods for normalization using functional data have been proposed, such as functional normalization (funnorm) by Fortin et al. (Fortin et al., 2014). This method uses simple additive models for quantile functions as FD objects. Unfortunately, coefficients near boundary values (i.e. at 0 and 1) may be subject to increased uncertainty in estimation when studying quantile functions. The funnorm approach also does not account for smoothness in the curves, but rather focuses on pointwise regression techniques in the context of gene expression distributions. We propose a novel FDA approach to density-valued data that uses a modified function-on-scalar regression applied to image-specific warping functions from a template density function. This method aims to increase the flexibility to capture more shape variation and reduce estimation uncertainty at boundary limits, while maintaining interpretability and improving statistical power for detecting group differences.

Our motivating example comes from the Vascular Inflammation in Psoriasis (VIP) Trial, a randomized, placebo-controlled study designed to test the effect of systemic therapy for psoriasis on systemic vascular inflammation as measured by PET/CT. Psoriasis is a common inflammatory disease that prominently manifests in the skin, but it is also independently associated with many cardiovascular comorbidities. These include myocardial infarctions (MI), stroke, and cardiovascular death (Gelfand et al., 2006; Ogdie et al., 2015). The underlying inflammatory mechanism in psoriasis is shared by other diseases that are associated with cardiovascular burden, and lowering systemic inflammation may mitigate the risk. Since inflammation involves biological processes with high metabolic activity (i.e. glucose consumption (Emami and Tawakol, 2014)), a radionuclide

Figure 3.1: Left: PET image. Intensity levels (shown in jet color spectrum) indicate the level of ^{18}F-FDG uptake in tissue. Right: Corresponding CT image in grayscale. The anatomical structures are more pronounced, and the intensity levels indicate tissue density.

glucose-analogue, ^{18}F-fludeoxyglucose (FDG), may be used as a biomarker to measure the level of metabolic activity. Figure 3.1 is a coronal (frontal) plane image view of a PET/CT scan of a study patient; red regions, including the brain, show tissue with high ^{18}F-FDG uptake, indicative of high metabolic activity. Our goal in this work is to improve quantification of systemic inflammation through improved normalization of the PET image. In the remainder of this chapter, we apply our normalization to the VIP data and compare the correlation between densities and cardiovascular biomarkers before and after normalization.

3.2. Methods

3.2.1. Density-valued data

In PET, the raw unit of measure is radioactivity concentration of becquerels per cubic centimeter (Bq/cc). This unit reflects the amount of radionuclide tracer concentration (i.e. ^{18}F-FDG) in tissue, and is the quantitative unit for a given voxel in a PET scan image. We denote the intensity at a given voxel v as $Y(v)$, which arises from an intensity distribution f. We make fundamental assumptions of the data that the intensity distribution of a scan for a patient is a realization of a stochastic process that generates these distributions. We further assume that the variations from a template density f_m are attributable to biological and non-biological factors, and can be modeled statistically.

3.2.2. Warping functions

For each subject i, we define the density-valued image intensity data as the empirical probability density function $f_i(x)$, where x represents image intensity values and ranges. Likewise, we let $F_i(x)$ represent corresponding cumulative distribution function. We denote the population template function by $F_m(x) = \int_0^x f_m(u)du$, and define the warping function $w_i(x)$ as the function that maps $F_m(x)$ to $F_i(x)$. That is, a warping function w_i remaps the domain such that

$$F_m\{w_i(x)\} = F_i(x). \tag{3.1}$$

The warping function is represented as a curve, with the identity function $w_I(x) = x$ indicating no warping. To define the template function, we use the depth measure of centrality of a given curve within a group of curves as defined by Fraiman and Muniz (2001). We use the functional median of a set of curves, which is defined as the function $F_m(x)$ having the greatest integrated depth.To assess the systematic effects of sources of variation that are not related to the biological processes of interest, we next propose to study the warping functions using functional regression models.

3.2.3. Restricted function-on-scalar regression

Following the assumption that the variations from the population mean or template intensity distributions can be explained by biological and non-biological factors, we can model the effects of these factors using the warpings as the functional responses. We denote by $\mathbf{V}_i = (V_{ij})_{j=1}^J$ the vector

of observed variables that are not of interest for the study; for example, the dose of radionuclide tracer administered only introduces unwanted variation into the measurements. We further denote by $\mathbf{X}_i = (X_{ik})_{k=1}^K$ the vector predictors of interest for subject i. Since CDFs range from 0 to 1, the warping functions at domain boundaries are restricted to the identity functions. This can be further simplified by subtracting the $w_I(x)$ from $w_i(x)$ to yield functional responses $r_i(x)$ that are restricted to 0 at the domain boundaries.

The $r_i(x)$ can be modeled as functional responses for regression models with scalar predictors (i.e. biological and non-biological factors). We employ regression models for functional responses and scalar predictors (Ramsay, 2005; Reiss, Huang, and Mennes, 2010):

$$\boldsymbol{r}(x) = \boldsymbol{Z}\boldsymbol{\beta}(x) + \boldsymbol{\epsilon}(x). \tag{3.2}$$

In this scenario, x ranges over some finite interval $X \subset \mathbb{R}$, and $\boldsymbol{r}(x)$ can be represented as an N-dimensional functional response vector. The design matrix $\boldsymbol{Z} = [\boldsymbol{X}\ \boldsymbol{V}]$ is $N \times q$ dimensional, $\boldsymbol{\beta}(x) = [\beta_1(x), \ldots, \beta_q(x)]^T$ is the functional coefficients vector, and $\boldsymbol{\epsilon}(x)$ is the functional error vector. Consider the b-spline basis function representation of $\boldsymbol{r}(x)$,

$$\boldsymbol{r}(x) = \boldsymbol{C}\boldsymbol{\theta}(x), \tag{3.3}$$

where $\boldsymbol{\theta}(x) = [\theta_1, \ldots, \theta_K]^T]$ is the vector of K b-spline functions and $\boldsymbol{C}$ is an $N \times K$ matrix of basis coefficients. The coefficient functions in (3.2) are then represented as

$$\beta_k(x) = \boldsymbol{b}_k\boldsymbol{\theta}(x), \tag{3.4}$$

where $\boldsymbol{b}_k$ is the basis coefficient vector. The problem reduces to estimating $\boldsymbol{B} = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_q)^T$ using this general form of (3.2):

$$\boldsymbol{r}(x) = \boldsymbol{Z}\boldsymbol{B}\boldsymbol{\theta}(x) + \boldsymbol{\epsilon}(x). \tag{3.5}$$

However, for the warping functions to yield proper CDFs, the aforementioned restriction to $\boldsymbol{r}(x)$ must be implemented. The basis function representation in (3.3) offers a simple solution to implement the restriction. Given sufficient K basis functions, we modify $\boldsymbol{B}$ and $\boldsymbol{\theta}(x)$ by removing the first and

last b-spline basis components:

$$\boldsymbol{B}^*_{N\times(K-2)} = \boldsymbol{B}_{N\times K} \times \begin{bmatrix} \mathbf{0}_{K-2} \\ \boldsymbol{I}_{K-2} \\ \mathbf{0}_{K-2} \end{bmatrix}_{K\times(K-2)}, \tag{3.6}$$

and

$$\boldsymbol{\theta}^*(x) = [\theta_2, \ldots, \theta_{K-1}]^T. \tag{3.7}$$

We estimate $\boldsymbol{B}^*$ by minimizing the following:

$$\int \|\boldsymbol{C\theta}(x) - \boldsymbol{ZB}^*\boldsymbol{\theta}^*(x)\|^2 dt + \sum_{k=1}^{q} \lambda_k \int [L(\boldsymbol{b}_k^*\boldsymbol{\theta}^*(x))]^2 dt. \tag{3.8}$$

The second term is a roughness penalty with a non-negative tuning parameter λ and a linear differential operator L. We use a second derivative operator for L.

Since the basis functions θ_1 and θ_K only contribute to the functional range at the boundaries of X, we first estimate $\boldsymbol{B}^*$ and then add the zero coefficients, i.e. $\hat{\boldsymbol{B}} = [\mathbf{0}_K \ \hat{\boldsymbol{B}}^* \ \mathbf{0}_K]$. We calculate the estimates of coefficient functions $\boldsymbol{\beta}_k(x)$ and their standard errors from $\hat{\boldsymbol{B}}$ using standard methods penalized likelihood estimating λ using generalized cross-validation. We implement the restricted function-on-scalar regression (`rfosr`) with the aforementioned modifications to the `fosr` function (Reiss, Huang, and Mennes, 2010) in the `refund` package for R (Huang et al., 2015). Appendix B is the R code for `rfosr`.

3.2.4. Normalization

We assume that the degree to which an intensity distribution of a particular scan differs from the population template is captured by the warping functions, and that the modeled estimated coefficient for various factors are additive in nature and can be adjusted for. Our overall approach to normalization is to regress out the effects of nuisance factors from the warping parameters, and ultimately estimate an intensity distribution $F_i^{norm}(x)$ adjusted for the nuisance factors for each image. Using functional regression, we separate $\boldsymbol{Z}$ and $\boldsymbol{\beta}(x)$ to K predictors of interest and J nuisance variables in (3.2):

$$\boldsymbol{r}(x) = \boldsymbol{X}_K\boldsymbol{\beta}_K(x) + \boldsymbol{V}_J\boldsymbol{\beta}_J(x) + \boldsymbol{\epsilon}(x). \tag{3.9}$$

We form the normalized functions,

$$\hat{\boldsymbol{r}}^{norm}(x) = \boldsymbol{r}(x) - \boldsymbol{V}_J\hat{\boldsymbol{\beta}}_J(x), \tag{3.10}$$

which we use to form the normalized warping function functions $\hat{\boldsymbol{w}}^{norm}(x)$. Following the warping function representation (3.1), the we normalize distributions using the normalized warping functions:

$$\hat{F}_i^{norm}(x) = F_m\{w_i^{norm}(x)\} \tag{3.11}$$

Figure 3.2 illustrates the normalization procedure. First, we convert empirical densities $f_i(x)$ (A) to $F_i(x)$ (B). We choose a template $F_m(x)$ (C) and calculate warping functions $w_i(x)$ (D). We then regress out the nuisance effects on the warping functions and estimate the normalized warping functions $w_i^{norm}(x)$ to obtain the normalized density estimates $\hat{F}_i^{norm}(x)$ (E) and $\tilde{f}_i^{norm}(x)$ (F).

3.3. VIP Trial data

3.3.1. Motivating Study

As a proof of concept, we apply our normalization to the VIP Trial PET/CT imaging data and assess sensitivity to associations between metabolic activity and lipoprotein particle biomarkers known to be associated with cardiovascular risk (Austin et al., 1988; Gordon et al., 1989). We study the PET scans, with intensities recorded in standardized uptake values (SUV), defined voxel-wise as

$$SUV(v) = \frac{I(v)}{C/W} \tag{3.12}$$

where $I(v)$ is the PET scan and W is the weight of the patient. C is the corrected radionuclide tracer activity, calculated as

$$C = T \cdot 2^{-\frac{t_S - t_I}{t_{1/2}}}, \tag{3.13}$$

where T is the injected tracer activity, t_S is the time of scan, t_I is the time of tracer injection, and $t_{1/2}$ is the half life. Base 10 logarithm transformation of SUVs are used as primary units due to the highly skewed distributions of the whole-body voxel SUVs.

