Inference in Perturbation Models, Finite Mixtures and Scan Statistics: The Volume-of-Tube Formula

Ramani S. Pilla[1] and Catherine Loader[2]

Department of Statistics, Case Western Reserve University, Cleveland, OH 44106

pilla@case.edu catherine@case.edu

Abstract

This research creates a general class of *perturbation models* which are described by an underlying *null* model that accounts for most of the structure in data and a perturbation that accounts for possible small localized departures. The perturbation models encompass finite mixture models and spatial scan process. In this article, (1) we propose a new test statistic to detect the presence of perturbation, including the case where the null model contains a set of nuisance parameters, and show that it is equivalent to the likelihood ratio test; (2) we establish that the asymptotic distribution of the test statistic is equivalent to the supremum of a Gaussian random field over a high-dimensional manifold (e.g., curve, surface etc.) with boundaries and singularities; (3) we derive a technique for approximating the quantiles of the test statistic using the Hotelling-Weyl-Naiman *volume-of-tube formula*; and (4) we solve the long-pending problem of testing for the order of a mixture model; in particular, derive the asymptotic null distribution for a general family of mixture models including the multivariate mixtures. The inferential theory developed in this article is applicable for a class of non-regular statistical problems involving loss of identifiability or when some of the parameters are on the boundary of the parametric space.

Keywords: Gaussian random field, Likelihood ratio test statistic, Multivariate Mixture Models, Nonparametric maximum likelihood estimator, Nuisance parameters, Score process, Volume-of-tube formula.

[1]Research supported in part by the National Science Foundation (NSF) grant DMS 02-39053 and Office of Naval Research (ONR) grants N00014-02-1-0316 and N00014-04-1-0481.

[2]Research supported in part by the NSF grant DMS 03-06202 and ONR grant N00014-04-1-0481.

1 Introduction and Motivation

A fundamental and yet a very challenging problem in finite mixtures is determining the *order of a mixture model* or *mixture complexity*. This problem has been under intense investigation for over thirty years (Wolfe, 1971; Roeder, 1990; Lindsay, 1995) with no practically feasible solution for a general class of mixture families. Establishing a valid large-sample theoretical framework along with a practically feasible machinery for testing the order of a mixture model formed from a broad class of densities remains an open problem and is the focus of this research. It has long been noted that testing for the number of mixture components is a non-regular problem (a) due to loss of identifiability of the null distribution (i.e., the parameters representing the null distribution are not unique) and (b) since the parameters under the null hypothesis are on the boundary of the parameter space, instead of its interior. Consequently, the likelihood ratio test (LRT) statistic does not have the standard asymptotic null distribution of chi-squared (Chernoff, 1954; Ghosh and Sen, 1985; Hartigan, 1985; Bickel and Chernoff, 1993). As noted by several authors, the asymptotic null distribution of the LRT statistic is highly complex and very difficult to simulate from in practice.

The main thrust of this research is to create a fundamental class of models referred to as *perturbation models* and derive large-sample theory to detect the presence of perturbation. These models play an instrumental role in the development of inferential theory for a class of important problems such as (1) testing for the order of a mixture model formed from smooth families of densities, including the multivariate case; (2) searching for an unusual activity or region in the context of spatial scan process; and (3) detecting a signal in the presence of noisy backgrounds (Pilla et al., 2005). The resulting theory has broad applications in astronomy, astrophysics, biology, medicine, particle physics and datamining, to name a few.

1.1 Perturbation Models

Let $\mathcal{P} = \{p(x; \eta, \boldsymbol{\lambda}, \boldsymbol{\theta}) \colon \boldsymbol{\lambda} \in \boldsymbol{\Lambda}, \boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathcal{R}^d\}$ be a family of probability density functions. Assume that $\mathbf{X} = (X_1, \ldots, X_n)$ is an independently and identically distributed (i.i.d.)

random sample from

$$p(x; \eta, \boldsymbol{\lambda}, \boldsymbol{\theta}) := (1 - \eta)\, f(x; \boldsymbol{\lambda}) + \eta\, \psi(x; \boldsymbol{\theta}), \tag{1.1}$$

where $f(\cdot; \boldsymbol{\lambda})$ is a *null density* for an unknown parameter vector $\boldsymbol{\lambda} \in \boldsymbol{\Lambda}$, $\psi(\cdot; \boldsymbol{\theta})$ is a *perturbation density* with an unknown *nuisance parameter* vector $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathcal{R}^d$, both defined on a sample space $\mathcal{X} \subset \mathcal{R}^s$ and $\eta \in [0, 1]$ is the size of the perturbation. In the context of finite mixture models, the null model represents a mixture with m component densities and the perturbation model represents additional component densities. In the spatial scan process scenario, the null density accounts for the background or noise whereas the perturbation searches for an unusual activity.

The central idea is to introduce a *perturbation parameter* η which creates a departure from the null model. There are two primary goals: (1) Estimation of the parameters in the perturbation model and (2) testing the hypothesis

$$\mathcal{H}_0 \colon \eta = 0 \quad \text{against} \quad \mathcal{H}_1 \colon \eta > 0. \tag{1.2}$$

Under $\mathcal{H}_0$, $p(\cdot; \eta, \boldsymbol{\lambda}, \boldsymbol{\theta}) = f(\cdot; \boldsymbol{\lambda})$ and the null model entirely describes the data. However, under $\mathcal{H}_1$, the term $\eta\, \psi(\cdot; \boldsymbol{\theta})$ represents a departure from the null model.

The perturbation model falls into a class of problems studied by Davies (1977, 1987) in which a vector of *nuisance parameters* (in our case $\boldsymbol{\theta}$) appears only under the alternative hypothesis and standard asymptotic theory for the LRT breaks down. In particular, the asymptotic behavior of the LRT for the testing problem (1.2) is very difficult to characterize due to the difficulties with the geometry of the parameter space (scenarios (a) and (b) discussed earlier). It is worth noting that these same set of problems occur in the context of testing for homogeneity in finite mixture models. The inferential theory developed in this article requires only mild smoothness conditions on the family of densities while being generic and applicable much more widely. The two most important and distinct statistical problems motivating this work are finite mixture models (Lindsay, 1995) and spatial scan analysis (Glaz et al., 2001).

1.2 Inference in Mixture Models

Let $\mathcal{F} = \{\psi(x; \boldsymbol{\theta}) : \boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathcal{R}^d\}$ be a family of probability densities with respect to a σ-finite dominating measure μ for an s-dimensional random vector $x \in \mathcal{X} \subset \mathcal{R}^s$ and let

$\mathcal{G}$ be the space of all probability measures on $\boldsymbol{\Theta}$ with the σ-field generated by its Borel subsets. Assume that the component density $\psi(\cdot;\boldsymbol{\theta})$ is bounded in $\boldsymbol{\theta} \in \boldsymbol{\Theta}$.

Suppose that given $\boldsymbol{\theta}$, a random variable X has a density $\psi(x;\boldsymbol{\theta})$ and that $\boldsymbol{\theta}$ follows a distribution $\mathcal{Q}$, referred to as *mixing distribution*. For a given $\mathcal{Q} \in \mathcal{G}$, assume that the sample arises from the marginal density $\mathrm{g}(x;\mathcal{Q}) := \int_{\boldsymbol{\theta}} \psi(x;\boldsymbol{\theta})\ d\,\mathcal{Q}(\boldsymbol{\theta})$ for $x \in \mathcal{X} \subset \mathcal{R}^s$ referred to as a *mixture density* with a corresponding *mixing measure* $\mathcal{Q}$. In the case of a discrete and finitely supported mixing measure, the *mixing distribution* can be expressed as $\mathcal{Q}_m = \sum_{j=1}^{m} \beta_j\, \varepsilon(\boldsymbol{\theta}_j)$, where $\varepsilon(\cdot)$ is a point mass function and $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_m$ are distinct *support point vectors* with a corresponding vector of *mixing weights* $\boldsymbol{\beta} := (\beta_1, \ldots, \beta_m)^T$ such that $\boldsymbol{\beta}$ belongs to the interior of the unit simplex $\{\boldsymbol{\beta} : \sum_{j=1}^{m} \beta_j = 1, \beta_j \geq 0, j = 1, \ldots, m\}$. Therefore, mixture density can be expressed as $\mathrm{g}(x;\mathcal{Q}_m) = \sum_{j=1}^{m} \beta_j\, \psi(x;\boldsymbol{\theta}_j)$, where the number of support points m becomes the order of the mixture model or mixture complexity. The probability distribution $\mathcal{Q}_m$ that maximizes the loglikelihood $l(\mathcal{Q}_m) = \sum_{i=1}^{n} \log\left[\mathrm{g}(x_i;\mathcal{Q}_m)\right]$ is the *nonparametric maximum likelihood estimator* (NPMLE) of $\mathcal{Q}_m$ (Lindsay, 1995).

