UNIVERSITY OF CALIFORNIA SAN DIEGO

Power and Probability Calculations in Longitudinal Outcome Measures and Image Analysis, with Applications to Biomedical Data

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy

in

Biostatistics

by

Yu Zhao

Committee in charge:

Professor Armin Schwartzman, Chair
Professor Steven Edland, Co-Chair
Professor Howard Feldman
Professor Jingjing Zou

2023

The Dissertation of Yu Zhao is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2023

DEDICATION

To my family and friends.

EPIGRAPH

The most important questions of life are indeed,
for the most part,
really only problems of probability.

Laplace, Pierre Simon

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

First of all, I would like to express my sincere gratitude to my committee chairs, Dr. Armin Schwartzman, and Dr. Steven Edland. I couldn't have finished this dissertation without their support and mentorship. I would also like to thank my other committee members, Dr. Howard Feldman and Dr. Jingjing Zou for providing valuable feedback on my work.

I grew my interest in Biostatistics from Dr. Schwartzman's class where you showed great passion and expertise for statistics. I really appreciate the way you encouraged questions and discussions as it strengthened my ability to think actively and critically. In terms of research, Dr. Schwartzman introduced me to the world of random fields theory which perfectly matches my strength and interest. You are a great advisor for guiding me to initiate ideas with my statistical sense and turn ideas into scientific papers. You have consistently been supportive of my career choices, sharing your own experience and offering valuable advice. I would also like to thank you for organizing various activities, such as hiking trips and the Thanksgiving dinner.

I am grateful to Dr. Edland for leading me into Biostatistics research and helping me to finish my first publications. I am deeply moved by your passion for research. I still remember your excitement when we discuss new research ideas. You provided valuable insights about the applications of Biostatistics in clinical trials to help me bridge the gap between academic research and industry. Your mentorship has also refined my communication and scientific writing which are valuable for my future career. Moreover, I want to express my gratitude for your kindness and care throughout my academic journey.

I also want to thank others that have helped me at the University of California, San Diego. I would like to thank Dr. Ian Abramson, Dr. Dimitris Politis, Dr. Ery Arias-Castro, Dr. Tianyi Zheng, and other professors from the math department for helping me to learn the fundamentals of Statistics theory. I want to thank the faculty members of the Division of Biostatistics and Bioinformatics for providing me with a collegial and stimulating academic environment. The courses and seminars have broadened my perspective and enriched my understanding of different research topics in Biostatistics. I want to thank the current and past staff of the Division of

Biostatistics and Bioinformatics, Sarah Dauchez, Alan Larson and Stella Tripp for your support.

I would like to thank the members of Dr. Schwartzman's research group. I had a great experience participating in the group meetings where we shared research ideas and learned from the insights of others. The opportunity to present my research projects and receive constructive feedback consolidated my understanding and helped me to improve my research works. I also want to thank Dr. Dan Cheng and Dr. Samuel Davenport for being patient and sharing your expertise in random field theory in our collaborative works. It's my great honor to collaborate with someone as knowledgeable and intellectually rigorous as you.

I want to thank my friends and peers in the department, Matt (Ruohui) Chen, Tuo Lin, Charles (Jiyu) Luo, Junting Ren, and Xinran Wang for helping me both in academics and in life. I would like to thank my girlfriend, Peiyu Si for your comfort and valuable information during moments of stress. You have filled my life with laughter and happiness. Finally, I want to thank my family, especially my parents, for your unconditional love and support. Your understanding and encouragement have been a constant source of motivation.

Chapter 2, in full, is a reprint of the material as it appears in the International Journal of Biostatistics. Zhao, Yu; Edland, Steven. Power formulas for mixed effects models with random slope and intercept comparing rate of change across groups, Int J Biostat, 18(1): 173-182, 2021. The dissertation author was the primary investigator and author of this material.

Chapter 3, in full, has been submitted for publication of the material as it may appear in Clinical Trials. Zhao, Yu; Edland, Steven. The chronic progressive repeated measures (CPRM) model for longitudinal data. The dissertation author was the primary investigator and author of this material.

Chapter 4, in full, has been submitted for publication of the material as it may appear in the Journal of Multivariate Analysis. Zhao, Yu; Cheng, Dan; Schwartzman, Armin. An approximation to peak detection power using Gaussian random field theory. The dissertation author was the primary investigator and author of this material.

Chapter 5, in full, is currently being prepared for submission for publication of the

material. Zhao, Yu; Cheng, Dan; Schwartzman, Armin. On the peak height distribution of non-stationary Gaussian random fields: 1D non-constant variance and scale space. The dissertation author was the primary investigator and author of this material.

VITA

2017	Bachelor of Science in Mathematics and Applied Mathematics, Beijing University of Posts and Telecommunications
2019	Master of Science in Statistics, University of California San Diego
2023	Doctor of Philosophy in Biostatistics, University of California San Diego

PUBLICATIONS

Zhao Yu, Edland SD. "Power formulas for mixed effects models with random slope and intercept comparing rate of change across groups" *Int J Biostat* 18(1):173-182, 2021.

Smirnov DS, Salmon DP, Galasko D, Goodwill VS, Hansen LA, Zhao Yu, Edland SD, Léger GC, Peavy GM, Jacobs DM, Rissman R, Pizzo DP, Hiniker A. "Association of Neurofibrillary Tangle Distribution With Age at Onset-Related Clinical Heterogeneity in Alzheimer Disease: An Autopsy Study" *Neurology* 98(5):e506-e517, 2022.

Zhao Yu, Edland SD. "The chronic progressive repeated measures (CPRM) model for longitudinal data" *Manuscript in Preparation*, 2023

Zhao Yu, Cheng D, Schwartzman A. "An approximation to peak detection power using Gaussian random field theory" *arXiv:2301.04830*, 2023.

Zhao Yu, Cheng D, Schwartzman A. "On the peak height distribution of non-stationary Gaussian random fields: 1D non-constant variance and scale space" *Manuscript in Preparation*, 2023.

ABSTRACT OF THE DISSERTATION

Power and Probability Calculations in Longitudinal Outcome Measures and Image Analysis, with Applications to Biomedical Data

by

Yu Zhao

Doctor of Philosophy in Biostatistics

University of California San Diego, 2023

Professor Armin Schwartzman, Chair
Professor Steven Edland, Co-Chair

In this dissertation, we aim to utilize modern statistical frameworks to perform power and probability calculations in longitudinal outcome measures and image analysis. Chapter 1 and Chapter 2 primarily focus on longitudinal outcome measures data. In practice, two different analysis plans are commonly applied to these data: linear mixed effects model and repeated measures analysis. In Chapter 1, we are interested in generalizing current power formulas for linear mixed effects models to accommodate missing data due to study subject attrition, and unequal sample size and variance parameters across groups.

For repeated measures analysis, a covariance structure needs to be specified when model-

ing the data and computing type I error and power. In Chapter 2, we describe a parsimonious covariance structure for repeated measures analysis that is useful for modeling longitudinal repeated measures of chronic progressive conditions and derive the power calculation formulas.

In image analysis, finding the peak height distribution and power for peak detection are known to be challenging due to the spatial aspect of the data. In Chapter 3, we propose a novel way to approximate the power for peak detection using Gaussian random field theory (RFT) and demonstrate scenarios where the approximation works well. We also apply our formulas to 2D and 3D simulated datasets, and the 3D data is induced by real fMRI data to measure performance in a realistic setting.

The main limitation of RFT-based image analysis is the model assumptions, and these assumptions are known to be difficult to check and even not appropriate in many application settings. In Chapter 4, we seek to relax the stationarity assumption and study the peak height distribution of non-stationary Gaussian random fields. The explicit formula for the peak height distribution is derived for 1D smooth Gaussian random fields and efficient numerical algorithms are proposed as a general solution for computing the peak height distribution in applications.

Chapter 1

Introduction

Accurate probability calculation is crucial in hypothesis testing and general statistical analysis. For example, calculating power, the probability that the test correctly rejects the null hypothesis when a specific alternative hypothesis is true, helps statisticians decide the minimum sample size required for an informative test, and thus reduce cost. Probability calculation formulas exist for common statistical and hypothesis testing problems, such as cumulative distribution function (CDF) of Gaussian distribution, and type I error/ power in t-test. However, due to the complexities of biomedical data, probability and power calculations can be challenging.

Longitudinal outcome measures and image data are two important types of biomedical data used in studying a certain disease like Alzheimer's disease. According to the National Institute on Aging, Alzheimer's disease is a leading cause of death and disability in the United States. A variety of treatments are being developed to manage the disease and slow the disease progression. Longitudinal outcome measures data collects outcome measures, such as cognitive assessment, of the same subject over time, and is widely used to evaluate the disease progression and treatment effect in clinical trials and cohort studies. Biomedical image data contains valuable information in the form of images that helps to fully understand the human body functions and gain insights into the complex changes that occur due to interventions. For example, we can measure the structural changes with scanning devices like computed tomography (CT) and magnetic resonance imaging (MRI), and functional changes with scanning devices like

functional magnetic resonance imaging (fMRI) and positron emission tomography (PET). With the rapid development of these devices, we are able to obtain high-quality images of the brain and other parts of the body, and it becomes suitable to perform statistical inference. Longitudinal data analysis and image analysis both play a crucial role in capturing the full range of disease progression to benefit the treatment development.

Although longitudinal outcome measures and image data are two different types of biomedical data, they share a common challenge for analysis: dealing with data correlation. Longitudinal data exhibits correlation among outcome measures taken from the same subject, and image data exhibits correlation between nearby pixels. Dealing with correlation in longitudinal outcome measures and image data is necessary for accurate power and probability calculations.

Ignoring the correlation in longitudinal data can bias the estimated variance of regression parameters leading to invalid statistical inferences. Two different analysis plans are commonly applied to longitudinal outcome measures data to address the issue of data correlation. Under the assumption of linear trajectories with random normal intercepts and slopes, the data can be fitted using the linear mixed effects model. In linear mixed effects model, the data correlation is induced by the subject specific random intercept and slope. The treatment effect can be seen as the difference in rate of progression between the treatment and placebo group and is estimated as the difference in fixed effect slope between the two groups. Alternatively, if the outcome measure is observed on a fixed schedule, data can be modeled using repeated measures analysis with arbitrary (non-linear) fixed effect estimates of level at each time point of observation. The repeated measures analysis allows the user to deal with data correlation by specifying a correlation structure such as compound symmetry (CS) and autoregressive (AR). For the repeated measures analysis, the treatment effect can be estimated using the difference-in-difference method.

Both the linear mixed effects model and repeated measures analysis are studied in this dissertation. For the linear mixed effects model, we generalize the existing power calculation formulas (e.g. Hedeker et al., 1999; Wang et al., 2015) to accommodate missing data due to study subject attrition. Missing data due to study subject dropout is common in clinical trials

and cohort studies and poses challenges to treatment effect estimates and power calculations. When there are missing observations, the effective sample size is reduced, leading to lower statistical power. Calculating the amount of power loss according to the missing pattern is necessary to make sure the study is adequately powered. The generalized power formulas also accommodate unequal sample size and variance-covariance parameters across groups. Differences in variance-covariance parameters are important but typically ignored in model fitting and power calculations. The variability in disease progression observed in the treatment group reflects the normal background variability in the control group plus variability in response to treatment, making it inappropriate to use the same set of variance-covariance parameters.

For the repeated measures analysis, we propose a novel covariance structure to model data under chronic progressive conditions. Mixed model repeated measures (MMRM) analysis with unstructured (UN) covariance (Mallinckrodt et al., 2008) is a common primary statistical analysis plan for clinical trials. This approach, however, can be problematic when the number of repeated measures is large and the model fails to converge. Alternatively, we may consider parsimonious covariance structures like compound symmetry (CS) and autoregressive (AR). We note that a substantial literature describes mixed model repeated measures analyses assuming CS or AR covariance structure (Rochon, 1991; Muller et al., 1992; Lu et al., 2008; Wang et al., 2015). Although CS and AR models have less number of parameters compared to UN, and are easier to converge, these models are not appropriate for repeated measures of chronic progressive conditions. A typical longitudinal pattern of disease progression for chronic progressive diseases is a fanning apart of trajectories over time, similar to the individual trajectories from the linear mixed effects model. The model we propose has non-linear fixed effect means like MMRM, but uses a pattern of dispersion of longitudinal repeated measures consistent with the covariance structure of the linear mixed effects model. Characterizing this covariance structure requires only four parameters for estimation and accurately describes the longitudinal pattern under chronic progressive conditions.

In image analysis, the data can be decomposed into a fixed signal component and a

random noise component. The signal and noise represent the underlying pattern and uncertainty of the data respectively. The goal is to search a large space of noisy data for a relatively small number of signals. This can be done by treating each data pixel or voxel as a statistic value and testing it against the null hypothesis of no signal anywhere. Due to the nature of biomedical image data, the number of pixels is usually large and the nearby pixels are correlated. As a result, we are conducting multiple correlated hypothesis tests simultaneously, and ignoring the spatial correlation leads to inaccurate estimation of the overall type I error and power. For example, a traditional multiple testing technique like the Bonferroni correction, which guards against correlation but does not model it explicitly, is too conservative and may miss detecting a true effect. In this dissertation, we model the random noise and account for the data correlation with the help of Gaussian random field theory (RFT). The basic idea is to treat the data as a realization of a random field, and use RFT to find a proper threshold to identify active pixels.

Peak detection is an important technique in image analysis for identifying the signals masked by random noise. It can be posed as a statistical testing problem intended to test whether the underlying signal has a peak at a given location. This is typically done by comparing the height of the observed peaks to a threshold computed from the distribution of peak height under the null hypothesis. RFT based peak detection has been studied in Cheng and Schwartzman (2017) and Schwartzman and Telschow (2019b), which provide methods to choose the threshold while controlling the overall type I error. Formulas for the type I error which is the probability of peak height falling above the threshold under the null hypothesis (zero mean) have been developed for stationary fields in 1D and isotropic fields in 2D and 3D (Cheng and Schwartzman, 2015; Cheng and Schwartzman, 2017; Cheng and Schwartzman, 2018). However, there is no exact formula to calculate power since it is challenging to deal with non-centered fields. Here we propose to approximate power by calculating the expected number of peaks above the threshold. We show that the approximation, which is also an upper bound, works well under certain scenarios.

The main limitation of peak detection using RFT is model assumptions like stationarity

and isotropy assumptions. In practice, these assumptions do not always hold. For example, kernel smoothing is common in brain imaging analysis to increase the signal-to-noise ratio, and the optimal smoothing bandwidth or scale of the kernel is usually unknown. One way to handle this is to treat the scale as a parameter just like the location and look for peaks in the scale-location space. The random field defined on this scale-location space is called the scale space random field (Siegmund and Worsley, 1995). The scale space field is non-stationary since the smoothness varies over the domain. To expand the usage of RFT in peak detection, we derive the peak height distribution of 1D centered non-stationary Gaussian random processes and study the statistical properties of the peak height distribution of the scale space field. These properties give us a better interpretation of the scale space field and simplify the computation of peak height distribution. Since the exact formula for multidimensional non-stationary Gaussian random fields remains difficult to derive, as a general solution in practice, we introduce two convenient and efficient numerical algorithms that can be used to evaluate the expected number of critical points and the peak height distribution without having to simulate the field itself. These two algorithms allow us to calculate the peak height distribution of any random field as long as it is sufficiently smooth.

1.1 Organization of this dissertation

In Chapter 2, we focus on the linear mixed effects model, and derive general power formulas for longitudinal studies with study subject dropout comparing mean slope across groups. The power formulas fully accommodate missing data due to study subject dropout, and allow for unequal missing patterns, sample size, and variance-covariance parameters across groups. We demonstrate how power formulas under this model can be used to power a future trial of arbitrary design (arbitrary number and interval between follow-up observations) regardless of the design of the pilot study informing power calculations, and how to estimate the required parameters from the pilot study through a real data example. We also perform simulations to evaluate the

performance of the power formulas using parameters estimated from the real data example.

In Chapter 3, we direct our efforts towards the repeated measures analysis approach and propose a novel covariance structure for the mixed model repeated measures (MMRM) analysis under the chronic progressive conditions. We call it the chronic progressive (CP) covariance structure, and the corresponding analysis plan the chronic progressive repeated measures (CPRM) analysis. We demonstrate empirically using data from completed clinical trials that the CP covariance assumption holds for longitudinal cognitive data in Alzheimer's disease and alternative parsimonious MMRM covariance structures including compound symmetry (CS) and autoregressive (AR) are not appropriate for modeling the covariance pattern of this type of data. We further demonstrate with computer simulations that CS and AR perform poorly for chronic progressive conditions, with the CS model being anticonservative, and the autoregressive model being poorly powered. Finally, the generalized power formulas for the CPRM model are derived, and similar to the mixed effects model, the advantage of using the CPRM parameterization is that the pilot study used to inform power calculations does not have to match the design of a future trial.

In Chapter 4, we study power approximation formulas for peak detection using Gaussian random field theory. The approximation, based on the expected number of local maxima above the threshold is proved to work well under three asymptotic scenarios: small domain, large threshold, and sharp signal. An adjusted version of this approximation is also proposed to improve accuracy when the expected number of local maxima exceeds one. The proposed approximation makes the problem more tractable, but in general, it does not have an explicit form. We further simplify the formula under the isotropy assumption and show its explicit form in 1D, 2D, and 3D. We also apply our formulas to 2D and 3D simulated datasets, and the 3D data is induced by real fMRI data from the Human Connectome Project to measure performance in a realistic setting.

In Chapter 5, our main focus is on the peak height distribution of non-stationary Gaussian random fields, so that we can expand the application of random field theory in peak inference.

We first derive the peak height distribution of 1D centered non-stationary Gaussian random fields by evaluating the Kac-Rice formula. Based on the results we have developed in 1D, we explore the properties of the peak height distribution of the scale space field. For multidimensional non-stationary Gaussian random fields like the scale space field, although the general formula is not available, we introduce two algorithms for evaluating the Kac-Rice formula numerically. We perform simulation studies to validate the theoretical properties of the non-stationary Gaussian random fields and evaluate the performance of the numerical algorithms.

Chapter 2

Power formulas for mixed effects models with random slope and intercept comparing rate of change across groups

2.1 Introduction

Lu et al. (2008) have previously described sample size formulas for longitudinal studies with study subject dropout for the mixed model repeated measures analysis comparing change from baseline to last visit across groups. Missing data due to study subject dropout in clinical trials and cohort studies is common and reduces statistical power to detect treatment effects or differences in change across groups. We here derive power formulas for longitudinal studies with study subject dropout for a different model, the mixed effects model with random slopes and intercepts comparing mean slope across groups. We demonstrate how power formulas under this model can be used to power a future trial of arbitrary design (arbitrary number and interval between follow-up observation) regardless of the design of pilot study informing power calculations. We expand and generalize previously published mixed effects model power formulas (e.g. Hedeker et al., 1999; Wang et al., 2015) to fully accommodate differences in length and interval between longitudinal observations, different allocation ratios, and different study subject attrition rates. We also derive a formula that accommodates different covariance structures across groups. Differences in covariance are typically ignored, but may be critical

to clinical trials, where changes over time in the treatment arm reflect the normal background variability in progression observed in the placebo control arm plus variability in response to treatment, meaning that power calculations based only on the placebo arm covariance structure may be anticonservative. To our knowledge, this is the first presentation of power formulas for the mixed effects model with random slopes and intercepts that accommodates differences in model variance parameters across groups. We note that a substantial literature describes many of these features for mixed model repeated measures analyses assuming compound symmetric or autoregressive covariance of repeated measures (Rochon, 1991; Muller et al., 1992; Lu et al., 2008; Wang et al., 2015). While compound symmetric and autoregressive covariance structures are mathematically more tractable, in our experience these models are not appropriate for repeated measures of chronic progressive conditions. We demonstrate by example that compound symmetric and autoregressive covariance structures typically are not appropriate for modeling chronic progressive conditions. In the interest of clarity, in this chapter, we focus exclusively on the model with covariance structure imposed by random slopes and intercepts most appropriate for chronic progressive outcome measures.

2.2 Background, the mixed effects model

The parameterization of the mixed effects model with random slopes and intercepts used in this derivation is the familiar Laird and Ware mixed effects model parameterization with estimation and hypothesis testing by restricted maximum likelihood (REML). We use the notation of Laird and Ware (1982) to represent within group longitudinal observations y_i on subject i as

$$y_i = X_i\alpha + Z_i b_i + e_i, \tag{2.1}$$

where α are the fixed effect intercept and slope describing the mean longitudinal trajectory, $b_i \sim \mathcal{N}(0, D)$ are random, subject-specific intercepts and slopes, and $e_i \sim \mathcal{N}(0, R_i)$ is residual variation about the individual trajectories. When convenient, we will represent the elements of D

as

$$\begin{pmatrix} \sigma_{b_0}^2 & \sigma_{b_0,b_1} \\ \sigma_{b_0,b_1} & \sigma_{b_1}^2 \end{pmatrix}.$$

In the derivation below, $X_i = Z_i$ are subject specific design matrices composed of a column of ones and a column of times at which measurements y_i were made. To simplify presentation we maintain large sample normality assumptions in all that follows, and we do not consider covariates beyond t_i. Consistent with prior literature (Hedeker et al., 1999; Wang et al., 2015), we assume that data are missing at random and that the covariance parameters are known.

Searle (1970) showed that $V(\hat{\alpha})$, the asymptotic variance of maximum likelihood estimates of α, is independent of $\hat{\alpha}$ and derived its value. Under model (2.1), y is normally distributed with mean $X\alpha$ and variance-covariance V. The likelihood function is

$$l = (2\pi)^{-\frac{1}{2}n}|V|^{\frac{1}{2}}\exp\left(-\frac{1}{2}(y-X\alpha)'V^{-1}(y-X\alpha)\right). \tag{2.2}$$

The log likelihood, apart from a constant is

$$L = -\frac{1}{2}\log|V| - \frac{1}{2}(y-X\alpha)'V^{-1}(y-X\alpha). \tag{2.3}$$

By the $\sqrt{n}$-consistency and asymptotic efficiency of MLE, $\hat{\alpha}$ the maximum likelihood estimate of α follows

$$\sqrt{n}(\hat{\alpha}-\alpha) \xrightarrow{d} \mathcal{N}(0, I^{-1}(\alpha)), \tag{2.4}$$

where $I(\alpha)$ is the information matrix which equals to $\mathbb{E}[\partial^2 L/\partial\alpha_h\partial\alpha_k]$. For the log likelihood (2.3), after taking the partial derivative and expectation,

$$I(\alpha) = X'V^{-1}X. \tag{2.5}$$

Thus the asymptotic variance of $\hat{\alpha}$ is

$$\text{Var}(\hat{\alpha}) = (X'V^{-1}X)^{-1}. \tag{2.6}$$

We can further simplify this as

$$\begin{aligned}
\text{Var}(\hat{\alpha}) &= (X'V^{-1}X)^{-1} \\
&= \left((X_1', X_2', \ldots, X_n') \begin{pmatrix} V_1^{-1} & 0 & 0 & 0 \\ 0 & V_2^{-1} & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & V_n^{-1} \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix} \right)^{-1} \\
&= \left(\sum_i (X_i' V_i^{-1} X_i) \right)^{-1},
\end{aligned} \tag{2.7}$$

where

$$V_i = \text{Cov}(y_i) = Z_i D Z_i' + R_i. \tag{2.8}$$

In particular, the lower right diagonal of $\text{Var}(\hat{\alpha})$ is the variance of the mean slope estimate which is required for sample size formulas to power clinical trials comparing mean slope in treatment versus control. The components of $\text{Var}(\hat{\alpha})$ can be estimated by REML (Laird and Ware, 1982).

Two specific cases of (2.7) are useful for illustrative purposes. If we are dealing with balanced data, then X_i and V_i are constant across subjects, and (2.7) reduces to simply

$$\text{Var}(\hat{\alpha}) = (nX_i' V_i^{-1} X_i)^{-1}. \tag{2.9}$$

A similar clinical trial with missing observations due to missed clinical exams or study subject dropout would not have constant V_i and X_i, but instead would have a finite set of design and variance matrix pairs. Letting k index this set, the variance of the fixed effect estimates for a

clinical trial with missing data is then equal to

$$\mathrm{Var}(\hat{\alpha}) = \left(\sum_k n_k (X_k' V_k^{-1} X_k) \right)^{-1} = \left(n \sum_k p_k (X_k' V_k^{-1} X_k) \right)^{-1}, \tag{2.10}$$

where the n_k are counts of subjects in each set and sum to n, and $p_k = n_k/n$.

2.3 Power formulas derived

2.3.1 Power formula, balanced design with no dropout

For the balanced design with no dropout, standard power formulas apply. E.g., for equal allocation to arms, sample size to detect a difference in mean slope Δ between treatment and control is

$$N/Arm = 2(z_{\alpha/2} + z_\beta)^2 [(X_i' V_i^{-1} X_i)^{-1}]_{22} / \Delta^2. \tag{2.11}$$

This formula can be used given an estimate of $V_i = \mathrm{Cov}(y_i)$ obtained from pilot data or a previously completed trial of comparable design.

A more generally applicable formula can be derived given the usual assumption of independent residual error ($R_i = \sigma_\varepsilon^2 I$). Under this assumption, it can be shown (see Appendix) that $[(X_i' V_i^{-1} X_i)^{-1}]_{22} = \sigma_{b_1}^2 + \sigma_\varepsilon^2 / \sum (t_j - \bar{t})^2$ (Snijders and Bosker, 1993), and (2.11) reduces to

$$N/Arm = 2(z_{\alpha/2} + z_\beta)^2 \left(\sigma_{b_1}^2 + \sigma_\varepsilon^2 / \sum (t_j - \bar{t})^2 \right) / \Delta^2, \tag{2.12}$$

where $\Sigma (t_j - \bar{t})^2$ is the sum over the measurement time vector $t = (t_1, t_2, ..., t_m)'$ of the squared differences t_j minus mean time.

(2.12) is more generally applicable because it only requires estimates of σ_ε^2 and $\sigma_{b_1}^2$, which can be obtained by REML fit to longitudinal pilot data of arbitrary design. That is, future studies can be powered using prior study data that do not necessarily have the same duration or interval between follow-up as the planned future study (Ard and Edland, 2011). Equation (2.12)

also provides a heuristic illustration of the influence of study design on power longer trials or trials with more longitudinal observations increase power by reducing the influence of σ_ε^2 on overall variance.

2.3.2 Power formula, balanced design with dropout

Another important example, following Lu et al. (2008), is the case of study subject dropout during a cohort study or clinical trial, also referred to as study subject attrition (SSA). SSA implies a subset of the dropout patterns indexed by k in (2.10), restricting to the $m-1$ longitudinal dropout patterns composed of subjects whose last visit is at t_k, $k = 2$ through m inclusive. Given the independent residual errors assumption and equal allocation to arms, under SSA the sample size is calculated by

$$N/Arm = 2(z_{\alpha/2}+z_\beta)^2([\left(\Sigma p_k(X_k'V_k^{-1}X_k)\right)^{-1}]_{22}/\Delta^2), \tag{2.13}$$

where the sum is over the $m-1$ dropout patterns defined by SSA, $p_k(X_k'V_k^{-1}X_k)$ are as in (2.10), and V_k are matrices with off diagonal elements u,v equal to $\sigma_{b_0}^2+(t_u+t_v)\sigma_{b_0,b_1}+t_ut_v\sigma_{b_1}^2$ and diagonal elements u,u equal to $\sigma_{b_0}^2+2t_u\sigma_{b_0,b_1}+t_u^2\sigma_{b_1}^2+\sigma_\varepsilon^2$. As before, the parameters $\sigma_{b_0}^2$, σ_{b_0,b_1}, and $\sigma_{b_1}^2$ of D and the residual error variance σ_ε^2 are estimated by REML fit to representative prior longitudinal data.

Power formulas accommodating study subject attrition such as (2.13) and Lu et al. (2008) are technically anticonservative because they ignore information lost by the occasional missed interim visit, although this bias is typically small. If missing interim visit data is a concern, then applying (2.13) over all sets of missing data patterns will ensure true nominal type I error rates are maintained.

2.3.3 Power formula, unequal allocation, unequal study subject attrition, and unequal variance across groups

Formulas (2.12) through (2.13) assume that variance parameters and study subject attrition rates are the same in the two groups being compared and the number of subjects in each group is equal. We may require a formula that accommodates different study subject attrition rates across groups, and/or unequal allocation to groups (Lu et al., 2008). It would also be useful to have a formula that accommodates different variance parameters across groups. Letting $Term_1$ and $Term_2$ indicate the values $[\left(\Sigma p_k (X_k' V_k^{-1} X_k)\right)^{-1}]_{22}$ calculated separately for $group_1$ and $group_2$, and given the independent identically distributed residual error assumption, sample size for $group_1$ can be calculated by

$$N_{group_1} = (z_{\alpha/2} + z_{\beta})^2 (Term_1 + \lambda Term_2)/\Delta^2), \tag{2.14}$$

where λ is the sample size ratio across groups ($N_{group_2} = N_{group_1}/\lambda$). The derivation of (2.14) is straightforward, and follows from the observation that the variance of the difference in fixed effects slope estimates equals the sum of the individual slope estimate variances. Factoring out $1/N_{group_1}$ from this sum leaves the quantity $(Term_1 + \lambda Term_2)$, and power as a function of N_{group_1} follows.

2.3.4 Modeling under the unequal variance across groups assumption

It is given that using (2.14) with unequal variance parameters to power a study presumes the analysis plan for the study explicitly models the covariance structure of the two groups. For most applications, including clinical trials, σ_ε^2 is assumed constant across groups. Sample syntax explicitly modeling the remaining, within group random effects parameters determining the covariance structure of repeated measures is included in Appendix.

2.4 Example

Given representative pilot data it is a simple matter to estimate the variance terms required for the power formulas. For example, Table 2.1 is the output from a mixed effect model fit to longitudinal ADAS-cog scores observed in the ADCS trial of a folic acid/B6/B12 compound to slow the progression of Alzheimer's disease (Aisen et al., 2008) (n = 330 subjects and m = 7 observations per subject) using the software provided with the standard mixed effects model text Mixed-Effects Models in S and S-PLUS (Pinheiro and Bates, 2000). The correlation of repeated measures estimated by the random slopes and random intercepts REML model fit (Table 2.2) mirrors the empirical correlation calculated from the same sample data, confirming that this model well represents the covariance structure of longitudinal repeated measures of a chronic progressive condition. In contrast, the commonly assumed compound symmetric and autoregressive covariance structures are constant on the diagonals and inconsistent with these longitudinal data of a chronic progressive condition.

Table 2.1. Sample model fit using the R package nlme and R function lme.

```
lme(y ~ time, random= ~ time|id)

Random effects formula: ~time | subject
            StdDev   Corr
(Intercept) 7.432548 (Intr)
time        3.964215 0.465
Residual    3.705466

Fixed effects: ADAS ~ time
                Value Std.Error
(Intercept) 17.745024 0.4321112
time         4.057879 0.2672020

Number of Observations: 2310  /  Number of Groups: 330
```

Table 2.2. Correlation matrices estimated using data from the ADCS Folate/B6/B12 clinical trial. The correlation matrix imposed by a random effect model fit (RE, bottom panel) closely mirrors the empirical correlation matrix (top panel).

Empirical correlation matrix

$$\begin{pmatrix} 1 & 0.83 & 0.82 & 0.79 & 0.80 & 0.80 & 0.76 \\ 0.83 & 1 & 0.84 & 0.82 & 0.82 & 0.85 & 0.80 \\ 0.82 & 0.84 & 1 & 0.85 & 0.85 & 0.85 & 0.84 \\ 0.79 & 0.82 & 0.85 & 1 & 0.88 & 0.85 & 0.84 \\ 0.80 & 0.82 & 0.85 & 0.88 & 1 & 0.90 & 0.88 \\ 0.80 & 0.85 & 0.85 & 0.85 & 0.90 & 1 & 0.89 \\ 0.76 & 0.80 & 0.84 & 0.84 & 0.88 & 0.89 & 1 \end{pmatrix}$$

Correlation matrix estimated assuming RE

$$\begin{pmatrix} 1 & 0.81 & 0.80 & 0.80 & 0.78 & 0.77 & 0.76 \\ 0.81 & 1 & 0.83 & 0.83 & 0.82 & 0.81 & 0.81 \\ 0.80 & 0.83 & 1 & 0.85 & 0.85 & 0.85 & 0.84 \\ 0.80 & 0.83 & 0.85 & 1 & 0.87 & 0.87 & 0.87 \\ 0.78 & 0.82 & 0.85 & 0.87 & 1 & 0.88 & 0.89 \\ 0.77 & 0.81 & 0.85 & 0.87 & 0.88 & 1 & 0.90 \\ 0.76 & 0.81 & 0.84 & 0.87 & 0.89 & 0.90 & 1 \end{pmatrix}$$

From Table 2.1, the estimated standard deviation of slopes $\hat{\sigma}_{b_1}$ is 3.964 and the estimated standard deviation of residual errors $\hat{\sigma}_\varepsilon$ is 3.705. Assuming equal variance across arms, and using these values in (2.12), the sample size required to detect a 25% slowing of cognitive decline ($\Delta = 0.25 * 4.06$) with 80% power and a type I error rate of 5% for an 18 month trial with observations every 3 months is 360 subjects/arm. For comparison, a 24 month trial with observations every 3 months would require 296 subjects per arm using (2.12). Note that it is not necessary for the design of the pilot study (i.e., the number of observations and interval between observations) to match the design of the future trial, we only require that there are sufficient pilot data to estimate the variance parameters $\sigma^2_{b_1}$ and σ^2_ε.

2.5 Validation by computer simulation

To evaluate the performance of equations (2.12) through (2.14) we have performed computer simulations assuming data following the model fit obtained in the Example above. We

first performed simulations assuming a clinical trial with balanced design with six post-baseline time points with no loss to follow-up and with equal variance within arms consistent with (2.12). Simulating a series of clinical trials with sample size from 200 to 600 subjects per arm with effect size equal to a 25% reduction in the mean rate of decline observed in placebo (25% of the mean 4.06 points per year rate of decline observed in the pilot data (Table 2.1)) with 10,000 simulations per sample size simulated, we found that simulated power closely tracks the power predicted by (2.12) (top line, Figure 2.1).

To validate the power formula for data with study subject attrition described in (2.13), we simulated data under equivalent conditions, except that for each simulation we randomly dropped 5% of the initial sample from each arm at t_2 through t_7. We similarly found that simulated power closely tracks the power predicted by the equation (2.13) power formula (bottom line, Figure 2.1). Study power decreases when there is study subject attrition (Figure 2.1).

To validate the power formula for data with unequal allocation to groups described in (2.14), we simulated data with 5% study subject attrition at each follow-up visit as above, but let the allocation ratio λ vary from one to two. Simulated power closely tracks the power predicted by the equation (2.14) power formula (Figure 2.2). Predictably (Meinert, 1986; Vozdolska et al., 2009), power is maximized when λ equals one, and declines as the allocation ratio deviates from one (Figure 2.2).

To validate (2.14) power formula when covariance structures differ across groups, we simulated data as done in the top line of Figure 2.1, but increased σ_{b_1} by 50% in one of the groups. Simulated power closely tracks the power predicted by (2.14) power formula (Figure 2.3). The top line from Figure 2.1 is included in Figure 2.3 for reference. Figure 2.3 illustrates the potential for anticonservative power calculations in the clinical trial setting when variance parameters used in power calculations are informed by prior placebo arm data and assumed to be constant across arms.