One of the trial objectives is to obtain full-body PET/CT scans, but due to scanner limitations,

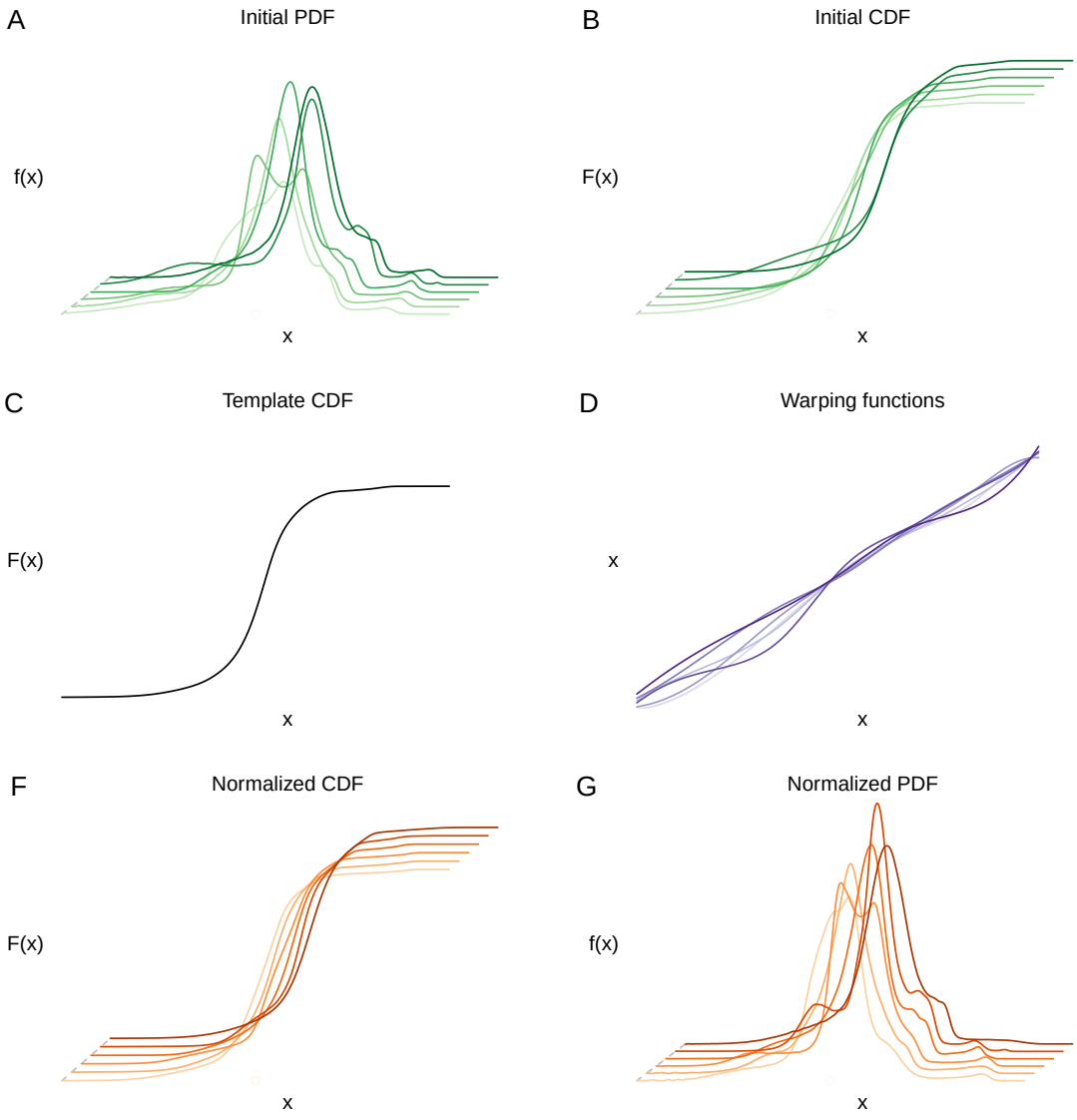

Figure 3.2: Normalization workflow. Empirical densities $f_i(x)$ (A) are transformed to $F_i(x)$ (B). A template density $F_m(x)$ (C) is chosen and warping functions $w_i(x)$ (D) are estimated. Using a modified functional regression, normalized warping functions $w_i^{norm}(x)$ are calculated, which are then used to estimate the normalized densities $F_i^{norm}(x)$ (E) and $f_i^{norm}(x)$ (F).

single-pass full-body scans are not available for many subjects and instead the torso and lower extremities are often scanned separately. To standardize the scanning protocol, we use single-pass PET/CT scans and truncate at the same anatomic landmark for each scan (the femoral head). Additionally, to eliminate scanner variability, we analyze scans from a single facility at the Hospital of the University of Pennsylvania. We obtain log SUV densities after removing background using active contour segmentation of tissue in ITK-SNAP (Yushkevich et al., 2006).

For the model, we choose variables that are known contributors to SUV variability, such as weight, drug incubation time (i.e. $t_S - t_I$), and radionuclide dose (Boellaard, 2009), as well as plausible biological factors such as age and sex. Since the underlying disease process is understood to be associated with systemic metabolic activity and thus disease, we also included the psoriasis area severity index (PASI), a clinical measure of psoriasis severity, in the model but do not residualize based on PASI during normalization. Finally, we compare the SUV means before and after normalization, and correlate the means with lipid and inflammatory biomarkers using each density's mean as a coarse measure of total systemic inflammation.

3.3.2. Results

Figure 3.3 shows the estimated functional coefficients of the covariates on $w_i(x)$, and figure 3.4 shows the densities before and after normalization (N=32). Of the six scalar covariates in the model, incubation time has the greatest effect on the warping function, whereas FDG dose has the least effect on the warping function.

Table 3.1 summarizes the association between the SUV density means and biomarkers pre- and post-normalization (N=30). We use spearman's ρ as a measure of correlation. None of the original density means are statistically associated with the various lipoprotein and inflammatory biomarkers. Employing the proposed normalization, we find LDL cholesterol concentration (LDLc), LDL particle size (LDLp), very large LDL particle size (vl LDLp), total cholesterol (Tc), and IL-6 to be statistically significantly associated density means.

3.4. Discussion

We introduce a new intensity normalization method for quantitative biomedical images using functional regression of density warping functions. Our method models the intensity distributions as

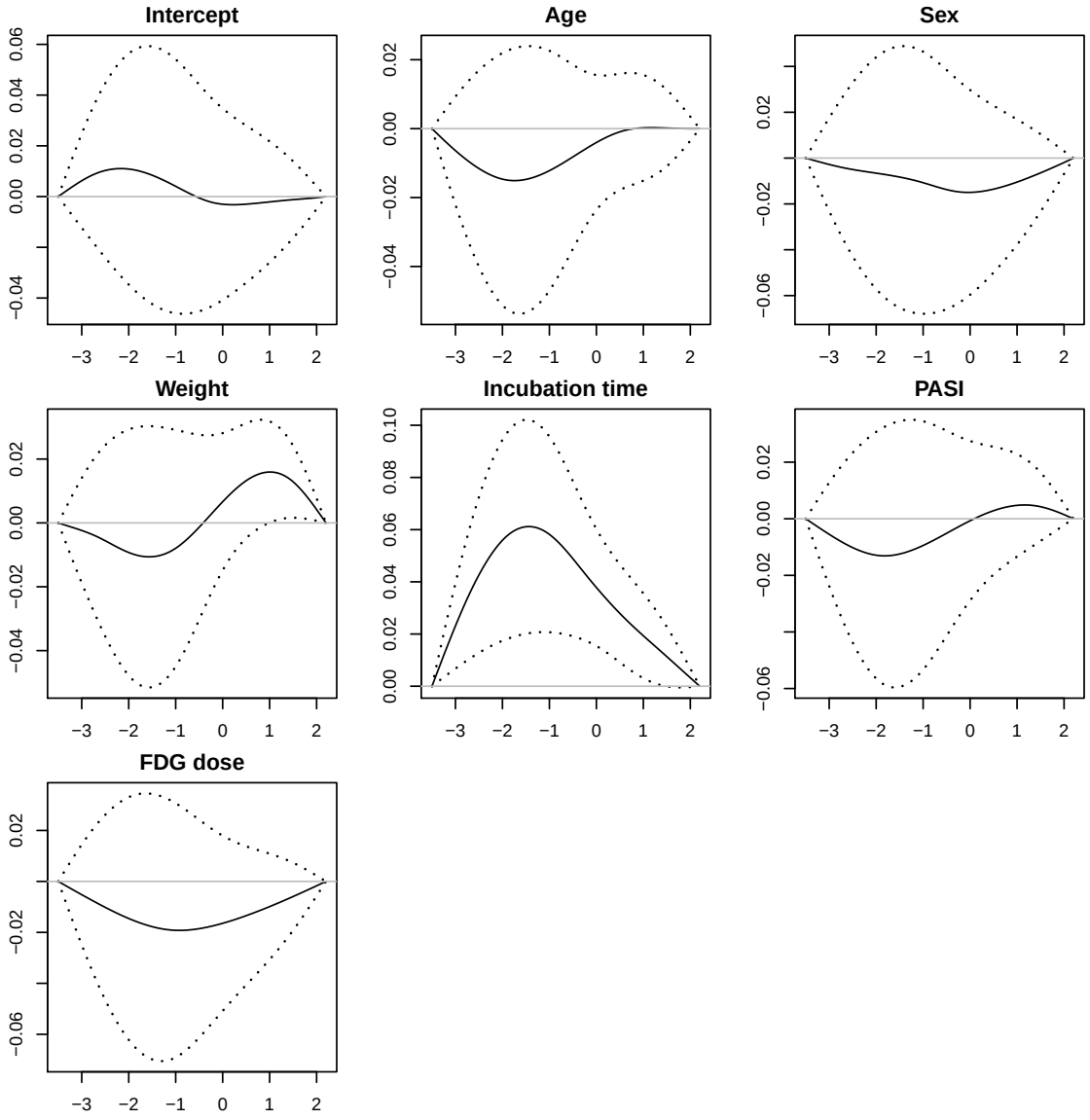

Figure 3.3: Coefficient functions of the restricted function-on-scalar regression using six covariates.

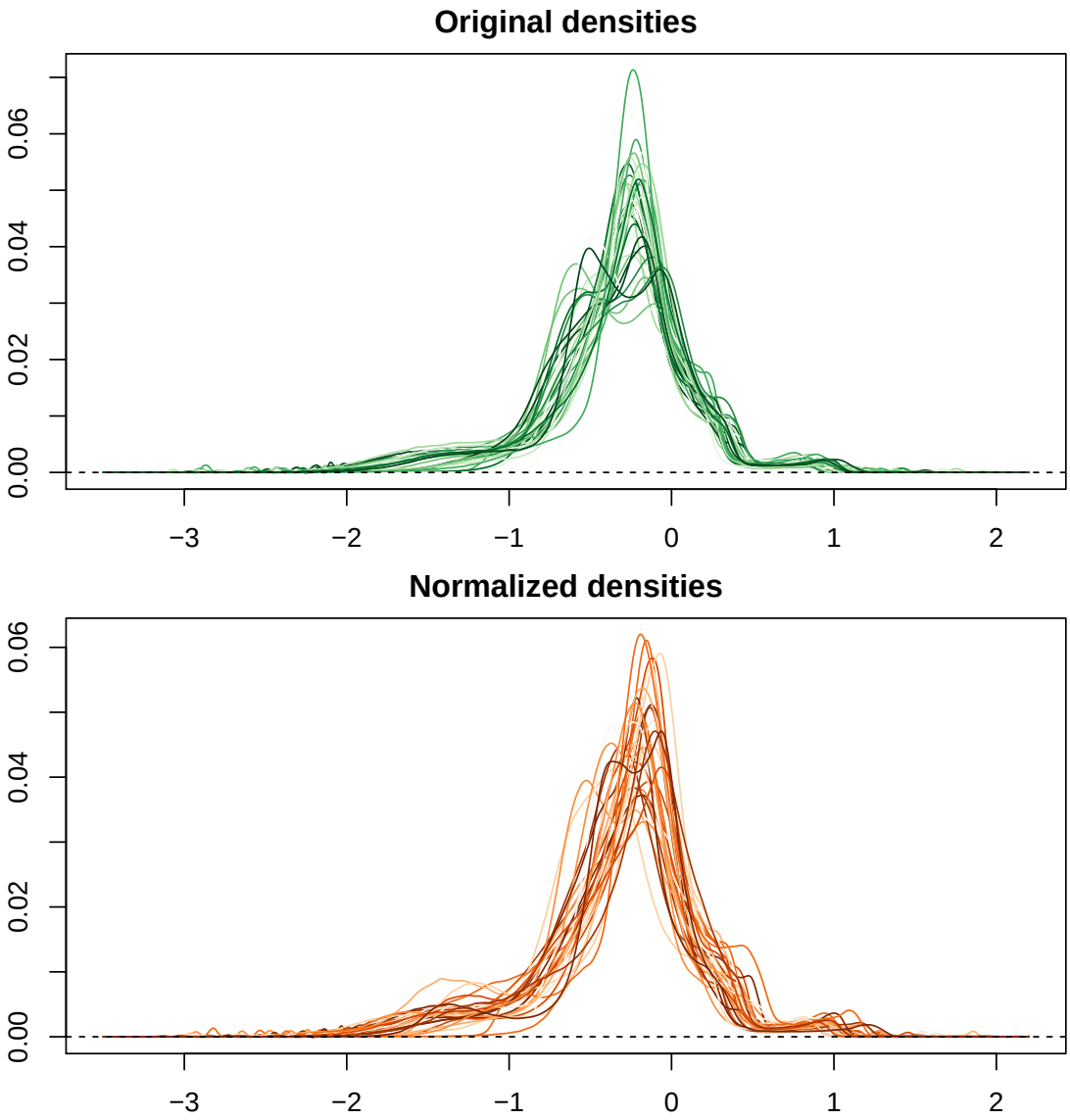

Figure 3.4: Empirical densities of PET intensities of 32 baseline scans. The top curves represent original densities, and the bottom represent densities after removing the effects of nuisance parameters.