A long-pending and very challenging problem is determining the order m of the mixture model. In the perturbation model framework, if $f(;\boldsymbol{\lambda})$ represents the m-component mixture density $\mathrm{g}(\cdot;\mathcal{Q}_m)$, then $\psi(\cdot;\boldsymbol{\theta}_{m+1})$ represents the $(m+1)$st component density. Therefore, inferential theory for perturbation models provides the machinery for testing the order of a mixture model. If m is fixed, the loglikelihood has multiple local maxima and the LRT has an unknown limiting distribution. In the case of normal mean mixtures and under severe identifiability conditions, Ghosh and Sen (1985) derived the asymptotic null distribution of the LRT as

$$\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left[Z(\boldsymbol{\theta})\right]^2 \mathbf{1}\left[Z(\boldsymbol{\theta}) \geq 0\right], \tag{1.3}$$

where $Z(\boldsymbol{\theta})$ is a zero mean Gaussian process indexed by a set $\boldsymbol{\theta}$ with a specified covariance function and $\mathbf{1}[\cdot]$ is the indicator function. When the support set of certain parameters in the model is unbounded (e.g., in normal and gamma mixtures), the LRT statistic can diverge to infinity as $n \to \infty$ instead of having a limiting distribution (Hartigan, 1985; Liu et al., 2003). This divergence of the LRT poses major difficulties in characterizing the distribution of the LRT and in obtaining reliable simulation results for the null distribution (Lindsay, 1995). For testing in multinomial mixture models, Lindsay (1995) derived approximation to the asymptotic distribution of the LRT based on the Hotelling-Weyl (Hotelling, 1939; Weyl, 1939) volume-of-tube formula.

Existing theoretical results have been obtained only for some special cases and many researchers have considered simulation and resampling based approaches to approximate the asymptotic null distribution of the LRT for simple models; see Lindsay (1995) and McLachlan and Peel (2000) for detailed discussion and other references. Dacunha-Castelle and Gassiat (1999) proposed a general theory for the asymptotic null distribution of the LRT in testing for $\mathcal{H}_0 \colon m = p$ mixtures against $\mathcal{H}_1 \colon m = q$ mixtures, where $q > p$ using a locally conic parameterization. Under certain stringent conditions, they showed that the asymptotic null distribution of the LRT statistic has a form similar to (1.3); however, tail probability calculations required for calibrating the LRT statistic are not derived. Unfortunately, analytic derivations of the distribution of supremum of the Gaussian process are difficult problems. Most importantly, the issue of "singularities of the process" (as described in Section 3.3) is of fundamental importance in the context of mixture testing problem and it has not been addressed in the existing literature, including by Dacunha-Castelle and Gassiat (1999).

The perturbation theory developed in this article, provides an elegant and flexible machinery for approximating the quantiles of the test statistic for the following class of fundamental problems: (1) testing problems in which the true parameter is on the boundary of the hypotheses regions; (2) testing $\mathcal{H}_0$: m-component mixture against $\mathcal{H}_1$: $(m + q)$-component mixture for $q = 1, 2, \ldots$ when mixtures are formed from any smooth families, including discrete, continuous and multivariate densities; and (3) testing for the presence of a signal when the probability density functions under the null and alternative hypotheses belong to different parametric families which occurs in physics applications (Pilla et al., 2005).

1.3 Inference in Spatial Scan Statistics

In the scan statistics problem, one observes a random field (such as a point process) in a region of interest. The goal is to detect unusual behavior in subregions, where the behavior of the field differs significantly from the background. Applications include mammography; automatic target recognition; disease clustering and minefield detection.

In the classical formulation of the scan statistic (see Glaz et al. (2001) and the references therein), a rectangular window is scanned across the data, with high values of the statistic indicating a local departure from uniformity. In contrast, the methods developed in this article are applicable to smooth scanning processes, where the window is tapered, rather than having sharp boundaries. The null density $f(\cdot;\boldsymbol{\lambda})$ represents the background model while the scan window $\psi(\cdot;\boldsymbol{\theta})$ represents departure from the background at location $\boldsymbol{\theta}$.

1.4 Main Results

We create a general family of models referred to as perturbation models that encompass a large class of statistical problems. Our treatment of the nuisance parameters under the null hypothesis is quite general. The inferential theory developed in this article provides a solution to an important class of statistical problems involving loss of identifiability and/or when some of the parameters are on the boundary of the parametric space. The main contributions of this article are as follows.

1. In Section 2, we propose a novel test statistic based on the score process, denoted by $\mathcal{T}$, for detecting the presence of perturbation and derive its fundamental properties. In particular, it is shown that the test statistic $\mathcal{T}$ based on the score process is asymptotically equivalent to the LRT statistic.
2. In Section 3, we derive a general inferential theory for approximating the asymptotic null distribution of $\mathcal{T}$. It is shown that the asymptotic distribution of $\mathcal{T}$ under $\mathcal{H}_0$ equals $\sup_{\boldsymbol{\theta}} Z(\boldsymbol{\theta})$, where $Z(\boldsymbol{\theta})$ is a differentiable Gaussian random field with continuous sample paths. Therefore, the goal becomes finding approximations for $\mathbb{P}(\sup_{\boldsymbol{\theta}} Z(\boldsymbol{\theta}) \geq c)$ for any large $c \in \mathcal{R}$ in order to determine the quantiles of $\mathcal{T}$. As eloquently pointed out by Adler (2000), this problem occurs in a large number of different applications including in image processing (Worsley, 1995). We describe a connection between $Z(\boldsymbol{\theta})$ and a differentiable manifold (curve, surface, etc.) through the *Karhunen-Loève expansion*. The Karhunen-Lòeve expansion converts the high-dimensional Gaussian probability problem into that of a chi-squared random variable and uniformly distributed random variables over the surfaces of spheres (Adler, 2000).

3. Our technique is based on the long-established and elegant geometric result known as the *volume-of-tube formula* (Hotelling, 1939; Weyl, 1939; Naiman, 1990). The problem of evaluating the Gaussian random field significance probabilities (i.e., tail probability for the asymptotic null distribution of $\mathcal{T}$) for testing the hypothesis (1.2) is reduced to that of determining the volume-of-tube about a manifold on the surface of a hypersphere (see Section 3.2). The novelty here lies in deriving explicit expressions for the geometric constants appearing in the volume-of-tube formula with boundaries; consequently, one can approximate the quantiles of the statistic $\mathcal{T}$ for detecting the presence of perturbation. We also address the difficult and yet important problem of presence of singularities in the score process.
4. In Section 4, the results of Section 3 are extended to the case where the null density is characterized by a vector of nuisance parameters.
5. An age old and fundamental question of determining the order of a mixture model is solved in Section 5. In particular, building on the perturbation theory, we develop inferential methods for approximating the quantiles of the test statistic for determining the mixture complexity. The flexibility and general applicability of the methodology is demonstrated through univariate and multivariate mixture families. Furthermore, it is shown that the results of Lindsay (1995), Lin (1997) and Chen and Chen (2001) become special cases of our general and broadly applicable theory.

The paper concludes with a discussion of the relative merits of the perturbation theory in Section 6. In Section 7, we derive the proofs of our general results. Explicit expressions for the geometric constants that appear in the volume-of-tube formula are derived in Appendix A.

2 A Score Process and its Fundamental Properties

In this section, we derive a score process and its fundamental properties that are required for the testing problem (1.2). As a first step, we assume that $\boldsymbol{\lambda}$ is fixed or known so that $f(;\boldsymbol{\lambda})$ is completely specified and the density (1.1) can be expressed simply as $p(;\eta,\boldsymbol{\theta})$; however, theory for the general case of an unknown $\boldsymbol{\lambda}$ will be derived in Section 4.

2.1 Loglikelihood Ratio Process

If $\boldsymbol{\theta}$ is fixed at a particular value, then the testing problem (1.2) becomes routine. However, the nuisance parameter vector $\boldsymbol{\theta}$ can assume any value under $\mathcal{H}_0$; therefore, the testing problem is non-regular. The loglikelihood function based on the perturbation model (1.1) is $l(\eta, \boldsymbol{\theta}|\mathbf{x}) = \sum_{i=1}^{n} \log\left[(1-\eta)\, f(x_i; \boldsymbol{\lambda}) + \eta\, \psi(x_i; \boldsymbol{\theta})\right]$. For a fixed $\boldsymbol{\theta}$, $l(\eta, \boldsymbol{\theta}|\mathbf{x})$ is a concave function of η and hence there exists a unique maximizer $\widehat{\eta}_{\boldsymbol{\theta}} \in [0,1]$. In general, there is no closed form solution for $\widehat{\eta}_{\boldsymbol{\theta}}$; however, the estimator can be found as a solution to

$$\sum_{i=1}^{n} \frac{[\psi(x_i; \boldsymbol{\theta}) - f(x_i; \boldsymbol{\lambda})]}{p(x_i; \eta, \boldsymbol{\theta})} = 0 \tag{2.1}$$

if a solution in $(0,1)$ exists; otherwise the estimator will be at one of the end-points. This leads to a corresponding *loglikelihood ratio process* $l^{\star}(\boldsymbol{\theta}|\mathbf{x}) = l(\widehat{\eta}_{\boldsymbol{\theta}}, \boldsymbol{\theta}|\mathbf{x}) - l(0, 0|\mathbf{x})$. Considered as a function of $\boldsymbol{\theta}$, the process $l^{\star}(\boldsymbol{\theta}|\mathbf{x})$ may be used as a diagnostic tool, with large values indicating the presence of perturbation. The maximum likelihood estimator (MLE) of $\boldsymbol{\theta}$ is the maximizer of $l^{\star}(\boldsymbol{\theta}|\mathbf{x})$. However, maximizing this process is computationally intensive, since $l^{\star}(\boldsymbol{\theta}|\mathbf{x})$ may have many local maxima. Any strategy for finding the global maximum has to involve an exhaustive search, which in turn requires solving (2.1) for each fixed $\boldsymbol{\theta}$. In the next section, we derive an alternative technique that will combat these difficulties.