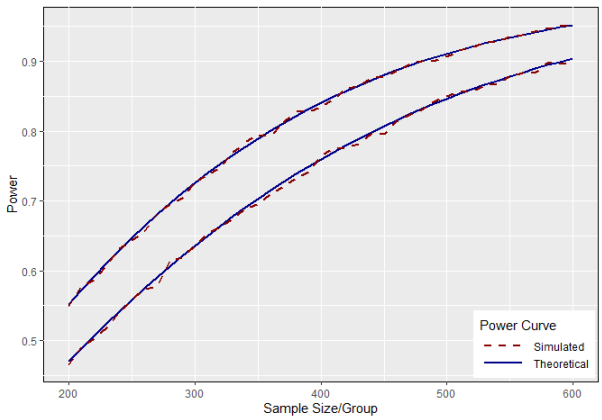


Figure 2.1. Theoretical power curves versus power estimated by computer simulation given no study subject attrition (top curve) and give 5% attrition per follow-up visit (bottom curve). (10,000 simulations per sample size, two-sided test, type I error $\alpha = 0.05$.)

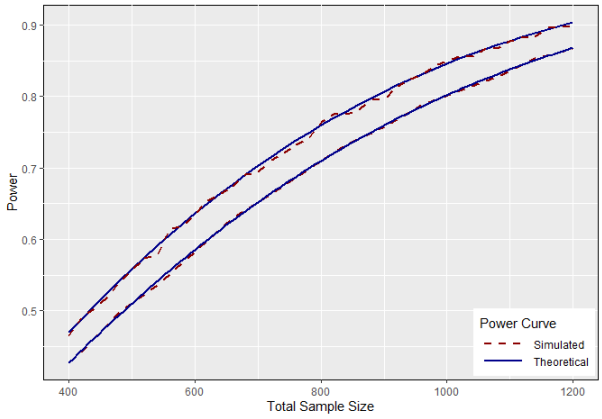


Figure 2.2. Theoretical powers curve versus power estimated by computer simulation given 5% study subject attrition per visit, and allocation ratio $\lambda = 1$ (top curve) and $\lambda = 2$ (bottom curve). (10,000 simulations per sample size, two-sided test, type I error $\alpha = 0.05$.)

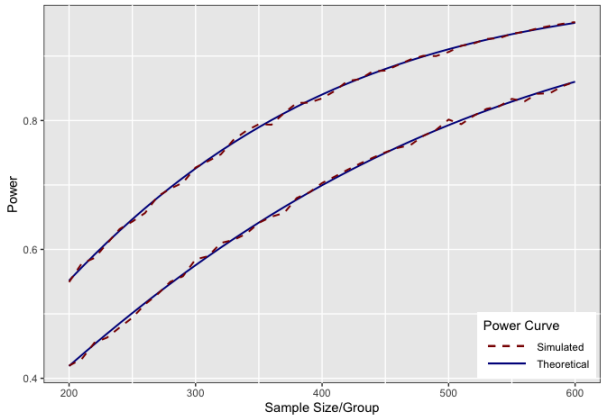


Figure 2.3. Theoretical power curves versus power estimated by computer simulation given equal variance of random slopes (top line) and given σ_{b_1} is increased by 50% in one of the groups (bottom line). (10,000 simulations per sample size, two-sided test, type I error $\alpha = 0.05$.)

2.6 Discussion

There are limitations to the Laird and Ware model as parameterized in (2.1), because this model depends on the assumption that mean trajectories are linear as a function of time. This assumption may be violated, particularly in clinical trials of treatments with potential acute treatments effect beyond simple alteration of rate of progress of disease. In this circumstance mixed model repeated measure analysis (Lu et al., 2008) or model robust alternatives such as generalized estimating equations (Jung and Ahn, 2003) would be preferred. In our experience the linearity assumption is often appropriate for chronic progression conditions, especially when the interval of observation under study is small relative to the full trajectory of disease.

We further note that the formulas presented here assume variance parameters are known, as is typical of the power formula literature (Rochon, 1991; Hedeker et al., 1999; Tu et al., 2007; Lu et al., 2008; Wang et al., 2015). However, variance parameters may be uncertain if sample size in pilot studies used to estimate the variance parameters is small or if pilot data are not perfectly representative of the future investigation being powered. There is a literature on characterizing power when variance parameter estimates are uncertain (e.g. Browne, 1995). However, these methods apply to narrow applications that do not include random effects models.

We recommend sensitivity analyses using a range of plausible variance parameters to ensure that planned future investigations are adequately powered. If the prior data informing power calculations are available, sensitivity analyses may be informed by bootstrap estimates of the uncertainty of variance parameter estimates (e.g., McEvoy et al., 2010). We have also used computer simulations to explore the adequacy of pilot study sample size to inform future trials in other applications (Edland et al., 2017).

The formulas derived here are useful for determining the relative efficiency of different study designs using the mixed effects model to test for differences in mean rate of change between groups. We have described how efficiency can vary by the number and interval between observations, the study subject attrition rate, the allocation ratio, and by differences in variance parameters between groups. Increasing the length of observation or number of observations increases statistical power, although with diminishing returns depending on the magnitude of residual error variance of the outcome measure under study (see (2.12)). Study subject attrition can also meaningfully impact statistical power and should be accounted for in study design (see (2.13) and, e.g., Figure 2.1).

Regarding recruitment allocation ratios, if all other conditions are equal across groups, then altering the allocation ratio from one-to-one reduces statistical power for given study sample size (Meinert, 1986). Altering the allocation ratio has been propose to improve statistical power when there is differential attrition rates across clinical trial arms (Lu et al., 2008). More commonly, allocation ratios are altered to increase the probability of randomization to the active treatment in the hope of increasing clinical trial recruitment rates. While this approach may increase recruitment rates, it also implies more subjects will have to be recruited to achieve target statistical power, and trade-offs between clinic trial cost and time to completion should be considered carefully when planning a trial with unequal randomization to arms (Vozdolska et al., 2009).

Finally, we describe how statistical power depends on variance parameters which may vary across groups (see (2.14)). This consideration is typically overlooked, but may be especially

relevant to clinical trials, where rate of progression in the active treatment arm is a function of both underlying variability in rate of progression and variability in response to treatment. Given that response to treatment is unlikely to be constant across subjects, we can anticipate that the variance of random slopes in the treatment arm will be larger than variance in the control arm if there is a treatment effect. Hence, power calculations based only on the covariance within placebo data will be anticonservative. Typically pilot data for clinical trials are from placebo arm data of a previous trial or registry trial with no treatment arm. A conservative power calculation assumption under these circumstances would be to use an inflation factor for $\sigma^2_{b_1}$ within the treatment arm in (2.14) to be more likely to achieve nominal power in the planned trial.

Formulas (2.12), (2.13), and (2.14) are implemented in the R package longpower (Donohue, 2019), and will be useful tools for planning future cohort studies and clinical trials as well as for comparing the influence of the many factors affecting the efficiency of such investigations. Areas of additional research include modifying power calculation methods in anticipation of evolving guidelines on statistical analysis plans for clinical trials in the presence of missing not at random data (ICH Working Group, 2019), and generalizing power formulas to more directly address the stochastic nature of covariance parameter estimates typically used in practice.

2.7 Acknowledgement

Chapter 2, in full, is a reprint of the material as it appears in the International Journal of Biostatistics. Zhao, Yu; Edland, Steven. Power formulas for mixed effects models with random slope and intercept comparing rate of change across groups, Int J Biostat, 18(1): 173-182, 2021. The dissertation author was the primary investigator and author of this material.

Chapter 3

The chronic progressive repeated measures (CPRM) model for longitudinal data

3.1 Introduction

Mixed model repeated measures (MMRM) assuming an unstructured covariance matrix (Mallinckrodt et al., 2008) is a standard analysis plan for clinical trials of chronic progressive conditions. For example, this was the statistical analysis plan for two recently approved treatments for early Alzheimer's disease (van Dyck et al., 2023; Budd Haeberlein et al., 2022). The number of covariance parameters in the MMRM model increases quadratically with the number of repeated measures, and MMRM models may fail to converge. For this reason, regulatory trial statistical analysis plans include contingency analysis plans, typically MMRM analysis assuming parsimonious compound symmetric (CS) or first order autoregression (AR1) covariance structures.

In this chapter, we demonstrate that the CS and AR1 models are not appropriate for chronic progressive data. We describe an alternative parsimonious covariance structure that is more appropriate for chronic progressive data. The model we propose (Section 3.2) has arbitrary fixed effect means like the MMRM model, but uses a pattern of dispersion of longitudinal repeated measures consistent with the covariance structure of a random slopes model. We call this alternative covariance structure the chronic progressive (CP) covariance structure, and the corresponding analysis the chronic progressive repeated measures (CPRM) analysis.

The CPRM analysis has a number of favorable features. The CP covariance model requires only four parameters for estimation. We demonstrate using data from completed clinical trials that the CP covariance assumption holds for longitudinal cognitive data in Alzheimer's disease (Section 3.3). We further demonstrate that the CS model is anticonservative, and the AR1 model is poorly powered (Section 3.4) when applied to chronic progressive data. Finally, we derive power formulas for the CPRM model (Section 3.5). A unique advantage of the CPRM parameterization is that power formulas are more flexible in that they can be used to power future clinical trials of arbitrary design (with arbitrary number and interval between followup visits) regardless of the design of the pilot study used to inform power calculations.

3.2 The chronic progressive repeated measures (CPRM) model

The CPRM model is a modification of the familiar application of the Laird and Ware (1982) mixed effects model of longitudinal data

$$y_{ij} = \beta_0 + \beta_1 t_j + b_{0i} + b_{1i} t_j + e_{ij}, \tag{3.1}$$

where y_{ij} is the response for subject i ($i = 1,2,...,n$) at time j ($j = 1,2,...,m$), $t_1, t_2, ..., t_m$ are times at which measurements y_i are made, $\beta = (\beta_0, \beta_1)$ are the fixed effect coefficients describing the mean longitudinal trajectory, $(b_{0i}, b_{1i}) \sim \mathcal{N}(0, D)$ are random, subject-specific intercepts and slopes, and $e_i \sim \mathcal{N}(0, R)$ is residual variation about the individual trajectories. Where convenient, we will represent the diagonal elements of D as σ^2_{b0} and σ^2_{b1}, and the off-diagonal elements as σ_{b_0,b_1}.

The CPRM model replaces the fixed effects intercept and slope in (3.1) with m means, one for each repeated measurement time, and can be written as

$$y_{ij} = \alpha_j + b_{0i} + b_{1i} t_j + e_{ij}, \tag{3.2}$$

where $\alpha_j, j = 1, ..., m$ are mean levels at each visit, $(b_{0i}, b_{1i}) \sim \mathcal{N}(0, D)$ as above are random intercepts and slopes modeling the dispersion of the longitudinal trajectories, and $e_i \sim \mathcal{N}(0, R)$ are residual errors. The alternative parameterizations of fixed effects by the two models is illustrated with a toy example in Figure 3.1.

Estimation of the parameters in (3.2) is by maximum likelihood or restricted maximum likelihood (REML). Derivation of the asymptotic variance of the fixed effect estimates for power calculation formulas proceeds analogously to the derivation of the variance of the fixed effects parameters under the random slopes model (see Chapter 2), and is presented here in outline only.

Writing (3.2) in matrix notation, we have

$$y_i = X_i \alpha + Z_i b_i + \varepsilon_i, \tag{3.3}$$

where identity matrix X_i is the fixed effects design matrix for subject i, and $Z_i = (1, t_i)$ is the random effects design matrix for subject i. More generally, X_i can include additional fixed effect covariates.

Under this model, the covariance V_i for subject i with data completion pattern t_i is

$$V_i = \mathrm{Cov}(y_i) = Z_i D Z_i' + R_i. \tag{3.4}$$

Assuming $i.i.d.$ residual error σ_ε^2, $R_i = \sigma_\varepsilon^2 I$, and the elements of V_i are a function of the pattern t_i of observations obtained for subject i, the residual error variance σ_ε^2, and the covariance parameters $\sigma_{b_0}^2$, $\sigma_{b_1}^2$, and σ_{b_0,b_1}. Specifically, V_i are matrices with off diagonal elements u, v equal to $\sigma_{b_0}^2 + (t_u + t_v)\sigma_{b_0,b_1} + t_u t_v \sigma_{b_1}^2$ and diagonal elements u, u equal to $\sigma_{b_0}^2 + 2t_u \sigma_{b_0,b_1} + t_u^2 \sigma_{b_1}^2 + \sigma_\varepsilon^2$.

Given V_i and X_i, when data are missing at random the asymptotic variance of REML estimates of the coefficients in (3.3) is

$$\mathrm{Var}(\hat{\alpha}) = \left(\sum_i (X_i' V_i^{-1} X_i) \right)^{-1}. \tag{3.5}$$

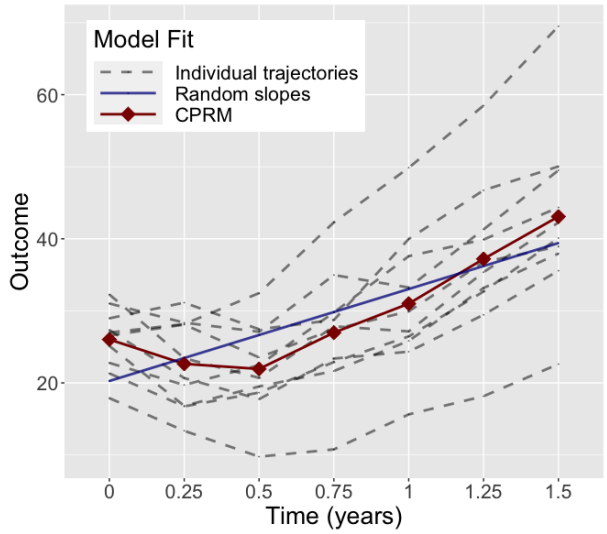


Figure 3.1. Fixed effect estimates by a repeated measures model (red) and a random slopes model (blue). The repeated measures model estimates the expect level at each time point while the random slopes model assumes a linear trajectory of fixed effect levels as a function of time.

Note that there is a finite set of missing value patterns defining X_i and V_i. Indexing these missing value patterns by k, (3.5) can be expressed as

$$\begin{aligned} \text{Var}(\hat{\alpha}) &= \left(\sum_k n_k (X_k' V_k^{-1} X_k) \right)^{-1} \\ &= \left(n \sum_k p_k (X_k' V_k^{-1} X_k) \right)^{-1}, \end{aligned} \tag{3.6}$$

where the n_k are counts of subjects in each set and sum to n, and $p_k = n_k/n$.

More generally, α and X in equations (3.2) through (3.6) can include covariate fixed effects beyond the visit level terms. (3.6) will be useful for power calculation formulas for the CPRM model (Section 3.5).

3.3 Empirical validation of the chronic progressive (CP) covariance structure for Alzheimer's disease data.

We explore the relevance of the CPRM model for modeling cognitive decline in

Alzheimer's disease using placebo arm data from representative clinical trials performed by the National Institute of Aging funded Alzheimer's Disease Cooperative Study (ADCS). Alzheimer's disease is a chronic progressive condition characterized by gradual loss of short term memory and other cognitive faculties. The primary outcome measures for Alzheimer's disease clinical trials are typically so called global cognitive measures, meaning neuropsychometric instruments querying multiple domains of cognitive function affected by the disease. We here report data from two representative clinical trials, an Alzheimer's treatment trial of vitamins to reduce homocysteine levels using the Alzheimer's Disease Assessment Scale - cognitive domain (the ADAS-cog) as the primary outcome (Quinn et al., 2010), and a prodromal Alzheimer's disease trial of vitamin E or donepezil using the Mini-Mental Status Exam (MMSE) as a secondary outcome (Petersen et al., 2005). The vitamin E arm was null in the prodromal Alzheimer trial (Petersen et al., 2005), and therefore we pooled the placebo and vitamin E arm data from this trial in analyses reported here.

Tables 3.1 and 3.2 report empirical covariance matrices and covariance matrices estimates by the CPRM model for the two trials. The empirical covariance matrices for these trials represent the covariance pattern one would expect for longitudinal trajectories that are fanning apart as a function of time. Under this pattern, we observe two phenomena. First, the variance increases over time, as reflected in the diagonal terms. Second, the covariance of neighboring observations increases over time, as reflected in the off-diagonal terms. The CPRM covariance model consistently recapitulates the empirical covariance observed in these data (Tables 3.1 and 3.2). For comparison, we have included covariance matrices for the heterogeneous CS and heterogeneous AR1 MMRM model fits to the ADCS homocysteine trial data (Table 3.1). In these data, the CS model overestimates covariance terms away from the diagonal and the AR1 model underestimates these terms. In particular, the covariance of the first and last observations, a critical component of the standard error of change first to last, is misrepresented by these models. Similar CS and AR1 covariance patterns were observed in the prodromal Alzheimer's disease trial (data not shown). We will use computer simulations (Section 3.4) to demonstrate

the effect of this model misspecification on statistical power and type I error rates.

Table 3.1. Covariance matrices estimated from the ADCS homocysteine trial (ADAS-cog data, n=330 subjects, quarterly observations for one and one half years).

Empirical covariance matrix

$$\begin{pmatrix} 68.6 & 61.8 & 60.6 & 66.0 & 67.7 & 73.9 & 78.4 \\ 61.8 & 80.4 & 67.2 & 74.8 & 75.4 & 84.5 & 89.6 \\ 60.6 & 67.2 & 79.5 & 76.9 & 77.4 & 84.6 & 93.2 \\ 66.0 & 74.8 & 76.9 & 102.7 & 91.0 & 96.1 & 106.1 \\ 67.7 & 75.4 & 77.4 & 91.0 & 104.9 & 102.8 & 112.4 \\ 73.9 & 84.5 & 84.6 & 96.1 & 102.8 & 123.7 & 123.7 \\ 78.4 & 89.6 & 93.2 & 106.1 & 112.4 & 123.7 & 155.6 \end{pmatrix}$$

Covariance assuming the CPRM model

$$\begin{pmatrix} 69.1 & 58.8 & 62.3 & 65.7 & 69.1 & 72.5 & 76.0 \\ 58.8 & 76.9 & 67.7 & 72.1 & 76.5 & 80.9 & 85.3 \\ 62.3 & 67.7 & 86.7 & 78.5 & 83.9 & 89.3 & 94.7 \\ 65.7 & 72.1 & 78.5 & 98.5 & 91.3 & 97.6 & 104.0 \\ 69.1 & 76.5 & 83.9 & 91.3 & 112.3 & 106.0 & 113.4 \\ 72.5 & 80.9 & 89.3 & 97.6 & 106.0 & 128.0 & 122.8 \\ 76.0 & 85.3 & 94.7 & 104.0 & 113.4 & 122.8 & 145.8 \end{pmatrix}$$

Covariance assuming heterogeneous CS

$$\begin{pmatrix} 72.7 & 64.4 & 63.5 & 72.2 & 72.2 & 78.2 & 89.2 \\ 64.4 & 81.6 & 67.2 & 76.5 & 76.5 & 82.9 & 94.5 \\ 63.5 & 67.2 & 79.2 & 75.4 & 75.3 & 81.6 & 93.1 \\ 72.2 & 76.5 & 75.4 & 102.5 & 85.7 & 92.9 & 105.9 \\ 72.2 & 76.5 & 75.3 & 85.7 & 102.4 & 92.8 & 105.9 \\ 78.2 & 82.9 & 81.6 & 92.9 & 92.8 & 120.2 & 114.7 \\ 89.2 & 94.5 & 93.1 & 105.9 & 105.9 & 114.7 & 156.3 \end{pmatrix}$$

Covariance assuming heterogeneous AR1

$$\begin{pmatrix} 78.1 & 73.6 & 62.1 & 58.8 & 49.3 & 45.5 & 44.6 \\ 73.6 & 92.1 & 77.7 & 73.5 & 61.7 & 57.0 & 55.8 \\ 62.1 & 77.7 & 87.1 & 82.4 & 69.1 & 63.9 & 62.6 \\ 58.8 & 73.5 & 82.4 & 103.4 & 86.7 & 80.1 & 78.6 \\ 49.3 & 61.7 & 69.1 & 86.7 & 96.6 & 89.2 & 87.5 \\ 45.5 & 57.0 & 63.9 & 80.1 & 89.2 & 109.5 & 107.3 \\ 44.6 & 55.8 & 62.6 & 78.6 & 87.5 & 107.3 & 139.6 \end{pmatrix}$$

3.4 Performance of the CPRM model

We use computer simulations to characterize the performance of the CPRM model relative to other models under null and alternative scenarios. We consider MMRM analyses assuming unstructured, compound symmetry (standard and heterogeneous), and first order autoregressive

Table 3.2. Covariance matrices estimated from the ADCS clinical trial of vitamin E and donepezil (MMSE data, n = 510, biannual observations for three years).

Empirical covariance matrix

$$\begin{pmatrix} 3.46 & 2.08 & 2.00 & 2.52 & 2.30 & 2.76 & 3.20 \\ 2.08 & 4.88 & 2.92 & 3.97 & 3.52 & 4.34 & 5.18 \\ 2.00 & 2.92 & 5.41 & 4.47 & 4.22 & 4.91 & 6.25 \\ 2.52 & 3.97 & 4.47 & 9.13 & 6.15 & 8.19 & 10.02 \\ 2.30 & 3.52 & 4.22 & 6.15 & 8.87 & 8.25 & 9.89 \\ 2.76 & 4.34 & 4.91 & 8.19 & 8.25 & 13.44 & 14.55 \\ 3.20 & 5.18 & 6.25 & 10.02 & 9.89 & 14.55 & 21.69 \end{pmatrix}$$

Covariance assuming the CPRM model

$$\begin{pmatrix} 5.13 & 2.20 & 2.50 & 2.81 & 3.12 & 3.42 & 3.73 \\ 2.20 & 6.12 & 3.56 & 4.24 & 4.93 & 5.61 & 6.29 \\ 2.50 & 3.56 & 7.86 & 5.68 & 6.74 & 7.80 & 8.85 \\ 2.81 & 4.24 & 5.68 & 10.36 & 8.55 & 9.98 & 11.42 \\ 3.12 & 4.93 & 6.74 & 8.55 & 13.60 & 12.17 & 13.98 \\ 3.42 & 5.61 & 7.80 & 9.98 & 12.17 & 17.60 & 16.54 \\ 3.73 & 6.29 & 8.85 & 11.42 & 13.98 & 16.54 & 22.34 \end{pmatrix}$$

(standard and heterogeneous) covariance structures. We also consider the random slopes analysis in the simulations. MMRM models were fit using the gls function within the R nlme package (Pinheiro et al., 2021). Random slopes and CPRM models were fit using the lme function within the same package. The random slopes models were constrained to have a single fixed effect intercept shared by both groups as recommended for randomized clinical trials (Hu et al. 2021).

3.4.1 Simulations under the null

Longitudinal repeat measures were generated following a CPRM model using covariance and residual error variance parameters estimated from the placebo arm of the ADCS homocysteine trial (Quinn et al., 2010), and using fixed effects means under the hypothetical scenario of a treatment with short term palliative effect that washes out by the end of the study period (Figure 3.2, top panel). A total of 10,000 simulated samples were performed (n=80 per group, 18 month trial with quarterly observations, and a nominal p-value for hypothesis testing of 0.05).

Type I error rate estimates under the different model fits are listed in Table 3.3. The CPRM model and unstructured MMRM model meet the nominal five percent type I error rate to within the accuracy of simulations. The compound symmetry and heterogeneous compound symmetry MMRM models had type I error rates of 13.4 percent and 9.6 percent respectively, meaning these two models are invalid and not appropriate for data that follow the chronic progressive pattern. The mixed effects model with random slopes was likewise anticonservative, with a type I error rate of nearly 15 percent (Table 3.3). This result clearly illustrates the concern of regulatory agencies that analyses imposing assumptions about the shape of individual trajectory, such as the random slopes model with linear trajectories illustrated here, can result in positive trial findings even when the treatment has no persistent efficacy. The AR1 and heterogeneous AR1 models (Table 3.3) had type I error rates much less than 0.05 (i.e., were substantively conservative).

3.4.2 Simulations under the alternative

We next simulated data following the CPRM model as above, but under the alternative scenario depicted in Figure 3.2, bottom panel, and with an effect size chosen to ensured an expected power of 80 percent under the CPRM analysis. The unstructured MMRM and the CPRM models acheived the expected 80 percent power, while the AR1 MMRM models were substantially underpowered (Table 3.3). We do not report power for the compound symmetry models because type I error rates for these models are substantially greater than 0.05, meaning these models are invalid for chronic progressive longitudinal data. Likewise we do not report power for the random slopes model based on the findings from the simulations under the null.

3.5 Sample size calculations for the CPRM model

Derivation of sample size formulas for the CPRM model follows directly from derivations for the random slopes model (see Chapter 2) and is reported here in outline only. Power is a function of the sample size in each arm, the covariance of repeated measures in each arm, the study design (the study length and interval between followups), the missing data pattern, and the effect size.

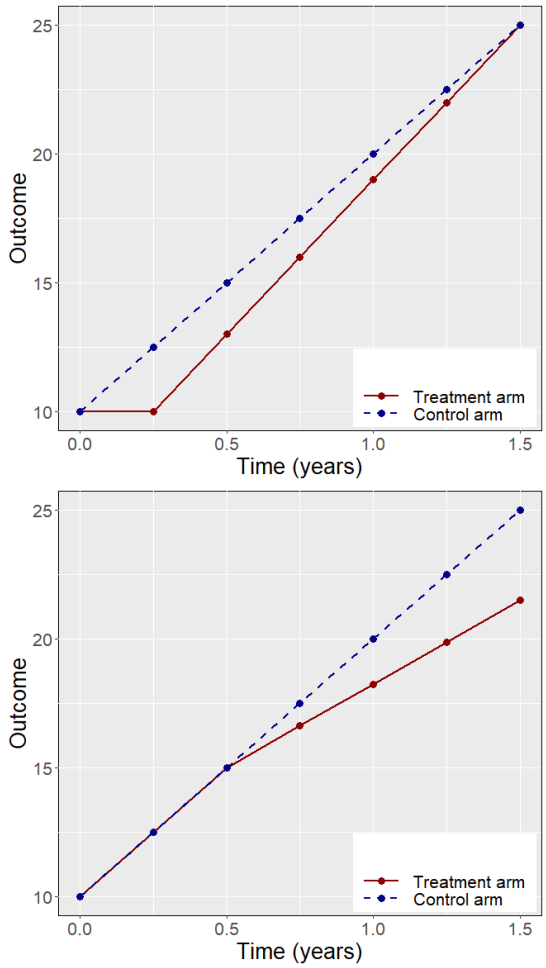


Figure 3.2. Top panel: Hypothetical trajectory of expected values by arm for a treatment with short term efficacy but no difference in treatment versus control at the end of the trial. The random slopes model is prone to spurious type I error under this scenario (Table 3.3). Bottom panel: Hypothetical trajectory of expected values for a treatment with positive effect starting after the third observation and persisting to the end of the trial.

Table 3.3. Type I error rate under the null (Figure 3.2, Top panel), and power under the alternative (Figure 3.2, Bottom panel). (10,000 simulations each, with effect size under the alternative chosen to achieve 80% power for the CPRM model.)

	α error rate	Power
CPRM	0.0536	0.7981
Random slopes	0.1499	-
MMRM, CS	0.1343	-
MMRM, hetCS	0.0955	-
MMRM, AR1	0.0069	0.5187
MMRM, hetAR1	0.0058	0.4997
MMRM, UN	0.0539	0.7989

CS = compound symmetric; hetCS = heterogeneous CS;
AR1 = autoregressive; hetAR1 = heterogeneous AR1;
UN = unstructured.

To simplify presentation, we begin by describing power formulas for the common circumstance of equal allocation to arms and equivalent covariance structure in the two arms. In this case, the variance of change first to last visit in each arm is $\text{Var}(\hat{\alpha_m} - \hat{\alpha_1})$, and the sample size required to detect a difference Δ in change scores between arms at last visit with power $1-\beta$ and type I error rate α is given by the familiar formula

$$N/\text{Arm} = 2(z_{\alpha/2} + z_\beta)^2 \sigma^2 / \Delta^2. \tag{3.7}$$

Two power calculation approaches are common when sizing a clinical trial, a conservative estimate of required sample size informed by the power of a completers-only analysis, and a less conservation estimate that explicitly adjusts for the anticipated missing data pattern to be obtained by the trial. We describe each of these in turn.

3.5.1 Completers-only approach

A conservative approach is to determine the sample size required to power a completers-only analysis and then increase the sample size to ensure this many subjects complete the trial. This method has the advantages of relying on a straightforward power calculation formula and

resulting in statistically conservative sample size estimates. For completers, there are no missing data and X_i and Z_i are equivalent full matrices for all subjects so that

$$\text{Var}(\hat{\alpha_m} - \hat{\alpha_1}) = \frac{2\sigma_\varepsilon^2 + (t_m - t_1)^2\sigma_{b1}^2}{n}. \tag{3.8}$$

See the Appendix for derivation of this result. (3.7) then reduces to

$$N_{\text{completers}}/\text{Arm} = 2(z_{\alpha/2} + z_\beta)^2[2\sigma_\varepsilon^2 + (t_m - t_1)^2\sigma_{b1}^2]/\Delta^2. \tag{3.9}$$

If p_m is the proportion of subjects who will complete the trial, then setting total N/Arm to $N_{completers}/p_m$ will ensure an expected $N_{completers}$ complete the planned trial.

3.5.2 Study subject attrition approach

Alternatively, one can use (3.6) to directly account for the anticipated dropout pattern expected in a study. Setting $W = n\text{Var}(\hat{\alpha})$, under equal allocation to study arm and assuming equivalent repeated measures covariance across arms, the sample size required to detect treatment effect Δ with power $1-\beta$ and type I error rate α is

$$N/\text{Arm} = 2(z_{\alpha/2} + z_\beta)^2 \left(W_{mm} + W_{11} - 2W_{1m}\right)/\Delta^2. \tag{3.10}$$

As a practical matter, investigators restrict to the m missing data patterns determined by study subject dropout (Lu et al., 2008). Given $i.i.d.$ residual error variance σ_ε^2, W and by extension (3.10) are simple linear functions of the variance parameters σ_ε^2, $\sigma_{b_0}^2$, $\sigma_{b_1}^2$, and σ_{b_0,b_1}, and the design vector t (see Appendix).

3.5.3 Implementation

Formula (3.9) and the study subject attrition approach formula (3.10) under the usual assumption of $i.i.d$ residual error are implemented in the cprm.power function within the R package longpower (Donohue, 2019). Generalizing these formulas to allow unequal allocation

and to the case where the covariance structure is different in the two groups is straightforward (Zhao and Edland, 2021), and has also been implemented for the CPRM model in the cprm.power function. Different covariance structures may be anticipated across groups. For example, in clinical trials a greater variance of change may occur within the treatment arm because the change observed in the treatment arm reflects both normal background variability in change and the variability in response to treatment (Mallinckrodt et al., 2008). The formulas provided in cprm.power can be used to perform sensitivity analyses of the potential magnitude of this effect on trial power.

Note that the variance parameters for these formulas can be estimated from prior data of arbitrary design. Stated differently, if we have variance parameter estimates from pilot studies or prior trials, we can use these values to power a future trial of arbitrary design (with arbitrary number and interval between followup visits). Furthermore, under CPRM these power calculation formulas are appropriate for trials using MMRM with unstructured covariance as the primary a priori analysis plan.

3.5.4 Validation of sample size formulas by computer simulation

We used computer simulations to evaluate the performance of (3.7). We simulated data following the random slopes model (3.1) using parameters estimated from the ADCS homocysteine trial as in Section 3.4, assuming an 18 month trial with quarterly observations and a 25% shift in mean change in treatment versus control. Power observed in simulations closely matches predicted power (Figure 3.3).

3.6 Discussion

We have introduced a novel parsimonious parameterization of the covariance structure of longitudinal repeated measures appropriate for chronic progressive conditions. We have demonstrated that alternative parsimonious parameterizations typically used in MMRM analysis are not appropriate for this pattern of longitudinal data that fan apart over time. In application, the

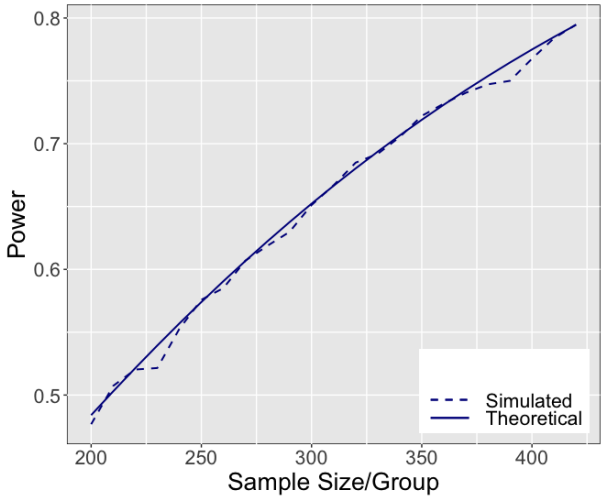


Figure 3.3. Theoretical power curve versus power estimated by computer simulation (10,000 simulations per sample size, two-sided test, type I error $\alpha = 0.05$).

MMRM analysis assuming compound symmetry is anticonservative, and the MMRM analysis assuming AR1 is underpowered for this type of data, while the CPRM analysis is both valid (maintaining its nominal type I error rate) and has equivalent power to MMRM with unstructured covariance when applied to chronic progressive data.

Further, we have derived power calculation formulas for the CPRM model that are independent of pilot study design. This is helpful when the design of pilot studies available to inform power calculations does not match the design of the future trial being powered.

Like MMRM, the CPRM model has the heuristic advantage of testing treatment efficacy based explicitly on differences in response at the end of the trial period without any assumptions about the shape of mean trajectories of response over time. Model results are therefore unambiguous and easier to describe to a non-technical audience. We illustrated with computer simulations the concern of regulators and clinical trialists that false positive findings are possible under different analysis plans (Figure 3.3 and Table 3.3). Recent advances in analytic methods, including natural cubic spline (Donohue et al., 2023) and progression repeated measures (Raket, 2022) models, may be less susceptible to type I error concerns, and this is an area of future research. However, the cost of using MMRM in terms of statistical efficiency is counterbal-

anced by the unambiguous characterization of treatment affects by this approach, and MMRM remains the de facto standard for Alzheimer's disease treatment trials (van Dyck et al., 2023; Budd Haeberlein et al., 2022).

The suitability of CPRM for data beyond the longitudinal Alzheimer data considered here will have to be examined on an individual basis. However, we note that the CPRM model assumptions hold for any scenario where the mixed effects model with random slopes is appropriate, so applications of CPRM are equally as broad as this common analytic approach.

3.7 Acknowledgement

Data made available by the Alzheimer's Disease Cooperative Study, NIA U19 AG10483. This work was supported by the following NIA grants: R01 AG049810 (Bondi/Edland - Co-PI); R01 AG066088 (Banks – PI); and P30 AG062429 (Brewer - PI).

Chapter 3, in full, has been submitted for publication of the material as it may appear in Clinical Trials. Zhao, Yu; Edland, Steven. The chronic progressive repeated measures (CPRM) model for longitudinal data. The dissertation author was the primary investigator and author of this material.

Chapter 4

An approximation to peak detection power using Gaussian random field theory

4.1 Introduction

Detection of peaks (local maxima) is an important topic in image analysis. For example, a fundamental goal in fMRI analysis is to identify the local hotspots of brain activity (see, for example, Genovese et al., 2002; Heller et al., 2006), which are typically captured by peaks in the fMRI signal. The detection of such peaks can be posed as a statistical testing problem intended to test whether the underlying signal has a peak at a given location. This is challenging because such tests are conducted only at locations of observed peaks, which depend on the data. Therefore, the height distribution of the observed peak is conditional on a peak being observed at that location. This is a nonstandard problem. Solutions exist using random field theory (RFT). RFT is a statistical framework that can be used to perform topological inference and modeling. RFT-based peak detection has been studied in Cheng and Schwartzman (2017) and Schwartzman and Telschow (2019b), which provide the peak height distribution for isotropic noise under the complete null hypothesis of no signal anywhere. However, particular challenges arise when we perform power calculations in peak detection settings. Due to the nature of imaging data, the number and location of the signal peaks are unknown. Besides, the power is affected by other spatial aspects of the problem, such as the shape of the peak function and the spatial autocorrelation of the noise. Considering these difficulties, it requires some extra effort to derive

a power formula for peak detection.