	Original		Normalized	
Biomarker	ρ^*	p-value	ρ^*	p-value
HDLc	17	0.38	10	0.61
HDLp	3	0.87	-1	0.97
HDLz	3	0.89	0	1.00
S HDLp	-23	0.22	-22	0.23
M HDLp	10	0.59	8	0.66
LM HDLp	16	0.40	13	0.49
L HDLp	20	0.30	16	0.39
IDLp	-11	0.55	9	0.63
LDLc	16	0.40	44	**0.02**
LDLp	10	0.59	46	**0.01**
LDLz	20	0.29	2	0.91
S LDLp	-1	0.94	29	0.11
L LDLp	31	0.10	28	0.13
VL LDLp	8	0.68	44	**0.01**
VLDLp	-23	0.22	1	0.97
VLDLz	-0.00	0.99	22	0.25
S VLDLp	-32	0.08	-22	0.24
M VLDLp	-9	0.62	17	0.37
LM VLDLp	-11	0.56	17	0.37
L VLDLp	-9	0.64	8	0.68
VLDLtg	-15	0.43	12	0.53
Efflux value	20	0.31	13	0.52
Tg	-14	0.47	16	0.41
Tc	16	0.40	39	**0.03**
LPIR	-2	0.92	17	0.37
CRP	-9	0.62	-6	0.77
IL-6	-30	0.10	-37	0.05

Table 3.1: Correlation between density means and various lipoprotein particle biomarkers. Abbreviations: HDL, high density lipoprotein; LDL, low density lipoprotein; VLDL, very low density lipoprotein; Tg, triglyceride; Tc, total cholesterol; LPIR, lipoprotein insulin resistance score; S, small; M, medium; LM, large-to-medium; L, large; VL, very large. CRP, C-reactive protein; IL-6, Interleukin 6. Suffixes: z, size; p, particle number; c, cholesterol concentration. $^*\rho \times 100$ are shown.

functions of nuisance imaging protocol parameters and removes their effect. By modifying the b-spline functional basis, we restrict the warping functions to yield proper density functions for normalization. One of the strengths of our method is that it allows the use of full scan intensity densities instead of using tissue-specific densities. Additionally, since we normalize the entire densities, the use of reference tissues is not required.

Our application to the VIP Trial data shows promise that normalization may reveal signals that may otherwise be obfuscated by nuisance parameters. This is extremely important for PET imaging, where numerous imaging and pharmacokinetic parameters are found to be contributing to SUV

variability. The regression model indicates that the standardization using the SUV did not adequately "standardize" the effects of the imaging parameters, in particular the FDG incubation time. Increases in SUV have been noted with increases in FDG incubation time (Basu et al., 2007), with standardizing the scan initiation time suggested as a corrective measure. In multi-site studies, these corrective measures such as these may fall short due to the complex nature of PET/CT acquisition and a statistical normalization may offer the best standardization.

While we have assessed our normalization in terms of association of the density means to lipoprotein biomarkers, another indicator for the usefulness of normalization would be to assess the primary outcomes of the VIP Trial, which are mean PET signals in the aorta. The density means may offer a measure of cardiovascular risk via systemic inflammation, but they may also oversimplify the relationship between metabolic activity of the entire body and an inflammatory disease process. Targeted quantification of the aorta using manual segmentations directly assesses the atherosclerotic burden, and the use of normalized images, i.e. applying w_i^{norm} and using $Y^{norm}(v)$, has the potential to improve dose-response and/or treatment effect signals. It is noteworthy, however, that LDL cholesterol concentration and particle size, which are associated with cardiovascular disease (Sacks and Campos, 2003), are correlated with normalized density means. Our findings are suggestive of stronger whole-body FDG signal throughout the body in patients with unhealthy lipoprotein profiles. The significance of the weak inverse correlation of IL-6 and whole-body FDG signal remains unclear, as IL-6 is implicated in pro- and anti-inflammatory processes (Scheller et al., 2011).

The overall goals of normalization should be the comparability of quantitative imaging units for population-level analysis, and the statistical principles of image normalization (Shinohara et al., 2014) provide guidelines for image normalization. Although the proposed methodology is theoretically consistent with these principles, further work is required to empirically assess the conformity of our normalization process to these guidelines in large multi-center studies. Some assessments are straightforward, such as testing the monotonicity of warping functions to assess the intensity rank preservation. Preserving a similar distributions for similar tissues of interest may be more challenging when a disease process may affect the distributions of every tissue class, such as the systemic inflammatory process associated with psoriasis. Future extensions of our work will be developed to normalize scans over time to assess treatment effects, especially when the treatments

may affect the signal densities. Disentangling treatment effects in a tissue of interest from its effect on the overall scan is paramount, and functional mixed effect-based models may be appropriate for modeling the warping functions these settings.

CHAPTER 4

NONLINEAR MIXED EFFECTS MODELING OF AMYLOID-β TRAJECTORIES IN PET IMAGING

4.1. Introduction

Alzheimer's disease (AD) is a neurodegenerative disease characterized by progressive dementia unrelated to normal aging. The disease onset, or time of first symptoms, varies from person to person, but the the disease progression is consistently characterized by accumulation of amyloid plaques in extracellular space between neurons and neurofibrillary tangles inside neurons. It is believed that these characteristic structural features disrupt normal brain function and lead to loss of neurons. Of interest is how these physical changes in the brain progress over time and correlate with progression of disease. Accurately measuring amyloid-beta (Aβ) protein levels, the components of amyloid plaques, is important in understanding the plaque burden as well as stage of disease progression. In-vivo imaging of these structures may enable earlier AD diagnosis and guide therapeutic regimens.

There is strong evidence that Aβ PET imaging is a promising biomarker of brain Aβ-plaque load (Kepe et al., 2013). Currently there are two PET imaging compounds for Aβ: 1) Pittsburgh-compound B (PiB) (Klunk et al., 2004), and 2) florbetapir (Clark et al., 2011), and both radionuclide tracers bind to Aβ plaques in the brain. While PET applications using PiB has been around since the early 2000's, the more recent florbetapir has a favorable half-life profile (109.8 min in ^{18}F florbetapir compared to 20.38 min in ^{11}C PiB), which allows an easier imaging protocol.

Several trajectory shapes of major AD biomarkers have been theorized, and they generally are based on the sigmoidal model (Caroli and Frisoni, 2010; Jack et al., 2010). However, their use in charactering Aβ PET biomarker trajectories is limited; linear mixed effects (LME) models are generally the basis of longitudinal trajectory modeling (Jack et al., 2012; Resnick et al., 2010, 2015). In the context of a sigmoidal trajectory model, the obvious limitation of LME model is the nonlinearity of the hypothesized trajectory; linearity is assumed once a threshold is reached (i.e. Aβ-positivity), but the baseline (bottom plateau) and Aβ saturation (upper plateau) are not properly

accounted for.

In this chapter, we propose a new method of modeling the Aβ trajectories based on a couple of biological assumptions. The first assumption is that each subject has a unique Aβ trajectory based on a common functional form (i.e. sigmoidal curve). The second assumption is that there are subjects for whom the accumulation is not observed. In Section 4.2, we incorporate these assumptions by combining a nonlinear mixed model framework with mixture of heterogeneous populations, and formulate the trajectory model in a Bayesian framework. In Section 4.3, we evaluate the performance of the Bayesian model in estimating the parameters using Monte Carlo simulation studies. In Section 4.4, we apply these methods to the florbetapir dataset from the Alzheimer's Disease Neuroimaging Initiative (ADNI) cohort for estimating group differences in Aβ trajectories.

4.1.1. ADNI data

The Alzheimer's Disease Neuroimaging Initiative (ADNI) is a consortium of researchers whose goal is to collect, validate, and utilize study data to define the progression of AD. MRI and PET study data are used, in addition to genetics, cognitive tests, and CSF/blood biomarkers, to understand and predict AD. Florbetapir PET imaging has been collected in a cohort of subjects with ADNI, which is publicly available (`adni.loni.usc.edu`).

The unit of Aβ florbetapir measure is the standardized uptake value ratio (SUVR), which is the nonweighted SUV average across four main cortical regions (frontal, anterior/poster cingulate, lateral parietal, lateral temporal) divided by the SUV average across the composite reference region (whole cerebellum, brainstem/pons, and eroded subcortical white matter) (Landau et al., 2015). Figure 4.1 is a spaghetti plot of the Aβ trajectories in the ADNI cohort as measured by PET imaging. To illustrate genetic differences, four color groups indicate the different AD risk profiles based on APOE gene alleles. Typically, there are two to three Aβ measurements per subject, with an average time between subsequent scans being 2 years. Due to radiation toxicity and cost of scans, repeated measurements throughout the course of Aβ accumulation are not performed.

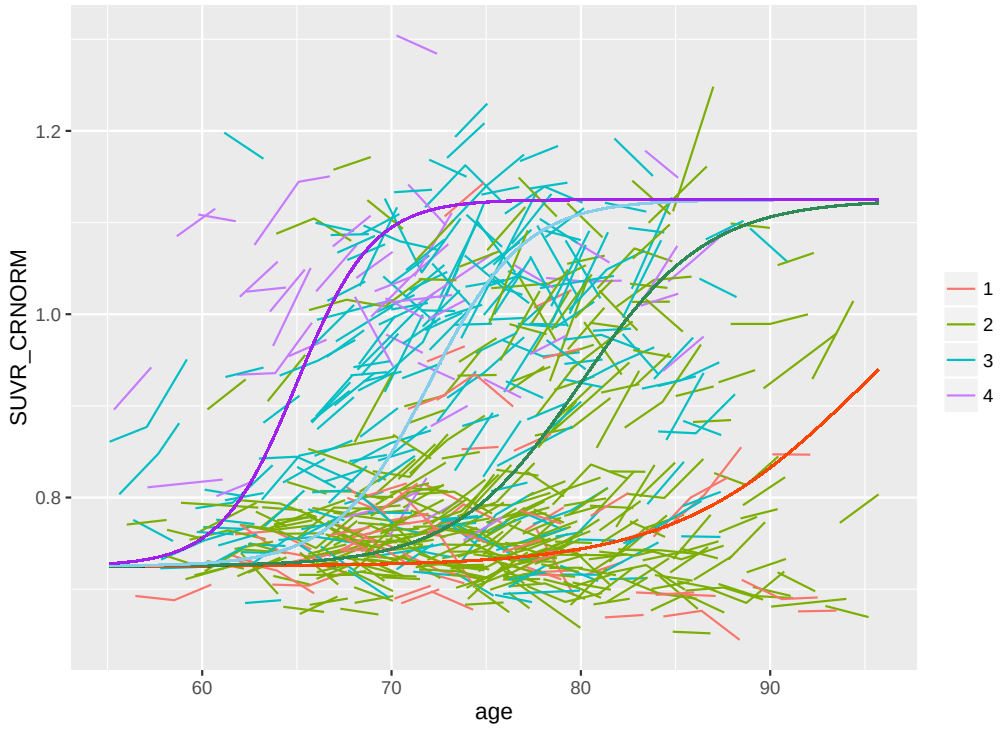

Figure 4.1: Trajectories by APOE risk category and overlay of proposed trajectories

4.2. Methods

4.2.1. Assumptions

As mentioned previously, we make several assumptions about Aβ trajectories:

1. An individual has a unique Aβ trajectory that follows a common trajectory shape
2. An individual may or may not exhibit Aβ accumulation

The first assumption is based on the literature of AD biomarkers that indicate Aβ accumulation is a common biological phenomenon in all subjects. The AD disease progression is associated with Aβ accumulation in the brain, from a baseline value corresponding to a healthy level to an upper value corresponding to a saturated Aβ level. However, healthy, nondemented individuals are subject to accumulation of Aβ (Braak and Braak, 1991), and amyloid plaque load does not correlate well with the degree of cognitive impairment (Hardy and Selkoe, 2002). Sections 4.2.2 and 4.2.3 describe the common trajectory and its model framework.