2.2 The Score Process: Theory

In this section, we propose a novel technique based on a *score process* defined as

$$S(\boldsymbol{\theta}) \quad := \quad \frac{\partial}{\partial \eta} l(\eta, \boldsymbol{\theta}, \boldsymbol{\lambda}|\mathbf{x})\bigg|_{\eta=0} = \sum_{i=1}^{n} \left[\frac{\psi(x_i; \boldsymbol{\theta})}{f(x_i; \boldsymbol{\lambda})} - 1\right]. \tag{2.2}$$

The interest is in the parameter vector $\boldsymbol{\theta}$ and since $\boldsymbol{\lambda}$ is fixed for now, for exposition, we drop $\boldsymbol{\lambda}$ from the expressions and simply write $S(\boldsymbol{\theta}), S^{\star}(\boldsymbol{\theta}), Z(\boldsymbol{\theta})$, etc.

The score process has several elegant features: (1) it is not as computationally intensive as the likelihood ratio process and (2) its explicit representation makes statistical inference tractable. It is shown in Theorem 1 (below) that the score process has mean zero when there is no perturbation (i.e., $\eta = 0$) and $\mathbb{E}[S(\boldsymbol{\theta})] > 0$ when there is a perturbation at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, the true parameter vector. This suggests that peaks in the score process provide evidence for the presence of perturbation. However, $S(\boldsymbol{\theta})$ can exhibit high random variability and the variance may have substantial dependence on $\boldsymbol{\theta}$. To combat this difficulty, we propose the *normalized score process* defined as

$$S^{\star}(\boldsymbol{\theta}) := \frac{S(\boldsymbol{\theta})}{\sqrt{n\,\mathbb{C}(\boldsymbol{\theta},\boldsymbol{\theta})}},$$

where the *covariance function* is defined as

$$\begin{aligned} \mathbb{C}(\boldsymbol{\theta},\boldsymbol{\theta}^{\dagger}) &:= \int \frac{[\psi(x;\boldsymbol{\theta}) - f(x;\boldsymbol{\lambda})]\,[\psi(x;\boldsymbol{\theta}^{\dagger}) - f(x;\boldsymbol{\lambda})]}{f(x;\boldsymbol{\lambda})}\,dx \\ &= \int \frac{\psi(x;\boldsymbol{\theta})\,\psi(x;\boldsymbol{\theta}^{\dagger})}{f(x;\boldsymbol{\lambda})}\,dx - 1. \end{aligned} \tag{2.3}$$

The covariance function $\mathbb{C}(\boldsymbol{\theta},\boldsymbol{\theta}^{\dagger})$ has an analytical expression for certain choices of $f(\cdot;\boldsymbol{\lambda})$ and $\psi(\cdot;\boldsymbol{\theta})$ while in other cases numerical integration is required.

The following conditions are assumed for deriving the large-sample theory.

A1: The parameter space $\boldsymbol{\Theta}$ is a compact and a convex subset of $\mathcal{R}^d$ for some integer d.

A2: The covariance function satisfies $\mathbb{C}(\boldsymbol{\theta},\boldsymbol{\theta}) < \infty$ for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$.

A3: For each $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, $\mathrm{supp}[\psi(\,\cdot\,;\boldsymbol{\theta})] \subset \mathrm{supp}[f(\,\cdot\,;\boldsymbol{\lambda})]$, where 'supp' refers to the support of a density.

In the following theorem, we characterize some fundamental properties of the score and normalized score processes.

Theorem 1 *Suppose assumptions A2 and A3 hold: (1) Under $\mathcal{H}_0$, the score process has mean $\mathbb{E}[S(\boldsymbol{\theta})] = 0$ for all $\boldsymbol{\theta}$ with a covariance function $\mathrm{cov}[S(\boldsymbol{\theta}), S(\boldsymbol{\theta}^{\dagger})] = n\,\mathbb{C}(\boldsymbol{\theta},\boldsymbol{\theta}^{\dagger})$, where $\mathbb{C}(\boldsymbol{\theta},\boldsymbol{\theta}^{\dagger})$ is defined in (2.3); (2) under $\mathcal{H}_1$,*

$$\mathbb{E}[S(\boldsymbol{\theta})] = n\,\eta\,\mathbb{C}(\boldsymbol{\theta},\boldsymbol{\theta}_0); \tag{2.4}$$

and (3) under $\mathcal{H}_1$, the expectation of the normalized score process is

$$\mathbb{E}[S^{\star}(\boldsymbol{\theta})] = \sqrt{n}\,\eta\,\frac{\mathbb{C}(\boldsymbol{\theta},\boldsymbol{\theta}_0)}{\sqrt{\mathbb{C}(\boldsymbol{\theta},\boldsymbol{\theta})}} \le \eta\sqrt{n\,\mathbb{C}(\boldsymbol{\theta}_0,\boldsymbol{\theta}_0)} \tag{2.5}$$

with equality at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$.

Proof. Under $\mathcal{H}_1$, it follows that

$$\begin{aligned}\mathbb{E}[S(\boldsymbol{\theta})] &= n\int\left[\frac{\psi(x;\boldsymbol{\theta})}{f(x;\boldsymbol{\lambda})}-1\right]p(x;\eta,\boldsymbol{\theta}_0)\,dx\\ &= n\,\eta\int\frac{\psi(x;\boldsymbol{\theta})\,\psi(x;\boldsymbol{\theta}^\dagger)}{f(x;\boldsymbol{\lambda})}\,dx = n\,\eta\,\mathbb{C}(\boldsymbol{\theta},\boldsymbol{\theta}_0)\end{aligned}$$

which yields the result (2.4). Similarly, one can derive the mean and covariance functions in part 1 of the theorem. The bound (2.5) is established by noting that $\mathbb{C}(\boldsymbol{\theta},\boldsymbol{\theta}_0)$ is a covariance function and therefore satisfies the Cauchy-Schwartz inequality $\mathbb{C}(\boldsymbol{\theta},\boldsymbol{\theta}_0)\leq\sqrt{\mathbb{C}(\boldsymbol{\theta},\boldsymbol{\theta})\,\mathbb{C}(\boldsymbol{\theta}_0,\boldsymbol{\theta}_0)}$. ■

The motivation for using the score processes lies in part 3 of Theorem 1: The expectation of $S^\star(\boldsymbol{\theta})$ is maximized at $\boldsymbol{\theta}_0$. Therefore, the supremum of the process $S^\star(\boldsymbol{\theta})$ can serve as a test statistic for the hypothesis (1.2). If $\mathcal{H}_0$ is rejected, then the maximizer of $S^\star(\boldsymbol{\theta})$ serves as a point estimator of $\boldsymbol{\theta}$. The final result of this section establishes the asymptotic equivalence between the score and loglikelihood processes; the proof is given in Section 7.

Theorem 2 *The score process and loglikelihood ratio process are asymptotically equivalent, in the sense that* $l^\star(\boldsymbol{\theta}|\mathbf{x})=\frac{1}{2}[\max\{0,S^\star(\boldsymbol{\theta})\}]^2+o_p(1)$ *as* $n\to\infty$.

3 Testing for the Presence of Perturbation

We first propose a statistic for the testing problem (1.2) and next derive its asymptotic null distribution. From the motivation presented in the previous section, it is natural to define a statistic for testing the hypothesis (1.2) as

$$\mathcal{T}:=\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}S^\star(\boldsymbol{\theta}). \tag{3.1}$$

Except in special cases, the distribution of $\mathcal{T}$ cannot be expressed analytically. Our next goal is to derive an asymptotic distribution of $\mathcal{T}$ under $\mathcal{H}_0$ for determining approximate quantiles of the test statistic. As a first step, we establish that under $\mathcal{H}_0$ the distribution of $\mathcal{T}$ is asymptotically equivalent to the distribution of the supremum of a Gaussian random field. Next, we derive approximations for the tail probability of the supremum of a Gaussian random field using the Karhunen-Loève expansion and the volume-of-tube formula.

The volume-of-tube problem for curves (i.e., $d = 1$) was first studied by Hotelling (1939) in the context of significance testing for nonlinear regression. In a second pioneering paper, Weyl (1939) extended the work of Hotelling to higher-dimensional manifolds (i.e., $d \geq 2$), deriving elegant expressions for the volume-of-tube of manifolds lying in a hypersphere. Naiman (1990) further extended the Hotelling-Weyl results to cases where the manifold has boundaries. Sun (1993) studied higher order terms for Gaussian processes and fields. Important statistical problems to which the volume-of-tube formula has been applied include non-linear regression (Hotelling, 1939; Knowles and Siegmund, 1989), projection pursuit (Johansen and Johnstone, 1990), testing for multinomial mixture models (Lindsay, 1995; Lin, 1997), simultaneous confidence bands [Naiman (1987), Sun and Loader (1994) and Chapter 9 of Loader (1999)] and inference under convex cone alternatives for correlated data (Pilla, 2006).

The following assumptions are required for the development of inferential theory.

A4: For all $x \in \mathcal{X}$, the perturbation density $\psi(x; \boldsymbol{\theta})$ is a twice differentiable, while

$$\int \frac{\psi'(x, \boldsymbol{\theta})^2}{f(x, \boldsymbol{\lambda})} dx < \infty \quad \text{and} \quad \int \frac{\psi''(x, \boldsymbol{\theta})^2}{f(x, \boldsymbol{\lambda})} dx < \infty,$$

where $'$ denotes differentiation with respect to $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. In the multi-parameter case, all first and second-order partial derivatives are assumed to satisfy the integrability condition as well.

A5: The covariance function $\mathbb{C}(\boldsymbol{\theta}, \boldsymbol{\theta})$ is positive in $\boldsymbol{\theta}$; equivalently, $f(\cdot; \boldsymbol{\lambda})$ is not identically equal to $\psi(\cdot; \boldsymbol{\theta})$ for any $\boldsymbol{\theta} \in \boldsymbol{\Theta}$.

The assumption A5 fails in several important problems including mixture models, leading to singularities in the score process. In Section 3.3, we derive modifications to our theory to handle this difficult but important problem.