A formal definition of power in peak detection is necessary to perform power calculations. In Cheng and Schwartzman (2017) and Durnez et al. (2016), the authors explored approaches to control the false discovery rate (FDR). For the entire domain, average peakwise power, i.e. power averaged over all non-null voxels, is a natural choice for these approaches. For a local domain where a single peak exists, the power can be defined as the probability of successfully detecting that peak. Following this idea, we describe the null and alternative hypothesis and the definition of detection power. We do so informally here for didactic purposes and present formal definitions in Section 4.2.

Consider a local domain where a single peak may exist, and consider the hypotheses

H_0 : "the signal is equal to 0 in the local domain." vs

H_1 : "the signal has at least one positive peak in the local domain."

Suppose we observe a random field to be used as test statistic at every location, typically as the result of statistical modeling of the data. For a fixed threshold u, the existence of observed peaks with height greater than u would lead to rejecting the null hypothesis. Therefore, we define the type I error and power as the probability of existing at least one local maximum above u under H_0 and H_1 respectively:

$$\begin{aligned} &\text{Type I error: } \mathbb{P}\{\exists \text{ a peak in the local domain with height} > u \text{ when } H_0 \text{ is true}\} \\ &\text{Power: } \mathbb{P}\{\exists \text{ a peak in the local domain with height} > u \text{ when } H_1 \text{ is true}\} \end{aligned} \tag{4.1}$$

Formulas for type I error have been developed for stationary fields in 1D and isotropic fields in 2D and 3D (Cheng and Schwartzman, 2015; Cheng and Schwartzman, 2017; Cheng and Schwartzman, 2018). However, there is no formula to calculate power. In order to get an appropriate estimate of power, we need to know the peak height distribution for non-centered

(the mean function is not 0) random fields. Generally speaking, it is very difficult to calculate the peak height distribution especially when the random field has non-zero mean. Durnez et al. (2016) suggests using Gaussian distribution to describe the non-null peaks and truncated Gaussian distribution to approximate the overshoot distribution. This approach is easy to implement but not very accurate because the peak height distribution is in reality always skewed and not close to any Gaussian distribution.

In this chapter, we propose to approximate the probability of an observed peak exceeding the detection threshold u by calculating the expected number of peaks above the threshold u. We show that the approximation, which is also an upper bound, works well under certain scenarios. For the entire domain, we can approximate the average peakwise power by taking the arithmetic mean of the approximation proposed in this chapter over non-null voxels.

The proposed approximation makes the problem more tractable, but in general, it does not have an explicit form. In order to make it applicable in practice, we further simplify the formula under the isotropy assumption and show its explicit form in 1D, 2D, and 3D. The explicit results are validated through 2D and 3D computer simulations carried out in MATLAB. The simulation also covers multiple scenarios by modifying the parameters used to generate the data. The performance of power approximation and its conservative adjustment under these scenarios are discussed.

Finally, to assess the real-data performance of our power approximation method, we apply it to a 3D simulation induced by a real brain imaging dataset, where the parameters are estimated from the Human Connectome Project (Van Essen et al., 2012) fMRI data. By testing the method in a realistic setting, we also demonstrate how effect size and other parameters affect the power.

4.2 Power approximation

4.2.1 Setup

Let $Y(s) = \sigma(s)Z(s) + \mu(s)$ where $Z = \{Z(s), s \in D\}$ representing the noise is a centered (zero-mean) smooth unit-variance Gaussian random field on an N dimensional non-empty domain $D \subset \mathbb{R}^N$, $\sigma(s)$ is the standard deviation of the noise and $\mu(s)$ is the mean function. Let $X(s) = Y(s)/\sigma(s) = Z(s) + \theta(s)$ where the ratio $\theta(s) = \mu(s)/\sigma(s)$ is the standardized mean function, which we assume to be C^2. Here C^3 is a sufficient smoothness condition for Z, and this will be clarified in Assumption 1 below.

Let

$$
\begin{aligned}
X_i(s) &= \frac{\partial X(s)}{\partial s_i}, \quad \nabla X(s) = (X_1(s), \ldots, X_N(s)), \\
X_{ij}(s) &= \frac{\partial^2 X(s)}{\partial s_i s_j}, \quad \nabla^2 X(s) = (X_{ij}(s))_{1 \le i,j \le N}, \\
Z_i(s) &= \frac{\partial Z(s)}{\partial s_i}, \quad \nabla Z(s) = (Z_1(s), \ldots, Z_N(s)), \\
Z_{ij}(s) &= \frac{\partial^2 Z(s)}{\partial s_i s_j}, \quad \nabla^2 Z(s) = (Z_{ij}(s))_{1 \le i,j \le N}.
\end{aligned}
$$

We will make use of the following assumptions:

Assumption 1. *$Z \in C^2(D)$ almost surely and its second derivatives satisfy the mean-square Hölder condition: for any $s_0 \in D$, there exists positive constants L, η and δ such that*

$$
\mathbb{E}[Z_{ij}(s) - Z_{ij}(t)]^2 \le L^2 \|s - t\|^{2\eta}, \quad \forall t, s \in U_{s_0}(\delta),\ i, j = 1, ..., N,
$$

where $U_{s_0}(\delta) = s_0 \oplus (-\delta/2, \delta/2)^N$ is the N dimensional open cube of side length δ centered at s_0. This condition is satisfied, for example, if Z is $C^3(D)$.

Assumption 2. *For every pair $(t,s) \in D \times D$ with $s \neq t$, the Gaussian random vector*

$$(Z(s), \nabla Z(s), Z_{ij}(s), Z(t), \nabla Z(t), Z_{ij}(t), 1 \leq i \leq j \leq N)$$

is non-degenerate, i.e. its covariance matrix has full rank.

4.2.2 Peak detection

Following the notation in the problem setup, the null and alternative hypothesis can be written as:

$$H_0 : \mu(s) = 0 \text{ for all } s \in D \quad \text{vs}$$
$$H_1 : \mu(s) > 0, \nabla\mu(s) = 0, \nabla^2\mu(s) \prec 0 \text{ for some } s \in D$$

The mean function $\mu(s)$ is not directly observed, so the hypothesis is tested based on the peak height of $X(s)$. For a peak detection procedure that aims to test this hypothesis, a threshold u for the peak height of $X(s)$ needs to be set in advance. If a local maximum with height greater than u is observed, we would choose to reject the null hypothesis due to the strong evidence against it. The probability that a peak of X exceeds u

$$\mathbb{P}\big(\exists\, s \in D\, s.t.\, X(s) > u | \nabla X(s) = 0 \text{ and } \nabla^2 X(s) \prec 0\big) \tag{4.2}$$

is the type I error under H_0 and power under H_1. The threshold u can be obtained based on the peak height distribution under H_0. A formula for peak height distribution of smooth isotropic Gaussian random fields has been derived in Cheng and Schwartzman (2018) and it can also be derived directly from a special case of the formulas presented in this chapter. Usually, u is set to be some quantile of the null distribution of peak height to maintain the nominal α type I error. More details about selecting the threshold will be discussed in the real data example. Selecting u is not the main focus of this Chapter and our method can be applied to any choice of u.

4.2.3 Power approximation

Let M_u be the number of local maxima of the random field X above u over the local domain D. The power defined in (4.2) can be represented as $\mathbb{P}[M_u \geq 1]$. We call this the *power function*, seen as a function of the threshold u. Note that

$$\mathbb{P}[M_u \geq 1] = \sum_{k=1}^{\infty} \mathbb{P}[M_u = k] \leq \sum_{k=1}^{\infty} k\mathbb{P}[M_u = k] = \mathbb{E}[M_u]. \tag{4.3}$$

On the other hand,

$$\mathbb{E}[M_u] - \mathbb{P}[M_u \geq 1] = \sum_{k=2}^{\infty} (k-1)\mathbb{P}[M_u = k] \leq \frac{1}{2}\sum_{k=2}^{\infty} k(k-1)\mathbb{P}[M_u = k] = \frac{1}{2}\mathbb{E}[M_u(M_u-1)]. \tag{4.4}$$

Thus, we have

$$\mathbb{E}[M_u] - \frac{1}{2}\mathbb{E}[M_u(M_u-1)] \leq \mathbb{P}[M_u \geq 1] \leq \mathbb{E}[M_u]. \tag{4.5}$$

This inequality tells us that for any fixed u, the power is bounded within an interval of length $\mathbb{E}[M_u(M_u-1)]/2$. Thus, $\mathbb{E}[M_u]$ is a good approximation of power if one of the two conditions below is satisfied:

1. The factorial moment $\mathbb{E}[M_u(M_u-1)]$ converges to 0 and $\mathbb{E}[M_u]$ does not.

2. They both converge to 0 and $\mathbb{E}[M_u(M_u-1)]$ converges faster than $\mathbb{E}[M_u]$.

The convergence above refers to conditions on the signal and noise parameters. In the rest of this section, we introduce four interesting results. The first result can be useful for simplifying the power function and the other three results give different scenarios where one of the conditions above holds.

4.2.4 Adjusted $\mathbb{E}[M_u]$

We have provided evidence of using $\mathbb{E}[M_u]$ to approximate power through (4.5). However, $\mathbb{E}[M_u]$ alone might not be sufficient for power approximation since it only gives an upper bound.

Also, unlike power, $\mathbb{E}[M_u]$ sometimes exceeds one. To correct for this, we define the adjusted $\mathbb{E}[M_u]$ as

$$\mathbb{E}[M_u]_{\text{adj}} = \mathbb{E}[M_u] / \max(1, \mathbb{E}[M_{-\infty}]). \tag{4.6}$$

The adjusted $\mathbb{E}[M_u]$ is the same as $\mathbb{E}[M_u]$ when the expected number of local maxima $\mathbb{E}[M_{-\infty}]$ is less or equal to 1. When $\mathbb{E}[M_{-\infty}]$ is greater than 1, we divide $\mathbb{E}[M_u]$ by $\mathbb{E}[M_{-\infty}]$ to make sure it never exceeds one. The adjusted $\mathbb{E}[M_u]$ is more conservative, and we conjecture that it is a lower bound of power when there exists at least one local maximum in the domain D. In applications, people are interested in a conservative estimator so that the test is guaranteed to have enough power. Combining $\mathbb{E}[M_u]$ and $\mathbb{E}[M_u]/\mathbb{E}[M_{-\infty}]$, we can get an approximate range of the true power. We will compare $\mathbb{E}[M_u]$ and adjusted $\mathbb{E}[M_u]$ in simulation studies.

4.2.5 Height equivariance

Our first result does not concern the approximation (4.5) yet, but it offers a simplification of the power function and $\mathbb{E}[M_u]$ that will be used later. The proposition below states that the power function and $\mathbb{E}[M_u]$ for peak detection are translation equivariant with respect to peak height.

Proposition 1. *Let $\theta(s) = h(s) + \theta_0$ be a peak signal with height θ_0, where $h(s)$ is a unimodal mean function with maximum equal to 0 at s_0 in D. Then the power function for peak detection and $\mathbb{E}[M_u]$ can be written in the form $F(u-\theta_0)$, where $F(u)$ is the power function or $\mathbb{E}[M_u]$ at $\theta_0 = 0$.*

Next, we give three scenarios where the equality in (4.5) can be achieved asymptotically: small domain size, large threshold, and sharp signal.

4.2.6 Small domain

If the size of the local domain D where a single peak exists is small enough, it can be shown that equality in (4.5) can be achieved asymptotically.

Theorem 1. *Consider a local domain $D_\varepsilon = U(s_0, \varepsilon)$ for any fixed $s_0 \in D$ where $U(s_0, \varepsilon) = s_0 \oplus (-\varepsilon/2, \varepsilon/2)^N$ is the N-dimensional open cube of side ε centered at s_0. For sufficiently small ε and fixed threshold u,*

$$\mathbb{P}[M_u \geq 1] = \mathbb{E}[M_u](1 - o(1)) = \mathbb{E}[M_u]_{\text{adj}}(1 - o(1)). \tag{4.7}$$

4.2.7 Large threshold

For large threshold u, the following asymptotic result shows power can be precisely approximated by $\mathbb{E}[M_u]$.

Theorem 2. *For any fixed domain D, as $u \to \infty$*

$$\mathbb{P}[M_u \geq 1] = \mathbb{E}[M_u](1 - o(e^{-\alpha u^2})), \tag{4.8}$$

where the error term $o(e^{-\alpha u^2})$ is non-negative and $\alpha > 0$ is some constant.

Notice that the threshold u does not affect the value of $\mathbb{E}[M_{-\infty}]$ which is part of the adjusted $\mathbb{E}[M_u]$. By (4.8)

$$\mathbb{P}[M_u \geq 1] = \mathbb{E}[M_u]_{\text{adj}}(1 - o(e^{-\alpha u^2})) \max(1, \mathbb{E}[M_{-\infty}]).$$

If $\mathbb{E}[M_{-\infty}] > 1$, the adjusted $\mathbb{E}[M_u]$ might be overly conservative for large threshold u. Therefore, we only recommend $\mathbb{E}[M_u]$ for this scenario.

4.2.8 Sharp signal

The following theorem provides an asymptotic power approximation when the signal is sharp. Interestingly, while the power function is generally non-Gaussian, it becomes closer to Gaussian as the signal peaks become sharper.

Theorem 3. *Let $\theta(s) = ah(s) + \theta_0$ where $h(s)$ is a unimodal mean function with maximum equal to 0 at s_0, $a > 0$, and θ_0 represents the height. For any fixed threshold u, as $a \to \infty$*

$$\mathbb{P}[M_u \geq 1] = \mathbb{E}[M_u] + o(1) = \mathbb{E}[M_u]_{\text{adj}} + o(1) = \Phi(\theta_0 - u)(1 + o(1)), \tag{4.9}$$

where $\Phi(x)$ is CDF of the standard Gaussian distribution.

4.3 Explicit formulas

We have showed that the power for peak detection can be approximated by the expected number of local maxima above u, $\mathbb{E}[M_u]$, under certain scenarios such as small domain and large threshold. Although we can apply the Kac-Rice formula to calculate $\mathbb{E}[M_u]$, it remains difficult to evaluate it explicitly for $N > 1$ without making any further assumptions. In this section, we focus on computing $\mathbb{E}[M_u]$ and show a general formula can be obtained if the noise field is isotropic. Furthermore, explicit formulas when $N = 1, 2, 3$ are derived for application purposes.

4.3.1 Isotropic Gaussian fields

Suppose Z is a zero-mean unit-variance isotropic random field. We can write the covariance function of Z as $\mathbb{E}\{Z(s)Z(t)\} = \rho(\|s-t\|^2)$ for an appropriate function $\rho(\cdot) : [0, \infty) \to \mathbb{R}$. Denote

$$\rho' = \rho'(0), \quad \rho'' = \rho''(0), \quad \kappa = -\rho'/\sqrt{\rho''}, \tag{4.10}$$

where ρ' and ρ'' are first and second derivative of function ρ respectively.

The following lemma comes from Cheng and Schwartzman (2018).

Lemma 1. *For each $s \in \mathbb{R}^N$ and $i, j, k, l \in \{1, \ldots, N\}$,*

$$\begin{aligned}
\mathbb{E}\{Z_i(s)Z(s)\} &= \mathbb{E}\{Z_i(s)Z_{jk}(s)\} = 0, \\
\mathbb{E}\{Z_i(s)Z_j(s)\} &= -\mathbb{E}\{Z_{ij}(s)Z(s)\} = -2\rho'\delta_{ij}, \\
\mathbb{E}\{Z_{ij}(s)Z_{kl}(s)\} &= 4\rho''(\delta_{ij}\delta_{kl} + \delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}),
\end{aligned}$$

where ρ' and ρ'' are defined in (4.10) *and δ_{ij} is the Kronecker delta function.*

In particular, it follows from Lemma 1 that $\mathrm{Var}(Z_i(s)) = -2\rho'$ and $\mathrm{Var}(Z_{ii}(s)) = 12\rho''$ for any $i \in \{1,\dots,N\}$, implying $\rho' < 0$ and $\rho'' > 0$ and hence $\kappa > 0$.

We can use theoretical results from Gaussian Orthogonally Invariant (GOI) matrices to make the calculation of $\mathbb{E}[M_u]$ easier. GOI matrices were first introduced in Schwartzman et al. (2008), and used for the first time in the context of random fields in Cheng and Schwartzman (2018). It is a class of Gaussian random matrices that are invariant under orthogonal transformations, and can be useful for computing the expected number of critical points of isotropic Gaussian fields. We call an $N \times N$ random matrix $G = (G_{ij})_{1 \le i,j \le N}$ GOI with covariance parameter c, denoted by $\mathrm{GOI}(c)$, if it is symmetric and all entries are centered Gaussian variables such that

$$\mathbb{E}[G_{ij}G_{kl}] = \frac{1}{2}(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}) + c\delta_{ij}\delta_{kl}. \tag{4.11}$$

The following lemma is Lemma 3.4 from Cheng and Schwartzman (2018).

Lemma 2. *Let the assumptions in Lemma 1 hold. Let $\widetilde{G}$ and G be* $\mathrm{GOI}(1/2)$ *and* $\mathrm{GOI}((1-\kappa^2)/2)$ *matrices respectively. I_N denotes $N \times N$ identity matrix.*

(i) The distribution of $\nabla^2 Z(s)$ is the same as that of $\sqrt{8\rho''}\widetilde{G}$.

(ii) The distribution of $(\nabla^2 Z(s) | Z(s) = z)$ is the same as that of $\sqrt{8\rho''}\left[G - \left(\kappa z/\sqrt{2}\right) I_N\right]$.

Lemma 2 shows the distribution and conditional distribution of the Hessian matrix of a centered random field $Z(s)$. Next, we establish the corresponding result for non-centered random field $X(s) = Z(s) + \theta(s)$.

Lemma 3. *Let $\widetilde{G}$ and G be* $\mathrm{GOI}(1/2)$ *and* $\mathrm{GOI}((1-\kappa^2)/2)$ *matrices respectively.*

(i) The distribution of $\nabla^2 X(s)$ is the same as that of

$$\sqrt{8\rho''}\widetilde{G} + \nabla^2\theta(s).$$

(ii) The distribution of $(\nabla^2 X(s)|X(s)=x)$ *is the same as that of*

$$\sqrt{8\rho''}\left[G-\frac{\kappa(x-\theta(s))}{\sqrt{2}}I_N\right]+\nabla^2\theta(s).$$

4.3.2 General formula under isotropy

Theorem 4. *Let* $X(s)=Z(s)+\theta(s)$*, where* $Z(s)$ *is a smooth zero-mean unit-variance isotropic Gaussian random field satisfying Assumption 1, 2. Let* $\theta(s)$ *a smooth* C^3 *mean function such that* $\nabla^2\theta(s)$ *is a non-singular matrix with ordered eigenvalues* $\theta_1''(s)...\theta_N''(s)$ *at all critical points s. Then for any domain D*

$$\mathbb{E}[M_u]=\left(\frac{2\rho''}{-\pi\rho'}\right)^{N/2}\int_D e^{\frac{\|\nabla\theta(s)\|^2}{4\rho'}}\int_u^\infty \phi\left(x-\theta(s)\right)\mathbb{E}\left[|\det(\text{Matrix}(s))|\,\mathbb{1}_{\{\text{Matrix}(s)\prec 0\}}\right]dx\,ds, \tag{4.12}$$

where $\phi(x)$ *is the PDF of the standard Gaussian distribution,* $\text{Matrix}(s)=G-\kappa(x-\theta(s))I_N/\sqrt{2}+\text{diag}\{\theta_1''(s),\ldots,\theta_N''(s)\}/\sqrt{8\rho''}$*, G as in Lemma 3 represents GOI((1-*κ^2*)/2), and* $\mathbb{1}_{\{.\}}$ *denotes the indicator function.*

The expression (4.12) can be simplified further if we further assume the mean function $\theta(s)$ to be a rotationally symmetric paraboloid centered at s_0. In this case, the Hessian of $\theta(s)$ is the identity matrix multiplied by a constant, i.e.

$$\theta''=\theta_1''(s)=\theta_2''(s)=...=\theta_N''(s).$$

Then we can write the mean function as $\theta(s)=\theta_0+\theta''\|s-s_0\|^2/2$. Define

$$\eta=\frac{\theta''}{2\kappa\sqrt{\rho''}}=\frac{\theta''}{-2\rho'}=\frac{\theta''}{\text{Var}(Z_1(s))}, \tag{4.13}$$

and

$$H(\tilde{x})=\mathbb{E}^N_{\text{GOI}((1-\kappa^2)/2)}\left[\prod_{j=1}^N\left|\lambda_j-\frac{\kappa\tilde{x}}{\sqrt{2}}\right|\mathbb{1}_{\{\lambda_N<\frac{\kappa\tilde{x}}{\sqrt{2}}\}}\right]. \tag{4.14}$$

$\mathbb{E}[M_u]$ can be simplified as

$$\mathbb{E}[M_u] = \left(\frac{2\rho''}{-\pi\rho'}\right)^{N/2} \int_D e^{\frac{\theta''^2\|s-s_0\|^2}{4\rho'}} \int_{\tilde{u}(s)}^{\infty} \phi\left(\tilde{x}+\eta\right) H(\tilde{x}) d\tilde{x} ds, \tag{4.15}$$

where we make a change of variable $\tilde{x} = x - \theta(s) - \eta$ and $\tilde{u}(s) = u - \theta(s) - \eta$. Note that the parameter κ depends on the correlation structure of $Z(s)$.

4.3.3 Explicit formulas in 1D, 2D and 3D

In (4.15), a general formula for $\mathbb{E}[M_u]$ under isotropy was derived. To make the formula easier to apply in practice, we have the following results for computing it in 1D, 2D, and 3D. When $N = 1$, the derivation is simple enough that we do not need additional assumptions on the mean function $\theta(s)$ except those in Theorem 4, and it follows directly from Kac-Rice formula. When $N = 2$ and 3, we assume the mean function $\theta(s)$ is a rotationally symmetric paraboloid centered at s_0. $\mathbb{E}[M_u]$ is calculated by first obtaining explicit formulas for $H(\tilde{x})$, and plugging H into (4.15).

Proposition 2. *Let $N = 1$, $X(s) = Z(s) + \theta(s)$, where $Z(s)$ is a smooth zero-mean unit-variance Gaussian process and $\theta(s)$ is a smooth mean function. Assume additionally that $Z(s)$ is stationary, then*

$$\mathbb{E}[M_u] = \int_D \frac{\sqrt{-2\rho'(3-\kappa^2)}}{\kappa} \phi\left(\frac{\theta'(s)}{\sqrt{-2\rho'}}\right) \int_u^{\infty} \phi(x-\theta(s)) \psi\left(\frac{\kappa[x-\theta(s)-\eta(s)]}{\sqrt{3-\kappa^2}}\right) dx\, ds, \tag{4.16}$$

where the function ψ is defined as

$$\psi(x) = \int_{-\infty}^{x} \Phi(y) dy = \phi(x) + x\Phi(x), \quad x \in \mathbb{R}.$$

Note that when $N=1$,

$$H(\tilde{x})=\phi\left(\frac{\kappa\tilde{x}}{\sqrt{3-\kappa^2}}\right)+\frac{\kappa\tilde{x}}{\sqrt{3-\kappa^2}}\Phi\left(\frac{\kappa\tilde{x}}{\sqrt{3-\kappa^2}}\right)=\psi\left(\frac{\kappa\tilde{x}}{\sqrt{3-\kappa^2}}\right). \tag{4.17}$$

We need the following lemmas to calculate $H(\tilde{x})$ explicitly when $N=2$ and $N=3$. They are direct calculation of integral by parts.

Lemma 4. *Let $N=2$, for constant $a>-\frac{1}{2}$ and $b\in\mathbb{R}$*

$$\begin{aligned}
&\int_{\mathbb{R}^2}\exp\left\{-\frac{1}{2}\sum_{i=1}^{2}\lambda_i^2-\frac{a}{2}\Big(\sum_{i=1}^{2}\lambda_i\Big)^2\right\}\left(\prod_{i=1}^{2}|\lambda_i-b|\right)|\lambda_1-\lambda_2|\mathbb{1}_{\{\lambda_1<\lambda_2<b\}}d\lambda_1\,d\lambda_2\\
=&\frac{\sqrt{2\pi}}{\sqrt{1+a}}e^{-\frac{1+2a}{2(1+a)}b^2}\Phi\left(\frac{1+2a}{\sqrt{1+a}}b\right)+\left(2b^2-\frac{1+4a}{1+2a}\right)\frac{\sqrt{\pi}}{\sqrt{1+2a}}\Phi(\sqrt{2(1+2a)}b)\\
&+\frac{b}{1+2a}e^{-(1+2a)b^2}.
\end{aligned} \tag{4.18}$$

Lemma 5. *Let $N=3$, for constant $a>0$ and $b\in\mathbb{R}$*

$$\begin{aligned}
&\int_{\mathbb{R}^3}\exp\left\{-\frac{1}{2}\sum_{i=1}^{3}\lambda_i^2+\frac{a}{2(2+3a)}\Big(\sum_{i=1}^{3}\lambda_i\Big)^2\right\}\\
&\times\left(\prod_{i=1}^{3}|\lambda_i-b|\right)\prod_{1\le i<j\le 3}|\lambda_i-\lambda_j|\mathbb{1}_{\{\lambda_1<\lambda_2<\lambda_3<b\}}d\lambda_1\,d\lambda_2\,d\lambda_3\\
=&\left[\frac{a^3+6a^2+12a+24}{2(a+2)^2}b^2+\frac{2a^3+3a^2+6a}{4(a+2)}+\frac{3}{2}\right]\frac{1}{\sqrt{\pi(a+2)}}e^{-\frac{b^2}{a+2}}\Phi\left(\frac{2\sqrt{2}b}{\sqrt{(a+2)(3a+2)}}\right)\\
&+\left[\frac{a+1}{2}b^2+\frac{a^2-a}{2}-1\right]\frac{1}{\sqrt{\pi(a+1)}}e^{-\frac{b^2}{a+1}}\Phi\left(\frac{\sqrt{2}b}{\sqrt{(a+1)(3a+2)}}\right)\\
&+\left(a+6+\frac{3a^3+12a^2+28a}{2(a+2)}\right)\frac{b}{2\pi(a+2)\sqrt{3a+2}}e^{-\frac{3b^2}{3a+2}}\\
&+b\left[b^2+\frac{3(a-1)}{2}\right][\Phi_{\Sigma_1}(0,b)+\Phi_{\Sigma_2}(0,b)],
\end{aligned}$$

where

$$\Sigma_1 = \begin{pmatrix} \frac{3}{2} & -1 \\ -1 & \frac{a+2}{2} \end{pmatrix}, \quad \Sigma_2 = \begin{pmatrix} \frac{3}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{a+1}{2} \end{pmatrix}.$$

Proposition 3. *Let $N = 2$, and assumptions in Theorem 4 hold. Then the function H defined in* (4.14) *can be written explicitly as*

$$\begin{aligned} H(\tilde{x}) =& \frac{\sqrt{2\pi}}{\sqrt{3-\kappa^2}}\phi\left(\frac{\kappa\tilde{x}}{\sqrt{3-\kappa^2}}\right)\Phi\left(\frac{\kappa\tilde{x}}{\sqrt{(2-\kappa^2)(3-\kappa^2)}}\right) + \frac{\kappa^2}{2}(\tilde{x}^2-1)\Phi\left(\frac{\kappa\tilde{x}}{\sqrt{2-\kappa^2}}\right) + \\ & \frac{\kappa\sqrt{2-\kappa^2}\tilde{x}}{2}\phi\left(\frac{\kappa\tilde{x}}{\sqrt{2-\kappa^2}}\right). \end{aligned} \tag{4.19}$$

Proposition 4. *When $N = 3$, let assumptions in Theorem 4 hold. Then the function H defined in* (4.14) *can be written explicitly as*

$$\begin{aligned} H(\tilde{x}) = & \left[\frac{\kappa^2\left[(1-\kappa^2)^3+6(1-\kappa^2)^2+12(1-\kappa^2)+24\right]}{4(3-\kappa^2)^2}\tilde{x}^2\right. \\ & \left. + \frac{2(1-\kappa^2)^3+3(1-\kappa^2)^2+6(1-\kappa^2)}{4(3-\kappa^2)} + \frac{3}{2}\right]\frac{\phi(\frac{\kappa\tilde{x}}{\sqrt{3-\kappa^2}})}{\sqrt{\pi(3-\kappa^2)}}\Phi\left(\frac{2\kappa\tilde{x}}{\sqrt{(3-\kappa^2)(5-3\kappa^2)}}\right) \\ & + \left[\frac{\kappa^2(2-\kappa^2)}{4}\tilde{x}^2 - \frac{\kappa^2(1-\kappa^2)}{2} - 1\right]\frac{\phi(\frac{\kappa\tilde{x}}{\sqrt{2-\kappa^2}})}{\sqrt{\pi(2-\kappa^2)}}\Phi\left(\frac{\kappa\tilde{x}}{\sqrt{(2-\kappa^2)(5-3\kappa^2)}}\right) \\ & + \left[7-\kappa^2+\frac{(1-\kappa^2)\left[3(1-\kappa^2)^2+12(1-\kappa^2)+28\right]}{2(3-\kappa^2)}\right]\frac{\kappa\tilde{x}\phi(\sqrt{\frac{3}{5-3\kappa^2}}\kappa\tilde{x})}{2\sqrt{2}\pi(3-\kappa^2)\sqrt{5-3\kappa^2}} \\ & + \frac{\kappa^3}{2\sqrt{2}}\tilde{x}(\tilde{x}^2-3)\left[\Phi_{\Sigma_1}(0,\kappa\tilde{x}/\sqrt{2})+\Phi_{\Sigma_2}(0,\kappa\tilde{x}/\sqrt{2})\right], \end{aligned}$$

where

$$\Sigma_1 = \begin{pmatrix} \frac{3}{2} & -1 \\ -1 & \frac{3-\kappa^2}{2} \end{pmatrix}, \quad \Sigma_2 = \begin{pmatrix} \frac{3}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{2-\kappa^2}{2} \end{pmatrix}.$$

Note that for $N = 2$ and $N = 3$, we need to solve an integral over the domain (see (4.15)) to get $\mathbb{E}[M_u]$. Although we can not derive the explicit form for the entire formula, this can be evaluated in applications with the help of numerical algorithms.

4.3.4 Isotropic unimodal mean function

We have calculated the explicit formulas assuming the mean function is a concave paraboloid. This is a very strong assumption. However, in a general setting, where the unimodal mean function is rotationally symmetric of any shape, we can apply a multivariate Taylor expansion at the peak and use the second-order approximation to estimate power. For example, suppose the shape of the mean function is proportional to a rotationally symmetric Gaussian density

$$\theta(s) = \theta_0 \exp\left(-\frac{\|s - s_0\|^2}{2\xi^2}\right), \tag{4.20}$$

where s_0 is the center of the mean function and ξ is the signal bandwidth. The Taylor expansion at the center is

$$\theta(s) = \theta_0 - \frac{\theta_0}{2\xi^2}\|s - s_0\|^2 + o(\|s - s_0\|^2). \tag{4.21}$$

When the domain size gets small, we neglect the remainder term, and use its quadratic approximation as the mean function. With quadratic mean function, it becomes convenient to use compute $\mathbb{E}[M_u]$. We will evaluate the performance of this approach for different domain sizes in the simulation study.

4.4 Simulations

In Section 4.2 above, we discussed power approximation under different scenarios. We showed the factorial moment $\mathbb{E}[M_u(M_u - 1)]$ decays faster than $\mathbb{E}[M_u]$ under some circumstances so that we can use $\mathbb{E}[M_u]$ or adjusted $\mathbb{E}[M_u]$ to approximate power. In this section, a simulation study is conducted to validate each scenario as well as visualize the power function, $\mathbb{E}[M_u]$, and adjusted $\mathbb{E}[M_u]$. Through simulation, we could also get a better sense of applying them to real

data.

4.4.1 Paraboloidal mean function

We generate $B = 100,000$ centered, unit-variance, smooth isotropic 2D Gaussian random fields over a grid of size 50×50 pixels as $Z(s)$, each field obtained as the convolution of white Gaussian noise with a Gaussian kernel of spatial standard deviation 5, and normalized to standard deviation $\sigma = 1$. For the mean function $\mu(s)$, we use a concave paraboloid centered at $s_0 = (25, 25)$. The equation of the paraboloid is

$$\theta(s) = \theta_0 - \frac{\|s - s_0\|^2}{2\xi^2}, \tag{4.22}$$

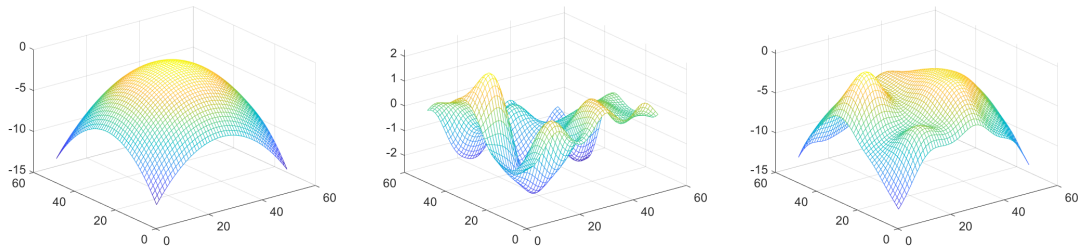
where ξ controls the sharpness of the mean function. The smaller ξ is, the sharper the paraboloid will be. θ_0 controls the height of the signal. To maintain the rotationally symmetric property of $\theta(s)$, we only consider those circles centered at s_0 as domain D. The size of D is measured by the radius $\mathrm{Rad}(D)$. The default value of each parameter is listed in Table 4.1 unless otherwise specified.

Table 4.1. 2D simulation: default value of each parameter

Parameter	Default value
$\mathrm{Rad}(D)$	10
ξ	7
θ_0	3

The first two panels of Figure 4.1 display two instances of $\theta(s)$ and $Z(s)$ respectively. The third panel displays the resulting sum $X(s)$ which is calculated by the signal-plus-noise model.

In the simulation, we validate and visualize the scenarios presented above, and check the effect of different choices of parameters on the power function, $\mathbb{E}[M_u]$ and adjusted $\mathbb{E}[M_u]$. Four different scenarios are considered as we discussed in Section 4.2:



(a) Mean function $\theta(s)$ **(b)** Noise $Z(s)$ **(c)** Data $X(s)$

Figure 4.1. 2D simulation: a single instance of $\theta(s)$, $Z(s)$ and their resulting $X(s)$.