The second assumption is based on the observation that many subject-level trajectories do no accumulate over time and stay at a baseline level (i.e. flat) far into the observation window, even in subjects who are at higher risk of AD. In the context of the first assumption, it is plausible that, given adequate follow-up without censoring due to death, an individual will exhibit the characteristic accumulation of Aβ. Section 4.2.4 incorporates the second assumption into the model framework.

4.2.2. Trajectory function

The prevailing trajectory shape for cortical Aβ accumulation is the sigmoidal curve (Jack et al., 2010). Under assumptions discussed by these authors, each subject's Aβ trajectory follows a common sigmoidal curve, and the general model for $Y(t)$, the SUVR measurement at time t, is of the form

$$Y(t) = A + \frac{K - A}{1 + e^{-B(t-M)}}, \tag{4.1}$$

which is the generalized logistic function. The parameters are as follows: A - lower asymptote; K - upper asymptote; B - shape/slope parameter; M - location parameter. In terms of Aβ trajectory, A describes the baseline (i.e. pre-accumulation) level of Aβ, and K is the limit of Aβ capacity of the brain. M describes the time of the inflection point and thereby is the parameter of onset of Aβ accumulation. It should be noted that M is not interpreted as the time of onset of accumulation; rather, M and B together determine the time of accumulation onset (i.e. start of the exponential growth phase).

4.2.3. Nonlinear mixed effects model

Nonlinear mixed effects models, or hierarchical nonlinear model, provide a flexible framework when linear mixed effects models are inadequate (Marie Davidian, 2003). We follow this framework to model the outcome data. Let Y_{ij} be a response variable for florbetapir SUVR, the biomarker for Aβ, at t_{ij}, the jth time measurement in age, for the ith subject for $i = 1, \ldots, n$. Additionally, let m_i and a_i be the subject-specific location and lower asymptote parameters for the ith subject. The subject-specific model of the response variable is as follows:

$$Y_{ij} = a_i + \frac{K - a_i}{1 + e^{-B(t_{ij}-m_i)}} + \epsilon_{ij}, \tag{4.2}$$

with

$$a_i \sim Unif(A_L, A_U),$$
$$m_i \sim \mathcal{N}(M, \sigma_M^2),$$
$$\epsilon_{ij} \sim \mathcal{N}(0, \sigma_\epsilon^2).$$

M is the population-level parameter of Aβ onset, and we assume subject-level parameters are normally distributed about M. Subject-level baseline of Aβ is assumed to be uniformly distributed within the clustered observations. Furthermore, this model assumes conditional, or serial, independence (i.e. $\epsilon_{ij} \perp \epsilon_{ik} \mid a_i, m_i$). We ignore serial correlation as its effects are often dominated by the combination of random effects and measurement error (Verbeke and Molenberghs, 2001).

4.2.4. Mixture distribution

We define accumulators ($G = 1$) as subjects whose observed Aβ trajectories have a positive slope, and non-accumulators ($G = 0$) as subjects whose observed Aβ trajectories are flat at baseline during the observation period. By our first assumption, all subjects have a common sigmoidal Aβ curve, so we address the second assumption of non-accumulation of certain subjects by introducing a shift in M such that the exponential growth portion (i.e. Aβ accumulation) is not observed in the age range. The mixture distribution of m_i is the following:

$$m_i \sim \pi \times \mathcal{N}(M_1, \sigma_{M_0}^2) + (1 - \pi) \times \mathcal{N}(M_0, \sigma_{M_1}^2), \tag{4.3}$$

where π is the probability of accumulator membership $P(G = 1)$, and M_1 and M_0 are the Aβ onset parameter group means of the accumulators and non-accumulators, respectively. Furthermore, π can be modeled based on subject-specific risk profiles such as a high-risk allele of a known gene, e.g.:

$$\text{logit}\Big\{\pi | X = x\Big\} = \beta_0 + \beta x, \tag{4.4}$$

where $X = \{0, 1\}$ is a binary variable for a high-risk allele.

4.2.5. *Group differences in Aβ onset*

Another implicit assumption is that there is a group difference in Aβ onset based on risk profiles (e.g. genetic). This is a widely accepted concept with strong biological foundations and empirical data. Let δ be the group difference in the population-level parameter M between two risk groups. Then the subject-specific random effects for Aβ onset is distributed as follows:

$$m_i \sim \mathcal{N}(M + \delta x_i, \sigma_M^2), \tag{4.5}$$

where $X = \{0, 1\}$ is a binary variable for risk group. The hypothesis testing of the group differences is simply:

$$\begin{aligned} H_0 &: \delta = 0 \\ &vs \\ H_a &: \delta \neq 0. \end{aligned} \tag{4.6}$$

4.2.6. *Bayesian estimation*

Bayesian formulation of the model offers a tractable estimation subject-level parameters (e.g. m_i) and inference (e.g. π, δ). Moreover, it allows incorporation of known constraints to model parameter values with prior distributions. Treating the parameters as random variables, we approximate a joint posterior density by sampling these parameters using MCMC techniques. We simulate the posterior densities using a Gibbs sampler implemented in `R` (*JAGS - Just Another Gibbs Sampler*; Plummer et al., 2006). Appendix C is the `BUGS` model specification (Lunn et al., 2000).

4.3. Simulation

We evaluate the performance of the proposed inference using simulation studies. In all simulations, the truth is simulated as in model XX to mimic the ADNI dataset. Unless otherwise stated, each simulation consists of M=500 simulated replicates with N=300. Each replicate is run with 11,000 iterations with 1,000 burn-in.

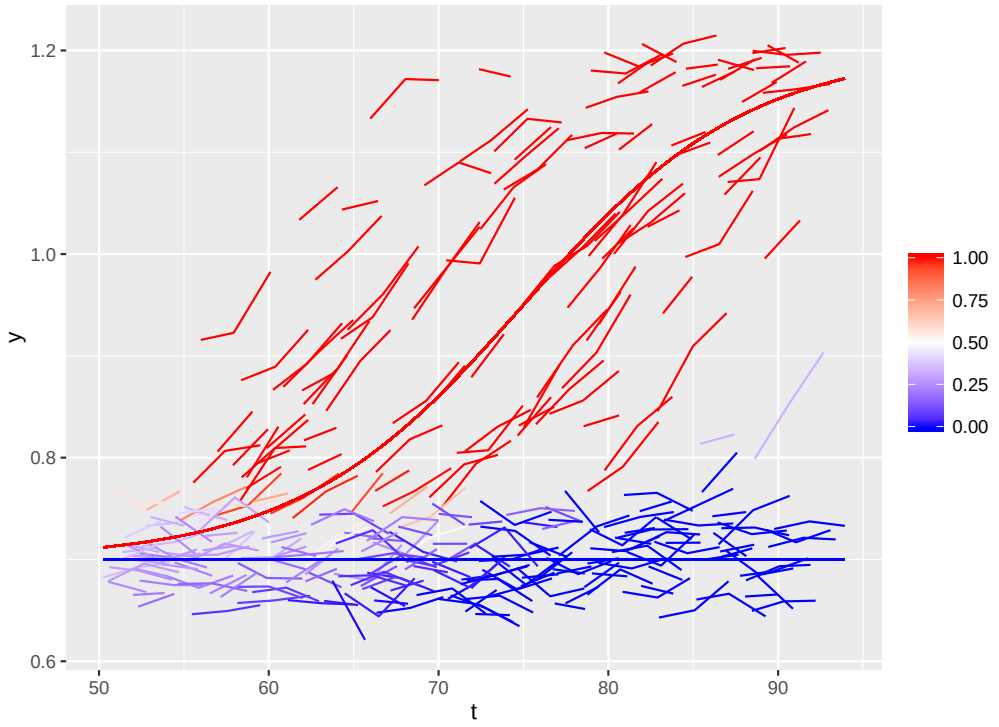

Figure 4.2: Simulated data per simulation 1a. The color gradient indicates the proportion of the subject being sampled as an accumulator.

4.3.1. Simulation 1 - Estimation of accumulation

The goal of these simulations is to evaluate the performance of the model in estimating the proportion of accumulators and predicting the accumulators. We simulate the truth model with normal random effects distributions, $m_{1i} \sim \mathcal{N}(75, 10^2)$, $m_{0i} \sim \mathcal{N}(500, 10^2)$, a uniform random effect distribution, $a_i \sim Unif(0.65, 0.75)$, a fixed π, $K = 1.2$, $B = 0.15$, and Y_{ij} based on (4.2). For each i, age at observation is sampled uniformly from 50 to 90, with 2 or 3 measurements generated with a probability of 0.50 each, and measurements intervals from $\mathcal{N}(2, 0.1^2)$. Figure 4.2 is a plot of the data simulated per simulation (a) in Table 4.1, with the color gradient indicating the proportion of the observation being sampled as an accumulator.

Table 4.1 summarizes the various performance measures with varying simulation conditions. In simulation 1a, the Bayesian model is specified as the simulation condition. In simulations 1b and 1c, we introduce misspecification to B and A in the model, where $B = 0.3$ and $a_i \sim Unif(0.55, 0.85)$, respectively. The proportion is altered in simulations 1d and 1e, and σ_ϵ is increased in 1f and 1g to

induce noise. In set 1h, the data are generated to reflect $K = 0.8$ and are modeled as 1a.

In nearly all simulation scenarios, the Bayesian posterior parameter means reflect the truth, with the exceptions in simulation 1h. π and σ_ϵ are slightly under- and overestimated, respectively. The AUC is for the classification of accumulators based on the probability score of being an accumulator, calculated as proportion a subject is sampled as such per replicate run (i.e. 10,000 random sampling). In all scenarios with the exception in 1h, there is good classification performance, with average AUC of 0.95.

4.3.2. Simulation 2 - Group differences

Next we consider a scenario where there are two groups, each with a different mean onset of Aβ accumulation. The data-generating model now consists of two groups which we denote by $X = \{0, 1\}$. We simulate data normal random effects distributions, $m_{1i} \sim \mathcal{N}(75 + \delta x_i, 10^2)$ and $m_{0i} \sim \mathcal{N}(500, 10^2)$, and g_i is sampled according to (4.4), with fixed effects β_0 and β. Additionally, we proceed as in the first simulation scenario with $a_i \sim Unif(0.65, 0.75)$, $K = 1.2$, $B = 0.15$, and Y_{ij} is sampled using (4.2). Again, for each i, age at observation is sampled uniformly from 50 to 90, with 2 or 3 measurements generated with a probability of 0.50 each, and measurements intervals from $\mathcal{N}(2, 0.1^2)$.

Table 4.2 summarizes various simulation scenarios. The simulated differences are at 15 years, 5 years, and 1 year, with varying parameters for β_0 and β_1. In all scenarios, the model estimates of δ correspond closely with the truth, indicating excellent model performance.

4.4. ADNI florbetapir data

The analytic ADNI florbetapir dataset contains a total of 1348 observations from 236 controls, 301 MCI, and 79 AD subjects, with each subject contributing from 2 to 3 observations over time. Data were processed and summarized as described in Landau et al. (2013). The mean age at baseline scan is 73.2 years, and the mean interval between scans is 2.1 years. Using the APOE gene allele data, four risk categories are assigned (Slooter et al., 1998), from lowest to highest: 1 - ε2/ε2 or ε2/ε3; 2 - ε3/ε3; 3 - ε2/ε4 or ε3/ε4; 4 - ε4/ε4. Figure 4.1 shows the data stratified by APOE risk status. Group 2 is at average risk of developing AD.

For the analysis, sixty one subjects in the below-average risk category (group 1) are excluded, leaving 1211 observations from 555 subjects. For each subject i, the risk classification is binarized to high risk (groups 3 and 4) and average risk (group 2), denoted by the variable $APOE_i$. We denote Y_{ij} the florbetapir SUVR measured at t_{ij}. We consider the hierarchical model based on (4.2), (4.3), and (4.4).