Let $\{Z(\boldsymbol{\theta}) : \boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathcal{R}^d\}$ be a d-dimensional differentiable Gaussian random field with continuous sample paths, with mean zero and covariance function

$$\rho(\boldsymbol{\theta}, \boldsymbol{\theta}^\dagger) := \mathbb{E}\left[Z(\boldsymbol{\theta}) Z(\boldsymbol{\theta}^\dagger)\right] = \frac{\mathbb{C}(\boldsymbol{\theta}, \boldsymbol{\theta}^\dagger)}{\sqrt{\mathbb{C}(\boldsymbol{\theta}, \boldsymbol{\theta})\, \mathbb{C}(\boldsymbol{\theta}^\dagger, \boldsymbol{\theta}^\dagger)}}. \tag{3.2}$$

Under assumptions A4 and A5, the asymptotic null distribution of $\mathcal{T}$ is the supremum of a Gaussian random field, expressed explicitly as

$$Z(\boldsymbol{\theta}) = [\mathbb{C}(\boldsymbol{\theta}, \boldsymbol{\theta})]^{-1/2} \int \left[\frac{\psi(x; \boldsymbol{\theta})}{f(x; \boldsymbol{\lambda})} - 1 \right] \sqrt{f(x, \boldsymbol{\lambda})}\, W(dx),$$

where W is the standard *Brownian sheet*.

Theorem 3 *Suppose that assumptions A1 to A5 hold. Under* $\mathcal{H}_0$*,*

$$\mathbb{P}\left(\mathcal{T} \geq c\right) \longrightarrow \mathbb{P}\left(\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} Z(\boldsymbol{\theta}) \geq c \right) \quad as \quad n \to \infty \quad for\ any \quad c \in \mathcal{R}. \tag{3.3}$$

Theorem 3 will be proved in Section 7. Generally, there is no exact result for finding $\mathbb{P}(\sup_{\boldsymbol{\theta}} Z(\boldsymbol{\theta}) \geq c)$ (Adler, 2000). The result of Theorem 3 holds even if we relax assumption A4. Our proof relies only on the assumption of first derivative of $\psi(\cdot; \boldsymbol{\theta})$; however, the second derivative conditions are required for the explicit probability approximations derived later using the volume-of-tube-formula.

The problem of approximating the distribution of the supremum of a smooth Gaussian random field (i.e., finding $\mathbb{P}(\sup_{\boldsymbol{\theta}} Z(\boldsymbol{\theta}) \geq c)$ for large c) can be addressed using several different techniques: (1) methods based on the Hotelling-Weyl (Hotelling, 1939; Weyl, 1939) volume-of-tube formula with boundary corrections (Naiman, 1990); (2) expected Euler characteristic methods (Siegmund and Worsley, 1995; Worsley, 2001); (3) approaches based on counting the local maxima and upcrossings; and (4) Rice formula (Siegmund and Zhang, 1993; Azaïs and Wschebor, 2005). All these techniques lead to similar results for practical purposes (see Adler (2000) for discussion). Some formal equivalence results between the tube formula and the expected Euler characteristic methods have been derived by Takemura and Kuriki (2002). In this article, for the development of inferential theory for perturbation models, we adopt the volume-of-tube formula technique for its relatively simple geometric interpretation and the flexibility to yield explicit results for higher-order boundary corrections. The disadvantage of the tube approach is that it is directly applicable only to processes that are Gaussian or Gaussian-like (Adler, 2000).

3.1 The Karhunen-Loève Expansion

In this section, we construct a sequence of finite-dimensional approximation to the Gaussian random field $Z(\boldsymbol{\theta})$ using the Karhunen-Loève expansion. Although Karhunen-Loève expansion is most convenient, any other uniformly convergent approximation, such as a cubic spline interpolant on a grid of $\boldsymbol{\Theta}$ is also applicable.

While some of the core ideas in this section are known, there does not exist a complete statement of the results in the form that are required for the general testing problem (1.2). In particular, addressing the following scenarios are of fundamental importance: (1) $\boldsymbol{\Theta}$ is a hyper-rectangle or a similar polygonal region with boundaries of various orders (edges, corners and so on) and (2) the score process $S(\boldsymbol{\theta})$ has singularities.

A concise presentation of the Karhunen-Loève expansion can be found in Section III.3 of Adler (1990). The Karhunen-Loève expansion of $Z(\boldsymbol{\theta})$ is the *uniformly convergent* series expansion

$$Z(\boldsymbol{\theta}) = \sum_{k=1}^{\infty} \mathfrak{z}_k \, \xi_k(\boldsymbol{\theta}) = \langle \boldsymbol{\mathfrak{z}}, \boldsymbol{\xi}(\boldsymbol{\theta}) \rangle \,, \tag{3.4}$$

where $\mathfrak{z}_k$ is an i.i.d. standard Gaussian random variable, $\{\xi_k(\boldsymbol{\theta})\}_{k=1}^{\infty}$ is a sequence of twice continuously differentiable functions, while $\boldsymbol{\mathfrak{z}}$ and $\boldsymbol{\xi}(\boldsymbol{\theta})$ are the corresponding vector counterparts. The covariance function (3.2) can be explicitly expressed as

$$\rho(\boldsymbol{\theta}, \boldsymbol{\theta}^{\dagger}) \;\; = \;\; \sum_{k=1}^{\infty} \xi_k(\boldsymbol{\theta}) \, \xi_k(\boldsymbol{\theta}^{\dagger}) \tag{3.5}$$

and $\mathfrak{z}_k = \mu_k^{-1} \int_{\boldsymbol{\Theta}} \xi_k(\boldsymbol{\theta}) \, Z(\boldsymbol{\theta}) \, d\boldsymbol{\theta}$, where $\mu_k = \int_{\boldsymbol{\theta}} \xi_k^2(\boldsymbol{\theta}) \, d\boldsymbol{\theta}$.

It is necessary for $Z(\boldsymbol{\theta})$ to have a finite Karhunen-Loève expansion for the application of the volume-of-tube formula. When the expansion is infinite, the series is truncated at J terms to yield

$$Z_J(\boldsymbol{\theta}) \;\; := \;\; \sum_{k=1}^{J-1} \mathfrak{z}_k \, \xi_k(\boldsymbol{\theta}) + \mathfrak{z}_0 \sqrt{\sum_{k=J+1}^{\infty} [\xi_k(\boldsymbol{\theta})]^2} = \langle \boldsymbol{\mathfrak{z}}_J, \boldsymbol{\xi}_J(\boldsymbol{\theta}) \rangle \,, \tag{3.6}$$

where $\mathfrak{z}_0 \sim N(0,1)$ and is independent of $\mathfrak{z}_1, \mathfrak{z}_2, \ldots$, $\boldsymbol{\mathfrak{z}}_J = (\mathfrak{z}_0, \ldots, \mathfrak{z}_{J-1})^T$ and $\boldsymbol{\xi}_J(\boldsymbol{\theta})$ is the corresponding truncated version of the sequence $\{\xi_k(\boldsymbol{\theta})\}_{k=1}^{\infty}$. The covariance function of $Z_J(\boldsymbol{\theta})$ can be expressed as

$$\rho_J(\boldsymbol{\theta}, \boldsymbol{\theta}^\dagger) = \sum_{k=1}^{J-1} \xi_k(\boldsymbol{\theta})\, \xi_k(\boldsymbol{\theta}^\dagger) + \sqrt{\sum_{k=J}^{\infty} \xi_k^2(\boldsymbol{\theta}) \sum_{k=J}^{\infty} \xi_k^2(\boldsymbol{\theta}^\dagger)} = \left\langle \boldsymbol{\xi}_J(\boldsymbol{\theta}), \boldsymbol{\xi}_J(\boldsymbol{\theta}^\dagger) \right\rangle . \tag{3.7}$$

The final term in (3.6) has been chosen to preserve unit variance; i.e., $\mathbb{V}[Z_J(\boldsymbol{\theta})] = \rho_J(\boldsymbol{\theta}, \boldsymbol{\theta}) = 1$.

3.2 Distribution of the Supremum of $Z(\boldsymbol{\theta})$

In this section, we provide an approximation to $\sup_{\boldsymbol{\theta}} Z(\boldsymbol{\theta})$ under a very general assumption that $\mathcal{M}$ is a manifold with a piecewise smooth boundary. This result, combined with Theorem 3 provides an elegant approximation to the asymptotic null distribution of the test statistic $\mathcal{T}$. The primary goal is to approximate the asymptotic probability in (3.3) when $c \in \mathcal{R}$ is large, $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathcal{R}^d$ and $d \geq 1$.

Conditioning on the length of the vector $\boldsymbol{\mathfrak{z}}_J$,

$$\begin{aligned} \mathbb{P}\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} Z_J(\boldsymbol{\theta}) \geq c\right) &= \mathbb{P}\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \langle \boldsymbol{\mathfrak{z}}_J, \boldsymbol{\xi}_J(\boldsymbol{\theta})\rangle \geq c\right) \\ &= \mathbb{P}\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\langle \frac{\boldsymbol{\mathfrak{z}}_J}{\|\boldsymbol{\mathfrak{z}}_J\|}, \boldsymbol{\xi}_J(\boldsymbol{\theta})\right\rangle \geq \frac{c}{\|\boldsymbol{\mathfrak{z}}_J\|}\right) \\ &= \int_{c^2}^{\infty} \mathbb{P}\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \langle \mathbf{U}_J, \boldsymbol{\xi}_J(\boldsymbol{\theta})\rangle \geq \frac{c}{\sqrt{y}}\right) h_J(y)\, dy, \end{aligned} \tag{3.8}$$

where the J-dimensional random vector $\mathbf{U}_J = (\mathfrak{z}_0/\|\boldsymbol{\mathfrak{z}}_J\|, \ldots, \mathfrak{z}_{J-1}/\|\boldsymbol{\mathfrak{z}}_J\|)^T$ is uniformly distributed on the unit sphere $\mathcal{S}^{(J-1)}$ embedded in $\mathcal{R}^J$, $\boldsymbol{\xi}(\boldsymbol{\theta})$ is a curve in $\mathcal{S}^{(J-1)}$ and $h_J(y)$ is the χ^2 density with J degrees of freedom. Consequently, the goal becomes evaluating the distribution of the supremum of a uniform process in (3.8).