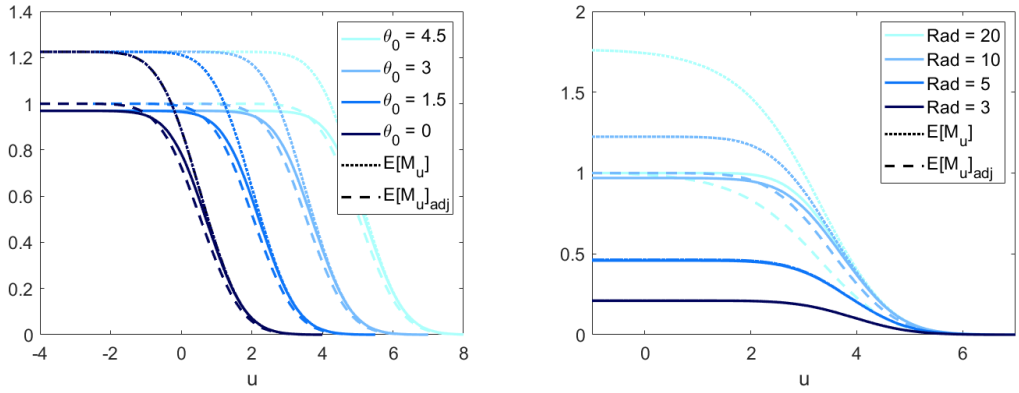
1. Height equivariance
2. Small domain size Rad(D)
3. Large threshold u
4. Sharp signal (small ξ)

For each simulated random field, we record the height of its highest peak if there exists at least one, and then for any threshold u, we calculate the empirical estimate of detection power (4.1) and $\mathbb{E}[M_u]$:

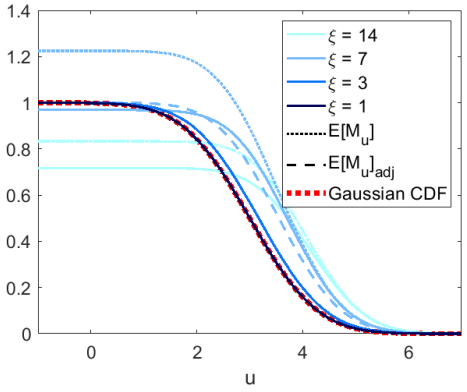
$$\hat{\mathbb{P}}[M_u \geq 1] = \frac{1}{B}\sum_{i=1}^{B} \mathbb{1}(\exists \text{ a peak in D with height } > u \text{ for } i\text{th simulated sample}), \tag{4.23}$$

$$\hat{\mathbb{E}}[M_u] = \frac{1}{B}\sum_{i=1}^{B} \#\text{ peaks in } D \text{ with height } > u \text{ for } i\text{th simulated sample}. \tag{4.24}$$

Figure 4.2 displays the power, $\mathbb{E}[M_u]$ and adjusted $\mathbb{E}[M_u]$ curves under the four scenarios. The first panel is to validate scenario 1 (height equivariance). As stipulated by Proposition 1, the power, $\mathbb{E}[M_u]$ and adjusted $\mathbb{E}[M_u]$ curves are parallel for different signal height h having other parameters remain the same. In the second panel, both the $\mathbb{E}[M_u]$ and adjusted $\mathbb{E}[M_u]$ curve are close to the power curve under scenario 2 (small domain) which indicates that for a smaller domain (quantified by Rad(D)), using $\mathbb{E}[M_u]$ and adjusted $\mathbb{E}[M_u]$ to approximate power becomes



(a) Height equivariance **(b)** Small domain size



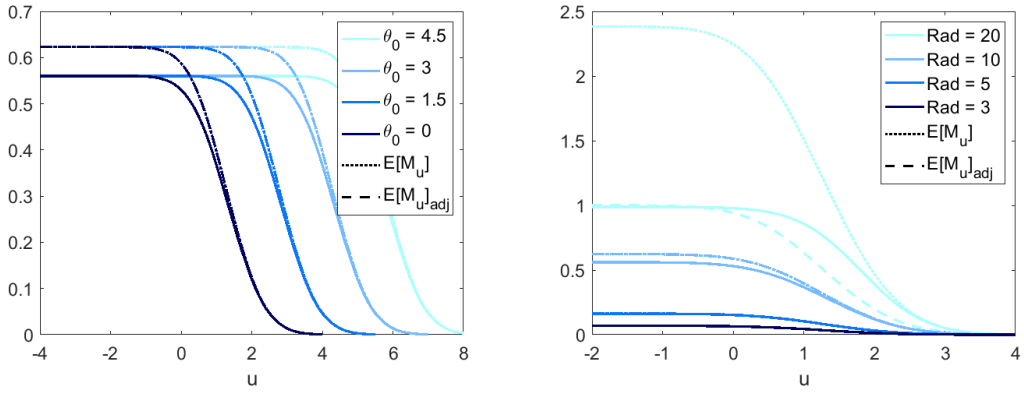
(c) Sharp signal

Figure 4.2. 2D simulation: Power approximation using $\mathbb{E}[M_u]$ under four different scenarios (scenario 3 is displayed in all three panels) when the mean function is quadratic.

more accurate as stipulated by Theorem 1. We can also find in all three panels that when u is large enough, the $\mathbb{E}[M_u]$ curve converges to the power curve as u increases, as stipulated by Theorem 2. The adjusted $\mathbb{E}[M_u]$ curve also converges to the power curve but with a slower rate compared to $\mathbb{E}[M_u]$. The third panel shows the power, $\mathbb{E}[M_u]$ and adjusted $\mathbb{E}[M_u]$ curve all converge to the Gaussian CDF for sharp signal (small ξ), as stipulated by Theorem 3.

4.4.2 Constant mean function

When the mean function $\theta(s)$ is constant, i.e. it does not depend on location s, $X(s)$ reduces to a centered isotropic Gaussian random field. Within the context of this Chapter, $\theta(s) = 0$ can be seen as the null hypothesis and the power function becomes the probability of



(a) Height equivariance **(b)** Small domain size

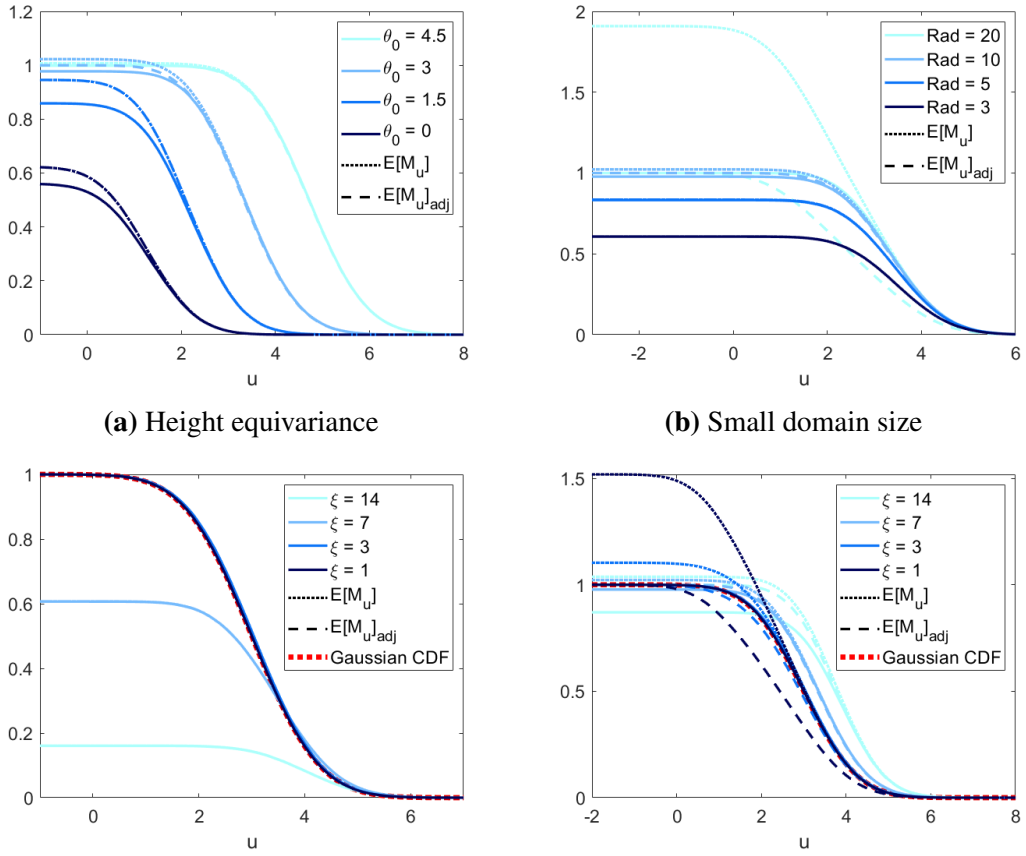
Figure 4.3. 2D simulation: Type I error approximation using $\mathbb{E}[M_u]$.

Type I error. We use the peak height distribution when $\theta(s) = 0$ (Cheng and Schwartzman, 2018) to decide the cutoff point such that the test meets the nominal type I error. The simulation result when $\theta(s) = 0$ is displayed in Figure 4.3.

The performance of Type I error approximation when the mean function is 0 is similar to what we find when the mean function is quadratic (scenario 4 is ignored since the shape parameter does not exist when the mean function is constant). The conclusion is that under large u (which is guaranteed in order to control the Type I error) or small domain, we have good Type I error approximation.

4.4.3 Gaussian mean function

The simulation results under Gaussian mean are displayed in Figure 4.4. For scenario 2 and 3, the results are consistent with those under quadratic mean. For scenario 1, since θ_0 controls both the signal height and sharpness, the power, $\mathbb{E}[M_u]$ and adjusted $\mathbb{E}[M_u]$ are no longer equivariant in terms of θ_0. For scenario 4, if we look at Figure 4.4c with Figure 4.4d, it can be seen that as the signal becomes sharper, the power, $\mathbb{E}[M_u]$ and adjusted $\mathbb{E}[M_u]$ curve converges to the Gaussian CDF only when the domain size (quantified by Rad(D)) is small. In this case, the asymptotic curve is a mixture of Gaussian CDF and $\mathbb{E}[M_u]$ under constant mean. This is due to the shape of Gaussian density as it converges to 0 if we expand the domain.



(a) Height equivariance

(b) Small domain size

(c) Sharp signal (Rad$(D) = 3$)

(d) Sharp signal (Rad$(D) = 20$)

Figure 4.4. 2D simulation: Power approximation using $\mathbb{E}[M_u]$ when the mean function is Gaussian.

In conclusion, for Gaussian mean function, we recommend applying our method to approximate power only when the domain size is small.

4.5 Estimation from data

To use our power approximation formula in real peak detection problems, we need to estimate the spatial covariance function of the noise as well as the mean function from the data. In this section, we demonstrate the 3D application setting and how to estimate the noise spatial covariance function and the mean function. Consider an imaging dataset with n subjects, and let $Y_i(s)$ represent the signal plus noise for subject i

$$Y_i(s) = \mu(s) + \sigma(s)\varepsilon_i(s),$$

where $s = (s_1, s_2, s_3)' \in \mathbb{R}^3$, the signal $\mu(s)$, standard deviation $\sigma(s)$ and noise $\varepsilon(s)$ are assumed to be smooth C^3 functions. If we compute the standardized mean of all n subjects, we will get a standardized random field

$$X(s) = \bar{Y}(s)/\mathrm{SE}(\bar{Y}(s)) = \sqrt{n}\mu(s)/\sigma(s) + \sqrt{n}\bar{\varepsilon}(s). \tag{4.25}$$

This standardized random field $X(s)$ has constant standard deviation of 1. We can treat $\sqrt{n}\mu(s)/\sigma(s)$ as the new signal and $\sqrt{n}\bar{\varepsilon}(s)$ as the new noise of the standardized field. Let $\theta(s) = \sqrt{n}\mu(s)/\sigma(s)$ and $Z(s) = \sqrt{n}\bar{\varepsilon}(s)$. We propose using the following method to estimate the new signal and noise.

4.5.1 Estimation of the noise spatial covariance function

We consider the noise $Z(s)$ to be constructed by convolving Gaussian white noise with a kernel:

$$Z(s) = \int_{\mathbb{R}^N} K(t-s)dB(t), \tag{4.26}$$

where $K(\cdot)$ is a N dimensional kernel function, and $dB(s)$ is Gaussian white noise. Assume that the kernel is rotationally symmetric so that the noise $Z(s)$ is isotropic. Under model (4.26), we would be able to simulate the noise if we were able to estimate the kernel function from the data.

It can be shown that the autocorrelation of $Z(s)$ is the convolution of the kernel with itself:

$$\mathrm{Cor}(Z(s), Z(s')) = \int_{\mathbb{R}^N} K(t-s)K(t-s')dt = \int_{\mathbb{R}^N} K(t-(s-s'))K(t)dt.$$

By the convolution theorem, convolution in the original domain equals point-wise multiplication in the Fourier-transformed domain. Thus the kernel function can be estimated

empirically using the following method:

1. Determine a location s_0 of interest (e.g. center of the peak), and calculate the empirical correlation vectors between $Y(s_0)$ and $Y(s)$ where s lies on the three orthogonal axes centered at s_0, and belongs to a subdomain of interest.

2. Take the average of the three estimated correlation vectors (forcing the noise to be isotropic) and perform Fourier transform.

3. Take the square root of the Fourier coefficients, then the estimated kernel function can be obtained by performing the inverse Fourier transform.

4.5.2 Estimation of the mean function

Our explicit formulas are derived assuming the Hessian of the mean function is a constant times the identity matrix. Therefore, we aim to find a rotationally symmetric paraboloid $\hat{\theta}(s)$ that best represents the mean function:

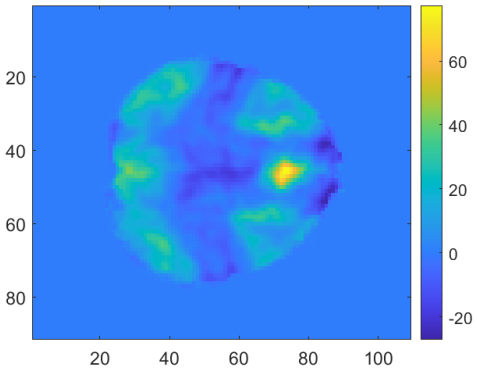
$$\theta(s) = \beta_0 + \beta_1 ||s||^2 + (\beta_2, \beta_3, \beta_4) \cdot s, \tag{4.27}$$

where the dot represents the vector inner product in $\mathbb{R}^3$. Note that all the quadratic terms share the same coefficient which is due to the rotational symmetry. To estimate (4.27), we can fit a linear regression using all $X(s)$ within the subdomain as outcome.

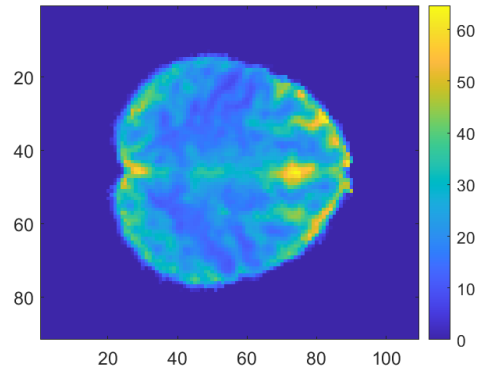
4.6 A 3D real data example

As an application, we illustrate the methods in a group analysis of fMRI data from the Human Connectome Project (HCP) (Van Essen et al., 2012). The data consists of the 2-back vs 0-back working memory contrast (Barch et al., 2013) from 80 unrelated subjects and is used here to get realistic 3D signal and noise parameters from which to do 3D simulations as well as evaluate the performance of our formulas in power approximation. For each subject, the

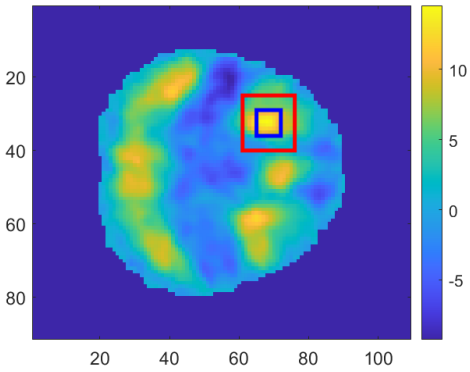
size of the fMRI image is $91 \times 109 \times 91$ voxels. The mean and standard deviation of the data are displayed in Figure 4.5a and 4.5b. Regions like the dorsal and ventral prefrontal cortex are activated in the contrast (Barch et al., 2013) as demonstrated by the standardized mean in Figure 4.5c.



(a) Mean of the data



(b) Standard deviation of the data



(c) Standardized mean of the smoothed data

Figure 4.5. HCP data: Mean, standard deviation of the data, and standardized mean of the smoothed data (transverse sliced at the peak of the image along the third dimension). The blue box represents the subdomain of the peak and the red box represents the subdomain we use to estimate the noise spatial covariance function.

4.6.1 Data preprocessing

Gaussian kernel smoothing is applied to the dataset to make the mean function unimodal around the peak and increase the signal-to-noise ratio. The standard deviation of the smoothing kernel we use in this example is 1 voxel which translates to full width at half maximum (FWHM)

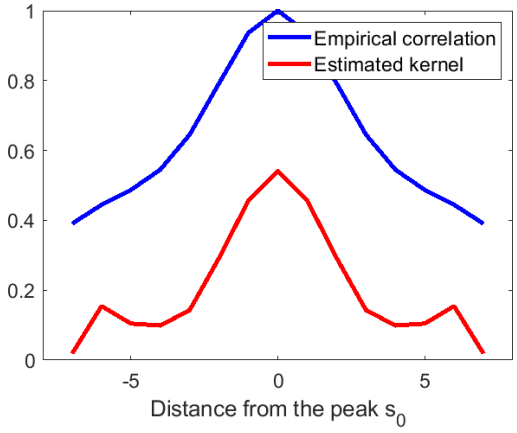


Figure 4.6. The empirical correlation after symmetrization and the estimated kernel from a subdomain of HCP data.

being around 2.235. It is obvious from Figure 4.5b that the standard deviation of the noise is not a constant for different locations, thus we use the transformation described in Section 4.5 to standardize the smoothed data before analyzing it. The standardized mean of the smoothed data $X(s)$ is displayed in Figure 4.5c.

4.6.2 Estimation of the autocorrelation and mean functions

After standardizing the data, our next step is to estimate the mean and kernel functions using the methods described in Section 4.5. Here a $15 \times 15 \times 15$ subdomain (the red box in Figure 4.5c) around the peak is taken to estimate the kernel. Since we assume the noise is isotropic, the correlation around the peak is supposed to be strictly symmetric along any dimension. However, this is not always true in real data. To tackle this, for each of the three dimensions, we first save the correlation $\mathrm{Cor}(X(s), X(s_0))$ around the peak s_0 as a vector and create a new vector by flipping the saved correlation vector. Then we take the mean of the two vectors so that it is guaranteed to be symmetric. The empirical correlation after such symmetrization and the corresponding estimated kernel function are displayed in Figure 4.6.

We consider two approaches to estimate the mean function, nonparametric and parametric.

The nonparametric mean estimation is obtained as a voxelwise average over subjects.

$$\hat{\theta}(s) = \sum_{i=1}^{n} X_i(s) = \bar{X}(s). \tag{4.28}$$

The parametric mean estimation is obtained by fitting a linear regression model (4.27) using all observed data $X(s)$ within the subdomain of size $6 \times 6 \times 6$ (the blue box in Figure 4.5c) as outcome and their corresponding location variables $||s||^2$, s as covariates. The least square estimate of the mean is

$$\hat{\theta}(s) = 13.03 - 0.26||s||^2 + (0.20, 0.11, 0.39) \cdot s. \tag{4.29}$$

We will compare the difference in simulated power and $\mathbb{E}[M_u]$ when the mean function is estimated by the nonparametric approach (4.28) vs the parametric approach (4.29).

4.6.3 3D Simulation induced by data

We have done several simulation studies under a well-designed 2D setting where the formulas are supposed to work well, but eventually, we want to apply the formulas to real-life data which is more complicated. Besides, in terms of fMRI data analysis, the image is always 3D by nature. Considering all these factors, a 3D simulation study induced by real data is necessary to validate the performance of the formulas under a more realistic setting,

In the previous two subsections, we have studied the signal and noise of the HCP data. For the simulation, we would like to generate 3D images using the estimated mean and kernel function. The noise field is generated by convolving the estimated kernel (displayed in Figure 4.6) and Gaussian white noise. For each simulation setting, 10,000 such noise fields are generated.

The signal from the standardized data is very strong (see Figure 4.5c). For illustrative purposes, we choose to weaken the signal by scaling down the estimated mean function (4.29) while maintaining the same shape. Here signal strength is measured by effect size, and the

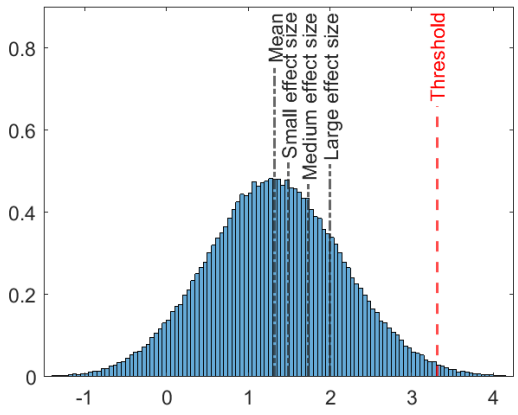


Figure 4.7. 3D Simulation induced by data: Simulated Peak height distribution under the null (zero mean) with different levels of effect sizes and threshold u.

amount of scaling is determined by different levels of effect size. In traditional t-test or z-test, Cohen's d values of 0.2, 0.5, and 0.8 (corresponding to 0.58^{th} 0.69^{th} and 0.79^{th} quantiles of the standard Gaussian distribution) are considered as small, medium, and large effect sizes (Cohen, 1988). The peak height distribution under the null hypothesis (zero mean function) is displayed in Figure 4.7, and does not follow a Gaussian distribution. Therefore, we take 0.58^{th}, 0.69^{th} and 0.79^{th} quantile of the null distribution minus the mean as small (0.16), medium (0.40), and large (0.65) effect size (see the black dash-dot lines in Figure 4.7). For simplicity, we see the peak height of the mean function as effect size in this simulation. However, this is not the most accurate way of defining effect size in the peak detection setting. More details will be discussed in 4.7.2. The threshold u for peak detection is chosen as the 0.99^{th} quantile of the peak height distribution under the null (≈ 3.42) according to Cheng and Schwartzman (2017) (see the red dashed line in Figure 4.7).

Similar to the 2D simulation, the search domain D is assumed to be a sphere centered at the true peak, and we use radius of D to control the domain size. Signal sharpness is fixed since it is estimated from the data. The empirical power and $\mathbb{E}[M_u]$ are computed using (4.23) and (4.24).

To derive the explicit formulas for $\mathbb{E}[M_u]$, we assume the mean function to be a rota-

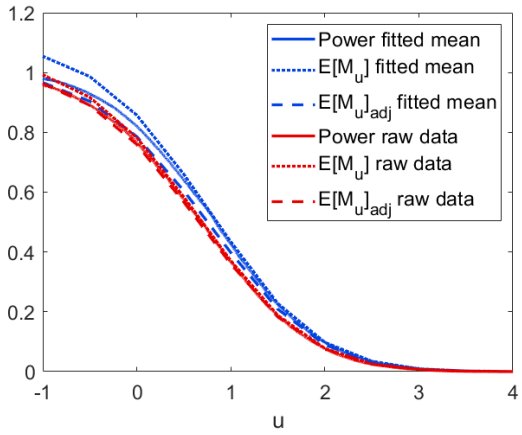


Figure 4.8. 3D Simulation induced by data ($\mathrm{Rad}(D) = 3$, medium effect size): Simulated power and $\mathbb{E}[M_u]$ when the mean function is obtained from raw data vs quadratic estimation (4.29).

tionally symmetric paraboloid, and this assumption might cause some bias in applications. In Figure 4.8, we demonstrate the difference in simulated power and $\mathbb{E}[M_u]$ when the mean function is estimated by the nonparametric approach (4.28) vs the parametric approach (4.29). It can be observed that the quadratic approximation only has a small impact on power and $\mathbb{E}[M_u]$ in this example. In Figure 4.9, we validate our theoretical formula for $\mathbb{E}[M_u]$ (4.15) as well as the adjusted $\mathbb{E}[M_u]$. As we can see, the theoretical curve for $\mathbb{E}[M_u]$ and adjusted $\mathbb{E}[M_u]$ closely mirrors the empirical curve. The figure also shows that the power approximation using $\mathbb{E}[M_u]$ is accurate for large u as stipulated by Theorem 2. Power curves using three different effect sizes, and comparisons between large and small domain sizes are displayed in Figure 4.10. We can see from the figure that the $\mathbb{E}[M_u]$ works well for small sample sizes and $\mathbb{E}[M_u]_{\mathrm{adj}}$ works better than $\mathbb{E}[M_u]$ for large sample sizes. We can also observe that the performance of power approximation using $\mathbb{E}[M_u]$ becomes better if the domain size is smaller as stipulated by Theorem 1.

4.7 Discussion

4.7.1 Explicit formulas and approximations

Calculating power for peak detection (4.1) has been a difficult problem in random field theory due to the lack of formula that can compute it directly. In this chapter, we have discussed

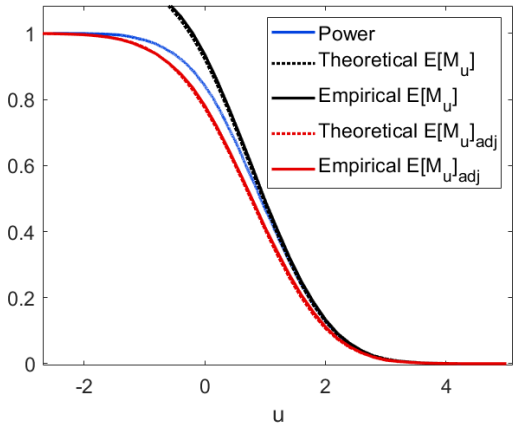
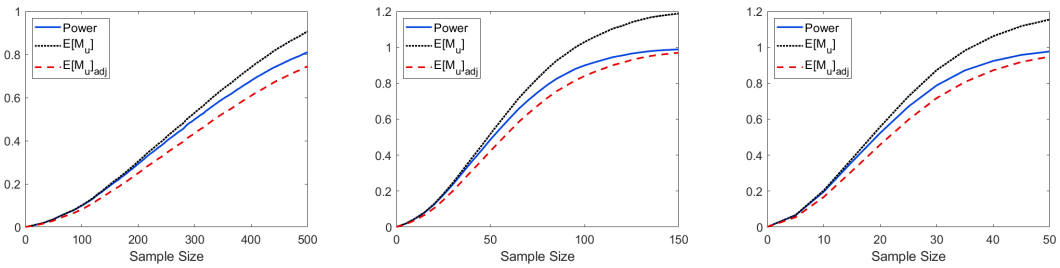
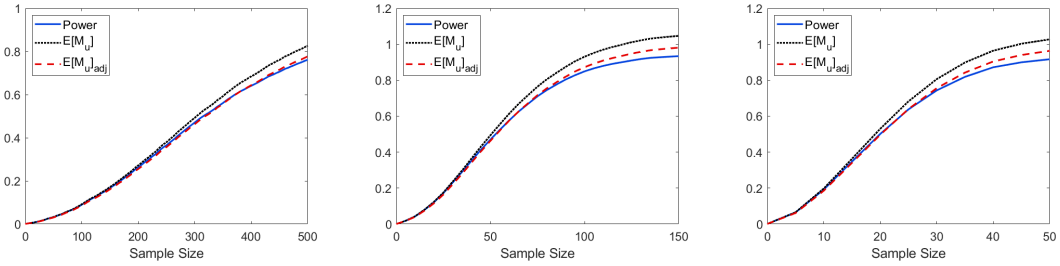


Figure 4.9. 3D Simulation induced by data ($\mathrm{Rad}(D) = 6$, medium effect size): Simulated vs theoretical $\mathbb{E}[M_u]$ and adjusted $\mathbb{E}[M_u]$.



(a) Small effect size ($\mathrm{Rad}(D) = 5$) **(b)** Medium effect size ($\mathrm{Rad}(D) = 5$) **(c)** Large effect size ($\mathrm{Rad}(D) = 5$)



(d) Small effect size ($\mathrm{Rad}(D) = 2$) **(e)** Medium effect size ($\mathrm{Rad}(D) = 2$) **(f)** Large effect size ($\mathrm{Rad}(D) = 2$)

Figure 4.10. 3D Simulation induced by data: Power curves when the signal has small, medium, and large effect size, and comparisons between large and small domain size.

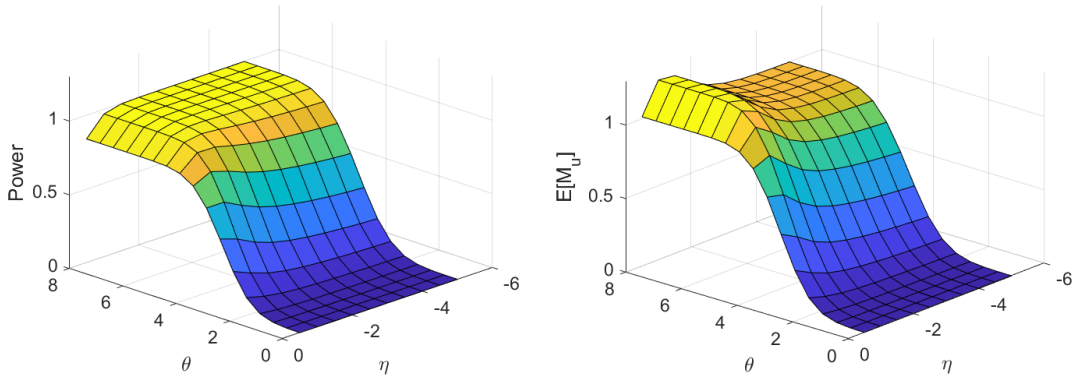
the rationale of using $\mathbb{E}[M_u]$ and $\mathbb{E}[M_u]_{\text{adj}}$ to approximate peak detection power under different scenarios and derived formulas to compute $\mathbb{E}[M_u]$ assuming isotropy. Isotropy is assumed so that we are able to use the GOI matrix (Cheng and Schwartzman, 2018) as a tool to calculate $\mathbb{E}[M_u]$ via the Kac-Rice formula.

We also showed explicit formulas for $H(\tilde{x})$ (defined as (4.14)) when $N = 1, 2, 3$ assuming the mean function is a paraboloid. Computing $H(\tilde{x})$ involves applying the probability density function for the eigenvalues of GOI matrices and details can be found in the proof of Proposition 2, 3 and 4. Then $\mathbb{E}[M_u]$ can be calculated by plugging $H(\tilde{x})$ to (4.15). The integration in (4.15), however, can not be evaluated explicitly. In practice, one may evaluate it numerically. For higher dimensions ($N > 3$), it remains difficult to get an explicit form of $H(\tilde{x})$ due to the fact inferred by Proposition 2, 3 and 4 that the integration becomes extremely complicated as N becomes large.

4.7.2 Effect size

We want to emphasize that the power depends on both the signal strength parameter θ_0 and shape parameter η. In a traditional z-test or t-test which tests a single null hypothesis that the mean value equal to 0, the detection power depends only on a single parameter we call effect size. Here the test is conditional on the point being a local maximum. Applying a simple z-test or t-test, one could reject the null hypothesis as long as the peak height θ_0 exceeds the pre-specified threshold. This approach is not accurate since the peak height does not follow a Gaussian or t distribution. To address this, the threshold can be determined by the null distribution of peak height (Cheng and Schwartzman, 2018) to control the type I error at a nominal level. However, power calculation based on the test over peak height is still biased since the true effect size depends both on the signal height and curvature. The height of the peak affects the likelihood of exceeding the threshold and the curvature affects the likelihood of existing such peak in the domain. It follows that a sharp and high peak is easier to detect compared to a flat and low peak, leading to a larger detection power.

For an interpretation of the parameter $\eta = \theta''/(-2\rho')$, we consider two types of mean



(a) Power **(b)** $\mathbb{E}[M_u]$

Figure 4.11. 2D simulation: Power and $\mathbb{E}[M_u]$ for different θ and η ($u = 3.92$ and $\text{Rad}(D) = 10$).

function: paraboloid and Gaussian. Suppose the noise is the result of the convolution of white noise with a Gaussian kernel with spatial std. dev. ν resulting in the covariance function with $\rho(r) = \exp(-r/(2\nu^2))$ as specified in Section 4.3.1. This is the same noise as we simulated in Section 4.5. When the mean function is paraboloid, consider $\theta(s) = -\|s\|^2/(2\xi^2) + \theta_0$ as in (4.22). Here we obtain $\theta'' = -1/\xi^2$ and $\rho' = -1/(2\nu^2)$, yielding $\eta = \theta''/(-2\rho') = -\nu^2/\xi^2$. Thus, η is a shape parameter representing the relative sharpness of the mean function with respect to the curvature of the noise. When the mean function is Gaussian, consider $\theta(s) = a\exp(-\|s\|^2/(2\tau^2))$. This expression is obtained, for example, if the signal is the result of the convolution of a delta function with a Gaussian kernel with spatial std. dev. τ. We obtain $\theta'' = -a/\tau^2$ and $\rho' = -1/(2\nu^2)$, yielding $\eta = \theta''/(-2\rho') = -a\nu^2/\tau^2$. Thus, η is the height of the signal a, scaled by the ratio of the spatial extent of the noise and signal filters. In both cases, the parameter η, and thus the power, are invariant under isotropic scaling of the domain, in a similar fashion to the peak height distribution under the null hypothesis (Cheng and Schwartzman, 2020).

Figure 4.11 illustrates how θ_0 and η affect power and $\mathbb{E}[M_u]$ in the 2D simulation described in Section 4.4. As we have explained, θ_0 and η together determine the effect size. Although deriving an explicit form of effect size as a function of θ_0 and η is difficult, we are able

to roughly show how the two parameters relate to power. θ_0 which can be seen as signal-to-noise ratio (SNR) plays a major role. Having η stay the same, the power monotonically increases with respect to θ_0. On the other hand, power monotonically decreases with respect to η having θ_0 stays the same. In this simulation example, the impact of θ_0 on power is about 10 times stronger than η if we fit a linear model of power using θ_0 and η. We can also observe from the figure that the effect of η on power is stronger for large θ_0 compared to that for small θ_0.

4.7.3 Application to data

To use our formula to calculate power in practice, one needs to assume the peak to be a certain type such as paraboloid or Gaussian. However, sometimes it might not be plausible to make such assumptions, leading to inaccurate power estimate.

Regarding the conjecture of $\mathbb{E}[M_u]_{\text{adj}}$ being a lower bound when there exists at least one local maximum in the domain D, it remains difficult to prove in general, but as we showed in the real data example, it seems to be correct in practice. When it comes to a real-life problem, we can take both $\mathbb{E}[M_u]$ and the $\mathbb{E}[M_u]_{\text{adj}}$ into consideration to get a better understanding of the true sample size. We suggest using $\mathbb{E}[M_u]$ as an approximation to power when the sample size is small, considering $\mathbb{E}[M_u]_{\text{adj}}$ when the sample size is large. $\mathbb{E}[M_u]_{\text{adj}}$ also gives a more conservative estimate of power compared to $\mathbb{E}[M_u]$ which is useful to guarantee that the test is powerful enough when we design future studies. Because of its difficulty, we leave further study of $\mathbb{E}[M_u]_{\text{adj}}$ for future work.

4.8 Acknowledgement

Yu Zhao, Dan Cheng, and Armin Schwartzman were partially supported by NIH grant R01EB026859 and NSF grant 1811659. Data were provided in part by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington

University.

Chapter 4, in full, has been submitted for publication of the material as it may appear in the Journal of Multivariate Analysis. Zhao, Yu; Cheng, Dan; Schwartzman, Armin. An approximation to peak detection power using Gaussian random field theory. The dissertation author was the primary investigator and author of this material.

Chapter 5

On the peak height distribution of non-stationary Gaussian random fields: 1D non-constant variance and scale space

5.1 Introduction

Peak inference is an important statistical problem in random field theory (RFT) and has been applied to multiple domains, such as neuroimaging (Worsley et al., 2004; Schwartzman and Telschow, 2019a) and statistical cosmology (Bertschinger, 2001). Researchers have made significant efforts to derive the formulas for the peak height distribution (Lindgren, 1972; Azaïs and Wschebor, 2008; Cheng and Schwartzman, 2015; Cheng and Schwartzman, 2018). The currently known exact formulas for the peak height distribution in 1D, 2D, and 3D require the stationarity assumption (Cheng and Schwartzman, 2018). However, the stationarity assumption is known to be difficult to check and even not appropriate in many application settings (Worsley et al., 1996a; Foody, 2004). To expand the application of RFT in peak inference, we investigate the peak height distribution of zero-mean Gaussian random fields and seek to develop a general solution.