We fit the Bayesian model by assuming a $\mathcal{N}(M_1, \sigma_M^2)$ distribution for m_{1i}, with hyperpriors $M_1 \sim \mathcal{N}(75, 10^4)$. For m_{0i}, we assume a fixed distribution of $\mathcal{N}(500, 1)$ to reinforce the non-accumulators. For β_0 and β_1, we assign $\mathcal{N}(0.5, 1)$ and $\mathcal{N}(0, 1)$ priors, respectively. Additionally, we specify $a_i \sim Unif(0.6, 0.8)$, $K = 1.1$, and $B = 0.15$.

The posterior means of the estimates are summarized in Table 4.3. The model estimates a mean difference of 21 years between the two APOE groups, with the lower risk group having a mean Aβ onset parameter of 90. The 95% credible interval of δ precludes the null hypothesis that $\delta = 0$. The parameters for (4.4) indicate that the Figure 4.3 show the data with the mean group trajectory for high risk (red line) and normal risk (orange line), with the color gradient indicating the proportion of the subject being sampled as an accumulator.

As an exploratory analysis, another gene associated with AD is investigated. Bridging Integrator 1 (BIN1) has been associated with AD, particularly with late-onset AD (Naj et al., 2011, 2014), and the SNP data for BIN1 (rs7561528) is readily available in a subset of ADNI florbetapir cohort. For the analysis, there are 675 observations from 297 subjects who have the SNP genotyped. For each subject i, the genotype is binarized to A/A and non-A/A (i.e. A/G, G/G). Denoted by the variable $BIN1_i$. The same model specifications are used as above.

The model estimates are summarized in Table 4.4. The model estimates a mean difference of 3.5 years between the two BIN1 groups, but the 95% credible interval includes the null value of 0, and the non-A/A group had an estimated mean Aβ onset parameter of 80.

4.5. Discussion

We propose a framework for Bayesian inference in nonlinear mixed models for characterizing biomarker trajectories and estimating group differences while incorporating a mixture distribution for random effects. This approach has several advantages over the more conventional linear mixed

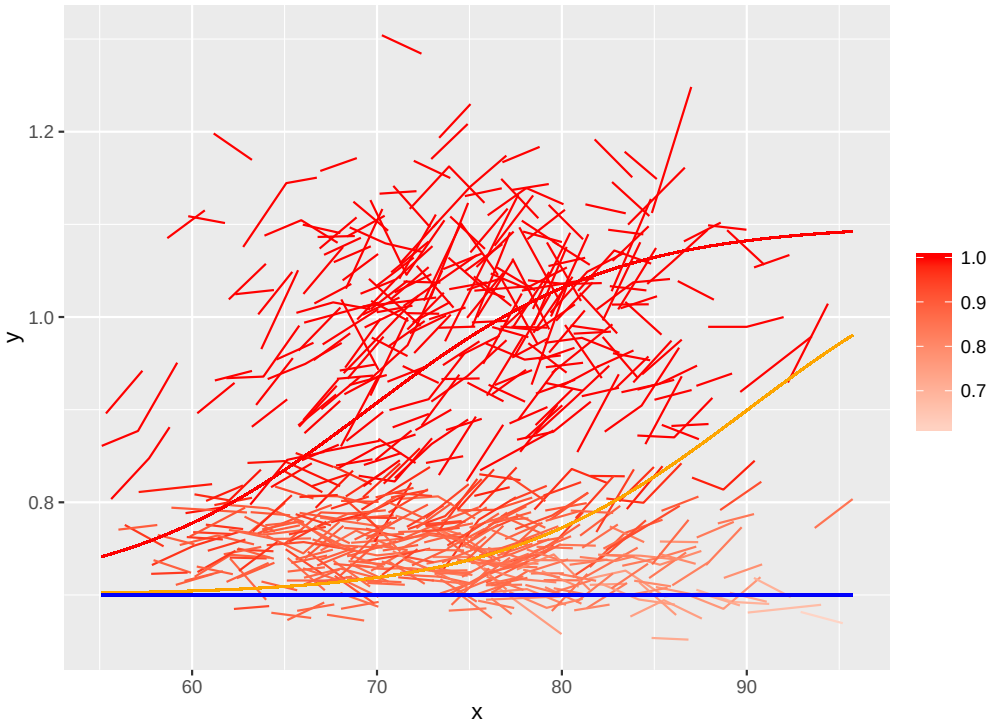

Figure 4.3: ADNI mean trajectories by APOE risk group. The red curve is the mean trajectory for the high risk group, and orange curve for the normal risk group. The color gradient indicates the proportion of the subject being sampled as an accumulator.

effect models including that the hypothesized longitudinal trajectory function is directly incorporated into the model and prior distributions can be used to inform known constraints on model parameters. By formulating the hierarchical models, this approach also identifies group differences based on functional parameters that have simple biological interpretation, such as the shift in the onset of biomarker accumulation.

To our knowledge, this is the first nonlinear modeling method applied to AD biomarkers based on Aβ accumulation dynamics. Simulation results show that this modeling approach may be able detect group differences in trajectories as small as one year in delayed accumulation. When applied to the ADNI data, we identify a difference of 21 years of average onset of Aβ accumulation between high- and average-risk group based on APOE genotype. An exploratory analysis of BIN1 indicated a difference average onset of 3.5 years but was not statistically significant.

Some limitations remain with our proposed approach. In our modeling framework, some parameters are fixed *a priori*, such as the sigmoidal slope parameter B. The prespecified value for B is based on the aggregate shape of the trajectories and previous reports. Due to the sparseness of individual measurements in the ADNI dataset (2 to 3 within a span of 4-6 years), model estimation of B may not accurate. While we show the misspecification of B from 0.15 to 0.3 did not affect the estimate of π in simulation scenario 1b, additional sensitivity analysis are necessary to determine effect of grossly misspecifying B. To estimate B, more individual measurements that span the baseline, growth, and plateau phases of the sigmoidal trajectory curve will be necessary. Sensitivity analyses are also necessary to determine the sparseness threshold for accurate estimation of the parameter. Furthermore, the assumption of a fixed slope parameter is a strong one, as individuals may exhibit different rates of accumulations. This may be addressed by incorporating random effects for the slope parameter.

The flexibility of our approach may be useful to model biomarkers with similar nonlinear functional trajectories, particularly in settings where limitations exist in data acquisition. While our examples are limited to two group differences, additional factors may be incorporated to explore biomarker trajectory interactions between APOE status and sex (Altmann et al., 2014). Other trajectory parameters such as the lower and upper asymptotes may also be modeled with appropriate link functions in a hierarchical manner.

Acknowledgement

Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (`adni.loni.usc.edu`). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: `http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf`.

Scenario	π	$\hat{\pi}$	σ_ϵ	$\hat{\sigma}_\epsilon$	RMSE($\hat{a}$)	RMSE($\hat{m}$)	AUC
Correctly specified							
(a)	0.500	0.462	0.010	0.011	0.019	7.709	0.956
B misspecified							
(b)	0.500	0.510	0.010	0.017	0.023	8.219	0.959
A misspecified							
(c)	0.500	0.447	0.010	0.014	0.051	8.624	0.931
More accumulators							
(d)	0.667	0.622	0.010	0.012	0.023	6.702	0.957
More non-accumulators							
(e)	0.250	0.230	0.010	0.011	0.014	8.983	0.955
Increased noise							
(f)	0.500	0.458	0.030	0.031	0.023	7.985	0.945
(g)	0.500	0.452	0.050	0.051	0.025	8.301	0.935
K Misspecified							
(h)	0.500	0.049	0.100	0.021	0.010	11.540	0.656

Table 4.1: Simulation 1 results. M=500 for each simulation set, with N=300 per simulation. RMSE: root-mean-square error; AUC: area under ROC curve.

Scenario	$\hat{M}$	δ	$\hat{\delta}$	β_0	$\hat{\beta}_0$	β	$\hat{\beta}$	$\hat{\sigma}_\epsilon$	RMSE($\hat{a}$)	RMSE($\hat{m}$)	AUC
(a)	75.038	15	14.743	1	1.050	1	0.936	0.010	0.023	6.046	0.913
(b)	75.007	5	4.919	1	1.020	1	0.957	0.010	0.024	5.450	0.954
(c)	75.317	15	14.540	-1	-0.894	1	0.883	0.010	0.016	8.665	0.925
(d)	75.161	15	14.482	1	0.968	0.2	0.286	0.010	0.022	6.668	0.915
(e)	75.061	15	14.498	1	1.053	1	0.967	0.010	0.024	5.274	0.941
(f)	74.973	5	4.855	1	0.947	0.2	0.269	0.010	0.023	5.902	0.958
(g)	74.982	1	1.048	1	1.044	1	0.932	0.010	0.024	5.243	0.966

Table 4.2: Simulation 2 results. M=500 for each simulation set, with N=300 per simulation.

	mean	sd	50%	2.5%	97.5%
M_1	90.318	1.711	90.359	86.791	93.538
σ_{M_1}	16.741	0.961	16.705	14.958	18.711
δ	-20.705	1.994	-20.708	-24.668	-16.766
β_0	2.492	0.677	2.471	1.236	3.870
β_1	1.453	0.746	1.449	0.016	2.966
σ_ϵ	0.027	0.001	0.027	0.026	0.029

Table 4.3: ADNI Bayesian model estimates using APOE

	mean	sd	50%	2.5%	97.5%
M_1	80.012	2.705	80.104	74.493	85.069
σ_{M_1}	19.929	1.972	19.852	16.165	23.960
δ	3.460	4.633	3.565	-5.866	12.353
β_0	1.655	0.558	1.600	0.700	2.921
β_1	0.081	0.821	0.041	-1.376	1.856
σ_ϵ	0.025	0.001	0.025	0.023	0.026

Table 4.4: ADNI Bayesian model estimates using BIN1

APPENDIX A

R CODE FOR NPE ESTIMATION OF OROC (INTERVAL CLASSIFICATION)

```
# function for nonparametric estimation of OROC and ML estimator of a and b
oroc.ll <- function(x,y,cut=100) {
  if (length(x)!=length(y)) stop('Lengths of X and Y differ.')
  # regularized cutoff points
  cutoffs <- rev(seq(min(x), max(x), length=cut))
  sum.y0<-sum(y==0)
  sum.y1<-sum(y==1)
  # Internal functions for FPR and TPR
  fun.fpr<-function(a,b) {
    fpr<-c(sum(x[y==0]>=a & x[y==0]<=b)/sum.y0)
    return(fpr)
  }
  fun.tpr<-function(a,b) {
    tpr<-c(sum(x[y==1]>=a & x[y==1]<=b)/sum.y1)
    return(tpr)
  }
  # function for likelihood for ME
  fun.ll<-function(a,b) {
    s <- (1*x>=a & x<=b)
    aic<-glm.fit(cbind(1,s),y,family=binomial())$aic
    return(-(aic-4)/2)
  }
  a<-b<-fpr<-tpr<-ll<- vector()
  # grid search
  for(cutoff in cutoffs) {
    fpr<-c(fpr,sapply(rev(cutoffs[cutoffs>=cutoff]),fun.fpr,a=cutoff))
    tpr<-c(tpr,sapply(rev(cutoffs[cutoffs>=cutoff]),fun.tpr,a=cutoff))
    ll<-c(ll,sapply(rev(cutoffs[cutoffs>=cutoff]),fun.ll,a=cutoff))
```

```
    a<-c(a,rep(cutoff,sum(cutoffs>=cutoff)))
    b<-c(b,rev(cutoffs[cutoffs>=cutoff]))
  }
  tpr[tpr<0 | is.na(tpr)] = 0
  fpr[fpr<0 | is.na(fpr)] = 0
  orcdata<-data.frame(a,b,tpr,fpr,ll)
  mle<-orcdata[sample(which(orcdata$ll==max(orcdata$ll)),1),1:2]
  # estimation of parameter path
  orcdata<-orcdata[order(orcdata$fpr,-orcdata$tpr),]
  orcset <- data.frame(orcdata[1,])
  for(i in 1:dim(orcdata)[1]) {
    if(orcdata[i,4]>orcset[dim(orcset)[1],4] &
       orcdata[i,3]>=orcset[dim(orcset)[1],3]) {
      orcset <- rbind(orcset,orcdata[i,])
    }
  }
  # Trapezoid Rule Numerical Integration
  auc<-trapz(c(0,orcset$fpr,1),c(0,orcset$tpr,1))
  return(list("orcset"=orcset,"auc"=auc,"mle"=mle))
}

# function for applying optimal cutoffs to validation set
oroc.validation <- function(x,y,rocdata) {
  sum.y0<-sum(y==0)
  sum.y1<-sum(y==1)
  # Internal functions for FPR and TPR
  fun.fpr<-function(a,b) {
    fpr<-c(sum(x[y==0]>=a & x[y==0]<=b)/sum.y0)
    return(fpr)
  }
  fun.tpr<-function(a,b) {
```

```
    tpr<-c(sum(x[y==1]>=a & x[y==1]<=b)/sum.y1)
    return(tpr)
  }
  fpr<-tpr<-vector()
  for(i in 1:length(rocdata$a)) {
    fpr<-c(fpr,fun.fpr(a=rocdata$a[i],b=rocdata$b[i]))
    tpr<-c(tpr,fun.tpr(a=rocdata$a[i],b=rocdata$b[i]))
  }
  tpr[tpr<0 | is.na(tpr)] = 0
  fpr[fpr<0 | is.na(fpr)] = 0
  fpr <- c(0,fpr,1)
  tpr <- c(0,tpr,1)
  rocdata<-data.frame(fpr,tpr)
  rocdata<-rocdata[order(rocdata$fpr,rocdata$tpr),]
  auc <- trapz(rocdata$fpr,rocdata$tpr)
  return(auc)
}
```