First, note that the inner product $\langle \mathbf{U}_J, \boldsymbol{\xi}_J(\boldsymbol{\theta})\rangle$ is bounded by 1 (using the Cauchy-Schwarz inequality) enabling the restriction of $c/\sqrt{y} < 1$ or $c^2 < y < \infty$. Since $\|\mathbf{U}_J - \boldsymbol{\xi}_J(\boldsymbol{\theta})\|^2 = \|\mathbf{U}_J\|^2 + \|\boldsymbol{\xi}_J(\boldsymbol{\theta})\|^2 - 2\,\langle \mathbf{U}_J, \boldsymbol{\xi}_J(\boldsymbol{\theta})\rangle = 2[1 - \langle \mathbf{U}_J, \boldsymbol{\xi}_J(\boldsymbol{\theta})\rangle]$, it follows that, for any $w \in (0,1)$, $\langle \mathbf{U}_J, \boldsymbol{\xi}_J(\boldsymbol{\theta})\rangle \geq w$ if and only if $\|\mathbf{U}_J - \boldsymbol{\xi}_J(\boldsymbol{\theta})\| \leq r := \sqrt{2(1-w)}$. Therefore,

$$\begin{aligned} \mathbb{P}\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \langle \mathbf{U}_J, \boldsymbol{\xi}_J(\boldsymbol{\theta})\rangle \geq w\right) &= \mathbb{P}\left(\inf_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\mathbf{U}_J - \boldsymbol{\xi}_J(\boldsymbol{\theta})\| \leq r\right) \\ &= \mathbb{P}[\mathbf{U}_J \in \mathfrak{T}(r, \mathcal{M})] = \frac{\vartheta(r, \mathcal{M})}{A_J}, \end{aligned} \tag{3.9}$$

where $\vartheta(r, \mathcal{M})$ denotes the volume of $\mathfrak{T}(r, \mathcal{M})$—a *tube* of radius r around the *manifold* $\mathcal{M} := \{\boldsymbol{\xi}_J(\boldsymbol{\theta})\colon \boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathcal{R}^d\}$, and $A_J = 2\pi^{J/2}/\Gamma(J/2)$ is the $(J-1)$-dimensional volume of the unit sphere $\mathcal{S}^{(J-1)}$. The last expression follows since $\mathbf{U}_J$ is uniformly distributed over $\mathcal{S}^{(J-1)}$.

Remark 1: Finding the distribution of the supremum of a Gaussian random field $Z(\boldsymbol{\theta})$ is now reduced to that of determining the volume-of-tube of the manifold $\mathcal{M}$. The solution to this problem depends on the geometry of $\mathcal{M}$. When the set $\boldsymbol{\Theta}$ is one-dimensional (i.e., $d = 1$) and $\boldsymbol{\xi}_J(\boldsymbol{\theta})$ is continuous, then $\mathcal{M}$ is a curve on the unit sphere $\mathcal{S}^1$ and the tube consists of a main "cylindrical" section plus the two boundary caps as shown in Fig. 1. In this case, results of Hotelling (1939) and Naiman (1990) yield the approximation

$$\vartheta(r, \mathcal{M}) \approx \kappa_0 \frac{A_J}{A_2} \mathbb{P}\left[B_{1,(J-2)/2} \geq w^2\right] + \ell_0 \frac{A_J}{2A_1} \mathbb{P}\left[B_{1/2,(J-1)/2} \geq w^2\right],$$

where κ_0 is the length of the manifold $\mathcal{M}$, $B_{a,b}$ is the beta density with parameters a and b and $\ell_0 = 2$ is the number of end-points. Introducing ℓ_0 allows us to treat cases where $\mathcal{M}$ consists of two or more disconnected segments (due to singularities in the score process), which is a common phenomena in the context of mixture models. The volume-of-tube formula is exact whenever r is less than a critical radius r_0 (equivalently, $w_0 \leq w \leq 1$) which depends on the curvature of $\mathcal{M}$.

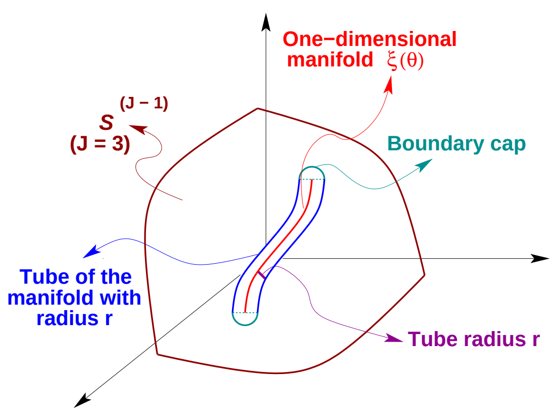

Figure 1: Tube of radius r around a one-dimensional manifold (curve) with boundaries embedded in $\mathcal{S}^2$.

Application of the volume-of-tube formula to a Gaussian random field leads to the main result of this section.

Theorem 4 *Under assumptions A1 to A5, the distribution of* $\sup_{\boldsymbol{\theta}} Z(\boldsymbol{\theta})$ *for a general* d *is given by*

$$\mathbb{P}\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} Z(\boldsymbol{\theta}) \geq c\right) = \sum_{t=0}^{d} \frac{\zeta_t}{A_{d+1-t}} \mathbb{P}\left(\chi^2_{d+1-t} \geq c^2\right) + o[c^{-1}\exp(-c^2/2)], \tag{3.10}$$

as $c \to \infty$*, where* $A_t = 2\pi^{t/2}/\Gamma(t/2)$ *is the* $(t-1)$*-dimensional volume of* $\mathcal{S}^{(t-1)}$ *in* $\mathcal{R}^t$ *and* ζ_t *are the geometric constants derived in Appendix A.*

Multinomial Mixture Problem: Equation (4.19) of Lindsay (1995), derived in the context of multinomial mixture models, is a special case of Theorem 4 (see also Lin (1997) for bounds). This connection is explored further in Section 5. It is important to note that for multinomial mixture models, the Karhunen-Loève expansion is finite.

Remark 2: Although the proof of Theorem 4, derived in Section 7, uses the Karhunen-Loève expansion, it is not necessary to find this expansion since one can determine the geometric constants ζ_ts appearing in (3.10) entirely from the covariance function $\mathbb{C}(\boldsymbol{\theta}, \boldsymbol{\theta}^\dagger)$. However, it is necessary to consider the geometry of the manifold $\mathcal{M}$ in order to treat the boundary corrections, particularly when $d > 2$.

3.3 Singularities in the Score Process

One of the conditions required for Theorem 4 is that $\mathbb{C}(\boldsymbol{\theta}, \boldsymbol{\theta})$ is positive for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. This condition is violated when $f(\cdot; \boldsymbol{\lambda}) = \psi(\cdot; \boldsymbol{\theta})$ for some $\boldsymbol{\theta}$. This is a commonly occurring phenomena in the context of finite mixture models. Therefore, we need to consider more carefully the behavior of the score process near $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. Let $S'(\boldsymbol{\theta}) = \partial S(\boldsymbol{\theta})/\partial \boldsymbol{\theta}$ and $\mathbb{V}[S'(\boldsymbol{\theta})]$ be the variance of $S'(\boldsymbol{\theta})$ so that $S(\boldsymbol{\theta}) = (\boldsymbol{\theta}-\boldsymbol{\theta}_0)S'(\boldsymbol{\theta}_0) + o(\boldsymbol{\theta}-\boldsymbol{\theta}_0)$, $n\,\mathbb{C}(\boldsymbol{\theta}, \boldsymbol{\theta}) = (\boldsymbol{\theta}-\boldsymbol{\theta}_0)^2\,\mathbb{V}[S'(\boldsymbol{\theta}_0)] + o[(\boldsymbol{\theta}-\boldsymbol{\theta}_0)^2]$ and $S^\star(\boldsymbol{\theta}) = \mathrm{sgn}(\boldsymbol{\theta}-\boldsymbol{\theta}_0)S'(\boldsymbol{\theta}_0)/\sqrt{\mathbb{V}[S'(\boldsymbol{\theta}_0)]} + o(\boldsymbol{\theta}-\boldsymbol{\theta}_0)$, where 'sgn' is the sign function. In particular, this implies that the process "flips" and

$$\lim_{\boldsymbol{\theta}\to\boldsymbol{\theta}_0^-} S^\star(\boldsymbol{\theta}) = -\lim_{\boldsymbol{\theta}\to\boldsymbol{\theta}_0^+} S^\star(\boldsymbol{\theta}). \tag{3.11}$$

Correspondingly, $\boldsymbol{\xi}(\boldsymbol{\theta}_0^-) = -\boldsymbol{\xi}(\boldsymbol{\theta}_0^+)$. In effect, the manifold $\mathcal{M}$ has two pieces and four boundary points. The result in Theorem 4 still holds; however, $\ell_0 = 4$.