Let $\{X(t) : t \in \mathbb{R}^N\}$ be a real-valued, N dimensional, C^2 Gaussian random field. Throughout this chapter, we always assume $X(t)$ satisfies the conditions (C1) and (C2) in Cheng and Schwartzman (2015) unless stated otherwise. For stationary random fields, the statistical prop-

erties, including the mean, variance, and covariance between points at a fixed distance remain constant across the domain. Any spatially varying statistical property leads to non-stationarity. For example, $X(t)$ is non-stationary if the mean function $\mu(t)$ is non-constant. This is known as the signal-plus-noise model, where $\mu(t)$ and $X(t) - \mu(t)$ represent the signal and noise respectively. Non-centered Gaussian random fields are frequently considered in data applications (Hayasaka et al., 2007; Cheng and Schwartzman, 2017), and the peak height distribution of non-centered Gaussian random fields has been studied in various RFT-based power calculation works (Durnez et al., 2016; Zhao et al., 2023). In this chapter, we only consider centered Gaussian random fields, and focus on non-stationary Gaussian random fields with non-constant variance, spatially varying correlation, or both.

The scale space random field (Siegmund and Worsley, 1995) is an example of non-stationary random field with spatially varying correlation and has many applications (Bertschinger, 2001; Worsley et al., 1996b). In peak detection, kernel smoothing is commonly applied to improve the signal-to-noise ratio. To maximize the detection power, the kernel bandwidth is chosen to match the shape of the kernel to that of the peak. However, when the peaks have different shapes and sizes, there does not exist a single smoothing bandwidth to match all the peak shapes. To handle peaks of different spatial extents, it is natural to consider the bandwidth as an extra parameter and search for peaks in the scale-location space. The random field defined in the scale-location space is called the scale space random field. Since the data smoothness varies over the search domain, the scale space random field is non-stationary even if it is stationary at any fixed scale. In this chapter, we study the scale space Gaussian field as a prototypical example of non-stationarity.

For Gaussian random fields, the Kac-Rice formula (see Chapter 11 of Adler and Taylor, 2007) is a powerful tool to calculate the expected number of critical points and the peak height distribution. With the Kac-Rice formula, it can be shown (see Cheng and Schwartzman, 2015) that the peak height distribution of $X(t)$ (the probability that the height of $X(t)$ at point t exceeds

the fixed threshold u, given that the point t is a local maximum of $X(t)$) can be computed as

$$F_t(u) = \frac{\mathbb{E}[|\det\nabla^2 X(t)|\mathbb{1}_{\{\nabla^2 X(t)\prec 0\}}\mathbb{1}_{\{X(t)>u\}}|\nabla X(t)=0]}{\mathbb{E}[|\det\nabla^2 X(t)|\mathbb{1}_{\{\nabla^2 X(t)\prec 0\}}|\nabla X(t)=0]}. \tag{5.1}$$

In this chapter, we first derive the exact formula for the peak height density (the derivative of (5.1)) of 1D centered, smooth Gaussian processes and compare it with that for stationary Gaussian processes (Cheng and Schwartzman, 2018). When the dimension $N > 1$, the Kac-Rice formula involves computing the conditional expectation of the determinant of the Hessian matrix of $X(t)$ which is a Gaussian random matrix. There is no analytical solution to this without making further assumptions. The explicit evaluation of the peak height distribution of isotropic Gaussian random fields has been studied in Cheng and Schwartzman (2018) with the help of Gaussian random matrix theory, but there is no exact formula for general anisotropic Gaussian random fields. Like the Gaussian CDF, lacking an exact formula does not hinder its practical applications. One approach to address this issue is to connect the random field of interest $X(t)$ to some other random field $Y(t)$ with known peak height distribution. For example, Cheng and Schwartzman (2020) have proved that the peak height distribution remains the same under diffeomorphic transformation, i.e. $X(t) = Y(f(t))$ where f is a diffeomorphic map. Another approach is to simplify the calculation by making practical assumptions, for example, assuming the field has constant variance. If we still find it difficult to evaluate explicitly or if the data indicates a violation of such assumptions, the peak height distribution can be obtained with numerical approximations. In this chapter, we introduce two convenient algorithms that compute the peak height distribution by evaluating the Kac-Rice formula numerically, and it works for any Gaussian random field as long as it is sufficiently smooth. Compared to the traditional simulation approach, which simulates the field itself, this algorithm achieves higher accuracy with much less computing time.

5.2 1D non-stationary Gaussian processes

In this section, we start from the simplest non-stationary Gaussian random field, the 1D non-stationary Gaussian process. Our aim is to derive the explicit form of the peak height distribution by evaluating the Kac-Rice formula (5.1), and look for connections and differences between the peak height distribution of stationary and non-stationary Gaussian processes.

Assuming that the random process $\{X(t) : t \in \mathbb{R}\}$ is non-stationary, the covariance between any two points may depend on both their location and the distance between them. As a generalization of the covariance function under the stationarity assumption (Azaïs and Wschebor, 2008), we can write the covariance function of a non-stationary Gaussian process as $\mathbb{E}[X(t-d/2)X(t+d/2)] = h(t, \tau = d^2)$ for an appropriate function $h(\cdot) : [0,\infty) \to \mathbb{R}$. Denote

$$h_0(t) = h(t,0), \quad h_1'(t) = \left.\frac{\partial h}{\partial t}\right|_{\tau=0}, \quad h_2'(t) = \left.\frac{\partial h}{\partial \tau}\right|_{\tau=0}, \quad h_{12}''(t) = \left.\frac{\partial^2 h}{\partial t \partial \tau}\right|_{\tau=0}, \text{ etc.} \tag{5.2}$$

The variance-covariance matrix of $(X(t), X'(t), X''(t))$ can be derived using (5.5.5) in Adler and Taylor (2007). The lemma below is a generalization of Lemma 3.2 in Cheng and Schwartzman (2018) for non-stationary Gaussian processes.

Lemma 6. *Let $\{X(t), t \in \mathbb{R}\}$ be a centered, smooth 1D Gaussian process. Then for each $t \in \mathbb{R}$,*

$$\begin{aligned}
\mathrm{Var}(X(t)) &= h_0(t), \\
\mathrm{Var}(X'(t)) &= \frac{1}{4}h_{11}''(t) - 2h_2'(t), \\
\mathrm{Var}(X''(t)) &= \frac{1}{16}h_{1111}''''(t) - h_{112}'''(t) + 12h_{22}''(t), \\
\mathbb{E}[X(t)X'(t)] &= \frac{1}{2}h_1'(t), \\
\mathbb{E}[X(t)X''(t)] &= \frac{1}{4}h_{11}''(t) + 2h_2'(t), \\
\mathbb{E}[X'(t)X''(t)] &= \frac{1}{8}h_{111}'''(t) - h_{12}''(t).
\end{aligned}$$

5.2.1 Peak height distribution of non-stationary Gaussian processes

The following theorem states the formula for the peak height distribution of centered smooth Gaussian processes.

Theorem 5. *The peak height density of a smooth Gaussian process* $\{X(t), t \in \mathbb{R}\}$ *with mean 0 and standard deviation* $\sigma(t)$ *is given by*

$$f_t(x) = \frac{1}{\tilde{\sigma}(t)} \phi\left(\frac{x}{\tilde{\sigma}(t)}\right) \sqrt{2\pi(1-\rho^2(t))}\, \psi\left(\frac{-\rho(t)x}{\sqrt{1-\rho^2(t)}\tilde{\sigma}(t)}\right), \tag{5.3}$$

where $\tilde{\sigma}^2(t) := \mathrm{Var}(X(t)|X'(t)=0)$, $\rho(t) := \mathrm{Cor}(X(t), X''(t)|X'(t)=0)$ *satisfying* $|\rho(t)| < 1$, *and the function* $\psi(\cdot)$ *is defined as*

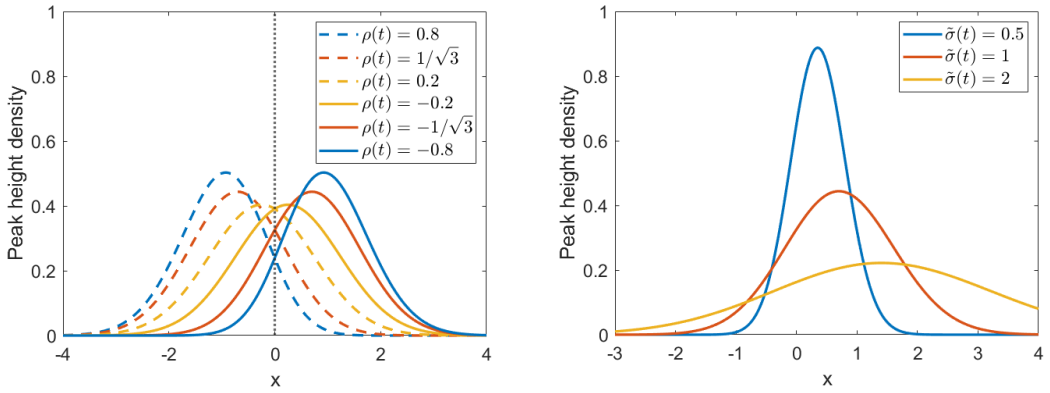
$$\psi(x) = \int_{-\infty}^{x} \Phi(y)dy = \phi(x) + x\Phi(x), \quad x \in \mathbb{R}. \tag{5.4}$$

The mean and variance of the peak height are $-\sqrt{\pi/2}\rho(t)\tilde{\sigma}(t)$ *and* $(1-(\pi/2-1)\rho^2(t))\tilde{\sigma}^2(t)$ *respectively.*

The density function (5.3) only depends on two parameters, $\rho(t)$ and $\tilde{\sigma}(t)$, and has the form of a Gaussian density multiplied by a tilting factor. By Lemma 6, $\rho(t)$ can be computed as

$$\begin{aligned}
\rho(t) =& \mathrm{Cor}(X(t), X''(t)|X'(t)=0) \\
=& \frac{\mathbb{E}[X(t)X''(t)]\mathrm{Var}(X'(t)) - \mathbb{E}[X(t)X'(t)]\mathbb{E}[X'(t)X''(t)]}{\sqrt{\mathrm{Var}(X(t))\mathrm{Var}(X'(t)) - \mathbb{E}[X(t)X'(t)]^2}} \\
& \times \frac{1}{\mathrm{Var}(X'(t))\mathrm{Var}(X''(t)) - \mathbb{E}[X'(t)X''(t)]^2} \qquad (5.5) \\
=& -\frac{2\left(2h_2'(t) - \frac{1}{4}h_{11}''(t)\right)\left(2h_2'(t) + \frac{1}{4}h_{11}''(t)\right) - h_1'(t)\left(h_{12}''(t) - \frac{1}{8}h_{111}'''\right)}{\sqrt{-\left(h_{12}''(t) - \frac{1}{8}h_{111}'''(t)\right)^2 - \left(2h_2'(t) - \frac{1}{4}h_{11}''(t)\right)\left(12h_{22}''(t) - h_{112}'''(t) + \frac{1}{16}h_{1111}''''(t)\right)}} \\
& \times \frac{1}{\sqrt{-h_1'(t)^2 - 4h_0(t)\left(2h_2'(t) - \frac{1}{4}h_{11}''(t)\right)}}. \qquad (5.6)
\end{aligned}$$

Both (5.5) and (5.6) can be used to compute $\rho(t)$, depending on which is more convenient, the



(a) $\tilde{\sigma}(t) = 1$ **(b)** $\rho(t) = -1/\sqrt{3}$

Figure 5.1. The peak height density (5.3) with different parameters. Left panel: Fixing $\tilde{\sigma}(t)$ at 1, the effect of $\rho(t)$ on the peak height density. Right panel: Fixing $\rho(t)$ at $-1/\sqrt{3}$, the effect of $\tilde{\sigma}(t)$ on the peak height density.

moments or the partial derivatives of the covariance function. In the analytical examples below, it is easier to use (5.5). In the numerical methods (Section 5.4), it is easier to use (5.6) because all the terms in (5.6) can be programmed as the partial derivatives of the covariance function. The same reasoning can be applied to the computation of $\tilde{\sigma}(t)$:

$$\tilde{\sigma}(t) = \text{Var}(X(t)|X'(t) = 0) = \text{Var}(X(t)) - \frac{\mathbb{E}[X(t)X'(t)]^2}{\text{Var}(X'(t))} = h_0(t) - \frac{h_1'(t)^2}{h_{11}''(t) - 8h_2'(t)}. \quad (5.7)$$

Compared to the parameter κ of the peak height density derived in Cheng and Schwartzman (2018), using $\rho(t)$ and $\tilde{\sigma}(t)$ to characterize the peak height density improves the understanding and interpretation of the density function, as they both have a clear statistical meaning. Figure 5.1 displays how these two parameters affect the peak height density. As we can see from (5.3) and Figure 5.1a, if we fix $\tilde{\sigma}(t)$, the density functions with $\rho(t)$ and $-\rho(t)$ are reflection symmetric with respect to $x = 0$. As $\rho(t)$ increases, the absolute value of the mean increases proportionally, and the variance decreases. The parameter $\tilde{\sigma}(t)$ is the scale parameter of the peak height distribution. The distribution is more spread out for large $\tilde{\sigma}(t)$ as displayed in Figure 5.1b. Increasing $\tilde{\sigma}(t)$ leads to a linear increase in both the absolute value of the mean and standard deviation.

Theorem 5 can be used to compute the peak height distribution of any centered smooth non-stationary Gaussian process, including the Gaussian process with non-constant variance. In general, Gaussian processes with non-constant variance are useful in modeling the heteroscedastic noise (i.e. the variance of noise depends on location), which is often present in real-world problems (see e.g. Le et al., 2005; Lázaro-Gredilla and Titsias, 2011). Understanding the peak height distribution of such Gaussian processes is crucial for peak detection given heteroscedastic noise. Figure 5.2a displays simulated instances of the Gaussian process with non-constant variance on a grid ranging from 0 to 1 with a step size of 0.005. The process is generated by multiplying a stationary unit-variance Gaussian process (convolution of Gaussian white noise with a Gaussian kernel of bandwidth 0.3) by a linear standard deviation function $\sigma(t) = t + 0.1$. Consequently, the spatial correlation is independent of the location t, as shown in Figure 5.2b. Figure 5.2c and Figure 5.2d display the two parameters $\rho(t)$ and $\tilde{\sigma}(t)$ that affect the peak height distribution. Peak height distribution of this process at a fixed location t can be computed by plugging $\rho(t)$ and $\tilde{\sigma}(t)$ into (5.3).

5.2.2 Special cases

In particular, if $X(t)$ has constant variance, $\rho(t)$ and $\tilde{\sigma}(t)$ can be simplified as

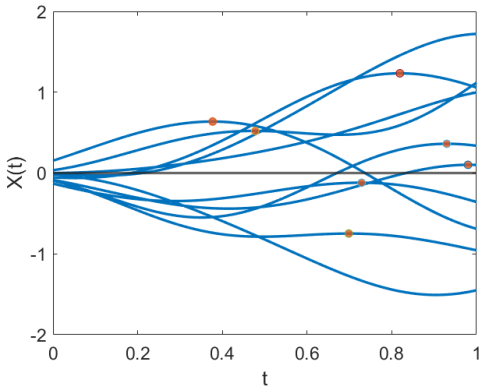
$$\rho(t) = -\frac{\mathrm{Var}(X'(t))}{\sqrt{\mathrm{Var}(X''(t)) - \frac{\mathbb{E}[X'(t)X''(t)]^2}{\mathrm{Var}(X'(t))}}} = \frac{2h_2'(t)}{\sqrt{-h_{112}'''(t) + 12h_{22}''(t) + \frac{h_{12}''(t)^2}{2h_2'(t)}}}, \tag{5.8}$$

$$\tilde{\sigma}(t) = \mathrm{Var}(X(t)) = h_0(t).$$

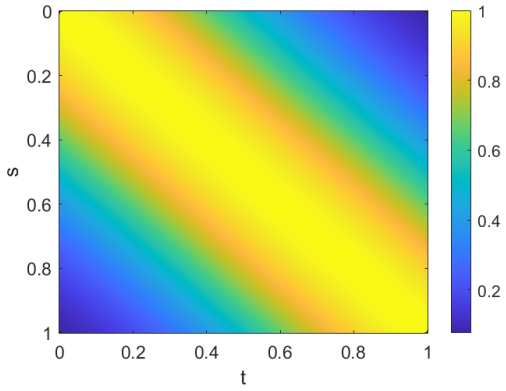
Note that (5.8) indicates $\rho(t) \leq 0$ when $X(t)$ has constant variance.

If we further assume stationarity and unit-variance, then $\rho(t)$ and $\tilde{\sigma}(t)$ become constant:

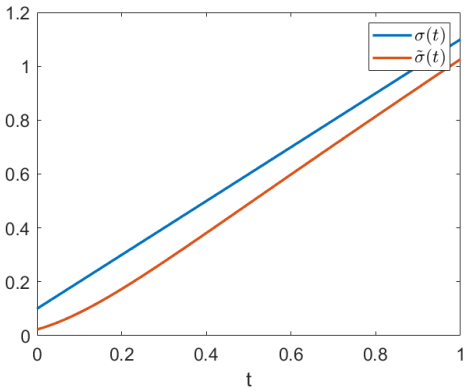
$$\rho = -\frac{\mathrm{Var}(X'(t))}{\sqrt{\mathrm{Var}(X''(t)) - \frac{\mathbb{E}[X'(t)X''(t)]^2}{\mathrm{Var}(X'(t))}}} = \frac{h_2'}{\sqrt{3h_{22}''}},$$



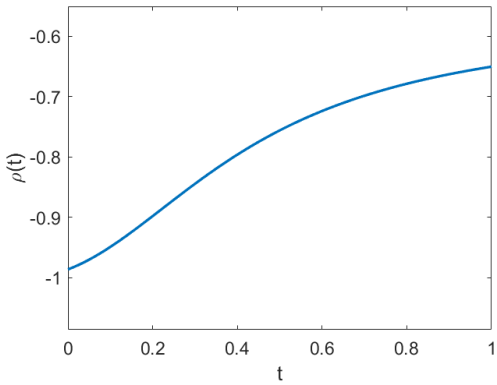
(a) Simulated instances with orange dots indicating the local maxima



(b) Correlation



(c) $\sigma(t)$ and $\tilde{\sigma}(t)$



(d) $\rho(t)$

Figure 5.2. Simulated instances of the Gaussian process with non-constant variance, its spatial correlation, and the parameters $\sigma(t)$, $\tilde{\sigma}(t)$, and $\rho(t)$.

$$\tilde{\sigma} = \mathrm{Var}(X(t)) = 1.$$

As a special case of (5.8), the parameter ρ is non-positive for stationary random processes.

The general peak height density (5.3) has a similar form as the peak height density of a unit-variance stationary Gaussian process derived in Cheng and Schwartzman (2018):

$$\begin{aligned} f_t(x) &= \frac{\sqrt{3-\kappa^2}}{\sqrt{3}}\phi\left(\frac{\sqrt{3}x}{\sqrt{3-\kappa^2}}\right) + \frac{\sqrt{2\pi}\kappa x}{\sqrt{3}}\Phi\left(\frac{\kappa x}{\sqrt{3-\kappa^2}}\right)\phi(x) \\ &= \phi(x)\sqrt{\frac{2\pi(3-\kappa^2)}{3}}\psi\left(\frac{\kappa x}{\sqrt{3-\kappa^2}}\right), \end{aligned} \tag{5.9}$$

where κ is defined as

$$\kappa := -h_2'/\sqrt{h_{22}''} = -\sqrt{3}\rho.$$

Note that $\kappa = 1$ if the covariance function is Gaussian (Cheng and Schwartzman, 2018), and the corresponding $\rho = -1/\sqrt{3}$.

5.2.3 The boundary case: $|\rho(t)| = 1$

The random vector $(X(t), X'(t), X''(t))$ is degenerate when $|\rho(t)| = 1$ violating the assumption (C2) in Cheng and Schwartzman (2015). Therefore, the technique used in the proof of Theorem 5 needs to be modified for the boundary case.

Proposition 5. *When $\rho(t) = -1$, the peak height distribution of $X(t)$ is* Rayleigh$(\tilde{\sigma}(t))$ *with density*

$$f_t(x) = \frac{\sqrt{2\pi}x}{\tilde{\sigma}^2(t)}\phi\left(\frac{x}{\tilde{\sigma}(t)}\right), \quad x \geq 0. \tag{5.10}$$

The mean and variance are $\sqrt{\pi/2}\tilde{\sigma}(t)$ and $(2-\pi/2)\tilde{\sigma}^2(t)$ respectively.

When $\rho(t) = 1$, the peak height distribution of $X(t)$ is $-$Rayleigh$(\tilde{\sigma}(t))$ *with density*

$$f_t(x) = -\frac{\sqrt{2\pi}x}{\tilde{\sigma}^2(t)}\phi\left(\frac{x}{\tilde{\sigma}(t)}\right), \quad x \leq 0. \tag{5.11}$$

The mean and variance are $-\sqrt{\pi/2}\tilde{\sigma}(t)$ and $(2-\pi/2)\tilde{\sigma}^2(t)$ respectively.

Example 1. Define the Cosine process

$$X(t) = c_1 z_1 \cos(\omega t) + c_2 z_2 \sin(\omega t), \tag{5.12}$$

where z_1 and z_2 are independent, standard Gaussian random variables and c_1, c_2, and ω are positive constants. $X(t)$ has non-constant variance and is therefore non-stationary. It can be derived by (5.5) and (5.7) that $\rho(t) = -1$ and $\tilde{\sigma}(t) = c_1 c_2/\sqrt{c_1^2\sin^2(\omega t) + c_2^2\cos^2(\omega t)} =$

$c_1c_2/\sigma(t)$. The peak height density is Rayleigh

$$f_t(x) = \frac{\sqrt{2\pi}[c_1^2\sin^2(\omega t) + c_2^2\cos^2(\omega t)]x}{c_1^2c_2^2}\phi\left(\frac{\sqrt{c_1^2\sin^2(\omega t) + c_2^2\cos^2(\omega t)}x}{c_1c_2}\right), \quad x \geq 0.$$

Figure 5.3a shows simulated instances of the cosine process ($\omega = 2$, $c_1 = 3$ and $c_2 = 4$) on a grid ranging from 0 to 2π with a step size of 0.002π. In Figure 5.3c, we compare the peak height of this process near $t = \pi/4$ (large $\tilde{\sigma}(t)$) and $t = \pi/2$ (small $\tilde{\sigma}(t)$). The parameter $\tilde{\sigma}(t)$ is the only parameter of the peak height distribution, and has a positive effect on both the mean and standard deviation, as displayed in Figure 5.3d.

5.2.4 The Gaussian process with non-constant bandwidth

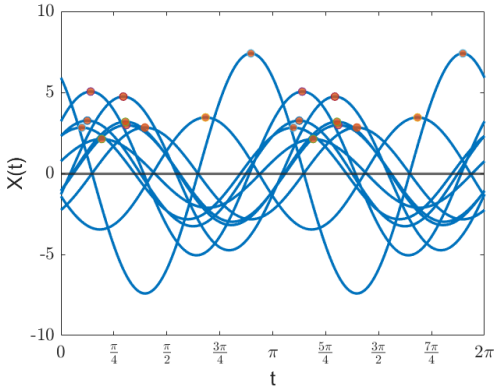
The parameter $\rho(t)$ is the only parameter that controls how the peak height distribution varies over the domain when the field $X(t)$ has constant variance. While constant $\rho(t)$ is necessary for stationarity, non-stationarity does not necessarily imply non-constant $\rho(t)$. Next, we provide an example of a non-stationary Gaussian process that has constant $\rho(t)$.

We have discussed the Gaussian process with non-constant variance. Another type of non-stationary Gaussian process is the Gaussian process with spatially varying corelation. Gaussian processes with spatially varying correlation are commonly used (see e.g. Paciorek and Schervish, 2003; Remes et al., 2017) to model data whose smoothness varies with t.

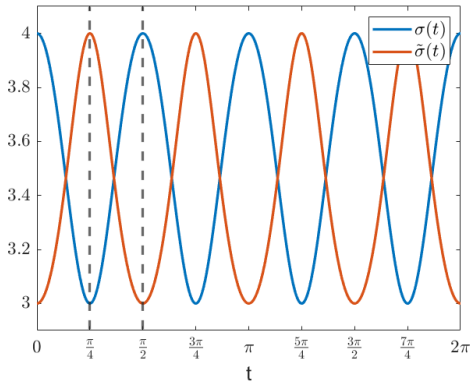
It is standard to define a stationary Gaussian process by convolving a white noise process with a kernel function that specifies the correlation structure:

$$Z(t) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{\nu}} k\left(\frac{t-s}{\nu}\right) dB(s), \tag{5.13}$$

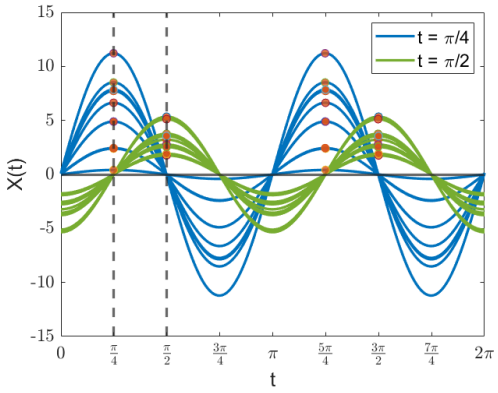
where $k(s)$, $s \in \mathbb{R}$, is a smooth kernel with $\int k(s)^2 ds = 1$ so that $Z(t)$ is unit-variance, $dB(s)$ is Gaussian white noise and ν is the bandwidth. In (5.13), the kernel bandwidth ν controls the smoothness and spatial correlation. We define the Gaussian process with non-constant bandwidth



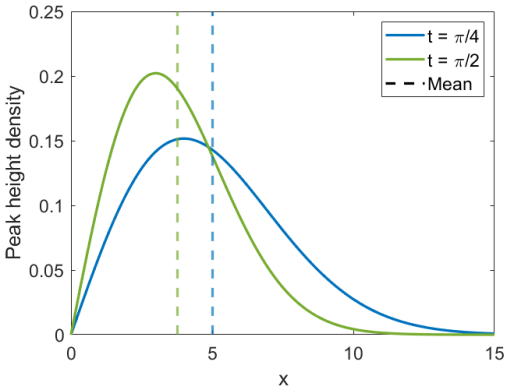
(a) Simulated instances with orange dots indicating the local maxima



(b) $\sigma(t)$ and $\tilde{\sigma}(t)$



(c) Selected simulated instances having a peak near $t = \pi/4$ versus those having a peak near $t = \pi/2$



(d) Peak height distribution at $t = \pi/4$ versus $t = \pi/2$

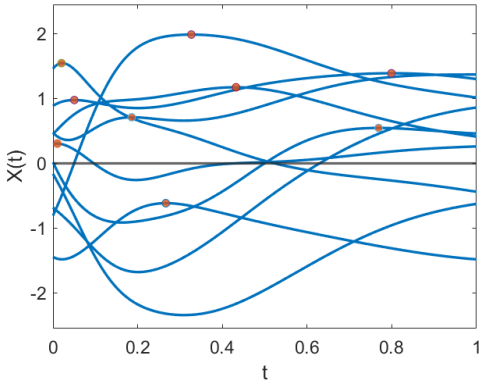
Figure 5.3. Simulated instances of the Cosine process in Example 1, $\sigma(t)$, $\tilde{\sigma}(t)$ and the peak height distribution at $t = \pi/4$ (large $\tilde{\sigma}(t)$) versus $t = \pi/2$ (small $\tilde{\sigma}(t)$).

as a generalization of (5.13)

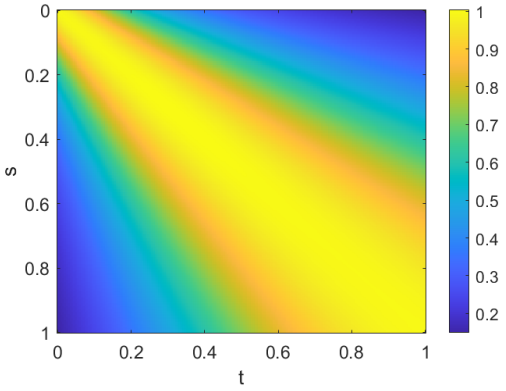
$$X(t) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{\nu(t)}} k\left(\frac{t-s}{\nu(t)}\right) dB(s), \tag{5.14}$$

where the bandwidth function $\nu(t)$, $t \in \mathbb{R}$, is a positive non-constant smooth function. The Gaussian process with non-constant bandwidth is non-stationary since the spatial correlation varies with t. Here, the peak height distribution depends solely on $\rho(t)$, since this process has unit variance.

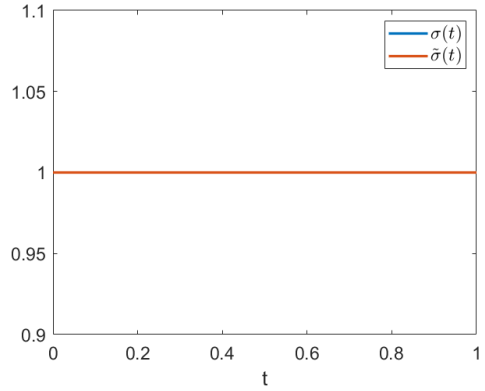
In Figure 5.4a, we show simulated instances of the Gaussian process with linear bandwidth function $\nu(t) = 0.5t + 0.1$ on a grid ranging from 0 to 1 with a step size of 0.005. From Figure 5.4b, we can clearly see the spatial correlation is dependent on the location. More specifically, since $\nu(t)$ is monotone increasing, the process becomes smoother as t increases, as illustrated in the simulated instances and the correlation plot. Furthermore, it is important to note that in this case, $\rho(t)$ is independent of t (see Figure 5.4d), even though the field is non-stationary. In the following result, we show that $\rho(t)$ is constant when $\nu(t)$ is linear.



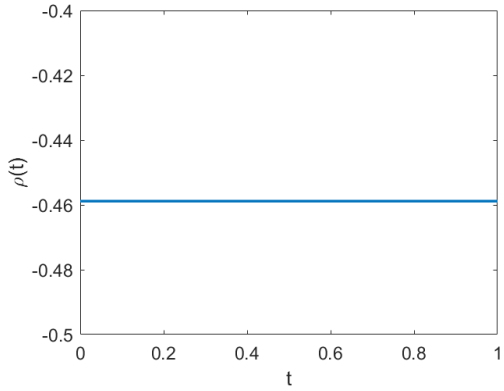
(a) Simulated instances with orange dots indicating the local maxima



(b) Correlation



(c) $\sigma(t)$ and $\tilde{\sigma}(t)$



(d) $\rho(t)$

Figure 5.4. Simulated instances of the Gaussian process with non-constant bandwidth, its spatial correlation, and the parameters $\sigma(t)$, $\tilde{\sigma}(t)$, and $\rho(t)$.

Theorem 6. *Consider the non-stationary Gaussian process*

$$X(t) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{v(t)}} k\left(\frac{t-s}{v(t)}\right) dB(s),$$

where $k(s)$, $s \in \mathbb{R}$, is a smooth kernel with $\int k(s)^2 ds = 1$. If the bandwidth function $v(t)$ is linear in t, then the parameter $\rho(t) = \mathrm{Cor}(X(t), X''(t) | X'(t) = 0)$ is independent of t.

Theorem 6 also indicates the peak height distribution is independent of t when the bandwidth function $v(t)$ is linear. This property is particularly interesting, considering the field $X(t)$ is non-stationary, and helps to reduce the computation time of evaluating the peak height distribution over a certain domain.

5.2.5 The Gaussian process with non-constant bandwidth: Gaussian kernel

As an example, we demonstrate the explicit calculation of $\rho(t)$ when the kernel function $k(\cdot)$ in (5.14) is Gaussian, i.e.

$$X(t) = \int_{-\infty}^{\infty} \frac{\sqrt{2}\pi^{1/4}}{\sqrt{v(t)}} \phi\left(\frac{t-s}{v(t)}\right) dB(s),$$

where $\phi(x)$ is the standard Gaussian density.