APPENDIX B

RESTRICTED FUNCTION-ON-SCALAR REGRESSION CODE

```
# Restricted function-on-scalar regression, modification of fosr (Reiss et al. 2010)
# Modification are highlighted
rfosr<-function(formula = NULL, Y = NULL, fdobj = NULL, data = NULL,
                X, con = NULL, argvals = NULL, method = c("OLS", "GLS", "mix"),
                gam.method = c("REML", "ML", "GCV.Cp", "GACV.Cp", "P-REML", "P-ML"),
                cov.method = c("naive", "mod.chol"), lambda = NULL, nbasis = 15,
                norder = 4, pen.order = 2, multi.sp = ifelse(method == "OLS", FALSE, TRUE),
                pve = 0.99, max.iter = 1, maxlam = NULL, cv1 = FALSE, scale = FALSE) {
  ## MODEL BASED ON FOMULA
  if (!is.null(formula)) {
    if (is.null(data))
      stop("Please specify the data.")
    tf <- terms.formula(formula)
    trmstrings <- attr(tf, "term.labels")
    terms <- sapply(trmstrings, function(trm) as.call(parse(text = trm))[[1]],
                    simplify = FALSE)
    responsename <- as.character(attr(tf, "variables")[2][[1]])
    Y = data[, responsename]
    X = model.matrix(formula, data = data)
  }
  ## DEFINE FUNCTIONAL DATA (RESP.TYPE)
  if (is.null(Y) == is.null(fdobj))
    stop("Please specify 'Y' or 'fdobj', but not both")
  resp.type <- if (is.null(Y))
    "fd"
  else "raw"
  ## ARGVALS
  if (is.null(argvals))
```

```
  argvals <- if (is.null(fdobj))
    seq(0, 1, length = ncol(Y))
else seq(min(fdobj$basis$range), max(fdobj$basis$range),
         length = 201)
## METHOD
method <- match.arg(method)
cov.method <- match.arg(cov.method)
gam.method <- match.arg(gam.method)
if (method != "OLS" & (length(lambda) > 1))
  stop("Vector-valued lambda allowed only if method = 'OLS'")
if (!is.null(lambda) & multi.sp)
  stop("Fixed lambda not implemented with multiple penalties")
if (method == "OLS" & multi.sp)
  stop("OLS not implemented with multiple penalties")
## BSS, BMAT, RESPMAT
if (resp.type == "raw") {
  bss = create.bspline.basis(range(argvals), nbasis = nbasis, norder = norder)
  Bmat <- Theta <- eval.basis(argvals, bss)
  respmat <- Y
}
else if (resp.type == "fd") {
  if (!is.fd(fdobj))
    stop("'fdobj' must be a functional data object")
  bss = fdobj$basis
  nbasis = bss$nbasis
  Theta <- eval.basis(argvals, bss)
  C = t(fdobj$coefs)
  J = getbasispenalty(bss, 0)
  svdJ = svd(J)
  Bmat <- J12 <- svdJ$u %*% diag(sqrt(svdJ$d)) %*% t(svdJ$u)
  respmat <- C %*% J12
```

```
}
## DEFINE INITIAL PARAMETERS AS NULL
newfit = U = pca.resid = NULL
## INITIAL COVARIATE DESIGN
X.sc = scale(X, center = FALSE, scale = scale)
q = ncol(X)
ncurve <- nrow(respmat)
## PEN
if (multi.sp) {
  pen = vector("list", q)
  for (j in 1:q) {
    one1 = matrix(0, q, q)
    one1[j, j] = 1
    pen[[j]] = one1 %x% getbasispenalty(bss, pen.order)
  }
}
else pen = list(diag(q) %x% getbasispenalty(bss, pen.order))
#################### <MODIFICATION> ####################
## MODIFY BMAT AND PEN
excl.basis <- 1
Bmat <- Bmat[,(excl.basis+1):(dim(Bmat)[2]-excl.basis)]
Theta <-Theta[,(excl.basis+1):(dim(Theta)[2]-excl.basis)]
pen = list(diag(q) %x% getbasispenalty(bss, pen.order)[2:9,2:9])
#################### <\MODIFICATION> ####################
## METHOD
constr = if (!is.null(con))
  con %x% diag(nbasis)
else NULL
cv = NULL
if (method == "OLS") {
  if (length(lambda) != 1 | cv1) {
```

```
################### <MODIFICATION> ###################
    lofo <- new.lofocv(respmat, X.sc %x% Bmat, S1 = pen[[1]], argvals = argvals,
                       lamvec = lambda, constr = constr, maxlam = maxlam)
################### <\MODIFICATION> ###################
    cv = if (is.null(lambda))
      lofo$objective
    else min(lofo[, 2])
    lambda = if (is.null(lambda))
      lofo$min
    else lofo[which.min(lofo[, 2]), 1]
  }
}
## FIRSTFIT (ADDITIVE MODEL WITH CONSTRAINTS)
## INITIAL COEFMAT
################### <MODIFICATION> ###################
 firstfit <- new.amc(as.vector(t(respmat)), X.sc %x% Bmat, gam.method = gam.method,
                     S = pen, C = constr, lambda = lambda)
################### <\MODIFICATION> ###################
coefmat = coefmat.ols = t(matrix( firstfit $coef, ncol = q))
se = NULL
## NON-OLS METHOD
if (method != "OLS") {
  iter = 0
  coefmat.old = 3 * coefmat.ols
  newfit = NULL
  if (!is.null(lambda) & max.iter > 0)
    warning("Given lambda used for initial fit only")
  ## CONVERGENCE CRITERION
  while (any(abs((coefmat - coefmat.old)/coefmat.old) > 0.001) & (iter < max.iter)) {
    iter = iter + 1
    if (max.iter > 1)
```

```
  cat("Refit", iter, "\n")
## SET LAST ESTIMATES AS OLD
 oldfit = if (!is.null(newfit))
   newfit
else  firstfit
coefmat.old = coefmat
## RESIDUAL VECTOR
residvec <- as.vector(t(respmat)) - (X.sc %x% Bmat) %*%
    oldfit $coef[1:(q * nbasis)]
residmat = t(matrix(residvec, ncol = ncurve))
if (method == "GLS") {
  if (cov.method == "mod.chol") {
    p = ncol(residmat)
    res.cent = scale(residmat, TRUE, FALSE)
    sqrt.prec.list = list()
    lwstat = lwpval = c()
    for (nband in 1:(p - 1)) {
      TT = diag(p)
      Ddiag = rep(0, p)
      Ddiag[1] = var(res.cent[, 1])
      for (k in 2:p) {
        qrResCent <- qr(res.cent[, max(1, k - nband):(k - 1)])
        TT[k, max(1, k - nband):(k - 1)] <- (-qr.coef(qrResCent, res.cent[, k]))
        Ddiag[k] <- var(qr.resid(qrResCent, res.cent[, k]))
      }
      prec = scale(t(TT), FALSE, Ddiag) %*% TT
      sqrt.prec.list [[nband]] = scale(t(TT), FALSE, sqrt(Ddiag))
      lwprec = lw.test(residmat %*% sqrt.prec.list[[nband]])
      lwstat[nband] = lwprec$stat
      lwpval[nband] = lwprec$pvalue
      if (lwstat[nband] < -5)
```

```
            break
          if (nband > 5) {
            if (lwstat[nband] > lwstat[1] && lwstat[nband] >
                lwstat[nband - 1])
              break
          }
        }
        nband.best = which.max(lwpval)
        cat("Using half-bandwidth", nband.best, "for precision matrix of residuals\n")
        sqrt.prec <- sqrt.prec.list[[nband.best]]
      }
      else if (cov.method == "naive") {
        if (nrow(residmat) < ncol(residmat))
          stop("Sample covariance matrix of residuals is singular.")
        svd.cov.mle <- svd(cov(residmat) * (ncurve - 1)/ncurve)
        sqrt.prec <- tcrossprod(scale(svd.cov.mle$u,
                                      FALSE, sqrt(svd.cov.mle$d)), svd.cov.mle$u)
      }
      newfit <- amc(as.vector(tcrossprod(sqrt.prec,respmat)),
                    X.sc %x% (sqrt.prec %*% Bmat), gam.method = gam.method,
                    S = pen, C = constr, start = if (is.null(con))
                    as.vector(t(coefmat))
                    else NULL)
      coefmat = t(matrix(newfit$coef, ncol = q))
    }
    else if (method == "mix") {
      if (resp.type == "fd") {
        resid.fd <- fd(solve(J12, t(residmat)), bss)
        if ( iter == 1) {
          pca.resid <- pca.fd(resid.fd, nharm = min(ncurve - 1, nbasis))
          npc <- min(which(cumsum(pca.resid$varprop) > pve))
```

```
  }
  else pca.resid <- pca.fd(resid.fd, nharm = npc)
  evalues <- pca.resid$values[1:npc]
  efuncmat.scaled <- Bmat %*% t(t(pca.resid$harmonics$coef[,1:npc]) * sqrt(evalues))
}
else if (resp.type == "raw") {
  if (iter == 1) {
    pca.resid <- fpca.sc(residmat, pve = pve)
    npc <- pca.resid$npc
  }
  else pca.resid <- fpca.sc(residmat, npc = npc)
  evalues <- pca.resid$evalues
  efuncmat.scaled <- t(t(pca.resid$efunctions) * sqrt(evalues))
}
if (iter == 1)
  cat("Using leading", npc, "PCs of residual functions for random effects\n")
npen <- length(pen)
pendim <- ncol(pen[[1]])
pen.aug = vector("list", npen + 1)
for (l in 1:npen) {
  pen.aug[[l]] <- matrix(0, pendim + npc * ncurve, pendim + npc * ncurve)
  pen.aug[[l]][1: pendim, 1:pendim] <- pen[[l]]
}
if (iter == 1)
  cat("Using leading", npc, "PCs of residual functions for random effects\n")
npen <- length(pen)
pendim <- ncol(pen[[1]])
pen.aug = vector("list", npen + 1)
for (l in 1:npen) {
  pen.aug[[l]] <- matrix(0, pendim + npc * ncurve, pendim + npc * ncurve)
  pen.aug[[l]][1: pendim, 1:pendim] <- pen[[l]]
```

```
        }
        pen.aug[[npen + 1]] <- diag(rep(0:1, c(pendim, npc * ncurve)))
        constr.aug <- if (is.null(constr))
          NULL
        else cbind(constr, matrix(0, nrow(constr), npc * ncurve))
        startB <- if ( iter == 1) {
          c(as.vector(t(coefmat)), rep(0, ncurve * npc))
        }
        else {
          newfit$coefficients
        }
        newfit <- amc(as.vector(t(respmat)),
                      cbind(X.sc %x% Bmat, diag(ncurve) %x% efuncmat.scaled),
                      gam.method = gam.method, S = pen.aug, C = constr.aug,
                      start = if (is.null(constr.aug)) startB
                      else NULL)
        vecBt = newfit$coef[1:(q * nbasis)]
        vecUt = newfit$coef[(q * nbasis + 1):(q * nbasis + npc * ncurve)]
        coefmat = t(matrix(vecBt, ncol = q))
        U <- t(matrix(vecUt, ncol = ncurve))
      }
    }
  }
  ## OLS METHOD
  if (method == "OLS" | max.iter == 0) {
    residvec <- as.vector(t(respmat)) - (X.sc %x% Bmat) %*% firstfit$coef
    covmat = ((ncurve - 1)/ncurve) * cov(t(matrix(residvec, ncol = ncurve)))
    var.b = firstfit $GinvXT %*% (diag(ncurve) %x% covmat) %*% t(firstfit$GinvXT)
  }
  else var.b = newfit$Vp
  ## STANDARD ERRORS
```

```
se.func = matrix(NA, length(argvals), q)
################### <MODIFICATION> ###################
for (j in 1:q) {
  se.func[, j] = sqrt(rowSums((Theta %*%
                 var.b[((nbasis-2*excl.basis) * (j - 1) + 1):((nbasis-2*excl.basis) * j),
                       ((nbasis-2*excl.basis) * (j - 1) + 1):((nbasis-2*excl.basis) * j)]) *
                      Theta))
}
for(i in 1:excl.basis) {
  coefmat <- cbind(0,coefmat,0)
}
################### <\MODIFICATION> ###################
fd = fd(t(coefmat), bss)
est.func = eval.fd(argvals, fd)
fit <- if (method == "mix" & max.iter > 0)
  newfit
else firstfit
roughness = diag(coefmat %*% getbasispenalty(bss, pen.order) %*% t(coefmat))
skale = attr(X.sc, "scaled:scale")
if (!is.null(skale)) {
  coefmat = t(scale(t(coefmat), center = FALSE, scale = skale))
  est.func = scale(est.func, center = FALSE, scale = skale)
  se.func = scale(se.func, center = FALSE, scale = skale)
  roughness = roughness/skale^2
}
yhat = if (resp.type == "raw")
  X %*% tcrossprod(coefmat, Theta)
else fd(t(X %*% coefmat), bss)
llist = list(fd = fd, pca.resid = pca.resid, U = U, yhat = yhat,
             resid = if (resp.type == "raw") Y - yhat else fdobj - yhat,
             est.func = est.func, se.func = se.func, argvals = argvals,
```

```
                fit = fit, edf = sum(fit$gam$edf),
                lambda = if (length(fit$gam$sp) > 0)
                  fit$gam$sp
                else fit$gam$full.sp, cv = cv, roughness = roughness,
                resp.type = resp.type)
  class(llist) = "fosr"
  llist
}

# Leave-one out cross-validation function for rfosr
# from fosr internal function lofocv (Reiss et al. 2010)
new.lofocv <- function(Y, X, S1, argvals, lamvec=NULL, constr=NULL, maxlam=NULL) {
  nn = nrow(X)
  N = NROW(Y); K = NCOL(Y)
  if (N*K!=nn) stop('Number of elements of Y must equal number of rows of X')
  y = as.vector(t(Y))
  if (!is.null(constr)) {
    # The following is based on Wood (2006), p. 186
    n.con = dim(constr)[1]
    Z. = qr.Q(qr(t(constr)), complete=TRUE)[, -(1:n.con)]
    X. = X %*% Z.
    S1. = crossprod(Z., S1 %*% Z.)
  }
  else {
    X. = X
    S1. = S1
  }
  qrX = qr(X.)
  Rinv = solve(qr.R(qrX))
  svd211 = svd(crossprod(Rinv, S1. %*% Rinv)) # see p. 211 of Wood
  QU = qr.Q(qrX) %*% svd211$u
```

```
# calculate the weight for the approx. integral using argvals
vecWeight = diff(argvals, 2)/2
vecWeight = c((argvals[2]-argvals[1])/2, vecWeight, (argvals[N]-argvals[N-1])/2)
cvfcn = function(lam) {
  A = tcrossprod(scale(QU, center=FALSE, scale=1+lam*svd211$d), QU)
  resmat = t(matrix(y - A %*% y, K))
  MSEp = 0
  for (i in 1:N) {
    ith = ((i-1)*K+1):(i*K)
    # when no argvals is used
    # MSEp = MSEp + crossprod(solve(diag(K)-A[ith,ith], resmat[i, ])) / N
    # when new argvals is implemented
    MSEp = MSEp + crossprod(solve(diag(K)-A[ith,ith], resmat[i, ])) * vecWeight[i]
  }
  MSEp
}
if (is.null(lamvec)) {  # minimize LOFO-CV criterion
  if (is.null(maxlam)) { # use GCV-minimizing lambda
    model.gcv = gam(y~X.-1, paraPen=list(X.=list(S1.)), method="GCV.Cp")
    maxlam = model.gcv$sp
  }
  cat("Finding optimal lambda by optimize()...\n")
  opt = optimize(cvfcn, c(0, maxlam), tol=.01)
  if (round(opt$minimum)==maxlam) warning("maxlam may be set too low")
  return(opt)
}
else {  # calculate LOFO-CV for given values
  cvvals = c()
  cat("Calculating CV for candidate smoothing parameter values...\n")
  for (i in 1:length(lamvec)) cvvals[i] = cvfcn(lamvec[i])
  cvtable = cbind(lamvec, cvvals)
```

```
    dimnames(cvtable)[[2]] = c('lambda', 'LOFO-CV')
    print(cvtable)
    return(cvtable)
    if (which.min(cvvals)==1) warning("CV minimized at lowest lambda considered")
    if (which.min(cvvals)==length(lamvec)) warning("CV minimized at highest lambda
        considered")
  }
}

# Internal function for additive models with constraints used in rfosr
# from fosr internal function amc (Reiss et al. 2010)
new.amc <- function(y, Xmat, S, gam.method='REML', C=NULL, lambda=NULL, ...) {
  n.p = length(S)
  if (!is.null(C)) {
    # The following is based on Wood (2006), p. 186
    n.con = dim(C)[1]
    Z. = qr.Q(qr(t(C)), complete=TRUE)[ , -(1:n.con)]
    Xmat. = Xmat %*% Z.
    S. = vector("list", n.p)
    for (i in 1:n.p) S.[[i]] = crossprod(Z., S[[i]] %*% Z.)
  }
  else {
    Z. = diag(ncol(Xmat))
    Xmat. = Xmat
    S. = S
  }
  fitter = if (length(y) > 10000) bam else gam
  if (is.null(lambda)) fitobj = fitter(y ~ Xmat.-1, method=gam.method, paraPen=list(Xmat.=
      S.), ...)
  else fitobj = fitter(y ~ Xmat.-1, paraPen=list(Xmat.=S.), sp=lambda, ...)
  lambdavec = if (!is.null(fitobj$full.sp)) fitobj$full.sp else fitobj$sp
```

```
  fullpen = 0
  for (i in 1:n.p) fullpen = lambdavec[i] * S.[[ i ]]
  list(gam = fitobj,
       coefficients = Z. %*% fitobj$coef,
       Vp = Z. %*% fitobj$Vp %*% t(Z.),
       GinvXT = Z. %*% solve(crossprod(Xmat.) + fullpen, t(Xmat.)),
       method = gam.method)
}
```