4 Nuisance Parameters under the Null Model

In this section, we derive general theory for the case of unknown nuisance parameter vector $\boldsymbol{\lambda}$. We derive a series of fundamental results that provide a "linearization of the score process" (defined below) to identify the correct covariance function (see Theorem 6 below) for this setting. We replace $\boldsymbol{\lambda}$ by $\widehat{\boldsymbol{\lambda}}$, the MLE of $\boldsymbol{\lambda}$, and assume that the MLE satisfies the necessary regularity conditions stated by Chernoff (1954). Our goal is to find an appropriate normalizing factor for the score process and in turn apply the volume-of-tube formula for approximating the asymptotic null distribution of $\mathcal{T}$.

In the context of finite mixture models, the null density $f(\cdot;\boldsymbol{\lambda})$ is equivalent to the mixture density $\mathrm{g}(\cdot;\mathcal{Q}_m)$ representing an m-component mixture model with $\boldsymbol{\lambda} \equiv \mathcal{Q}_m$ containing a vector of support points and the corresponding mixing weights. The score process is searching for an $(m+1)$st component.

If $\boldsymbol{\lambda}$ is estimated via the ML method, then under $\mathcal{H}_0$, the score process can be expressed as

$$S(\boldsymbol{\theta}|\widehat{\boldsymbol{\lambda}}) := \sum_{i=1}^{n} \left[\frac{\psi(x_i;\boldsymbol{\theta})}{f(x_i;\widehat{\boldsymbol{\lambda}})} - 1 \right].$$

The statistic $\mathcal{T}$ will still be the supremum (over $\boldsymbol{\theta}$) of the normalized score process; however, estimating the nuisance parameter vector $\boldsymbol{\lambda}$ means that the covariance function $\mathbb{C}(\boldsymbol{\theta},\boldsymbol{\theta}^{\dagger})$ defined in (2.3) is no longer appropriate for normalizing the score process.

As a first step, it is assumed that the MLE $\widehat{\boldsymbol{\lambda}}$ under $\mathcal{H}_0$ satisfies the required conditions for the second-order asymptotic theory (Lehmann, 1999). Hence, the following results hold:

$$(\widehat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}_0) \quad = \quad [n\mathbf{I}(\boldsymbol{\lambda}_0)]^{-1} \sum_{i=1}^{n} \nabla\, l(\boldsymbol{\lambda}|x_i) + o_p(n^{-1/2}) \tag{4.1}$$

and $n^{-1/2}\sum_{i=1}^{n} \nabla\, l(\boldsymbol{\lambda}_0|x_i) \rightsquigarrow N[0,\mathbf{I}(\boldsymbol{\lambda}_0)]$ as $n \to \infty$, where $\boldsymbol{\lambda}_0$ is the true null parameter vector, $\rightsquigarrow$ indicates convergence in distribution, $\mathbf{I}(\boldsymbol{\lambda}_0)$ is the Fisher information matrix and $\nabla\, l(\boldsymbol{\lambda}|x)$ is the vector of partial derivatives of $l(\boldsymbol{\lambda}|x) = \log f(x;\boldsymbol{\lambda})$ with respect to $\boldsymbol{\lambda}$.

Theorem 5 *Suppose that assumptions A2 to A5 hold. Under $\mathcal{H}_0$ with the true null* *parameter vector $\boldsymbol{\lambda}_0$, the score process has the asymptotic representation of*

$$S(\boldsymbol{\theta}|\widehat{\boldsymbol{\lambda}}) = S(\boldsymbol{\theta}|\boldsymbol{\lambda}_0) - \mathbb{C}^T(\boldsymbol{\theta}|\boldsymbol{\lambda}_0)\,[\mathbf{I}(\boldsymbol{\lambda}_0)]^{-1}\sum_{i=1}^{n}\nabla\, l(\boldsymbol{\lambda}_0|x_i) + o_p(n^{1/2}),$$

where $o_p(n^{1/2})$ is uniform in $\boldsymbol{\theta}$ and $\mathbb{C}(\boldsymbol{\theta}|\boldsymbol{\lambda}_0)$ is the covariance vector defined as

$$\mathbb{C}(\boldsymbol{\theta}|\boldsymbol{\lambda}_0) := cov\left[\left(\frac{\psi(x_1;\boldsymbol{\theta})}{f(x_1;\boldsymbol{\lambda}_0)} - 1\right),\ \nabla\, l(\boldsymbol{\lambda}_0|x_1)\right] = \int \psi(x;\boldsymbol{\theta})\,\nabla\, l(\boldsymbol{\lambda}_0|x)\,dx.$$

Proof. By expanding the score process in a Taylor series around $\boldsymbol{\lambda}_0$, we obtain

$$S(\boldsymbol{\theta}|\widehat{\boldsymbol{\lambda}}) = S(\boldsymbol{\theta}|\boldsymbol{\lambda}_0) + (\widehat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}_0)^T\,\frac{\partial}{\partial\boldsymbol{\lambda}}S(\boldsymbol{\theta}|\boldsymbol{\lambda})\bigg|_{\boldsymbol{\lambda}=\tilde{\boldsymbol{\lambda}}},$$

where $\widetilde{\boldsymbol{\lambda}} \in [\boldsymbol{\lambda}_0, \widehat{\boldsymbol{\lambda}}]$. Direct calculation shows that

$$\frac{1}{n}\frac{\partial}{\partial\boldsymbol{\lambda}}S(\boldsymbol{\theta}|\boldsymbol{\lambda})\bigg|_{\boldsymbol{\lambda}=\widetilde{\boldsymbol{\lambda}}} = -\frac{1}{n}\sum_{i=1}^{n}\frac{\psi(x_i;\boldsymbol{\theta})}{f(x_i;\boldsymbol{\lambda}_0)}\,\nabla\, l(\boldsymbol{\lambda}|x_i).$$

From the uniform strong law of large numbers and the fact that $\widetilde{\lambda} \xrightarrow{a.s.} \lambda_0$, it follows that

$$\begin{aligned}\frac{1}{n}\frac{\partial}{\partial\boldsymbol{\lambda}}S(\boldsymbol{\theta}|\boldsymbol{\lambda})\bigg|_{\boldsymbol{\lambda}=\widetilde{\boldsymbol{\lambda}}} &\xrightarrow{a.s.} -\,\mathrm{cov}\left[\frac{\psi(x_1;\boldsymbol{\theta})}{f(x_1;\boldsymbol{\lambda}_0)} - 1,\ \nabla\, l(\boldsymbol{\lambda}_0|x_1)\right]\\ &= -\,\mathbb{C}(\boldsymbol{\theta}|\boldsymbol{\lambda}_0) \quad \text{as} \quad n \to \infty.\end{aligned}$$

It follows from assumption A1 and the continuity of $\boldsymbol{\Theta}$ that the convergence is uniform in $\boldsymbol{\theta}$. Combining this result with (4.1) completes the proof. ∎

Theorem 6 *The process*

$$n^{-1/2}\,S(\boldsymbol{\theta}|\boldsymbol{\lambda}_0) - n^{-1/2}\,\mathbb{C}^T(\boldsymbol{\theta}|\boldsymbol{\lambda}_0)\,[\mathbf{I}(\boldsymbol{\lambda}_0)]^{-1}\sum_{i=1}^{n}\nabla\, l(\boldsymbol{\lambda}_0|x_i) \tag{4.2}$$

has the covariance function

$$\mathbb{C}^\star(\boldsymbol{\theta},\boldsymbol{\theta}^\dagger) = \mathbb{C}(\boldsymbol{\theta},\boldsymbol{\theta}^\dagger) - [\mathbb{C}^T(\boldsymbol{\theta}|\boldsymbol{\lambda}_0)]\,[\mathbf{I}(\boldsymbol{\lambda}_0)]^{-1}\,\mathbb{C}(\boldsymbol{\theta}^\dagger|\boldsymbol{\lambda}_0), \tag{4.3}$$

where $\mathbb{C}(\boldsymbol{\theta},\boldsymbol{\theta}^\dagger)$ is defined in (2.3) with $f(\cdot;\boldsymbol{\lambda}_0)$ replacing $f(\cdot;\boldsymbol{\lambda})$.

Proof. The result follows immediately from the observations that

$$\begin{aligned} n^{-1}\,\mathrm{cov}\left[S(\boldsymbol{\theta}|\boldsymbol{\lambda}_0),\, S(\boldsymbol{\theta}^\dagger|\boldsymbol{\lambda}_0)\right] &= \mathbb{C}(\boldsymbol{\theta},\boldsymbol{\theta}^\dagger), \\ n^{-1}\,\mathrm{cov}\left[\sum_{i=1}^n \nabla\, l(\boldsymbol{\lambda}_0|x_i), \sum_{i=1}^n \nabla\, l(\boldsymbol{\lambda}_0|x_i)\right] &= \mathbf{I}(\boldsymbol{\lambda}_0) \\ \text{and}\quad n^{-1}\,\mathrm{cov}\left[S(\boldsymbol{\theta}|\boldsymbol{\lambda}_0), \sum_{i=1}^n \nabla\, l(\boldsymbol{\lambda}_0|x_i)\right] &= \mathbb{C}(\boldsymbol{\theta}|\boldsymbol{\lambda}_0). \quad \blacksquare \end{aligned}$$

A6: Suppose $\mathbb{C}^\star(\boldsymbol{\theta},\boldsymbol{\theta}^\dagger)$ is continuous and $0 < \mathbb{C}^\star(\boldsymbol{\theta},\boldsymbol{\theta}) < \infty$ for all $\boldsymbol{\theta}\in\boldsymbol{\Theta}$.