We first compute $\mathrm{Var}(X'(t))$, $\mathrm{Var}(X''(t))$ and $\mathbb{E}[X'(t)X''(t)]$ and use (5.8) to get $\rho(t)$. Taking the first and second order derivative of $X(t)$, we obtain

$$X'(t) = \sqrt{2}\pi^{1/4} \int_{-\infty}^{\infty} \left(-\frac{v'(t)}{2v(t)^{3/2}} - \frac{t-s}{v(t)^{5/2}} + \frac{3(t-s)^2 v'(t)}{v(t)^{7/2}} \right) \phi\left(\frac{t-s}{v(t)}\right) dB(s),$$

and

$$X''(t) = \sqrt{2}\pi^{1/4} \int_{-\infty}^{\infty} \left(\frac{v'(t)^2}{v(t)^{13/2}}(t-s)^4 - \frac{2v'(t)}{v(t)^{11/2}}(t-s)^3 + \left(\frac{v''(t)}{v^{7/2}} + \frac{2-9v'(t)^2}{2v(t)^{9/2}} \right) (t-s)^2 \right.$$

$$+\frac{5v'(t)}{v(t)^{7/2}}(t-s)-\frac{v''(t)}{2v(t)^{3/2}}+\frac{3v'(t)^2}{4v(t)^{5/2}}-\frac{1}{v(t)^{5/2}}\Bigg)\phi\left(\frac{t-s}{v(t)}\right)dB(s).$$

$\mathrm{Var}(X'(t))$, $\mathrm{Var}(X''(t))$ and $\mathbb{E}[X'(t)X''(t)]$ can be computed using the property of Wiener integral:

$$\begin{aligned}\mathrm{Var}(X'(t)) &= \frac{1}{\pi^{1/2}}\int_{-\infty}^{\infty}\left(-\frac{v'(t)}{2v(t)^{3/2}}-\frac{t-s}{v(t)^{5/2}}+\frac{3(t-s)^2v'(t)}{v(t)^{7/2}}\right)^2 e^{-\frac{(t-s)^2}{v(t)^2}}ds\\ &= \frac{1+v'(t)^2}{2v(t)^2}.\end{aligned}$$

Similarly

$$\mathrm{Var}(X''(t)) = \frac{51v'(t)^4+68v'(t)^2+12}{16v(t)^4}-\frac{v''(t)(3v'(t)^2-2)}{2v(t)^3}+\frac{v''(t)^2}{2v(t)^2},$$

and

$$\mathbb{E}[X'(t)X''(t)] = -\frac{v'(t)^3+v'(t)}{2v(t)^3}+\frac{v'(t)v''(t)}{2v(t)^2}.$$

Applying (5.8), we have

$$\begin{aligned}\rho(t) &= \frac{-1}{2}\frac{v'(t)^2+1}{\sqrt{\frac{51}{16}v'(t)^4+(\frac{68}{16}-\frac{3v''(t)v(t)}{2})v'(t)^2+\frac{3}{4}+v''(t)v(t)+\frac{1}{2}v''(t)^2v(t)^2-\frac{(-v'(t)^3+(v''(t)v(t)-1)v'(t))^2}{2(v'(t)^2+1)}}}\\ &= \frac{-1}{2}\sqrt{\frac{(v'(t)^2+1)^3}{\frac{43}{16}v'(t)^6+(\frac{103}{16}-\frac{1}{2}v''(t)v(t))v'(t)^4+\frac{9}{2}v'(t)^2+\frac{3}{4}+v''(t)v(t)+\frac{1}{2}v''(t)^2v(t)^2}}.\end{aligned}\tag{5.15}$$

It is straightforward from (5.15) that when $v(t)$ is linear in t, say $v(t)=v't+v_0$, $\rho(t)$ is constant:

$$\rho(t) = \frac{-1}{2}\sqrt{\frac{(v'^2+1)^3}{\frac{43}{16}v'^6+\frac{103}{16}v'^4+\frac{9}{2}v'^2+\frac{3}{4}}}.$$

5.2.6 The Gaussian process with non-constant bandwidth and variance

The parameter $\rho(t)$ is non-positive when the field has constant variance, but can be positive in general. To give an example of positive $\rho(t)$, we generalize (5.14) to have non-constant variance:

$$\begin{aligned} Y(t) &= \sigma(t)X(t) \\ &= \sigma(t)\int_{-\infty}^{\infty} \frac{1}{\sqrt{\nu(t)}} k\left(\frac{t-s}{\nu(t)}\right) dB(s), \end{aligned} \tag{5.16}$$

where the standard deviation function $\sigma(t)$, $t \in \mathbb{R}$, is a positive non-constant smooth function so that $\mathrm{Var}(Y(t)) = \sigma(t)^2$. We consider the following specification of $Y(t)$

$$\begin{aligned} k(s) &= \sqrt{2}\pi^{1/4}\phi(s), \\ \nu(t) &= 0.5t + 0.1, \\ \sigma(t) &= 8t^2 - 10t + 6. \end{aligned} \tag{5.17}$$

Simulated instances of this process on a grid ranging from 0 to 1 with a step size of 0.005 are displayed in Figure 5.5a. Taking first and second derivatives on both sides of (5.16), we have $Y'(t) = \sigma'(t)X(t) + \sigma(t)X'(t)$ and $Y''(t) = \sigma''(t)X(t) + 2\sigma'(t)X'(t) + \sigma(t)X''(t)$. We use the variances and covariances of the first and second order derivatives of $X(t)$ (see 5.2.5) as a bridge for computing those of $Y(t)$:

$$\begin{aligned} \mathrm{Var}(Y'(t)) =& \sigma'(t)^2 + \sigma(t)^2\mathrm{Var}(X'(t)), \\ \mathrm{Var}(Y''(t)) =& \sigma''(t)^2 + (4\sigma'(t)^2 - 2\sigma(t)\sigma''(t))\mathrm{Var}(X'(t)) \\ &+ \sigma(t)^2\mathrm{Var}(X''(t)) + 4\sigma(t)\sigma'(t)\mathbb{E}[X'(t)X''(t)], \\ \mathbb{E}[Y(t)Y'(t)] =& \sigma(t)\sigma'(t), \\ \mathbb{E}[Y(t)Y''(t)] =& \sigma(t)\sigma''(t) - \sigma(t)^2\mathrm{Var}(X'(t)), \end{aligned}$$

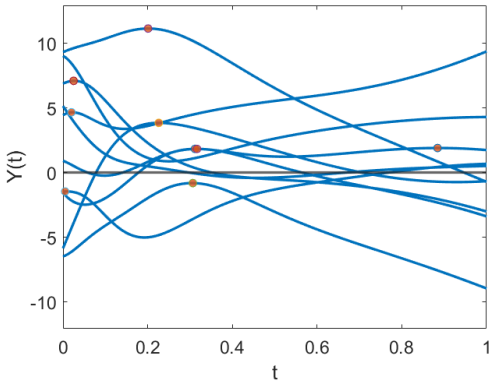
$$\mathbb{E}[Y'(t)Y''(t)] = \sigma'(t)\sigma''(t) + \sigma(t)\sigma'(t)\mathrm{Var}(X'(t)) + \sigma(t)^2\mathbb{E}[X'(t)X''(t)].$$

Then $\rho(t)$ and $\tilde{\sigma}(t)$ can be obtained by applying (5.5) and (5.7). As displayed in Figure 5.5c, $\rho(t)$ reaches its minimum -0.54 at around $t = 0.2$ and maximum 0.38 at around $t = 0.7$. Taking a closer look at $t = 0.2$ and 0.7 (Figure 5.6a), we compare the peak height distribution at $t = 0.2$ (negative $\rho(t)$) versus $t = 0.7$ (positive $\rho(t)$) in Figure 5.6b. Given that the field has zero mean for all t, and in particular at $t = 0.7$, it is quite surprising that both the mean and median of the peak height distribution are negative. In conclusion, for a general Gaussian process, the parameter $\rho(t)$ can have either a positive or negative value depending on the specification of the field and the location t.

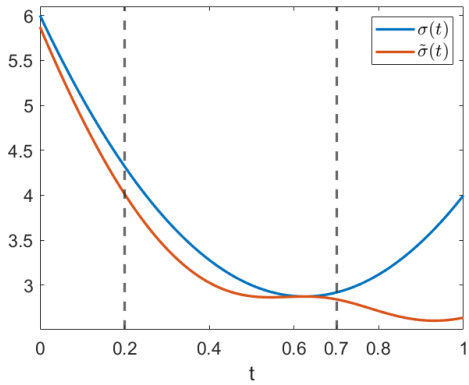
5.3 Multidimensional non-stationary Gaussian random fields

We have explored the peak height distribution of 1D non-stationary Gaussian processes. However, many real-world problems exist in higher dimensions. For example, the RFT-based brain imaging analysis requires the model to be 3D due to the nature of fMRI data (see Worsley et al., 1992; Worsley et al., 1996b). As we step into the realm of higher dimensional non-stationary Gaussian random fields, the complexity of the Kac-Rice formula increases drastically, making direct evaluation difficult.

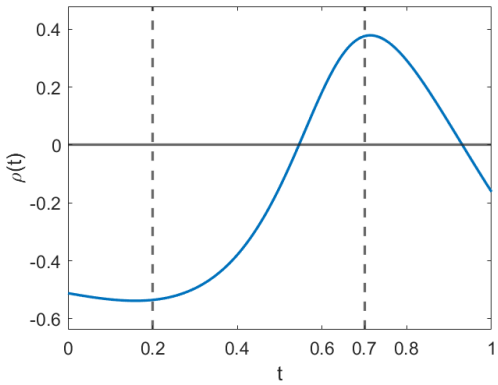
In this section, with the results we have developed in 1D as a foundation, we explore the properties of the peak height distribution of an important multidimensional non-stationary Gaussian random field: the scale space Gaussian random field. Although these properties are not enough to allow us to derive the explicit formula for the peak height distribution, they are helpful for a better understanding of the spatial structure of these non-stationary random fields. Moreover, these properties effectively increase the efficiency of the numerical methods we highlight later in this chapter (see Section 5.5).



(a) Simulated instances with orange dots indicating the local maxima



(b) $\sigma(t)$ and $\tilde{\sigma}(t)$



(c) $\rho(t)$

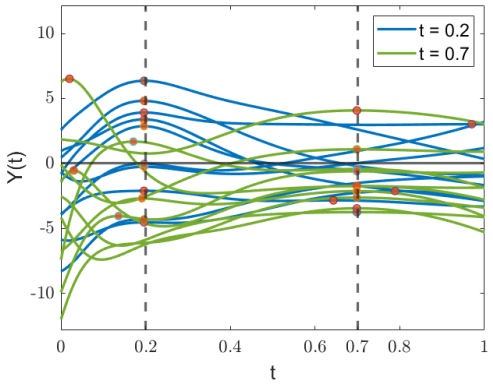
Figure 5.5. Simulated instances of the Gaussian process with non-constant bandwidth and variance, and the parameters $\sigma(t)$, $\tilde{\sigma}(t)$, and $\rho(t)$.

5.3.1 The scale space Gaussian random field

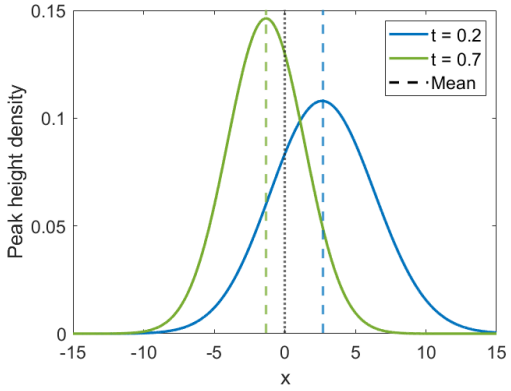
First, we consider the following zero-mean unit-variance 2D scale space Gaussian random field

$$X(t,v) = \int \frac{1}{\sqrt{v}} k\left(\frac{s-t}{v}\right) dB(s), \tag{5.18}$$

where $k(s)$, $s \in R$, is a smooth kernel with $\int k(s)^2 ds = 1$. In the scale space field, the smoothing bandwidth is not fixed and is treated as a parameter, so that the parameter space (t,v) is 2D. Solving the peak height distribution of a scale space field helps to detect peaks of unknown location and scale.



(a) Selected simulated instances having a peak near $t = 0.2$ versus those having a peak near $t = 0.7$



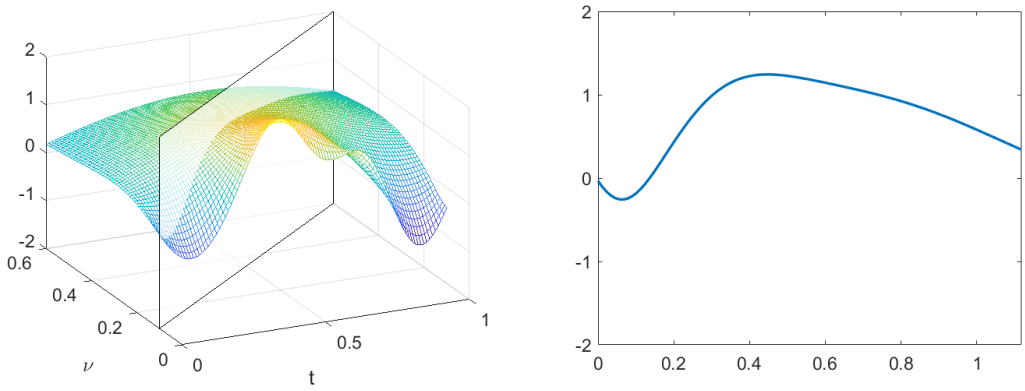
(b) Peak height distribution at $t = 0.2$ versus $t = 0.7$

Figure 5.6. The peak height distribution of the Gaussian process with non-constant bandwidth and variance at $t = 0.2$ (negative $\rho(t)$) versus $t = 0.7$ (positive $\rho(t)$).

If we draw a slice plane $v = v(t)$ for the 2D scale space field, the slice along the plane is a 1D Gaussian process with non-constant bandwidth, the same process studied in Section 5.2.4 above. An example of the 2D scale space field and the slice plane is displayed in Figure 5.7a, and the slice along the plane is displayed in Figure 5.7b. The 2D scale space field can be seen as an extension of the Gaussian process with non-constant bandwidth.

For the Gaussian process with non-constant bandwidth, we have demonstrated that when the bandwidth $v(t)$ is a linear function of t, the peak height distribution remains the same over the domain. Therefore, the peak height distribution of the 1D process obtained by slicing a 2D scale space field with a flat plane is independent of t. Next, we generalize this result to 2D scale space fields and, even further, to scale space fields of any dimension.

To evaluate the Kac-Rice formula and compute the peak height distribution of the scale space field, we once again need the variances and covariances of the first and second order derivatives. To simplify the computation, we let $v = -\log v$, so that $X(t,v)$ is stationary in v for any fixed t, and $X(t,v)$ is also stationary in t for any fixed v (Siegmund and Worsley, 1995). We



(a) $X(t,v)$ and a slice plane $v(t) = 0.5t + 0.1$ **(b)** Slice along the plane

Figure 5.7. A single instance of the 2D scale space field and a slice plane $v(t) = 0.5t + 0.1$.

replace v in (5.18) with e^{-v} and the covariance of X after the change of variable is

$$\mathbb{E}[X(t_1,v_1)X(t_2,v_2)] = e^{(v_1+v_2)/2} \int k((s-t_1)e^{v_1})k((s-t_2)e^{v_2})ds. \tag{5.19}$$

Define

$$\nabla X = \left(\frac{\partial X}{\partial t}, \frac{\partial X}{\partial v}\right)^{\mathrm{T}}, \quad \nabla^2 X = \begin{pmatrix} \frac{\partial^2 X}{\partial t^2} & \frac{\partial^2 X}{\partial t \partial v} \\ \frac{\partial^2 X}{\partial t \partial v} & \frac{\partial^2 X}{\partial v^2} \end{pmatrix},$$

and their joint distribution

$$\begin{pmatrix} \nabla X \\ \mathrm{vec}(\nabla^2 X) \end{pmatrix} \sim N\left(0, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}\right).$$

The variances and covariances can be obtained using (5.5.5) in Adler and Taylor (2007).

$$\Sigma_{11} = \begin{pmatrix} e^{2v}\int k'(s)^2 ds & 0 \\ 0 & \int (\frac{1}{2}k(s) + sk'(s))^2 ds \end{pmatrix},$$

$$\Sigma_{22} = \begin{pmatrix} e^{4v}\int k''(s)^2 ds & e^{2v}\int k''(s)\eta(s)ds & 0 \\ e^{2v}\int k''(s)\eta(s)ds & \int \eta(s)^2 ds & 0 \\ 0 & 0 & e^{2v}\int (\frac{3}{2}k'(s)+sk''(s))^2 ds \end{pmatrix},$$

$$\Sigma_{21} = \begin{pmatrix} 0 & e^{2v}\int k''(s)(\frac{1}{2}k(s)+sk'(s))ds \\ 0 & \int (\frac{1}{2}k(s)+sk'(s))\eta(s)ds \\ e^{2v}\int (\frac{3}{2}k'(s)+sk''(s))(\frac{1}{2}k(s)+sk'(s))ds & 0 \end{pmatrix},$$

where $\eta(s) = 1/4k(s)+2sk'(s)+s^2k''(s)$. Using the properties of conditional normal distribution, we obtain

$$\nabla^2 X|\nabla X = 0 \sim N(0, \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}),$$

and

$$\nabla^2 X|X = x, \nabla X = 0 \sim N(\tilde{\Sigma}_{21}\tilde{\Sigma}_{11}^{-1}(x,0,0)', \Sigma_{22} - \tilde{\Sigma}_{21}\tilde{\Sigma}_{11}^{-1}\tilde{\Sigma}_{12}),$$

where

$$\tilde{\Sigma}_{11} = \begin{pmatrix} 1 & 0 \\ 0 & \Sigma_{11} \end{pmatrix},$$

$$\tilde{\Sigma}_{21} = \begin{pmatrix} -e^{2v}\int k'(s)^2 ds & \\ \frac{1}{4} - \int s^2 k'(s)^2 ds & \Sigma_{21} \\ 0 & \end{pmatrix}.$$

5.3.2 The scale space Gaussian random field: Gaussian kernel

Consider the special case when the kernel function is Gaussian, i.e.

$$k(s) = \sqrt{2}\pi^{1/4}\phi(s).$$

The covariance matrices can be further simplified as

$$\Sigma_{11} = \begin{pmatrix} \frac{1}{2}e^{2v} & 0 \\ 0 & \frac{1}{2} \end{pmatrix}, \quad \Sigma_{22} = \begin{pmatrix} \frac{3}{4}e^{4v} & \frac{1}{4}e^{2v} & 0 \\ \frac{1}{4}e^{2v} & \frac{7}{4} & 0 \\ 0 & 0 & \frac{5}{4}e^{2v} \end{pmatrix}, \quad \Sigma_{21} = \begin{pmatrix} 0 & -\frac{1}{2}e^{2v} \\ 0 & 0 \\ \frac{1}{2}e^{2v} & 0 \end{pmatrix},$$

and

$$\nabla^2 X | \nabla X = 0 \sim N\left(0, \begin{pmatrix} \frac{1}{4}e^{4v} & \frac{1}{4}e^{2v} & 0 \\ \frac{1}{4}e^{2v} & \frac{7}{4} & 0 \\ 0 & 0 & \frac{3}{4}e^{2v} \end{pmatrix}\right).$$

A pattern can be observed from these matrices: the variance of the derivative with respect to v is constant, and the variance of the derivative with respect to t has the form of a constant multiplied by e^{2v}. The second-order derivatives show a similar pattern. This pattern is crucial for proving the peak height distribution of the scale space field independent of t and v.

5.3.3 Peak height distribution of the scale space Gaussian random field

The following result is a generalization of Theorem 6 for the scale space field.

Theorem 7. *Consider the (N+1)-dimensional scale space Gaussian random field*

$$X(t, v) = \int \frac{1}{\sqrt{v^N}} k\left(\frac{s-t}{v}\right) dB(s), \tag{5.20}$$

where $k(s)$, $s \in R^N$, is an isotropic smooth kernel with $\int k(s)^2 ds = 1$. The peak height distribution of $X(t, v)$ is independent of (t, v).

Theorem 7 demonstrates the peak height distribution of the scale space field remains the same across the domain although it is non-stationary. This discovery is helpful both in theoretical calculation of the peak height distribution and in application. In theoretical calculation of the peak height distribution, we can choose some specific v that helps to simplify the evaluation of the Kac-Rice formula, as the final result is independent of v. In application, this property

vastly reduces the running time, particularly when dealing with multidimensional data like fMRI images.

5.4 Numerical implementation of the Kac-Rice formula

Gaussian random field theory is widely used as a statistical framework in the analysis of fMRI images especially when we are dealing with thresholding problems (Worsley et al., 1996b; Genovese et al., 2002). As we mentioned, the main bottleneck for peak inference in fMRI analysis using RFT is the model assumptions (Chumbley et al., 2010) like the stationarity assumption. Making the stationarity assumption is common in practice (Cheng and Schwartzman, 2017) to simplify the derivation of the analytical formula for the peak height distribution, but sometimes, for example, if the images involve different regions of the brain, the stationarity assumption would be unsuitable. For datasets exhibiting signs of violating the stationarity assumption, instead of making false assumptions, we should explore ways to solve the problem more generally.

5.4.1 Algorithm

In Section 5.2, we have derived the peak height distribution formula for 1D non-stationary Gaussian processes. When it comes to multidimensional anisotropic and non-stationary Gaussian random fields, evaluating the Kac-Rice formula is challenging. We here present two numerical methods: Algorithm 1 and 2 for evaluating the Kac-Rice formula (5.1). Although in practice, the numerical methods are convenient and accurate for a sufficiently large number of iterations, we still recommend exploring the properties of the peak height distribution first to fully understand its spatial structure and increase the algorithm efficiency. Theorem 7 is a good example. The peak height distribution for the scale space field is independent of the location and scale parameter. This allows us to compute the peak height distribution for a single point and generalize it to the entire domain.

Algorithm 1 is a faster method compared to Algorithm 2, and is recommended for

Algorithm 1. Numerical Kac-Rice

Input

fv — the variance of the field

dv — the d by d variance-covariance matrix of the first derivatives of the field

d2v — the $d*(d+1)/2$ by $d*(d+1)/2$ variance-covariance matrix of the second derivatives of the field

fdcov — the 1 by d covariance matrix between the field and its first derivatives

fd2cov — the 1 by $d*(d+1)/2$ covariance matrix between the field and its second derivatives

dd2cov — the d by $d*(d+1)/2$ covariance matrix between the first and second derivatives of the field

u — a vector giving the thresholds

niters — the number of iterations used to calculate the expectation

Ensure: the variance-covariance matrix is positive definite

Compute the covariance matrix $\mathbf{\Sigma}$ of $(X, \nabla^2 X)|\nabla X = 0$

Compute the probability density $p_{\nabla X}$ of ∇X

Generate i.i.d. sample (X_i, HessianMat$_i$) (i = 1...niters) from the multivariate Gaussian distribution $N(\mathbf{0}, \mathbf{\Sigma})$

for `i = 1...niters` **do**

Compute det(HessianMat$_i$)

end for

for `j = 1...length(u)` **do**

Compute KRexpectation[j] $= \frac{1}{\text{niters}} \sum |\det(\text{HessianMat}_i)| \mathbb{1}_{\text{HessianMat}_i \prec 0} \mathbb{1}_{X_i > u[j]}$

end for

Update KRexpectation by multiplying $p_{\nabla X}(0)$

return KRexpectation

computing the peak height distribution over a range of u. Algorithm 1, however, is not optimal for large threshold u which is common in the peak detection setting. This is due to the indicator term $\mathbb{1}_{\{X(t)>u\}}$ in (5.1). As $u \to -\infty$, the indicator has no effect, so that all samples of $X(t)$ contribute to the numerical approximation. However, as $u \to \infty$, the probability of $X(t) > u$ shrinks to zero, leading to a significantly smaller effective sample size compared to the number of simulated instances. If we take the 3D scale space field as an example and set $u = 3.91$ which is the 99$^{\text{th}}$ percentile of the peak height distribution. It can be shown that the effective sample size is only about $1 - \Phi(3.91) = 0.005\%$ of the total number of simulated instances. In order to achieve the same accuracy, we need to sample about $1/0.005\% = 20000$ times more data for $u = 3.91$ compared to $u = -\infty$ which is computationally intensive. In Algorithm 2, we solve this

issue by computing the numerator of (5.1) as

$$\begin{aligned}&\mathbb{E}[|\det\nabla^2 X(t)|\mathbb{1}_{\{\nabla^2 X(t)\prec 0\}}\mathbb{1}_{\{X(t)>u\}}|\nabla X(t)=0]\\ =&\mathbb{E}[|\det\nabla^2 X(t)|\mathbb{1}_{\{\nabla^2 X(t)\prec 0\}}|X(t)>u,\nabla X(t)=0]p(X(t)>u|\nabla X(t)=0).\end{aligned}$$

We generate $X(t)$ from a truncated normal distribution to bound $X(t)$ from below by u, and therefore $X(t)$ has to be resampled for different u making it slower compared to Algorithm 1 but having a better accuracy for large u.

Algorithm 2. Numerical Kac-Rice for large u

Input
Same as Algorithm 1
Compute the covariance matrix $\mathbf{\Sigma}$ of $\nabla^2 X|X,\nabla X=(0,0)$
Compute the probability density $p_{\nabla X}$ of ∇X
Compute the tail distribution $P_{X|\nabla X=0}$ of $X|\nabla X=0$
Generate i.i.d. sample HessianMat$_i$ (i = 1...niters) from the multivariate Gaussian distribution $N(\mathbf{0},\mathbf{\Sigma})$
for `j = 1...length(u)` **do**
Generate i.i.d. sample XTruncated$_i$ (i = 1...niters) from a truncated Gaussian distribution $X|X>u[j]$.
for `i = 1...niters` **do**
Update HessianMat$_i$ by adding a correction term (the conditional mean $\mathbb{E}[\nabla^2 X|X,\nabla X=(\text{XTruncated}_i,0)]$) so that the updated term HessianMatNew$_i$ has the same distribution as $\nabla^2 X|X,\nabla X=(\text{XTruncated}_i,0)$
Compute det(HessianMatNew$_i$)
end for
Compute KRexpectation[j] $=\frac{1}{\text{niters}}\sum|\det(\text{HessianMatNew}_i)|\mathbb{1}_{\text{HessianMatNew}_i\prec 0}$
Update KRexpectation[j] by multiplying $p_{\nabla X}(0)P_{X|\nabla X=0}(u[j])$
end for
return KRexpectation

The most difficult part of the Kac-Rice formula (5.1) is to compute the expectation of the determinant of the Hessian matrix $\nabla^2 X(t)$ given the field $X(t)$ and the gradient $\nabla X(t)$. Deriving the joint distribution of ($X(t)$, $\nabla X(t)$, $\nabla^2 X(t)$) is usually easy especially when the field is defined with convolution. Our numerical algorithms solely require the computation of the joint distribution of ($X(t)$, $\nabla X(t)$, $\nabla^2 X(t)$), and evaluate the expectation by simulation.

5.4.2 Relative efficiency

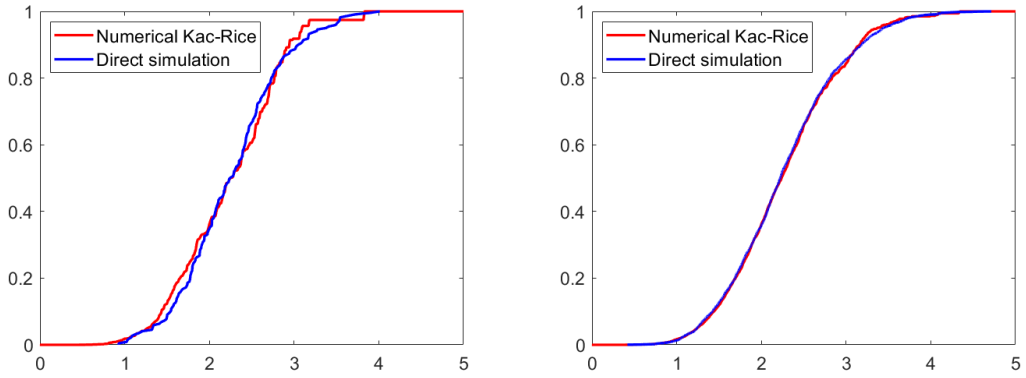
A common and straightforward way of estimating the peak height distribution is direct simulation. This is achieved by generating a sufficiently large number of instances of the random field of interest on a discrete lattice, and recording the height of all observed local maxima from these simulated instances. The empirical distribution of the simulated peak heights can be used to approximate the true peak height distribution. Direct simulation is easy to implement but requires a significant amount of time to run. The low efficiency can be attributed to two primary factors. First, generating the field itself provides more information than we actually need to estimate the peak height distribution. Second, the performance heavily depends on the granularity and the size of the grid and when we use a discrete lattice to approximate the continuous domain.

To compare the performance between direct simulation and the numerical Kac-Rice algorithm, we apply these two approaches to estimate the peak height distribution of a scale space field while controlling the number of simulations. In the direct simulation, the 3D scale space field is generated from (5.20) over a grid of size $20 \times 20 \times 20$ pixels with scale parameter ν ranging from 0.2 to 1.2. For the numerical Kac-Rice method, we apply Algorithm 1. The comparison of runtime for the two methods is presented in Table 5.1, and the simulated CDF is displayed in Figure 5.8.

Table 5.1. Runtime of direct simulation versus numerical Kac-Rice algorithm for estimating the peak height distribution of a scale space field.

# of simulations	10,000	100,000
Direct simulation	1,306.42s (221 peaks)	13,293.85s (2441 peaks)
Numerical Kac-Rice	0.20s	1.50s

As demonstrated in Table 5.1 and Figure 5.8, our numerical Kac-Rice algorithm performs comparably to direct simulation with much less runtime. Neither of the two methods can produce a smooth simulated CDF curve for 10,000 simulations, and when the number of simulations increases to 100,000, they both perform well. Therefore, if estimating the peak height distribution



(a) 10,000 simulations **(b)** 100,000 simulations

Figure 5.8. Simulated CDF of the 3D scale space field, direct simulation versus numerical Kac-Rice.

is the only goal, the numerical Kac-Rice algorithm is preferable. This is the case when we explore the theoretical properties of the peak height distribution, such as how different parameters affect the CDF.

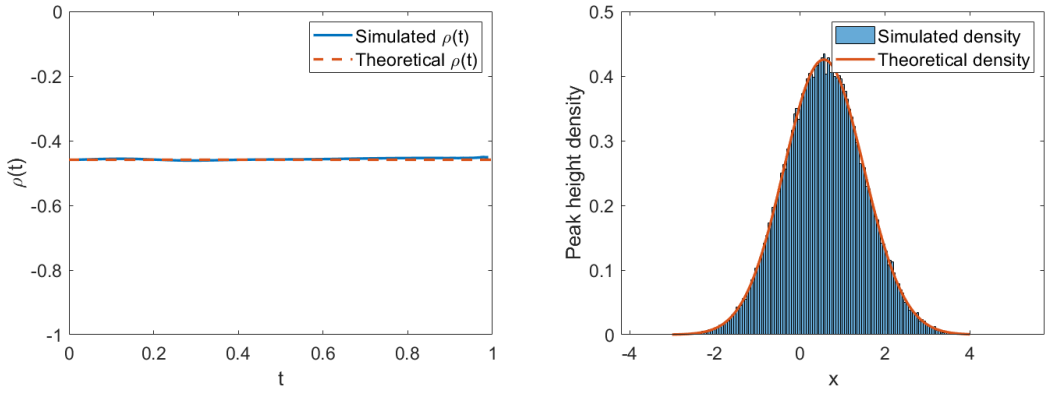
5.5 Validation via simulations

In this section, we present simulation results to validate the theoretical properties of the Gaussian process with non-constant bandwidth and the scale space field.

5.5.1 The Gaussian process with non-constant bandwidth

To validate Theorem 5 and 6, we generate $100,000$ Gaussian processes with non-constant bandwidth. t is sampled from a grid of $[0,1]$ containing 200 points. As defined in (5.14), each process is obtained as the convolution of Gaussian white noise with a Gaussian kernel of non-constant bandwidth function $v(t) = 0.5t + 0.1$. Simulated instances of the process are displayed in Figure 5.4a.

Since $v(t) = 0.5t + 0.1$ is a linear function of t, according to Theorem 6, both the parameter $\rho(t)$ and the peak height distribution should be independent of t. Figure 5.9a shows strong evidence that $\rho(t)$ is constant and the simulated value matches the theoretical value.



(a) $\rho(t)$ **(b)** Peak height density

Figure 5.9. 1D simulation: $\rho(t)$ and the peak height density of the Gaussian process with non-constant bandwidth.

Figure 5.9b demonstrates that the theoretical density curve computed using (5.3) at a single location closely resembles the empirical peak height density over the entire domain.

5.5.2 3D scale space Gaussian random field

Theorem 7 states that the peak height distribution of the scale space field does not vary with respect to the location parameter t and scale parameter v. To validate this, we apply the numerical Kac-Rice algorithm to compare the peak height distribution of a 3D scale space field with different t and v. The simulation setup considers Gaussian kernel and three different scenarios, each with a different set of parameters. For each scenario, we perform 1,000,000 simulations. The input of the algorithm, which are the variances and covariances of the first and second order derivatives, is derived in 5.3.2. The parameter choices for the three scenarios are listed in Table 5.2.

Table 5.2. Parameter choices for the three simulation scenarios to validate Theorem 7.

Parameter	Scenario 1	Scenario 2	Scenario 3
t_1	0.5	0.5	1
t_2	0.5	0.5	0.5
v	0.7	0.2	0.7

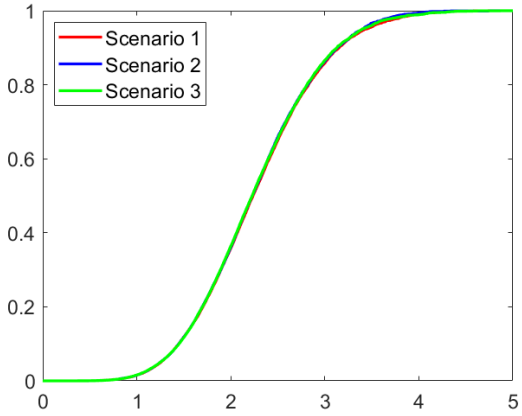


Figure 5.10. Simulated CDF of the 3D scale space field under the three scenarios described in Table 5.2.

As we can see from Figure 5.10, all three scenarios lead to the same peak height distribution. Given the large number of simulations, this provides strong evidence to support Theorem 7.

5.6 Discussion

5.6.1 Peak height distribution in 1D

In this chapter, we have derived the explicit formula for the peak height distribution of non-stationary Gaussian processes in 1D, and demonstrated the formula has a similar form as that for stationary Gaussian processes while using a different parameterization. Cheng and Schwartzman (2018) defined the parameter κ for isotropic Gaussian random fields because it has the nice property of being invariant to the scaling of the parameter space and equal to 1 when the covariance is Gaussian. However, here we discovered $\rho(t)$ is a better parameterization, as it has a clear statistical meaning (the conditional correlation of the process and its second derivative given the gradient is 0) and allows characterizing a larger range of possible processes.

5.6.2 Peak height distribution of the scale space field

With the help of the peak height density function in 1D, we discover an interesting property of the Gaussian process with non-constant bandwidth and generalize it to the multidimensional scale space field. This property shows the peak height distribution of the scale space field does not vary over the domain. Studying the peak height distribution of the scale space field helps to choose the proper threshold to control the type I error when the signal peaks we aim to detect have different shapes. However, the scale space field is not optimal for detecting peaks that are not rotationally symmetric. In future work, we plan to investigate the Gaussian random field defined on the scale-rotation space and explore the properties of the peak height distribution.

5.6.3 Numerical Kac-Rice algorithm

Apart from the theoretical discoveries, we also provide numerical algorithms for computing the peak height distribution of any smooth Gaussian random field. One inconvenience of the numerical algorithms is that the user needs to compute the variances and covariances of the field and its first and second order derivatives. Since deriving the analytical form for the covariance matrices, like what we have done for the scale space field, involves only taking derivatives of the covariance function (5.19), the slight inconvenience of extra computation is usually not a big issue in practice. Also, the derivation can be automated by employing symbolic computation tools (e.g. Symbolic Math Toolbox in Matlab and WolframAlpha) as such tools are effective in performing and simplifying complicated differentiation. Alternatively, it is feasible to estimate these variances and covariances empirically from the real or simulated data, as it typically requires a smaller number of data instances to obtain accurate estimates for them compared to estimating the peak height distribution considering not all data instances have local maximum.

The amount of time required to generate the field and search for local maxima has

always been an issue for evaluating the theoretical properties of the peak height distribution by simulation, especially when the field is multidimensional. We anticipate that the numerical Kac-Rice algorithm will facilitate and accelerate the exploration of the peak height distribution of multidimensional Gaussian random fields.

5.7 Acknowledgement

Yu Zhao, Dan Cheng, and Armin Schwartzman were partially supported by NIH grant R01EB026859 and NSF grant 1811659.

Chapter 5, in full, is currently being prepared for submission for publication of the material. Zhao, Yu; Cheng, Dan; Schwartzman, Armin. On the peak height distribution of non-stationary Gaussian random fields: 1D non-constant variance and scale space. The dissertation author was the primary investigator and author of this material.

Appendix A

Supplementary materials for Chapter 2

A.1 Proofs

Derivation of equation (2.12). To derive the variance term in equation (2.12), we need to find the bottom right corner of $(X_i' V_i^{-1} X_i)^{-1}$. By equation (15) in de Leeuw and Kreft (1986),

$$\begin{aligned} V_i^{-1} &= (X_i D X_i^T + \sigma_\varepsilon^2 I)^{-1} \\ &= \sigma_\varepsilon^{-2} I - \sigma_\varepsilon^{-2} X_i (X_i' X_i)^{-1} X_i' + X_i (X_i' X_i)^{-1} (\sigma_\varepsilon^2 (X_i' X_i)^{-1} + D)^{-1} (X_i' X_i)^{-1} X_i'. \end{aligned}$$

Substituting and collecting terms,

$$(X_i' V_i^{-1} X_i)^{-1} = \sigma_\varepsilon^2 (X_i' X_i)^{-1} + D,$$

and

$$[(X_i' V_i^{-1} X_i)^{-1}]_{22} = \sigma_\varepsilon^2 / \sum (t_j - \bar{t})^2 + \sigma_{b_1}^2.$$

□

A.2 Implementation of the random effects model with random slopes and intercepts in R

The random effects model with random slopes and intercepts can be performed with the lmer function within the R package lmerTest (Kuznetsova et al., 2017). To test for differences in slopes between groups under the assumption of equal covariance structure in the two groups, the lmer model call is

```
lmer(Y ~ GROUP*TIME+(TIME|ID))
```

where ID indexes individual subjects, GROUP is a 0, 1 variable indicating placebo (0) and active treatment (1), and TIME are times of repeated observations on the dependent variable Y.