APPENDIX C

BUGS CODE FOR NONLINEAR MIXED EFFECTS MODELING

C.1. Specification for Simulation 1a

```
model {
    # Priors for global parameters
    sigma~dunif(0, 100) # Residual standard deviation
    tau <- 1/(sigma*sigma)
    # Hyperpriors for subject-specific parameters
    p~dbeta(1,1)
    for (i in 1:ngroups) {
    A[i]~dunif(0.65,0.75)
    group[i]~dbern(p)
    m1[i]~dunif(50,90)
    m2[i]~dnorm(500,1/sqrt(10))
    M[i]<-m1[i]*group[i]+m2[i]*(1-group[i])
    }
    # Likelihood
    for (i in 1:n) {
    # Expectation
    mu[i] <- A[id[i]] + (1.2-A[id[i]]) /(1 + exp(-0.15*(x[i]-M[id[i]]) ))
    # Actual response
    y[i] ~ dnorm(mu[i], tau)
    }
    }
```

C.2. Specification for Simulation 2a

```
model {
    # Priors for global parameters
    sigma~dunif(0, 100) # Residual standard deviation
```

```
tau <- 1/(sigma*sigma)
# Hyperpriors for subject-specific parameters
beta0 ~ dnorm(0,1)
beta1 ~ dnorm(0,1)
M1 ~ dnorm(75,0.01)
delta ~ dnorm(0,0.01)
for (i in 1:ngroups) {
logit (p[i]) <- beta0 + beta1*cat[i]
A[i]~dunif(0.65,0.75)
group[i]~dbern(p[i])
m1[i]~dnorm(M1*cat[i] + (M1+delta)*(1-cat[i]),0.01)
m2[i]~dnorm(500,1/sqrt(10))
M[i]<-m1[i]*group[i]+m2[i]*(1-group[i])
}
# Likelihood
for (i in 1:n) {
# Expectation
mu[i] <- A[id[i]] + (1.2-A[id[i]]) /(1 + exp(-0.15*(x[i]-M[id[i]]) ))
# Actual response
y[i] ~ dnorm(mu[i], tau)
}
}
```

C.3. Specification for ADNI data

```
model {
# Priors for global parameters
sigma~dunif(0, 100) # Residual standard deviation
tau <- 1/(sigma*sigma)
sigma_M1~dunif(0, 100) # M1 standard deviation
tau_M1 <- 1/(sigma_M1*sigma_M1)
# Hyperpriors for subject-specific parameters
```

```
beta0 ~ dnorm(0.5,1)
beta1 ~ dnorm(0,1)
M1 ~ dnorm(75,0.01)
delta ~ dnorm(0,0.01)
for (i in 1:ngroups) {
logit(p[i]) <- beta0 + beta1*apoe[i]
A[i]~dunif(0.6,0.8)
group[i]~dbern(p[i])
m1[i]~dnorm(M1+delta*apoe[i],tau_M1)
m2[i]~dnorm(500,1/sqrt(10))
M[i]<-m1[i]*group[i]+m2[i]*(1-group[i])
}
# Likelihood
for (i in 1:n) {
# Expectation
mu[i] <- A[id[i]] + (1.1-A[id[i]])/(1 + exp(-0.15*(x[i]-M[id[i]])))
# Actual response
y[i] ~ dnorm(mu[i], tau)
}
}
```

BIBLIOGRAPHY

Alois Kneip, KJU (2001). Inference for Density Families Using Functional Principal Component Analysis. *Journal of the American Statistical Association* 96.454, 519–532. ISSN: 01621459. URL: http://www.jstor.org/stable/2670290.

Alonzo, TA and Nakas, CT (2007). Comparison of ROC umbrella volumes with an application to the assessment of lung cancer diagnostic markers. *Biom J* 49.5, 654–664.

Altmann, A, Tian, L, Henderson, VW, and Greicius, MD (2014). Sex modifies the APOE-related risk of developing Alzheimer disease. *Ann. Neurol.* 75.4, 563–573.

Austin, MA, Breslow, JL, Hennekens, CH, Buring, JE, Willett, WC, and Krauss, RM (1988). Low-density lipoprotein subclass patterns and risk of myocardial infarction. *JAMA* 260.13, 1917–1921.

Basu, S, Zaidi, H, Houseni, M, Bural, G, Udupa, J, Acton, P, Torigian, DA, and Alavi, A (2007). Novel quantitative techniques for assessing regional and global function and structure based on modern imaging modalities: implications for normal variation, aging and diseased states. *Semin Nucl Med* 37.3, 223–239.

Boellaard, R (2009). Standards for PET image acquisition and quantitative data analysis. *J. Nucl. Med.* 50 Suppl 1, 11S–20S.

Braak, H and Braak, E (1991). Neuropathological stageing of Alzheimer-related changes. *Acta Neuropathol.* 82.4, 239–259.

Caroli, A and Frisoni, GB (2010). The dynamics of Alzheimer's disease biomarkers in the Alzheimer's Disease Neuroimaging Initiative cohort. *Neurobiol. Aging* 31.8, 1263–1274.

Chen, Z (2008). Histogram partition and interval thresholding for volumetric breast tissue segmentation. *Comput Med Imaging Graph* 32.1, 1–10.

Clark, CM, Schneider, JA, Bedell, BJ, Beach, TG, Bilker, WB, Mintun, MA, Pontecorvo, MJ, Hefti, F, Carpenter, AP, Flitter, ML, Krautkramer, MJ, Kung, HF, Coleman, RE, Doraiswamy, PM, Fleisher, AS, Sabbagh, MN, Sadowsky, CH, Reiman, EP, Reiman, PE, Zehntner, SP, and Skovronsky, DM (2011). Use of florbetapir-PET for imaging beta-amyloid pathology. *JAMA* 305.3, 275–283.