Theorem 7 *Under assumptions A2 through A6,*

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \frac{S(\boldsymbol{\theta}|\widehat{\boldsymbol{\lambda}})}{\sqrt{n\,\mathbb{C}^\star(\boldsymbol{\theta},\boldsymbol{\theta})}} \rightsquigarrow \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} Z^\star(\boldsymbol{\theta}) \quad \textit{as} \quad n\to\infty,$$

where $Z^\star(\boldsymbol{\theta})$ is a Gaussian random field with the covariance function

$$\rho^\star(\boldsymbol{\theta},\boldsymbol{\theta}^\dagger) := \frac{\mathbb{C}^\star(\boldsymbol{\theta},\boldsymbol{\theta}^\dagger)}{\sqrt{\mathbb{C}^\star(\boldsymbol{\theta},\boldsymbol{\theta})\,\mathbb{C}^\star(\boldsymbol{\theta}^\dagger,\boldsymbol{\theta}^\dagger)}}.$$

Proof. First, the result holds for the process (4.2) (which is similar to Theorem 3). Next, the result follows from Theorem 6. ■

We apply the results of Theorem 4 to the case of one-dimensional $\boldsymbol{\Theta}$:

Theorem 8 *The tail probability is expressed as* $\mathbb{P}\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} Z(\boldsymbol{\theta}) \ge c\right) = \kappa_0/(2\pi)\,\mathbb{P}(\chi_2^2 \ge c^2) + (\ell_0/4)\,\mathbb{P}(\chi_1^2 \ge c^2) + o[c^{-1}\exp(-c^2/2)]$ *with*

$$\kappa_0 = \int_{\boldsymbol{\Theta}} \left[\frac{\partial^2}{\partial\boldsymbol{\theta}\,\partial\boldsymbol{\theta}^\dagger}\rho^\star(\boldsymbol{\theta},\boldsymbol{\theta}^\dagger)\right]^{1/2}\Bigg|_{\boldsymbol{\theta}^\dagger=\boldsymbol{\theta}} d\boldsymbol{\theta}$$

and $\ell_0 = 2$.

The covariance function and κ_0 depend on $\boldsymbol{\lambda}_0$; hence, cannot be evaluated directly. However, replacing $\boldsymbol{\lambda}_0$ by $\widehat{\boldsymbol{\lambda}}$ yields a consistent estimator for $\boldsymbol{\lambda}_0$. Just as in the case of a fixed $\boldsymbol{\lambda}$, the condition $\mathbb{C}^\star(\boldsymbol{\theta},\boldsymbol{\theta}) > 0$ for all $\boldsymbol{\theta}\in\boldsymbol{\Theta}$ (part of assumption A6) will be violated in the context of finite mixture models. However, one cannot handle the singularities in a nice fashion and they are best treated on a case-by-case basis. In particular, (1) there may be multiple singularities, corresponding to each component of the mixture model under $\mathcal{H}_0$ and (2) in some cases the singularities lead to discontinuities (as described earlier) while in other cases the singularities are removable.

5 Testing for the Order of a Mixture Model

In this section, building on the perturbation theory, we derive results for the long-pending problem of testing for the order of a mixture model while achieving the following goals: (1) Demonstrating how the existing results for a special class of mixtures can be derived from our general theory, (2) obtaining explicit and flexible expressions for the geometric constants in the asymptotic tail probability and (3) a careful examination of the singularities of the score process that routinely occur in mixture models.

5.1 Mixtures of Binomial Distributions

Discrete mixtures for a random variable X assuming a finite set of values (e.g., $0, \ldots, b$) are of special interest, since the data can be summarized by the bin counts $N_0, \ldots, N_b$. The loglikelihood and the score process $S(\theta)$ depend on the data only through these values. After appropriate centering and scaling, it is easy to verify that the bin counts have an asymptotic b-variate multivariate normal distribution. Consequently, the score process $S(\theta)$ must have a finite Karhunen-Loève expansion.

Consider the case of $b = 2$ and a mixture of Binomial$(2, \theta)$ distributions with $\theta \in [0, 1]$. That is, our interest is in testing $\mathcal{H}_0 : \eta = 0$ against $\mathcal{H}_1 : \eta > 0$ and $\psi(x, \theta)$ is assumed to have a Binomial$(2, \theta)$ distribution expressed as

$$\psi(x; \theta) = \begin{cases} (1-\theta)^2 & \text{if} \quad x = 0 \\ 2\theta\,(1-\theta) & \text{if} \quad x = 1 \\ \theta^2 & \text{if} \quad x = 2 \end{cases}$$

with the null density $\psi(\cdot; \lambda)$ for some $\lambda \in [0, 1]$. Therefore, the perturbation model can be expressed as $p(x; \eta, \lambda, \theta) = (1 - \eta)\, \psi(x, \lambda) + \eta\, \psi(x, \theta)$.

Case 1: Assume λ is known and θ is unknown. The score process

$$S(\theta) = N_0 \, \frac{(1-\theta)^2}{(1-\lambda)^2} + N_1 \, \frac{\theta\,(1-\theta)}{\lambda(1-\lambda)} + N_2 \, \frac{\theta^2}{\lambda^2} - n.$$

Since $N_1 = (n - N_0 - N_2)$, the score process reduces to

$$n^{-1/2}\, S(\theta) = Z_0 \, \frac{(1-\theta)(\lambda-\theta)}{(1-\lambda)^2\lambda} + Z_2 \, \frac{\theta(\theta-\lambda)}{\lambda^2\,(1-\lambda)} = \mathfrak{c}_0(\theta)\, Z_0 + \mathfrak{c}_2(\theta)\, Z_2,$$

where $Z_0 = n^{-1/2}[N_0 - n\,(1-\lambda^2)]$ and $Z_2 = n^{-1/2}(N_2 - n\,\lambda^2)$. The vector $[\mathfrak{c}_0(\theta),\, \mathfrak{c}_2(\theta)]^T$ traces a smooth curve through the origin at $\theta = \lambda$. The normalized score process $S^\star(\theta)$ has the flip property discussed earlier.

The random variables Z_0 and Z_2 are correlated; hence, explicit representation of $S(\theta)$ in terms of the uncorrelated random variables is quite messy. However, the corresponding manifold $\mathcal{M}$ consists of two arcs on the unit circle and the one-dimensional volume of the tube is $\kappa_0 = \cos^{-1}(\mathfrak{r}_0) + \cos^{-1}(\mathfrak{r}_1)$, where $\mathfrak{r}_0 = \mathrm{cor}[S(0), -S'(\lambda)]$, $\mathfrak{r}_1 = \mathrm{cor}[S(1), S'(\lambda)]$ and $\ell_0 = 4$. Note that $\mathfrak{r}_0$ and $\mathfrak{r}_1$ can be evaluated explicitly based on $\mathbb{V}(Z_0) = (1-\lambda)^2\,\lambda\,(2-\lambda), \mathrm{cov}(Z_0, Z_2) = -\lambda^2\,(1-\lambda)^2$ and $\mathbb{V}(Z_2) = \lambda^2\,(1-\lambda)\,(1+\lambda)$. After some algebra, it is easy to verify that $\mathfrak{r}_0 = \sqrt{2\lambda/(1+\lambda)}$ and $\mathfrak{r}_1 = \sqrt{2(1-\lambda)/(2-\lambda)}$. Since $\mathcal{M}$ consists of two arcs on a unit circle, the exact asymptotic null distribution of $\mathcal{T}$ is obtained using the method of Uusipaikka (1983).

Case 2: Assume that both λ and θ are unknown. Consider the MLE of λ, $\widehat{\lambda} = (N_1 + 2N_2)/(2n) = (n + N_2 - N_0)/(2n)$, so that $Z_0 = Z_2 = (N_0 + N_2)/2 - n/4 - (N_2 - N_0)^2/(4n)$ and $S(\theta|\widehat{\lambda}) = Z_0(\theta - \widehat{\lambda})^2/[\widehat{\lambda}^2\,(1-\widehat{\lambda})^2]$. In this case, the normalized score process is constant and hence the manifold $\mathcal{M}$ consists of a single point. Therefore, $\kappa_0 = 0$ and $\ell_0 = 2$ resulting in a distribution of $(0.5\,\chi^2_0 + 0.5\,\chi^2_1)$, where χ^2_0 is a degenerate distribution with all its mass at zero. This is the special case derived by Lindsay (1995, p. 95). Shapiro (1985) referred to this mixture of chi-square distributions with differing degrees of freedom as *chi-bar* distribution.

5.2 Mixtures of Exponential Family of Densities

Suppose that $\psi(x;\boldsymbol{\theta})$ belongs to an exponential family of densities so that $\psi(x;\boldsymbol{\theta}) = \exp[\boldsymbol{\theta}^T x - \varphi(\boldsymbol{\theta})]\,\psi_0(x)$. The null density is $f(\cdot;\boldsymbol{\lambda})$ for some $\boldsymbol{\lambda}$.