To test for differences in slopes between groups under the assumption of unequal covariance structure in the two groups, as implemented in power formula (2.14), the lmer model call is

```
TIME_0 <- ifelse(GROUP==0,TIME,0)
TIME_1 <- ifelse(GROUP==1,TIME,0)
lmer(Y ~ TIME*GROUP+(TIME_0|ID)+(TIME_1|ID))
```

Appendix B

Supplementary materials for Chapter 3

B.1 Proofs

Derivation of equation (3.8). Restricting to completers only, there is only one drop-out pattern in equation (3.6) (the complete data pattern, with design matrix X_k, an m by m identity matrix). Hence, the asymptotic variance of $\hat{\alpha}$ reduces to

$$(nX_k'V_k^{-1}X_k)^{-1} = V_k/n,$$

where

$$\begin{aligned} V_k &= Z_k D Z_k' + \sigma_\varepsilon^2 I \\ &= \begin{pmatrix} 1 & t_1 \\ \vdots & \ddots \\ 1 & t_m \end{pmatrix} \begin{pmatrix} \sigma_{b_0}^2 & \sigma_{b_0,b_1} \\ \sigma_{b_0,b_1} & \sigma_{b_1}^2 \end{pmatrix} \begin{pmatrix} 1 & \dots & 1 \\ t_1 & \dots & t_m \end{pmatrix} + \sigma_\varepsilon^2 I \\ &= \left(\sigma_{b_0}^2 + t_i \sigma_{b_0,b_1} + t_j \sigma_{b_0,b_1} + t_i t_j \sigma_{b_1}^2 \right)_{i,j=1\dots m} + \sigma_\varepsilon^2 I. \end{aligned}$$

Applying elements of this matrix to the variance of $\hat{\alpha_m} - \hat{\alpha_1}$, we obtain

$$\text{Var}(\hat{\alpha_m} - \hat{\alpha_1}) = \text{Var}(\hat{\alpha_m}) + \text{Var}(\hat{\alpha_1}) - 2\text{Cov}(\hat{\alpha_m}, \hat{\alpha_1})$$

$$
\begin{aligned}
=&\frac{1}{n}[\sigma_{b_0}^2 + 2t_m\sigma_{b_0,b_1} + t_m^2\sigma_{b_1}^2 + \sigma_\varepsilon^2 \\
&+ \sigma_{b_0}^2 + 2t_1\sigma_{b_0,b_1} + t_1^2\sigma_{b_0}^2 + \sigma_\varepsilon^2 \\
&- 2(\sigma_{b_0}^2 + t_1\sigma_{b_0,b_1} + t_m\sigma_{b_0,b_1} + t_1t_m\sigma_{b_1}^2)] \\
=&\frac{1}{n}[2\sigma_\varepsilon^2 + (t_m - t_1)^2\sigma_{b_1}^2].
\end{aligned}
$$

This completes the proof. □

B.2 Explicit expression of W in equation (3.10)

Let V_m be the covariance of repeated measures of completers defined above. Study subject dropout defines $m-1$ additional covariance matrices

$$
V_k = \begin{pmatrix} U_k & 0 \\ 0 & 0 \end{pmatrix},
$$

$k = 1, ..., m-1$ Where U_k are the k by k upper left submatrices of the completers' covariance matrix V_m. Then by equation (3.6)

$$
W = nV(\hat{\alpha}) = \left(\sum_k p_k V_k^{-1}\right)^{-1}.
$$

Appendix C

Supplementary materials for Chapter 4

C.1 Proofs

Proof of Proposition 1. Let $\tilde{\theta}(s) = \theta(s) - \theta_0 = h(s) + 0$ and $\tilde{M}_u$ be the number of local maxima of the random field $\tilde{X}(s) = Z(s) + \tilde{\theta}(s)$ above u over D. Considering the definition of power, we have

$$F(u - \theta_0) = \mathbb{P}[\tilde{M}_{u-\theta_0} \geq 1] = \mathbb{P}[M_u \geq 1].$$

Given that $\mathbb{E}[\tilde{M}_{u-\theta_0}] = \mathbb{E}[M_u]$, is is also straightforward to show $\mathbb{E}[M_u]$ is translation equivariant with respect to θ_0. □

Proof of Theorem 1. The proof is based on the proof of Lemma 3 in Piterbarg (1996) and Lemma 4.1 in Cheng and Schwartzman (2015).

$$\mathbb{E}[M_u(M_u-1)] = \int_{D_\varepsilon}\int_{D_\varepsilon}\int_u^\infty\int_u^\infty \mathbb{E}\left[|\det\nabla^2 X(s)||\det\nabla^2 X(t)|\,\middle|\begin{matrix} X(s)=x_1, X(t)=x_2 \\ \nabla X(s) = \nabla X(t) = 0\end{matrix}\right] \tag{C.1}$$
$$P_{X(s),X(t),\nabla X(s),\nabla X(t)}(x_1,x_2,0,0)\,dx_1\,dx_2\,ds\,dt.$$

Let

$$\mathbb{E}_1(s,t) = \mathbb{E}\left[|\det\nabla^2 X(s)||\det\nabla^2 X(t)|\,\middle|\begin{matrix} X(s)=x \\ \nabla X(s) = \nabla X(t) = 0\end{matrix}\right],$$

and replace one of the integration limits in (C.1) by $-\infty$, we have

$$\mathbb{E}[M_u(M_u-1)] \leq \int_{D_\varepsilon}\int_{D_\varepsilon} P_{\nabla X(s),\nabla X(t)}(0,0)\,ds\,dt \int_u^\infty \mathbb{E}_1(s,t)P_{X(s)}\left(x|\nabla X(s)=\nabla X(t)=0\right)dx.$$

Then we can take the Taylor expansion

$$\nabla X(t) = \nabla X(s) + \nabla^2 X(s)(t-s) + ||t-s||^{1+\alpha} Y_{s,t},$$

where $Y_{s,t} = (Y^1_{s,t},...,Y^N_{s,t})^T$ is a Gaussian vector field. Note that the determinant of $\nabla^2 X(s)$ is equal to the determinant of

$$\begin{pmatrix} 1 & -(t_1-s_1) & \dots & -(t_N-s_N) \\ 0 & & & \\ \vdots & & \nabla^2 X(s) & \\ 0 & & & \end{pmatrix}. \tag{C.2}$$

For $i = 2,...,N+1$, multiply the ith column of this matrix by $(t_i-s_i)/||t_i-s_i||^2$, take the sum of all such columns and add the result to the first column. Since $\nabla X(s) = \nabla X(t) = 0$, we can derive $\nabla^2 X(s)(t-s) = -||t-s||^{1+\alpha} Y_{s,t}$, and obtain the matrix below with the same determinant as (C.2)

$$\begin{pmatrix} 0 & -(t_1-s_1) & \dots & -(t_N-s_N) \\ -||t-s||^{-1+\alpha} Y^1_{s,t} & & & \\ \vdots & & \nabla^2 X(s) & \\ -||s-t||^{-1+\alpha} Y^N_{s,t} & & & \end{pmatrix}.$$

Let $r = \max_{1\le i\le N} |t_i - s_i|$,

$$A_{s,t} = \begin{pmatrix} 0 & -(t_1-s_1)/r & \dots & -(t_N-s_N)/r \\ Y^1_{s,t} & & & \\ \vdots & & \nabla^2 X(s) & \\ Y^N_{s,t} & & & \end{pmatrix}.$$

So we have

$$\mathbb{E}_1(s,t) \le ||t-s||^{\alpha}\mathbb{E}_2(s,t),$$

where

$$\mathbb{E}_2(s,t) = \mathbb{E}\left[|\det A_{s,t}||\det\nabla^2 X(t)| \,\middle|\, \begin{matrix} X(s)=x, \nabla X(s)=0 \\ \nabla^2 X(s)(t-s) = -||t-s||^{1+\alpha} Y_{s,t} \end{matrix} \right].$$

Using the inequality of arithmetic and geometric means, we can bound the determinant

$$|\det\nabla^2 X(t)| \le N^{2N-2}\sum_{i,j}|X_{ij}(t)|^N,$$
$$|\det A_{s,t}| \le (N+1)^{2N}\sum_{i,j}|a_{ij}|^{N+1},$$

where a_{ij} is the i,j entry of $A_{s,t}$. Apply the inequality again

$$|\det\nabla^2 X(t)||\det A_{s,t}| \le \frac{1}{2}N^{2N-2}(N+1)^{2N+1}\left(\sum_{i,j}|X_{ij}(t)|^{2N} + \sum_{i,j}|a_{ij}|^{2N+2}\right).$$

For any Gaussian variable X and integer $N \ge 0$, the following inequality holds

$$\mathbb{E}[X^{2N}] \le 2^{2N}(\mathbb{E}[X]^{2N} + C_N \mathrm{Var}(X)^{2N}),$$

where C_N is a constant depending on N. Next, we can focus on the conditional expectation and conditional variance of $X_{ij}(t)$ and $Y_{s,t}$.

By Assumption 1 and 2 and the fact that the conditional variance of a Gaussian variable is less or equal to the unconditional variance, we can conclude that the conditional variance of $X_{ij}(t)$ and $Y_{s,t}$ are bounded above by some constant.

Summarizing the results above,

$$\sup_{s,t\in D_\varepsilon, s\neq t} |\mathbb{E}_2(s,t)| \leq C_1$$

for some constant $C_1 > 0$ and

$$\mathbb{E}_1(s,t) \leq ||t-s||^\alpha \mathbb{E}_2(s,t) \leq C_1 ||t-s||^\alpha.$$

Combine the results above and with a fixed threshold u

$$\begin{aligned}
&\int_u^\infty \mathbb{E}_1(s,t) P_{X(s)}\left(x|\nabla X(s) = \nabla X(t) = 0\right) dx \\
&\leq C_1 ||t-s||^\alpha \int_u^\infty P_{X(s)}\left(x|\nabla X(s) = \nabla X(t) = 0\right) dx \\
&= C_1 ||t-s||^\alpha \int_u^\infty \exp(-(Ax-B)^2) dx \quad \text{for some constant } A, B \\
&= C_2 ||t-s||^\alpha
\end{aligned}$$

for some constant $C_2 > 0$.

Next, by the proof of Lemma 4.1 in Cheng and Schwartzman (2015)

$$p_{\nabla X(s),\nabla X(t)}(0,0) \leq C_3 ||t-s||^{-N}$$

for some constant $C_3 > 0$.

Therefore, there exists C_4 such that

$$\mathbb{E}[M_u(M_u-1)] \leq C_4 \int_{D_\varepsilon}\int_{D_\varepsilon} \frac{1}{||t-s||^{N-\alpha}} dt\, ds = o(\varepsilon^N).$$

For $\mathbb{E}[M_u]$, by Kac-Rice formula in Adler and Taylor (2007)

$$\mathbb{E}[M_u] = \int_{D_\varepsilon} p_{\nabla X(s)}(0)\mathbb{E}\left[|\det\nabla^2 X(s)|\mathbb{1}_{\{\nabla^2 X(s)\prec 0\}}\mathbb{1}_{\{X(s)>u\}}|\nabla X(s)=0\right]ds.$$

Denote the integrand by $g(s)$. The function $g(s)$ is continuous and positive over the compact domain D_ε. Thus $\inf_{s\in D_\varepsilon} g(s) \geq g_0 > 0$, implying

$$\mathbb{E}[M_u] \geq g_0\varepsilon^N.$$

Then (4.7) is an immediate consequence of (4.5).

For $\mathbb{E}[M_{-\infty}]$, by Kac-Rice formula

$$\mathbb{E}[M_{-\infty}] = \int_{D_\varepsilon} p_{\nabla X(s)}(0)\mathbb{E}\left[|\det\nabla^2 X(s)|\mathbb{1}_{\{\nabla^2 X(s)\prec 0\}}|\nabla X(s)=0\right]ds.$$

The integrand is also continuous and positive over the compact domain D_ε indicating $\mathbb{E}[M_{-\infty}] = o(1)$ for small ε. Thus we have

$$\mathbb{E}[M_u]_{\text{adj}} = \mathbb{E}[M_u]/\max(1,\mathbb{E}[M_{-\infty}]) = \mathbb{E}[M_u]/\max(1,o(1)) = \mathbb{E}[M_u]$$

for sufficiently small ε. □

Proof of Theorem 2. By lemma 3 of Piterbarg (1996), as $u\to\infty$, the factorial moment is super-exponentially small. That means $\exists\alpha>0$ s.t.

$$\mathbb{E}[M_u(M_u-1)] = o(e^{-\frac{u^2}{2}-\alpha u^2}).$$

Also

$$\mathbb{E}[M_u] \geq \mathbb{P}[M_u\geq 1] \geq \mathbb{P}[\sup X(s)\geq u] = O(e^{-\frac{u^2}{2}}).$$

Thus, the factorial moment decays exponentially faster than $\mathbb{E}[M_u]$. The result is an immediate

consequence of (4.5). □

Proof of Theorem 3. By lemma A.1 of Cheng and Schwartzman (2017), as $a \to \infty$

$$\mathbb{P}(M_{-\infty} = 1) \geq 1 - O(\exp(-ca^2)),$$

where $c > 0$ is some constant. Therefore $M_{-\infty} \xrightarrow{p} 1$.

Since $M_u \leq M_{-\infty}$ and both of them only take non-negative integer values, $|M_u(M_u - 1)|$ and $|M_{-\infty}(M_{-\infty} - 1)|$ are bounded above by $|M(M-1)|$ where M is the number of critical points of the random field X. Apply Kac-Rice formula

$$\mathbb{E}[M(M-1)] = \int_D \int_D \mathbb{E}\left[|\det\nabla^2 X(s)||\det\nabla^2 X(t)|\middle|\nabla X(s) = \nabla X(t) = 0\right] P_{\nabla X(s),\nabla X(t)}(0,0)\,ds\,dt.$$

Denote the integrand by $g(s,t,a)$. The function $g(s,t,a)$ is continuous and positive over the compact domain D and $M(M-1) \xrightarrow{p} 0$ as $a \to \infty$. Thus there exists $g_0 > 0$ such that $\mathbb{E}[M(M-1)] \leq g_0$. Then by dominated convergence theorem

$$\mathbb{E}[M_u(M_u - 1)] \to 0$$

as $a \to \infty$. Since $M_{-\infty} \xrightarrow{p} 1$, the adjusted $\mathbb{E}[M_u]$

$$\mathbb{E}[M_u]_{\text{adj}} = \mathbb{E}[M_u]\max(1, \mathbb{E}[M_{-\infty}]) = \mathbb{E}[M_u](1 + o(1)) = \mathbb{E}[M_u] + o(1).$$

To calculate $\mathbb{E}[M_u]$, apply Kac-Rice formula

$$\begin{aligned}
\mathbb{E}[M_u] &= \int_D p_{\nabla X(s)}(0)\mathbb{E}\left[|\det\nabla^2 X(s)|\mathbb{1}_{\{\nabla^2 X(s) \prec 0\}}\mathbb{1}_{\{X(s) > u\}}|\nabla X(s) = 0\right] ds \\
&= \int_D p_{\nabla X(s)}(0)\mathbb{E}\left[|\det\nabla^2 X(s)|\mathbb{1}_{\{\nabla^2 X(s) \prec 0\}}\mathbb{1}_{\{X(s) > u\}}|\nabla X(s) = 0\right] ds \\
&= \int_D \frac{1}{(2\pi)^{N/2}\sqrt{\det(\Lambda)}} \exp(-a^2(\nabla h(s))^T \Lambda^{-1} \nabla h(s)/2)
\end{aligned}$$

$$\mathbb{E}\left[|\det(\nabla^2 Z(s)+a\nabla^2 h(s))|\mathbb{1}_{\{\nabla^2 X(s)\prec 0\}}\mathbb{1}_{\{X(s)>u\}}|\nabla X(s)=0\right]ds, \tag{C.3}$$

where Λ is the covariance matrix of $\nabla h(s)$. Let $f(s)=(\nabla h(s))^T\Lambda^{-1}\nabla h(s)/2$ which attains its minimum 0 only at s_0. Similar to the proof of A.4 in Cheng and Schwartzman (2017), as $a\to\infty$, (C.3) can be approximated by applying Laplace's method

$$\begin{aligned}\mathbb{E}[M_u] =& \frac{\det(a\nabla^2 h(s_0))}{(2\pi)^{N/2}\sqrt{\det(\Lambda)}}\left(\frac{(2\pi)^N\det(\Lambda)}{a^{2N}\det(\nabla^2 h(s_0))}\right)^{1/2}\Phi(\theta_0-u)+O(a^{-2})\\ =&\Phi(\theta_0-u)+O(a^{-2}).\end{aligned}$$

This finishes the proof. □

Proof of Lemma 3. Part (i) is a direct consequence of Lemma 2. For part (ii), note that $(\nabla^2 X(s)|X(s)=x)$ is equivalent to $(\nabla^2 Z(s)|Z(s)=x-\theta(s))+\nabla^2\theta(s)$, and the result follows immediately from Lemma 2.

□

Proof of Theorem 4. By the Kac-Rice formula

$$\begin{aligned}\mathbb{E}[M_u] &= \int_D p_{\nabla X(s)}(0)\mathbb{E}\left[|\det\nabla^2 X(s)|\mathbb{1}_{\{\nabla^2 X(s)\prec 0\}}\mathbb{1}_{\{X(s)>u\}}|\nabla X(s)=0\right]ds\\ &= \int_D p_{\nabla Z(s)+\nabla\theta(s)}(0)\mathbb{E}\left[|\det\nabla^2 X(s)|\mathbb{1}_{\{\nabla^2 X(s)\prec 0\}}\mathbb{1}_{\{X(s)>u\}}|\nabla X(s)=0\right]ds\\ &= \int_D \frac{1}{(2\pi)^{N/2}(-2\rho')^{N/2}}e^{\frac{\|\nabla\theta(s)\|^2}{4\rho'}}\mathbb{E}\left[|\det\nabla^2 X(s)|\mathbb{1}_{\{\nabla^2 X(s)\prec 0\}}\mathbb{1}_{\{X(s)>u\}}|\nabla X(s)=0\right]ds\\ &= \int_D \frac{(8\rho''^2)^{N/2}}{(2\pi)^{N/2}(-2\rho')^{N/2}}e^{\frac{\|\nabla\theta(s)\|^2}{4\rho'}}\int_u^\infty \phi\left(x-\theta(s)\right)\mathbb{E}\left[|\det(\text{Matrix}(s))|\mathbb{1}_{\{\text{Matrix}(s)\prec 0\}}\right]dxds\\ &= \int_D \left(\frac{2\rho''}{-\pi\rho'}\right)^{N/2}e^{\frac{\|\nabla\theta(s)\|^2}{4\rho'}}\int_u^\infty \phi\left(x-\theta(s)\right)\mathbb{E}\left[|\det(\text{Matrix}(s))|\mathbb{1}_{\{\text{Matrix}(s)\prec 0\}}\right]dxds\\ &= \left(\frac{2\rho''}{-\pi\rho'}\right)^{N/2}\int_D e^{\frac{\|\nabla\theta(s)\|^2}{4\rho'}}\int_u^\infty \phi\left(x-\theta(s)\right)\mathbb{E}\left[|\det(\text{Matrix}(s))|\mathbb{1}_{\{\text{Matrix}(s)\prec 0\}}\right]dxds.\end{aligned}$$

Next, we show the derivation from the third to the fourth line in the equation above. Since we assume $\nabla^2\theta(s)$ is a non-singular matrix at all critical points, then there exists an orthonormal matrix, denoted by $A(s)$, such that $A(s)^T\nabla^2\theta(s)A(s) = \mathrm{diag}\{\theta_1''(s), \theta_2'', \ldots, \theta_N''(s)\}$, where $\theta_1'' \leq \ldots \leq \theta_N''(s)$ are ordered eigenvalues of $\nabla^2\theta(s)$. On the other hand, GOI matrices are invariant under orthonormal transformations. By Lemma 3, the conditional expectation $\mathbb{E}[|\det(\nabla^2 X(s))|\mathbb{1}_{\{\nabla^2 X(s)\prec 0\}}|X(s) = x]$ is therefore

$$
\begin{aligned}
&= \mathbb{E}\left[\left|\det\left(\sqrt{8\rho''}\left[G - \frac{\kappa(x-\theta(s))}{\sqrt{2}}I_N\right] + \nabla^2\theta(s)\right)\right| \mathbb{1}_{\{\mathrm{Matrix}(s)\prec 0\}}\right] \\
&= \mathbb{E}\left[\left|\det\left(\sqrt{8\rho''}\left[G - \frac{\kappa(x-\theta(s))}{\sqrt{2}}I_N\right] + A(s)^T\nabla^2\theta(s)A(s)\right)\right| \mathbb{1}_{\{\mathrm{Matrix}(s)\prec 0\}}\right] \\
&= (\sqrt{8\rho''})^N\mathbb{E}\left[\left|\det\left(\left[G - \frac{\kappa(x-\theta(s))}{\sqrt{2}}I_N\right] + A(s)^T\nabla^2\theta(s)A(s)/\sqrt{8\rho''}\right)\right| \mathbb{1}_{\{\mathrm{Matrix}(s)\prec 0\}}\right] \\
&= (\sqrt{8\rho''})^N\mathbb{E}\left[\left|\det\left(G - \frac{\kappa(x-\theta(s))}{\sqrt{2}}I_N + \mathrm{diag}\{\theta_1''(s), \ldots, \theta_N''(s)\}/\sqrt{8\rho''}\right)\right| \mathbb{1}_{\{\mathrm{Matrix}(s)\prec 0\}}\right].
\end{aligned}
$$

□

Proof of Proposition 2. Since we assume that $Z(s)$ is stationary, $Z'(s)$ is independent of $Z(s)$ and $Z''(s)$, and $\rho' = -\mathrm{Var}(Z'(s))/2 = \mathbb{E}[Z(s)Z''(s)]/2$ and $\rho'' = \mathrm{Var}(Z''(s))/12$ do not depend on s. Therefore,

$$
\mathrm{Var}(X(s)) = 1, \quad \mathrm{Var}(X'(s)) = -\mathrm{Cov}(X(s)X''(s)) = -2\rho' \quad \text{and} \quad \mathrm{Var}(X''(s)) = 12\rho''.
$$

Note that, by the formula of conditional Gaussian distributions,

$$
X''(s)|X(s) = x \sim N(\theta''(s) + 2\rho'(x-\theta(s)), 12\rho'' - 4\rho'^2).
$$

By the Kac-Rice formula

$$
\mathbb{E}[M_u] = \int_D p_{X'(s)}(0)\mathbb{E}[|X''(s)|\mathbb{1}_{\{X(s)>u\}}\mathbb{1}_{\{X''(s)<0\}}|X'(s) = 0]ds
$$

$$
\begin{aligned}
&= \int_D p_{X'(s)}(0) \int_u^\infty \phi(x-\theta(s)) \int_{-\infty}^0 (-x'') \frac{1}{\sqrt{12\rho'' - 4\rho'^2}} \\
&\quad \times \phi\left[\frac{x'' - \theta''(s) - 2\rho'(x-\theta(s))}{\sqrt{12\rho'' - 4\rho'^2}} \right] dx'' dx ds \\
&= \int_D p_{X'(s)}(0) \sqrt{12\rho'' - 4\rho'^2} \int_u^\infty \phi(x-\theta(s)) \psi\left(\frac{-2\rho'(x-\theta(s)) - \theta''(s)}{\sqrt{12\rho'' - 4\rho'^2}} \right) dx ds \\
&= \int_D \frac{1}{\sqrt{-2\rho'}} \phi\left(\frac{\theta'(s)}{\sqrt{-2\rho'}} \right) \sqrt{12\rho'' - 4\rho'^2} \int_u^\infty \phi(x-\theta(s)) \\
&\quad \times \psi\left(\frac{-2\rho'(x-\theta(s)) - \theta''(s)}{\sqrt{12\rho'' - 4\rho'^2}} \right) dx ds \\
&= \int_D \frac{\sqrt{12\rho'' - 4\rho'^2}}{\sqrt{-2\rho'}} \phi\left(\frac{\theta'(s)}{\sqrt{-2\rho'}} \right) \int_u^\infty \phi(x-\theta(s)) \psi\left(\frac{-2\rho'(x-\theta(s)) - \theta''(s)}{\sqrt{12\rho'' - 4\rho'^2}} \right) dx ds.
\end{aligned}
$$

The second to third line is due to the fact that

$$
\int_{-\infty}^0 (-x) \frac{1}{b} \phi\left(\frac{x+a}{b} \right) dx = \int_{-\infty}^0 \Phi\left(\frac{x+a}{b} \right) dx = b \int_{-\infty}^{a/b} \Phi(y) dy = b\psi\left(\frac{a}{b} \right).
$$

Recall the κ (4.10) and η (4.13) parameters defined before. We can rewrite $\mathbb{E}[M_u]$ as

$$
\int_D \frac{\sqrt{-2\rho'(3-\kappa^2)}}{\kappa} \phi\left(\frac{\theta'(s)}{\sqrt{-2\rho'}} \right) \int_u^\infty \phi(x-\theta(s)) \psi\left(\frac{\kappa[x-\theta(s)-\eta(s)]}{\sqrt{3-\kappa^2}} \right) dx ds.
$$

This finishes the proof. □

Proof of Proposition 3. By Lemma 4 above with $a = (\kappa^2-1)/(4-2\kappa^2)$ and $b = \kappa\tilde{x}/\sqrt{2}$, and Lemma 2.2 in Cheng and Schwartzman (2018),

$$
\begin{aligned}
&\mathbb{E}^N_{\mathrm{GOI}((1-\kappa^2)/2)} \left[\prod_{j=1}^N \left| \lambda_j - \frac{\kappa\tilde{x}}{\sqrt{2}} \right| I\left(\lambda_N < -\frac{\kappa\tilde{x}}{\sqrt{2}} \right) \right] \\
&= \frac{1}{2\sqrt{\pi(2-\kappa^2)}} \int_{\mathbb{R}^2} \exp\left\{ -\frac{1}{2}(\lambda_1^2 + \lambda_2^2) + \frac{1-\kappa^2}{4(2-\kappa^2)} (\lambda_1 + \lambda_2)^2 \right\} (\lambda_2 - \lambda_1)
\end{aligned}
$$

$$
\begin{aligned}
&\times \left|\lambda_1 - \frac{\kappa\tilde{x}}{\sqrt{2}}\right| \left|\lambda_2 - \frac{\kappa\tilde{x}}{\sqrt{2}}\right| \mathbb{1}_{\{\lambda_2 < \frac{\kappa\tilde{x}}{\sqrt{2}}\}} d\lambda_1\, d\lambda_2 \\
=&\frac{1}{\sqrt{3-\kappa^2}} e^{-\frac{\kappa^2\tilde{x}^2}{2(3-\kappa^2)}} \Phi\left(\frac{\kappa\tilde{x}}{\sqrt{(2-\kappa^2)(3-\kappa^2)}}\right) + \quad \frac{\kappa^2}{2}\left(\tilde{x}^2-1\right)\Phi\left(\frac{\kappa\tilde{x}}{\sqrt{2-\kappa^2}}\right) \\
&+ \frac{\kappa\sqrt{2-\kappa^2}\tilde{x}}{2\sqrt{2\pi}} e^{-\frac{\kappa^2\tilde{x}^2}{2(2-\kappa^2)}}.
\end{aligned} \tag{C.4}
$$

This simplifies to (4.19).

□

Proof of Proposition 4. This is a direct result of Lemma 5 above with $a = 1-\kappa^2$ and $b = \kappa\tilde{x}/\sqrt{2}$. □

Appendix D

Supplementary materials for Chapter 5

D.1 Proofs

Proof of Lemma 6. Following (5.2), the covariance function of X is

$$\mathbb{E}[X(t)X(s)] = h\left(\frac{t+s}{2}, (t-s)^2\right).$$

Taking partial derivatives on both sides, we have

$$\begin{aligned}
\frac{\partial}{\partial t}\mathbb{E}[X(t)X(s)] =& \frac{1}{2}h_1'\left(\frac{t+s}{2}, (t-s)^2\right) + 2(t-s)h_2'\left(\frac{t+s}{2}, (t-s)^2\right), \\
\frac{\partial^2}{\partial t^2}\mathbb{E}[X(t)X(s)] =& \frac{1}{4}h_{11}''\left(\frac{t+s}{2}, (t-s)^2\right) + 2(t-s)h_{12}''\left(\frac{t+s}{2}, (t-s)^2\right) \\
& + 4(t-s)^2 h_{22}''\left(\frac{t+s}{2}, (t-s)^2\right) + 2h_2'\left(\frac{t+s}{2}, (t-s)^2\right), \\
\frac{\partial^2}{\partial t \partial s}\mathbb{E}[X(t)X(s)] =& \frac{1}{4}h_{11}''\left(\frac{t+s}{2}, (t-s)^2\right) - 4(t-s)^2 h_{22}''\left(\frac{t+s}{2}, (t-s)^2\right) \\
& - 2h_2'\left(\frac{t+s}{2}, (t-s)^2\right), \\
\frac{\partial^3}{\partial t^2 \partial s}\mathbb{E}[X(t)X(s)] =& \frac{1}{8}h_{111}'''\left(\frac{t+s}{2}, (t-s)^2\right) + \frac{1}{2}(t-s)h_{112}'''\left(\frac{t+s}{2}, (t-s)^2\right) \\
& - 2(t-s)^2 h_{122}'''\left(\frac{t+s}{2}, (t-s)^2\right) - 8(t-s)^3 h_{222}'''\left(\frac{t+s}{2}, (t-s)^2\right) \\
& - h_{12}''\left(\frac{t+s}{2}, (t-s)^2\right) - 12(t-s)h_{22}''\left(\frac{t+s}{2}, (t-s)^2\right),
\end{aligned}$$

$$\begin{aligned}\frac{\partial^4}{\partial t^2 \partial s^2}\mathbb{E}[X(t)X(s)] =& \frac{1}{16}h''''_{1111}\left(\frac{t+s}{2},(t-s)^2\right) - 2(t-s)^2 h''''_{1122}\left(\frac{t+s}{2},(t-s)^2\right) \\ &+ 16(t-s)^4 h''''_{2222}\left(\frac{t+s}{2},(t-s)^2\right) - h'''_{112}\left(\frac{t+s}{2},(t-s)^2\right) \\ &+ 48(t-s)^2 h'''_{222}\left(\frac{t+s}{2},(t-s)^2\right) + 12h''_{22}\left(\frac{t+s}{2},(t-s)^2\right).\end{aligned}$$

By (5.5.5) in Adler and Taylor (2007),

$$\begin{aligned}\mathrm{Var}(X(t)) &= \mathbb{E}[X(t)X(s)]|_{s=t} = h_0(t), \\ \mathrm{Var}(X'(t)) &= \frac{\partial^2}{\partial t \partial s}\mathbb{E}[X(t)X(s)]\bigg|_{s=t} = \frac{1}{4}h''_{11}(t) - 2h'_2(t), \\ \mathrm{Var}(X''(t)) &= \frac{\partial^4}{\partial t^2 \partial s^2}\mathbb{E}[X(t)X(s)]\bigg|_{s=t} = \frac{1}{16}h''''_{1111}(t) - h'''_{112}(t) + 12h''_{22}(t), \\ \mathbb{E}[X(t)X'(t)] &= \frac{\partial}{\partial t}\mathbb{E}[X(t)X(s)]\bigg|_{s=t} = \frac{1}{2}h'_1(t), \\ \mathbb{E}[X(t)X''(t)] &= \frac{\partial^2}{\partial t^2}\mathbb{E}[X(t)X(s)]\bigg|_{s=t} = \frac{1}{4}h''_{11}(t) + 2h'_2(t), \\ \mathbb{E}[X'(t)X''(t)] &= \frac{\partial^3}{\partial t^2 \partial s}\mathbb{E}[X(t)X(s)]\bigg|_{s=t} = \frac{1}{8}h'''_{111}(t) - h''_{12}(t).\end{aligned}$$

□

Proof of Theorem 5. Let $Z(t) = X(t)/\sigma(t)$ so that $Z(t)$ has unit variance. To apply the Kac-Rice formula (5.1) to compute the peak height distribution, we need the variances and covariances of the first and second order derivatives. Suppose the random vector $(Z(t), Z'(t), Z''(t))$ has the covariance structure

$$\begin{pmatrix} \mathrm{Var}(Z(t)) & \mathbb{E}[Z(t)Z'(t)] & \mathbb{E}[Z(t)Z''(t)] \\ \mathbb{E}[Z(t)Z'(t)] & \mathrm{Var}(Z'(t)) & \mathbb{E}[Z'(t)Z''(t)] \\ \mathbb{E}[Z(t)Z''(t)] & \mathbb{E}[Z'(t)Z''(t)] & \mathrm{Var}(Z''(t)) \end{pmatrix} = \begin{pmatrix} 1 & 0 & -\lambda_1(t) \\ 0 & \lambda_1(t) & r_1(t) \\ -\lambda_1(t) & r_1(t) & \lambda_2(t) \end{pmatrix},$$

where $Z'(t)$ and $Z''(t)$ are the first and second order derivatives of $Z(t)$. $\mathbb{E}[Z(t)Z'(t)] = 0$ and $\mathrm{Var}(Z'(t)) = -\mathbb{E}[Z(t)Z''(t)]$ can be derived from taking the first and second derivatives on both

sides of $\mathrm{Var}(Z(t)) = 1$.