Coxson, HO (2013). Sources of variation in quantitative computed tomography of the lung. *J Thorac Imaging* 28.5, 272–279.

Delicado, P (2011). Dimensionality Reduction when Data Are Density Functions. *Comput. Stat. Data Anal.* 55.1, 401–420. ISSN: 0167-9473. DOI: 10.1016/j.csda.2010.05.008. URL: http://dx.doi.org/10.1016/j.csda.2010.05.008.

Dreiseitl, S, Ohno-Machado, L, and Binder, M (2000). Comparing three-class diagnostic tests by three-way ROC analysis. *Med Decis Making* 20.3, 323–331.

Emami, H and Tawakol, A (2014). Noninvasive imaging of arterial inflammation using FDG-PET/CT. *Curr. Opin. Lipidol.* 25.6, 431–437.

Fortin, JP, Labbe, A, Lemire, M, Zanke, B, Hudson, T, Fertig, E, Greenwood, C, and Hansen, K (2014). Functional normalization of 450k methylation array data improves replication in large cancer studies. *Genome Biology* 15.11, 503. ISSN: 1465-6906. DOI: 10.1186/s13059-014-0503-2. URL: http://genomebiology.com/2014/15/11/503.

Fraiman, R and Muniz, G (2001). Trimmed means for functional data. *Test* 10.2, 419–440. ISSN: 1863-8260. DOI: 10.1007/BF02595706. URL: http://dx.doi.org/10.1007/BF02595706.

Gelfand, JM, Neimann, AL, Shin, DB, Wang, X, Margolis, DJ, and Troxel, AB (2006). Risk of myocardial infarction in patients with psoriasis. *JAMA* 296.14, 1735–1741.

Gordon, DJ, Probstfield, JL, Garrison, RJ, Neaton, JD, Castelli, WP, Knoke, JD, Jacobs, DR, Bangdiwala, S, and Tyroler, HA (1989). High-density lipoprotein cholesterol and cardiovascular disease. Four prospective American studies. *Circulation* 79.1, 8–15.

Hall, P, Hyndman, RJ, and Fan, Y (2004). Nonparametric Confidence Intervals for Receiver Operating Characteristic Curves. English. *Biometrika* 91.3, pp. 743–750. ISSN: 00063444.

Hanley, JA and McNeil, BJ (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 143.1, 29–36.

Hardy, J and Selkoe, DJ (2002). The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics. *Science* 297.5580, 353–356.

Hastie, T and Tibshirani, R (1986). Generalized Additive Models. English. *Statistical Science* 1.3, pp. 297–310. ISSN: 08834237.

Huang, L, Scheipl, F, Goldsmith, J, Gellar, J, Harezlak, J, McLean, MW, Swihart, B, Xiao, L, Crainiceanu, C, and Reiss, P (2015). *refund: Regression with Functional Data*. R package version 0.1-13. URL: https://CRAN.R-project.org/package=refund.

Huet, P, Burg, S, Le Guludec, D, Hyafil, F, and Buvat, I (2015). Variability and uncertainty of 18F-FDG PET imaging protocols for assessing inflammation in atherosclerosis: suggestions for improvement. *J. Nucl. Med.* 56.4, 552–559.

Jack, CR, Knopman, DS, Jagust, WJ, Shaw, LM, Aisen, PS, Weiner, MW, Petersen, RC, and Trojanowski, JQ (2010). Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. *Lancet Neurol* 9.1, 119–128.

Jack, CR, Vemuri, P, Wiste, HJ, Weigand, SD, Lesnick, TG, Lowe, V, Kantarci, K, Bernstein, MA, Senjem, ML, Gunter, JL, Boeve, BF, Trojanowski, JQ, Shaw, LM, Aisen, PS, Weiner, MW, Petersen, RC, and Knopman, DS (2012). Shapes of the trajectories of 5 major biomarkers of Alzheimer disease. *Arch. Neurol.* 69.7, 856–867.

JAGS - Just Another Gibbs Sampler. http://mcmc-jags.sourceforge.net/.

Janes, H and Pepe, MS (2009). Adjusting for covariate effects on classification accuracy using the covariate-adjusted receiver operating characteristic curve. *Biometrika* 96.2, 371–382.

Kepe, V, Moghbel, MC, Langstrom, B, Zaidi, H, Vinters, HV, Huang, SC, Satyamurthy, N, Doudet, D, Mishani, E, Cohen, RM, H?ilund-Carlsen, PF, Alavi, A, and Barrio, JR (2013). Amyloid- positron emission tomography imaging probes: a critical review. *J. Alzheimers Dis.* 36.4, 613–631.

Keyes, JW (1995). SUV: standard uptake or silly useless value? *J. Nucl. Med.* 36.10, 1836–1839.

Klunk, WE, Engler, H, Nordberg, A, Wang, Y, Blomqvist, G, Holt, DP, Bergstrom, M, Savitcheva, I, Huang, GF, Estrada, S, Ausen, B, Debnath, ML, Barletta, J, Price, JC, Sandell, J, Lopresti, BJ, Wall, A, Koivisto, P, Antoni, G, Mathis, CA, and Langstrom, B (2004). Imaging brain amyloid in Alzheimer's disease with Pittsburgh Compound-B. *Ann. Neurol.* 55.3, 306–319.

Kratz, A, Ferraro, M, Sluss, PM, and Lewandrowski, KB (2004). Case records of the Massachusetts General Hospital. Weekly clinicopathological exercises. Laboratory reference values. *N. Engl. J. Med.* 351.15, 1548–1563.

Landau, SM, Breault, C, Joshi, AD, Pontecorvo, M, Mathis, CA, Jagust, WJ, and Mintun, MA (2013). Amyloid- imaging with Pittsburgh compound B and florbetapir: comparing radiotracers and quantification methods. *J. Nucl. Med.* 54.1, 70–77.

Landau, SM, Fero, A, Baker, SL, Koeppe, R, Mintun, M, Chen, K, Reiman, EM, and Jagust, WJ (2015). Measurement of longitudinal -amyloid change with 18F-florbetapir PET and standardized uptake value ratios. *J. Nucl. Med.* 56.4, 567–574.

Lunn, DJ, Thomas, A, Best, N, and Spiegelhalter, D (2000). WinBUGS - A Bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing* 10.4, 325–337. ISSN: 1573-1375. DOI: 10.1023/A:1008929526011. URL: http://dx.doi.org/10.1023/A:1008929526011.

Marie Davidian, DMG (2003). Nonlinear Models for Repeated Measurement Data: An Overview and Update. *Journal of Agricultural, Biological, and Environmental Statistics* 8.4, 387–419. ISSN: 10857117. URL: http://www.jstor.org/stable/1400665.

Metz, CE (1978). Basic principles of ROC analysis. *Semin Nucl Med* 8.4, 283–298.

Mossman, D (1999). Three-way ROCs. *Med Decis Making* 19.1, 78–89.

Muralidharan, R, Mateen, FJ, and Rabinstein, AA (2014). Outcome of fulminant bacterial meningitis in adult patients. *Eur. J. Neurol.* 21.3, 447–453.

Naj, AC et al. (2011). Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. *Nat. Genet.* 43.5, 436–441.

Naj, AC et al. (2014). Effects of multiple genetic loci on age at onset in late-onset Alzheimer disease: a genome-wide association study. *JAMA Neurol* 71.11, 1394–1404.

Nakas, CT and Yiannoutsos, CT (2004). Ordered multiple-class ROC analysis with continuous measurements. *Stat Med* 23.22, 3437–3449.

Nakas, CT and Alonzo, TA (2007). ROC Graphs for Assessing the Ability of a Diagnostic Marker to Detect Three Disease Classes with an Umbrella Ordering. English. *Biometrics* 63.2, pp. 603–609. ISSN: 0006341X.

Nyul, LG and Udupa, JK (1999). On standardizing the MR image intensity scale. *Magn Reson Med* 42.6, 1072–1081.

Ogdie, A, Yu, Y, Haynes, K, Love, TJ, Maliha, S, Jiang, Y, Troxel, AB, Hennessy, S, Kimmel, SE, Margolis, DJ, Choi, H, Mehta, NN, and Gelfand, JM (2015). Risk of major cardiovascular events in patients with psoriatic arthritis, psoriasis and rheumatoid arthritis: a population-based cohort study. *Ann. Rheum. Dis.* 74.2, 326–332.

Pepe, MS (2003). *The Statistical Evaluation of Medical Tests for Classification and Prediction (Oxford Statistical Science Series)*. 1st ed. Oxford University Press. ISBN: 9780198509844.

Petersen, A and Mller, HG (2016). Functional data analysis for density functions by transformation to a Hilbert space. *Ann. Statist.* 44.1, 183–218. DOI: 10.1214/15-AOS1363. URL: http://dx.doi.org/10.1214/15-AOS1363.

Plummer, M, Best, N, Cowles, K, and Vines, K (2006). CODA: Convergence Diagnosis and Output Analysis for MCMC. *R News* 6.1, 7–11. URL: http://CRAN.R-project.org/doc/Rnews/.

R Core Team (2014). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: http://www.R-project.org/.

Ramsay J. O. Silverman, BW (2005). *Functional Data Analysis*. 2nd ed. New York: Springer.

Reiss, PT, Huang, L, and Mennes, M (2010). Fast function-on-scalar regression with penalized basis expansions. *Int J Biostat* 6.1, Article 28.

Resnick, SM, Sojkova, J, Zhou, Y, An, Y, Ye, W, Holt, DP, Dannals, RF, Mathis, CA, Klunk, WE, Ferrucci, L, Kraut, MA, and Wong, DF (2010). Longitudinal cognitive decline is associated with fibrillar amyloid-beta measured by [11C]PiB. *Neurology* 74.10, 807–815.

Resnick, SM, Bilgel, M, Moghekar, A, An, Y, Cai, Q, Wang, MC, Thambisetty, M, Prince, JL, Zhou, Y, Soldan, A, Wong, DF, O'Brien, RJ, Ferrucci, L, and Albert, MS (2015). Changes in AÎ biomarkers and associations with APOE genotype in 2 longitudinal cohorts. *Neurobiol. Aging* 36.8, 2333–2339.

Sacks, FM and Campos, H (2003). Clinical review 163: Cardiovascular endocrinology: Low-density lipoprotein size and cardiovascular disease: a reappraisal. *J. Clin. Endocrinol. Metab.* 88.10, 4525–4532.

Scheller, J, Chalaris, A, Schmidt-Arras, D, and Rose-John, S (2011). The pro- and anti-inflammatory properties of the cytokine interleukin-6. *Biochim. Biophys. Acta* 1813.5, 878–888.

Shinohara, RT, Sweeney, EM, Goldsmith, J, Shiee, N, Mateen, FJ, Calabresi, PA, Jarso, S, Pham, DL, Reich, DS, and Crainiceanu, CM (2014). Statistical normalization techniques for magnetic resonance imaging. *Neuroimage Clin* 6, 9–19.

Slooter, AJ, Cruts, M, Kalmijn, S, Hofman, A, Breteler, MM, Van Broeckhoven, C, and Duijn, CM van (1998). Risk estimates of dementia by apolipoprotein E genotypes from a population-based incidence study: the Rotterdam Study. *Arch. Neurol.* 55.7, 964–968.

Verbeke, G and Molenberghs, G (2001). *Linear Mixed Models for Longitudinal Data*. Springer Series in Statistics. Springer New York. ISBN: 9780387950273. URL: https://books.google.com/books?id=nIoME3vzND0C.

Wood, SN (2000). Modelling and smoothing parameter estimation with multiple quadratic penalties. *Journal of the Royal Statistical Society (B)* 62.2, 413–428.

Wood, SN (2003). Thin plate regression splines. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 65.1, 95–114. ISSN: 1467-9868.

Wood, SN (2004). Stable and efficient multiple smoothing parameter estimation for generalized additive models. *Journal of the American Statistical Association* 99.467, 673–686.

Wood, SN (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society (B)* 73.1, 3–36.

Yushkevich, PA, Piven, J, Cody Hazlett, H, Gimpel Smith, R, Ho, S, Gee, JC, and Gerig, G (2006). User-Guided 3D Active Contour Segmentation of Anatomical Structures: Significantly Improved Efficiency and Reliability. *Neuroimage* 31.3, 1116–1128.