With a minor abuse of notation, we use ρ, λ_1, λ_2, r_1 to represent $\rho(t)$, $\lambda_1(t)$, $\lambda_2(t)$ and $r_1(t)$, and σ_t, σ_t', σ_t'', $\tilde{\sigma}_t$ to represent $\sigma(t)$, $\sigma'(t)$, $\sigma''(t)$, $\tilde{\sigma}(t)$. Following this notation, we have $X'(t) = \sigma_t' Z(t) + \sigma_t Z'(t)$ and $X''(t) = \sigma_t'' Z(t) + 2\sigma_t' Z'(t) + \sigma_t Z''(t)$. $X(t)$, $X'(t)$ and $X''(t)$ are Gaussian distributed with mean 0, and variance-covariance matrix

$$
\begin{aligned}
&\begin{pmatrix}
\mathrm{Var}(X(t)) & \mathbb{E}[X(t)X'(t)] & \mathbb{E}[X(t)X''(t)] \\
\mathbb{E}[X(t)X'(t)] & \mathrm{Var}(X'(t)) & \mathbb{E}[X'(t)X''(t)] \\
\mathbb{E}[X(t)X''(t)] & \mathbb{E}[X'(t)X''(t)] & \mathrm{Var}(X''(t))
\end{pmatrix} \\
&= \begin{pmatrix}
\sigma_t^2 & \sigma_t\sigma_t' & \sigma_t\sigma_t'' - \sigma_t^2\lambda_1 \\
\sigma_t\sigma_t' & \sigma_t'^2 + \sigma_t^2\lambda_1 & \sigma_t'\sigma_t'' + \sigma_t\sigma_t'\lambda_1 + \sigma_t^2 r_1 \\
\sigma_t\sigma_t'' - \sigma_t^2\lambda_1 & \sigma_t'\sigma_t'' + \sigma_t\sigma_t'\lambda_1 + \sigma_t^2 r_1 & \sigma_t''^2 + (4\sigma_t'^2 - 2\sigma_t\sigma_t'')\lambda_1 + \sigma_t^2\lambda_2 + 4\sigma_t\sigma_t' r_1
\end{pmatrix} \\
&= \begin{pmatrix}
a & d & f \\
d & b & e \\
f & e & c
\end{pmatrix},
\end{aligned}
$$

where a, b, c, d, e, f are some constants for fixed t. Therefore

$$\mathbb{E}[X(t), X''(t) | X'(t) = 0] = (0,0),$$

and

$$\mathrm{Var}(X(t), X''(t) | X'(t) = 0) = \begin{pmatrix} a & f \\ f & c \end{pmatrix} - \begin{pmatrix} d \\ e \end{pmatrix} \frac{1}{b} \begin{pmatrix} d & e \end{pmatrix} = \begin{pmatrix} a - \frac{d^2}{b} & f - \frac{de}{b} \\ f - \frac{de}{b} & c - \frac{e^2}{b} \end{pmatrix}.$$

According the definition, we have

$$\rho := \mathrm{Cor}(X(t), X''(t) | X'(t) = 0)$$

$$= \frac{fb-de}{\sqrt{(ab-d^2)(bc-e^2)}}$$
$$= -\frac{\sigma_t^2\lambda_1^2 + (2\sigma_t'^2 - \sigma_t\sigma_t'')\lambda_1 + \sigma_t\sigma_t' r_1}{\sqrt{\lambda_1(\sigma_t^2(3\sigma_t'^2 - 2\sigma_t\sigma_t'')\lambda_1^2 + (\sigma_t\sigma_t'' - 2\sigma_t'^2)^2\lambda_1 + \sigma_t^4\lambda_1\lambda_2 + 2\sigma_t^3\sigma_t'\lambda_1 r_1 + \sigma_t^2\sigma_t'^2\lambda_2 - \sigma_t^4 r_1^2 + (4\sigma_t\sigma_t'^3 - 2\sigma_t^2\sigma_t'\sigma_t'')r_1)}},$$

and

$$\tilde{\sigma}_t^2 := \mathrm{Var}(X(t)|X'(t)=0) = a - \frac{d^2}{b} = \frac{\sigma_t^4\lambda_1}{\sigma_t'^2 + \sigma_t^2\lambda_1}.$$

The peak height density is given by the derivative of (5.1)

$$f_t(x) = \frac{\mathbb{E}[|X''(t)|\mathbb{1}_{\{X''(t)<0\}}|X(t)=x, X'(t)=0]p_{X(t)}(x|X'(t)=0)}{\mathbb{E}[|X''(t)|\mathbb{1}_{\{X''(t)<0\}}|X'(t)=0]}.$$

First, we compute the denominator

$$\mathbb{E}[|X''(t)|\mathbb{1}_{\{X''(t)<0\}}|X'(t)=0] = -\int_{-\infty}^{0} \frac{x}{\sqrt{c-\frac{e^2}{b}}}\phi\left(\frac{x}{\sqrt{c-\frac{e^2}{b}}}\right)dx = \frac{\sqrt{c-\frac{e^2}{b}}}{\sqrt{2\pi}}.$$

Similarly, $X''(t)|X(t)=x, X'(t)=0$ is Gaussian distributed with mean

$$\mathbb{E}[X''(t)|X(t)=x, X'(t)=0] = \begin{pmatrix} f & e \end{pmatrix}\begin{pmatrix} a & d \\ d & b \end{pmatrix}^{-1}\begin{pmatrix} x \\ 0 \end{pmatrix} = \frac{fb-ed}{ab-d^2}x = -\eta x,$$

and variance

$$\begin{aligned}\mathrm{Var}(X''(t)|X(t)=x, X'(t)=0) =& c - \begin{pmatrix} f & e \end{pmatrix}\begin{pmatrix} a & d \\ d & a \end{pmatrix}^{-1}\begin{pmatrix} f \\ e \end{pmatrix} \\ =& c - \frac{f^2 b - 2fed + ae^2}{ab-d^2} = c - \xi,\end{aligned}$$

where $\eta := (ed - fb)/(ab - d^2)$ and $\xi := (f^2 b - 2fed + ae^2)/(ab - d^2)$. Now the numerator of the Kac-Rice formula becomes

$$\begin{aligned}
&\mathbb{E}[|X''(t)|\mathbb{1}_{\{X''(t)<0\}}|X(t) = x, X'(t) = 0]\frac{1}{\tilde{\sigma}_t}\phi\left(\frac{x}{\tilde{\sigma}_t}\right) \\
=&-\frac{1}{\tilde{\sigma}_t}\phi\left(\frac{x}{\tilde{\sigma}_t}\right)\int_{-\infty}^{0}\frac{y}{\sqrt{c-\xi}}\phi\left(\frac{y+\eta x}{\sqrt{c-\xi}}\right)dy \\
=&\eta x\Phi\left(\frac{\eta x}{\sqrt{c-\xi}}\right)\frac{1}{\tilde{\sigma}_t}\phi\left(\frac{x}{\tilde{\sigma}_t}\right) + \sqrt{c-\xi}\,\phi\left(\frac{\eta x}{\sqrt{c-\xi}}\right)\frac{1}{\tilde{\sigma}_t}\phi\left(\frac{x}{\tilde{\sigma}_t}\right).
\end{aligned}$$

Combine the results above and consider (5.4), we have

$$\begin{aligned}
f_t(x) &= \frac{\eta x\Phi\left(\frac{\eta x}{\sqrt{c-\xi}}\right)\frac{1}{\tilde{\sigma}_t}\phi\left(\frac{x}{\tilde{\sigma}_t}\right) + \sqrt{c-\xi}\,\phi\left(\frac{\eta x}{\sqrt{c-\xi}}\right)\frac{1}{\tilde{\sigma}_t}\phi\left(\frac{x}{\tilde{\sigma}_t}\right)}{\frac{\sqrt{c-\frac{e^2}{b}}}{\sqrt{2\pi}}} \\
&= \frac{1}{\tilde{\sigma}_t}\phi\left(\frac{x}{\tilde{\sigma}_t}\right)\sqrt{\frac{2\pi(c-\xi)}{c-\frac{e^2}{b}}}\left[\phi\left(\frac{\eta x}{\sqrt{c-\xi}}\right) + \frac{\eta x}{\sqrt{c-\xi}}\Phi\left(\frac{\eta x}{\sqrt{c-\xi}}\right)\right] \\
&= \frac{1}{\tilde{\sigma}_t}\phi\left(\frac{x}{\tilde{\sigma}_t}\right)\sqrt{\frac{2\pi(c-\xi)}{c-\frac{e^2}{b}}}\psi\left(\frac{\eta x}{\sqrt{c-\xi}}\right).
\end{aligned}$$

Using the fact that

$$\xi = \tilde{\sigma}_t^2\eta^2 + \frac{e^2}{b} = \sigma_t^2\lambda_1^2 - 2(\sigma_t\sigma_t'' - 2\sigma_t'^2)\lambda_1 + \frac{r_1^2\sigma_t^2}{\lambda_1} + 4\sigma_t\sigma_t' + \sigma_t''^2,$$

we have

$$1 - \frac{(c-\xi)}{c-\frac{e^2}{b}} = \frac{\xi - \frac{e^2}{b}}{c-\frac{e^2}{b}} = \frac{\tilde{\sigma}_t^2\eta^2}{c-\frac{e^2}{b}} = \frac{(fb-dc)^2}{(ab-d^2)(bc-e^2)} = \rho^2,$$

and

$$1 + \tilde{\sigma}_t^2\frac{\eta^2}{c-\xi} = \frac{c-\frac{e^2}{b}}{c-\xi} = \frac{1}{1-\rho^2}.$$

Therefore, the peak height density can be further simplified as

$$f_t(x) = \frac{1}{\tilde{\sigma}_t}\phi\left(\frac{x}{\tilde{\sigma}_t}\right)\sqrt{2\pi(1-\rho^2)}\psi\left(\frac{-\rho x}{\sqrt{1-\rho^2}\tilde{\sigma}_t}\right).$$

The mean and variance of the peak height can be derived from its density function. Using the properties of Gaussian integrals (Owen, 1980), we have

$$\begin{aligned}
&\int_{-\infty}^{\infty} x f_t(x)dx \\
=&\int_{-\infty}^{\infty}\sqrt{2\pi(1-\rho^2(t))}\frac{x}{\tilde{\sigma}(t)}\phi\left(\frac{x}{\tilde{\sigma}(t)}\right)\phi\left(\frac{-\rho(t)x}{\sqrt{1-\rho^2(t)}\tilde{\sigma}(t)}\right)dx \\
&-\int_{-\infty}^{\infty}\sqrt{2\pi(1-\rho^2(t))}\frac{x}{\tilde{\sigma}(t)}\phi\left(\frac{x}{\tilde{\sigma}(t)}\right)\frac{\rho(t)x}{\sqrt{1-\rho^2(t)}\tilde{\sigma}(t)}\Phi\left(\frac{-\rho(t)x}{\sqrt{1-\rho^2(t)}\tilde{\sigma}(t)}\right)dx \\
=&-\sqrt{2\pi}\rho(t)\tilde{\sigma}(t)\int_{-\infty}^{\infty}x^2\phi(x)\Phi\left(\frac{-\rho(t)x}{\sqrt{1-\rho^2(t)}}\right)dx \\
=&-\sqrt{\frac{\pi}{2}}\rho(t)\tilde{\sigma}(t),
\end{aligned}$$

$$\begin{aligned}
&\int_{-\infty}^{\infty} x^2 f_t(x)dx \\
=&\sqrt{2\pi(1-\rho^2(t))}\tilde{\sigma}^2(t)\int_{-\infty}^{\infty}x^2\phi(x)\phi\left(\frac{-\rho(t)x}{\sqrt{1-\rho^2(t)}}\right)dx \\
&-\sqrt{2\pi}\rho(t)\tilde{\sigma}^2(t)\int_{-\infty}^{\infty}x^3\phi(x)\Phi\left(\frac{-\rho(t)x}{\sqrt{1-\rho^2(t)}}\right)dx \\
=&(1-\rho^2(t))^2\tilde{\sigma}^2(t)-\rho^2(t)(\rho^2(t)-3)\tilde{\sigma}^2(t) \\
=&(\rho^2(t)+1)\tilde{\sigma}^2(t).
\end{aligned}$$

The mean and variance are direct results of the first and second moments.

□

Proof of Proposition 5. Since the derivation for $\rho(t) = -1$ and $\rho(t) = 1$ follows a similar ap-

proach, here we only illustrate the method for $\rho(t) = -1$. With $\rho(t) = \mathrm{Cor}(X(t), X''(t)|X'(t) = 0) = -1$, the numerator of the Kac-Rice formula (5.1) becomes

$$\begin{aligned}
&\mathbb{E}[|X''(t)|\mathbb{1}_{\{X''(t)<0\}}\mathbb{1}_{\{X(t)>u\}}|X'(t)=0] \\
=&\mathbb{E}[|X''(t)|\mathbb{1}_{\{X''(t)<-\sqrt{\mathrm{Var}(X''(t)|X'(t)=0)/\mathrm{Var}(X(t)|X'(t)=0)}u\}}|X'(t)=0] \\
=&\int_{-\infty}^{-\sqrt{\mathrm{Var}(X''(t)|X'(t)=0)/\tilde{\sigma}(t)}u} \frac{x}{\sqrt{\mathrm{Var}(X''(t)|X'(t)=0)}}\phi\left(\frac{x}{\sqrt{\mathrm{Var}(X''(t)|X'(t)=0)}}\right)dx \\
=&\sqrt{\mathrm{Var}(X''(t)|X'(t)=0)}\phi\left(\frac{u}{\tilde{\sigma}(t)}\right).
\end{aligned}$$

The denominator is the same as the non-boundary case

$$\mathbb{E}[|X''(t)|\mathbb{1}_{\{X''(t)<0\}}|X'(t)=0] = \frac{\sqrt{\mathrm{Var}(X''(t)|X'(t)=0)}}{\sqrt{2\pi}}.$$

The peak height distribution

$$F_t(u) = \frac{\mathbb{E}[|X''(t)|\mathbb{1}_{\{X''(t)<0\}}\mathbb{1}_{\{X(t)>u\}}|X'(t)=0]}{\mathbb{E}[|X''(t)|\mathbb{1}_{\{X''(t)<0\}}|X'(t)=0]} = 2\pi\phi\left(\frac{u}{\tilde{\sigma}(t)}\right).$$

Taking derivative, we obtain

$$f_t(x) = \frac{\sqrt{2\pi}x}{\tilde{\sigma}^2(t)}\phi\left(\frac{x}{\tilde{\sigma}(t)}\right).$$

□

Proof of Theorem 6. Take the derivative on both sides of (5.14), we obtain

$$X'(t) = \int_{-\infty}^{\infty} -\frac{\nu'(t)}{2\nu(t)^{3/2}}k\left(\frac{t-s}{\nu(t)}\right) + k'\left(\frac{t-s}{\nu(t)}\right)\frac{\nu(t)-(t-s)\nu'(t)}{\nu(t)^{5/2}}dB(s).$$

Using the property of Wiener Integral and applying the change of variable $u = (t-s)/\nu(t)$, we

have

$$\begin{aligned}
\mathrm{Var}(X'(t)) &= \int_{-\infty}^{\infty} \left(-\frac{\nu'(t)}{2\nu(t)^{3/2}} k\left(\frac{t-s}{\nu(t)}\right) + k'\left(\frac{t-s}{\nu(t)}\right) \frac{\nu(t)-(t-s)\nu'(t)}{\nu(t)^{5/2}} \right)^2 ds \\
&= \int_{-\infty}^{\infty} \left(-\frac{\nu'(t)}{2\nu(t)} k(u) + \frac{1-u\nu'(t)}{\nu(t)} k'(u) \right)^2 du \\
&= \frac{1}{\nu(t)^2} \int_{-\infty}^{\infty} \left(-\frac{1}{2}\nu'(t)k(u) + (1-u\nu'(t))k'(u) \right)^2 du \\
&= \frac{1}{\nu(t)^2} C_1, \qquad \text{(D.1)}
\end{aligned}$$

where C_1 is a constant when $\nu(t)$ is a linear function of t.

Take the derivative again, we have

$$\begin{aligned}
X''(t) = \int_{-\infty}^{\infty} -\frac{1}{2} \Bigg(& \frac{2\nu''(t)\nu(t)^{3/2} - 3\nu'(t)^2\nu(t)^{1/2}}{2\nu(t)^3} k\left(\frac{t-s}{\nu(t)}\right) \\
& + \frac{\nu'(t)\nu(t) - (t-s)\nu'(t)^2}{\nu(t)^{7/2}} k'\left(\frac{t-s}{\nu(t)}\right) \Bigg) + \frac{(\nu(t)-(t-s)\nu'(t))^2}{\nu(t)^{9/2}} k''\left(\frac{t-s}{\nu(t)}\right) \\
& - \frac{2(t-s)\nu''(t)\nu(t)^{5/2} + 5\nu'(t)\nu(t)^{5/2} - 5(t-s)\nu'(t)^2\nu(t)^{3/2}}{\nu(t)^5} dB(s).
\end{aligned}$$

Again, let $u = (t-s)/\nu(t)$ and use the fact $\nu''(t) = 0$ and $\nu'(t)$ independent of t when $\nu(t)$ is linear,

$$\mathrm{Var}(X''(t)) = \frac{1}{\nu(t)^4} C_2 \qquad \text{(D.2)}$$

for some constant C_2. Similar to (D.1) and (D.2),

$$\mathbb{E}[X'(t)X''(t)] = \frac{1}{\nu(t)^3} C_3, \qquad \text{(D.3)}$$

where C_3 is a constant when $\nu(t)$ is linear. Combining (D.1), (D.2) and (D.3),

$$\rho(t) = -\frac{\mathrm{Var}(X'(t))}{\sqrt{\mathrm{Var}(X''(t)) - \frac{\mathbb{E}[X'(t)X''(t)]^2}{\mathrm{Var}(X'(t))}}} = \frac{C_1}{\sqrt{C_2 - \frac{C_3^2}{C_1}}}.$$

This finishes the proof.

□

Proof of Theorem 7. For convenience, apply the change of variable $v = -log\nu$. The covariance function

$$\mathbb{E}[X(t,v)X(\tilde{t},\tilde{v})] = e^{N(v+\tilde{v})/2} \int k((s-t)e^{v})k((s-\tilde{t})e^{\tilde{v}})ds.$$

At a fixed point, the first-order partial derivatives of X are Gaussian distributed with mean 0. The variances and covariances can be derived by applying (5.5.5) in Adler and Taylor (2007)

$$\begin{aligned}
\mathrm{Var}\left(\frac{\partial X}{\partial t}\right) &= e^{2v} \int \dot{k}(s)\dot{k}(s)'dh, \\
\mathrm{Var}\left(\frac{\partial X}{\partial v}\right) &= \int \left(s'\dot{k}(s) + \frac{N}{2}k(s)\right)^2 ds, \\
\mathbb{E}\left[\frac{\partial X}{\partial t}\frac{\partial X}{\partial v}\right] &= 0.
\end{aligned}$$

We define a random vector $G = (G_i)_{1 \le i \le N+1}$ such that

$$G_i = \begin{cases} \frac{1}{e^v}\frac{\partial X}{\partial t_i} & i \le N \\ \frac{\partial X}{\partial v} & i = N+1 \end{cases}.$$

Note that the distribution of G is independent of (t,v).

The second-order partial derivatives of X, i.e. entries of the Hessian matrix $\nabla^2 X$ are also Gaussian distributed with mean 0.

$$\frac{\partial^2 X}{\partial t \partial t'} = e^{2v}A, \quad \frac{\partial^2 X}{\partial t \partial v} = e^{v}B, \quad \frac{\partial^2 X}{\partial v^2} = C,$$

where the distribution of A is proportional to $\mathrm{GOI}(1/2)$ (Cheng and Schwartzman, 2018). The variance covariance matrix for vector B and scalar C are

$$\mathrm{Var}(B) = \int \left(\ddot{k}(s)s + \left(\frac{N}{2}+1\right)\dot{k}(s)\right)\left(\ddot{k}(s)s + \left(\frac{N}{2}+1\right)\dot{k}(s)\right)' ds,$$
$$\mathrm{Var}(C) = \int \left(s'\ddot{k}(s)s + (N+1)s'\dot{k}(s) + \frac{N^2}{4}k(s)\right)^2 ds,$$

and the covariance between A, B and C does not depend on the parameter (t,v).

$$\nabla^2 X(t,v) = \begin{bmatrix} (\frac{\partial^2 X}{\partial t_i \partial t_j})_{1\le i,j\le N} & (\frac{\partial^2 X}{\partial t_i \partial v})_{1\le i\le N} \\ (\frac{\partial^2 X}{\partial v \partial t_j})_{1\le j\le N} & \frac{\partial^2 X}{\partial v^2} \end{bmatrix} = \begin{bmatrix} e^{2v}A & e^v B \\ e^v B & C \end{bmatrix},$$

and define the matrix H as

$$H = \begin{bmatrix} A & B \\ B & C \end{bmatrix}.$$

The distribution of H is again independent of (t,v) like G. Similarly, the covariance between X and the entries of G, the covariance between X and the entries of H, and the covariance between the entries of G and the entries of H are all independent of (t,v). Therefore, the joint distribution of (X,G,H) is independent of (t,v).

Next, we have

$$\begin{aligned} \det \nabla^2 X(t,v) &= \det(e^{2v}A)\det(C - e^v B' e^{-2v} A^{-1} e^v B) \\ &= e^{2Nv}\det(A)\det(C - B'A^{-1}B) \\ &= e^{2Nv}\det(H). \end{aligned}$$

Note that the determinant of the kth principal minor of $\nabla^2 X(t,v)$ is equal to a positive number times the determinant of the kth principal minor of H. By Sylvester's criterion, $\nabla^2 X(t,v)$ being negative definite is equivalent to H being negative definite. Applying the Kac-Rice formula (5.1),

the peak height distribution of $X(t,v)$ for any fixed point

$$\begin{aligned}
F_{t,v}(u) &= \frac{\mathbb{E}[|\det\nabla^2 X(t,v)|\mathbb{1}_{\{\nabla^2 X(t,v)\prec 0\}}\mathbb{1}_{\{X(t,v)>u\}}|\nabla X(t,v)=0\}}{\mathbb{E}\{|\det\nabla^2 X(t,v)|\mathbb{1}_{\{\nabla^2 X(t,v)\prec 0\}}|\nabla X(t,v)=0]} \\
&= \frac{\int_x^\infty \mathbb{E}[|\det\nabla^2 X(t,v)|\mathbb{1}_{\{\nabla^2 X(t,v)\prec 0\}}|X(t,v)=x,\nabla X(t,v)=0]\phi(x)dx}{\mathbb{E}[|\det\nabla^2 X(t,v)|\mathbb{1}_{\{\nabla^2 X(t,v)\prec 0\}}|\nabla X(t,v)=0]} \\
&= \frac{e^{2Nv}\mathbb{E}[|\det H|\mathbb{1}_{\{H\prec 0\}}\mathbb{1}_{\{X>u\}}|G=0]}{e^{2Nv}\mathbb{E}[|\det H|\mathbb{1}_{\{H\prec 0\}}|G=0]} \\
&= \frac{\mathbb{E}[|\det H|\mathbb{1}_{\{H\prec 0\}}\mathbb{1}_{\{X>u\}}|G=0]}{\mathbb{E}[|\det H|\mathbb{1}_{\{H\prec 0\}}|G=0]}.
\end{aligned}$$

This finishes the proof. □

Bibliography

R. J. Adler and J. E. Taylor. *Random Fields and Geometry.* New York: Springer, 2007.

P. S. Aisen, L. S. Schneider, M. Sano, R. Diaz-Arrastia, C. H. van Dyck, M. F. Weiner, T. Bottiglieri, S. Jin, K. T. Stokes, R. G. Thomas, L. J. Thal, and Alzheimer Disease Cooperative Study. High-dose B vitamin supplementation and cognitive decline in alzheimer disease: a randomized controlled trial. *JAMA*, 300(15):1774–1783, October 2008.

M. C. Ard and S. D. Edland. Power calculations for clinical trials in alzheimer's disease. *J. Alzheimers. Dis.*, 26 Suppl 3(s3):369–377, 2011.

J.-M. Azaïs and M. Wschebor. A general expression for the distribution of the maximum of a gaussian field and the approximation of the tail. *Stochastic Processes and their Applications*, 118(7):1190–1218, 2008. ISSN 0304-4149. doi: https://doi.org/10.1016/j.spa.2007.07.016. URL https://www.sciencedirect.com/science/article/pii/S030441490700138X.

D. M. Barch, G. C. Burgess, M. P. Harms, S. E. Petersen, B. L. Schlaggar, M. Corbetta, M. F. Glasser, S. Curtiss, S. Dixit, C. Feldt, D. Nolan, E. Bryant, T. Hartley, O. Footer, J. M. Bjork, R. Poldrack, S. Smith, H. Johansen-Berg, A. Z. Snyder, D. C. Van Essen, and WU-Minn HCP Consortium. Function in the human connectome: task-fMRI and individual differences in behavior. *Neuroimage*, 80:169–189, October 2013.

E. Bertschinger. Multiscale gaussian random fields and their application to cosmological simulations. *The Astrophysical Journal Supplement Series*, 137:1–20, November 2001.

R. H. Browne. On the use of a pilot sample for sample size determination. *Statistics in Medicine*, 14:1933–1940, 1995.

S. Budd Haeberlein, P. S. Aisen, F. Barkhof, S. Chalkias, T. Chen, S. Cohen, G. Dent, O. Hansson, K. Harrison, C. von Hehn, T. Iwatsubo, C. Mallinckrodt, C. J. Mummery, K. K. Muralidharan, I. Nestorov, L. Nisenbaum, R. Rajagovindan, L. Skordos, Y. Tian, C. H. van Dyck, B. Vellas, S. Wu, Y. Zhu, and A. Sandrock. Two randomized phase 3 studies of aducanumab in early Alzheimer's disease. *J Prev Alzheimer's Dis*, 9(2):197–210, 2022.

D. Cheng and A. Schwartzman. Distribution of the height of local maxima of gaussian random

fields. *Extremes*, 18:213–240, 2015.

D. Cheng and A. Schwartzman. Multiple testing of local maxima for detection of peaks in random fields. *Annals of Statistics*, 45(2):529–556, 2017.

D. Cheng and A. Schwartzman. Expected number and height distribution of critical points of smooth isotropic gaussian random fields. *Bernoulli*, 24(4B):3422–3446, 2018.

D. Cheng and A. Schwartzman. On critical points of gaussian random fields under diffeomorphic transformations. *Statistics & Probability Letters*, 158:108672, 2020. ISSN 0167-7152. doi: https://doi.org/10.1016/j.spl.2019.108672. URL https://www.sciencedirect.com/science/article/pii/S0167715219303189.

J. Chumbley, K. Worsley, G. Flandin, and K. Friston. Topological FDR for neuroimaging. *Neuroimage*, 49(4):3057–3064, February 2010.

J. Cohen. *Statistical Power Analysis for the Behavioral Sciences*. Routledge, 1988. ISBN 978-1-134-74270-7.

J. de Leeuw and I. Kreft. Random coefficient models for multilevel analysis. *Journal of Educational Statistics*, 11(1):57–85, 1986. ISSN 03629791. URL http://www.jstor.org/stable/1164848.

M. C. Donohue. *longpower: Power and sample size calculations for longitudinal data*, 2019. R package version 1.0-19.

M. C. Donohue, O. Langford, P. S. Insel, C. H. van Dyck, R. C. Petersen, S. Craft, G. Sethuraman, R. Raman, P. S. Aisen, and Alzheimer's Disease Neuroimaging Initiative. Natural cubic splines for the analysis of Alzheimer's clinical trials. *Pharm Stat*, January 2023. doi: 10.1002/pst.2285. URL 10.1002/pst.2285.

J. Durnez, J. Degryse, B. Moerkerke, R. Seurinck, V. Sochat, R. A. Poldrack, and T. E. Nichols. Power and sample size calculations for fmri studies based on the prevalence of active peaks. 2016. URL https://www.biorxiv.org/content/biorxiv/early/2016/04/20/049429.full.pdf.

S. D. Edland, M. C. Ard, W. Li, and L. Jiang. Design of pilot studies to inform the construction of composite outcome measures. *Alzheimers Dement. (N. Y.)*, 3(2):213–218, June 2017.

G. M. Foody. Spatial nonstationarity and scale-dependency in the relationship between species richness and environmental determinants for the sub-saharan endemic avifauna. *Global Ecology and Biogeography*, 13(4):315–320, 2004.

C. R. Genovese, N. A. Lazar, and T. Nichols. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage*, 15(4):870–878, April 2002.

S. Hayasaka, A. M. Peiffer, C. E. Hugenschmidt, and P. J. Laurienti. Power and sample size calculation for neuroimaging studies by non-central random field theory. *Neuroimage*, 37(3): 721–730, September 2007.

D. Hedeker, R. D. Gibbons, and C. Waternaux. Sample size estimation for longitudinal designs with attrition: Comparing time-related contrasts between two groups. *Journal of Educational and Behavioral Statistics*, 24(1):70–93, 1999. doi: 10.3102/10769986024001070. URL https://doi.org/10.3102/10769986024001070.

R. Heller, D. Stanley, D. Yekutieli, N. Rubin, and Y. Benjamini. Cluster-based analysis of FMRI data. *Neuroimage*, 33(2):599–608, November 2006.

ICH Working Group. *Addendum On Estimands And Sensitivity Analysis in Clinical Trials to the Guideline on Statistical Principles for Clinical Trials E9(R1)*. European Medical Agency, 2019. URL https://database.ich.org/sites/default/files/E9-R1_Step4_Guidel\ine_2019_1203.pdf.

S.-H. Jung and C. Ahn. Sample size estimation for GEE method for comparing slopes in repeated measurements data. *Stat. Med.*, 22(8):1305–1315, April 2003.

A. Kuznetsova, P. B. Brockhoff, and R. H. B. Christensen. lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13):1–26, 2017. doi: 10.18637/jss.v082.i13.

N. M. Laird and J. H. Ware. Random-effects models for longitudinal data. *Biometrics*, 38(4): 963–974, December 1982.

M. Lázaro-Gredilla and M. K. Titsias. Variational heteroscedastic gaussian process regression. *Proceedings of the 28th International Conference on Machine Learning*, pages 841–848, 2011.

Q. V. Le, A. J. Smola, and S. Canu. Heteroscedastic gaussian process regression. *Proceedings of the 22nd International Conference on Machine Learning*, pages 489–496, 2005.

G. Lindgren. Local maxima of Gaussian fields. *Arkiv för Matematik*, 10(1-2):195 – 218, 1972. doi: 10.1007/BF02384809. URL https://doi.org/10.1007/BF02384809.

K. Lu, X. Luo, and P.-Y. Chen. Sample size estimation for repeated measures analysis in randomized clinical trials with missing data. *Int. J. Biostat.*, 4(1):Article 9, 2008.

C. H. Mallinckrodt, P. W. Lane, D. Schnell, Y. Peng, and J. P. Mancuso. Recommendations for the primary analysis of continuous endpoints in longitudinal clinical trials. *Drug Information Journal*, 42:303–319, 2008.

L. K. McEvoy, S. D. Edland, D. Holland, D. J. Hagler, Jr, J. C. Roddey, C. Fennema-Notestine, D. P. Salmon, A. K. Koyama, P. S. Aisen, J. B. Brewer, A. M. Dale, and Alzheimer's Disease

Neuroimaging Initiative. Neuroimaging enrichment strategy for secondary prevention trials in alzheimer disease. *Alzheimer Dis. Assoc. Disord.*, 24(3):269–277, July 2010.

C. L. Meinert. *Clinical Trials Design, Conduct, and Analysis*. Oxford University Press, New York, 1986.

K. E. Muller, L. M. Lavange, S. L. Ramey, and C. T. Ramey. Power calculations for general linear multivariate models including repeated measures applications. *J. Am. Stat. Assoc.*, 87 (420):1209–1226, December 1992.

D. B. Owen. A table of normal integrals. *Communications in Statistics - Simulation and Computation*, 9(4):389–419, 1980. doi: 10.1080/03610918008812164. URL https://doi.org/10.1080/03610918008812164.

C. Paciorek and M. Schervish. Nonstationary covariance functions for gaussian process regression. *Advances in Neural Information Processing Systems*, 16, 2003.

R. C. Petersen, R. G. Thomas, M. Grundman, D. Bennett, R. Doody, S. Ferris, D. Galasko, S. Jin, J. Kaye, A. Levey, E. Pfeiffer, M. Sano, C. H. van Dyck, L. J. Thal, and Alzheimer's Disease Cooperative Study Group. Vitamin E and donepezil for the treatment of mild cognitive impairment. *N. Engl. J. Med.*, 352(23):2379–2388, June 2005.

J. C. Pinheiro and D. M. Bates. *Multivariate Statistical Modelling Based on Generalized Linear Models*. Springer-Verlag, New York, 2000.

J. Pinheiro, D. Bates, S. DebRoy, D. Sarkar, and R Core Team. *nlme: Linear and Nonlinear Mixed Effects Models*, 2021. URL https://CRAN.R-project.org/package=nlme. R package version 3.1-152.

V. I. Piterbarg. Rice's method for large excursions of gaussian random fields. *Technical Report NO 478. Center for Stochastic Processes, Univ. North Carolina*, 1996.

J. F. Quinn, R. Raman, R. G. Thomas, K. Yurko-Mauro, E. B. Nelson, C. Van Dyck, J. E. Galvin, J. Emond, C. R. Jack, Jr, M. Weiner, L. Shinto, and P. S. Aisen. Docosahexaenoic acid supplementation and cognitive decline in alzheimer disease: a randomized trial. *JAMA*, 304 (17):1903–1911, November 2010.

L. L. Raket. Progression models for repeated measures: Estimating novel treatment effects in progressive diseases. *Stat Med*, 41(28):5537–5557, December 2022.

S. Remes, M. Heinonen, and S. Kaski. Non-stationary spectral kernels. *Advances in Neural Information Processing Systems*, 30, 2017.

J. Rochon. Sample size calculations for two-group repeated-measures experiments. *Biometrics*,

47(4):1383–1398, 1991. ISSN 0006341X, 15410420. URL http://www.jstor.org/stable/2532393.

A. Schwartzman and F. Telschow. Peak p-values and false discovery rate inference in neuroimaging. *Neuroimage*, 197:402–413, August 2019a.

A. Schwartzman and F. Telschow. Peak p-values and false discovery rate inference in neuroimaging. *NeuroImage*, 197:402–413, 2019b. doi: doi:10.1016/j.neuroimage.2019.04.041.

A. Schwartzman, W. F. Mascarenhas, and J. E. Taylor. Inference for eigenvalues and eigenvectors of Gaussian symmetric matrices. *The Annals of Statistics*, 36(6):2886 – 2919, 2008. doi: 10.1214/08-AOS628. URL https://doi.org/10.1214/08-AOS628.

S. R. Searle. Large sample variances of maximum likelihood estimators of variance components using unbalanced data. *Biometrics*, 26(3):505–524, 1970. ISSN 0006341X, 15410420. URL http://www.jstor.org/stable/2529105.

D. O. Siegmund and K. J. Worsley. Testing for a Signal with Unknown Location and Scale in a Stationary Gaussian Random Field. *The Annals of Statistics*, 23(2):608 – 639, 1995. doi: 10.1214/aos/1176324539. URL https://doi.org/10.1214/aos/1176324539.

T. A. B. Snijders and R. J. Bosker. Standard errors and sample sizes for two-level research. *Journal of Educational Statistics*, 18(3):237–259, 1993. ISSN 03629791. URL http://www.jstor.org/stable/1165134.

X. M. Tu, J. Zhang, J. Kowalski, J. Shults, C. Feng, W. Sun, and W. Tang. Power analyses for longitudinal study designs with missing data. *Stat. Med.*, 26(15):2958–2981, July 2007.

C. H. van Dyck, C. J. Swanson, P. Aisen, R. J. Bateman, C. Chen, M. Gee, M. Kanekiyo, D. Li, L. Reyderman, S. Cohen, L. Froelich, S. Katayama, M. Sabbagh, B. Vellas, D. Watson, S. Dhadda, M. Irizarry, L. D. Kramer, and T. Iwatsubo. Lecanemab in early Alzheimer's disease. *N Engl J Med*, 388(1):9–21, January 2023.

D. C. Van Essen, K. Ugurbil, E. Auerbach, D. Barch, T. E. J. Behrens, R. Bucholz, A. Chang, L. Chen, M. Corbetta, S. W. Curtiss, S. Della Penna, D. Feinberg, M. F. Glasser, N. Harel, A. C. Heath, L. Larson-Prior, D. Marcus, G. Michalareas, S. Moeller, R. Oostenveld, S. E. Petersen, F. Prior, B. L. Schlaggar, S. M. Smith, A. Z. Snyder, J. Xu, E. Yacoub, and WU-Minn HCP Consortium. The human connectome project: a data acquisition perspective. *Neuroimage*, 62 (4):2222–2231, October 2012.

R. Vozdolska, M. Sano, P. Aisen, and S. D. Edland. The net effect of alternative allocation ratios on recruitment time and trial cost. *Clin. Trials*, 6(2):126–132, April 2009.

C. Wang, C. B. Hall, and M. Kim. A comparison of power analysis methods for evaluating

effects of a predictor on slopes in longitudinal designs with missing data. *Stat. Methods Med. Res.*, 24(6):1009–1029, December 2015.

K. J. Worsley, A. C. Evans, S. Marrett, and P. Neelin. A three-dimensional statistical analysis for CBF activation studies in human brain. *J. Cereb. Blood Flow Metab.*, 12(6):900–918, November 1992.

K. J. Worsley, S. Marrett, P. Neelin, and A. C. Evans. Searching scale space for activation in PET images. *Hum. Brain Mapp.*, 4(1):74–90, 1996a.

K. J. Worsley, S. Marrett, P. Neelin, A. C. Vandal, K. J. Friston, and A. C. Evans. A unified statistical approach for determining significant signals in images of cerebral activation. *Hum. Brain Mapp.*, 4(1):58–73, 1996b.

K. J. Worsley, J. E. Taylor, F. Tomaiuolo, and J. Lerch. Unified univariate and multivariate random field theory. *Neuroimage*, 23 Suppl 1:S189–95, 2004.

Y. Zhao and S. D. Edland. Power formulas for mixed effects models with random slope and intercept comparing rate of change across groups. *Int J Biostat*, 18(1):173–182, January 2021.

Y. Zhao, D. Cheng, and A. Schwartzman. An approximation to peak detection power using gaussian random field theory. 2023.