Case of Fixed $\boldsymbol{\lambda}$*:* The covariance function becomes

$$\begin{aligned}\mathbb{C}(\boldsymbol{\theta},\boldsymbol{\theta}^\dagger) &= \int \exp[(\boldsymbol{\theta}+\boldsymbol{\theta}^\dagger-\boldsymbol{\lambda})^T x + \varphi(\boldsymbol{\lambda}) - \varphi(\boldsymbol{\theta}) - \varphi(\boldsymbol{\theta}^\dagger)]\,\psi_0(x)\,dx - 1\\ &= \exp[\varphi(\boldsymbol{\theta}+\boldsymbol{\theta}^\dagger-\boldsymbol{\lambda}) + \varphi(\boldsymbol{\lambda}) - \varphi(\boldsymbol{\theta}) - \varphi(\boldsymbol{\theta}^\dagger)] - 1.\end{aligned}$$

If $\psi(\cdot;\boldsymbol{\theta})$ has a multivariate normal distribution with a mean vector $\boldsymbol{\theta}$ and an identity variance covariance matrix, it follows that $\varphi(\boldsymbol{\theta}) = \|\boldsymbol{\theta}\|^2/2$ and

$$\mathbb{C}(\boldsymbol{\theta},\boldsymbol{\theta}^\dagger) = \exp[\langle\boldsymbol{\theta}-\boldsymbol{\lambda}, \boldsymbol{\theta}^\dagger-\boldsymbol{\lambda}\rangle] - 1. \tag{5.1}$$

Consider the special case of $d = 1$. The critical values are obtained using Theorem 4 and the one-dimensional volume of $\mathcal{M}$ has the following explicit expression when $\lambda = 0$:

$$\kappa_0 = \int_\Theta \frac{\left[\exp(2\theta^2) - (1+\theta^2)\exp(\theta^2)\right]^{1/2}}{\left[\exp(\theta^2) - 1\right]} \, d\theta.$$

The normalized score process again has the flip property (3.11) and $\ell_0 = 4$.

Case of Unknown $\boldsymbol{\lambda}$*:* Straightforward calculations show that the covariance function (4.3) in Theorem 6 becomes

$$\mathbb{C}^\star(\boldsymbol{\theta}, \boldsymbol{\theta}^\dagger) \;=\; \mathbb{C}(\boldsymbol{\theta}, \boldsymbol{\theta}^\dagger) - [\varphi'(\boldsymbol{\theta}) - \varphi'(\boldsymbol{\lambda})]^T \, [\varphi''(\boldsymbol{\lambda})]^{-1} \, [\varphi'(\boldsymbol{\theta}^\dagger) - \varphi'(\boldsymbol{\lambda})],$$

since $\mathbb{C}(\boldsymbol{\theta}|\boldsymbol{\lambda}) = \mathbb{E}_{\boldsymbol{\theta}}[X - \varphi'(\boldsymbol{\lambda})] = \varphi'(\boldsymbol{\theta}) - \varphi'(\boldsymbol{\lambda})$ and $\mathbf{I}(\boldsymbol{\lambda}) = \varphi''(\boldsymbol{\lambda})$.

In the case of a univariate normal distribution, the volume of the one-dimensional manifold becomes

$$\kappa_0 = \int_\Theta \frac{\left[\exp\{2(\theta-\lambda)^2\} + 1 - \exp\{(\theta-\lambda)^2\}\,\{2 + (\theta-\lambda)^4\}\right]^{1/2}}{\left[\exp\{(\theta-\lambda)^2\} - 1 - (\theta-\lambda)^2\right]} \, d\theta.$$

The normalized score process has a singularity at $\theta = \widehat{\theta}$; however, the precise behavior at this point needs careful consideration, which is presented next. In the neighborhood of $\widehat{\theta}$, we have

$$S(\theta) = S(\widehat{\theta}) + (\theta - \widehat{\theta})\, S'(\widehat{\theta}) + \frac{1}{2}\,(\theta - \widehat{\theta})^2\, S''(\widehat{\theta}) + o\left[(\theta - \widehat{\theta})^2\right]. \tag{5.2}$$

Note that $S(\widehat{\theta}) = S'(\widehat{\theta}) = 0$ (since the latter is simply the score equation defining $\widehat{\theta}$). By continuity, $S''(\widehat{\theta}) = \varphi''(\lambda) + o(1)$; hence, the normalized score process becomes $S''(\lambda)/\sqrt{\mathbb{V}[S''(\lambda)]} + o(1)$ in the neighborhood of λ. This is continuous so there is no flip at $\theta = \widehat{\theta}$. The manifold $\mathcal{M}$ for this process is a single segment and $\ell_0 = 2$.

5.3 Testing for m versus $(m+q)$ Component Mixture Model

One of the important applications of the perturbation theory is in building finite mixture models formed from a broad class of smooth densities. First, consider testing

$$\mathcal{H}_0\colon m\text{-component mixture} \quad \text{against} \quad \mathcal{H}_1\colon (m+1)\text{-component mixture}$$

when mixtures are formed from any smooth families, including discrete, continuous and multivariate densities. Under the mixture model framework, the null model $f(\cdot;\boldsymbol{\lambda})$ is the m component mixture $\mathrm{g}(x;\mathcal{Q}_m) = \sum_{j=1}^{m} \beta_j \,\psi(x;\boldsymbol{\theta}_j)$, where $\mathcal{Q}_m = (\boldsymbol{\theta}^T, \boldsymbol{\beta}^T)^T$ while the alternative is the $(m+1)$st component. We consider two cases: (1) The support point vectors $\boldsymbol{\theta}$s are fixed and only the mixing weight vector $\boldsymbol{\beta}$ is estimated and (2) $\boldsymbol{\theta}$s and $\boldsymbol{\beta}$ are estimated.

Case 1: Assume $\boldsymbol{\theta}$ is fixed and the goal is to estimate $\boldsymbol{\beta}$. The likelihood surface is concave in $\boldsymbol{\beta}$ and the MLEs satisfy

$$S(\boldsymbol{\theta}_j|\widehat{\boldsymbol{\beta}}) = \sum_{i=1}^{n} \frac{\psi(x_i;\boldsymbol{\theta}_j)}{\mathrm{g}(x_i;\widehat{\mathcal{Q}})} - n = 0 \quad \text{for all} \quad j = 1,\ldots,m \tag{5.3}$$

provided that the solution satisfies $0 < \widehat{\beta}_j < 1$ (otherwise, some components are set to zero). The MLE satisfies the conditions of Section 4, provided that $\beta_j > 0$ for each j. The covariance function is determined based on the result in Theorem 6.

The set of equations in (5.3) implies that the normalized score process has a singularity at each $\boldsymbol{\theta}_j$. Using an argument similar to (3.11), the process flips at each of these points.

Case 2: The goal is to estimate both $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$. Note that each support point is of dimension d. The equations defining the MLEs become

$$S(\boldsymbol{\theta}_j|\widehat{\mathcal{Q}}_m)\Big|_{\boldsymbol{\theta}_j=\widehat{\boldsymbol{\theta}}_j} = 0 \quad \text{and} \quad S'(\boldsymbol{\theta}_j|\widehat{\mathcal{Q}}_m)\Big|_{\boldsymbol{\theta}_j=\widehat{\boldsymbol{\theta}}_j} = \mathbf{0} \tag{5.4}$$

for all $j = 1,\ldots,m$. Note that for $d > 1$, the above equation is a vector. Using an expansion similar to (5.2), around each of the true support points, it is easy to verify that all the singularities in the normalized score process are removable.

Consistent estimators of the nuisance parameters are required to apply the results of Section 4. This is achieved by imposing an order constraint on the support point vectors $\boldsymbol{\theta}_j$ and a corresponding constraint on the estimators. Under these constraints, the approximate critical values are obtained from Theorems 6 and 8.

General case: Consider the more general problem of testing $\mathcal{H}_0$: m-component mixture against $\mathcal{H}_1$: $(m+q)$-component mixture for $q = 1, 2, \ldots$. For this case, Theorem 4 is still applicable and the score process is easy to derive (see (Pilla and Loader, 2005) for details). Suppose $d = 1$ and $\boldsymbol{\Theta}$ is an interval, then the manifold has two corner points and two edges with two boundary faces as shown in Fig. 2.

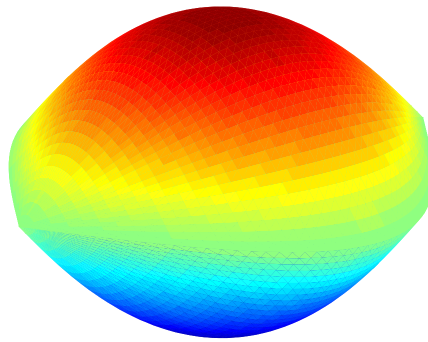

Figure 2: Manifold for testing m versus $(m+2)$ components in mixture models. The manifold has two corners, two edges and two boundary faces.

5.4 Mixtures of Bivariate Normal Distributions

In this section, we consider the bivariate mixture testing problem so that $d = 2, \mathbf{x} = (x_1, x_2)^T$ and $\boldsymbol{\theta} = (\theta_1, \theta_2)^T$. To the best of the authors' knowledge, this is the first attempt at testing for mixtures of multivariate distributions. Assume $f(\cdot; \boldsymbol{\lambda})$ is a bivariate standard normal density and $\psi(\cdot; \boldsymbol{\theta})$ is a bivariate normal density with mean $\boldsymbol{\theta}$ and an identity covariance matrix. From equation (5.1), it is easy to verify that the covariance function can be explicitly expressed as $\mathbb{C}(\boldsymbol{\theta}, \boldsymbol{\theta}^\dagger) = \exp[\langle \boldsymbol{\theta}, \boldsymbol{\theta}^\dagger \rangle] - 1$. Suppose $\boldsymbol{\Theta}$ is a disk of radius $\varrho_1 > 0$, so that

$$\mathcal{T} = \sup_{0 < \|\boldsymbol{\theta}\| \leq \varrho_1} \frac{S(\boldsymbol{\theta})}{\sqrt{n\, \mathbb{C}(\boldsymbol{\theta}, \boldsymbol{\theta})}}.$$

In order to address the singularity at $\|\boldsymbol{\theta}\| = 0$, first consider the supremum over $\varrho_0 \leq \|\boldsymbol{\theta}\| \leq \varrho_1$, where $0 < \varrho_0 < \varrho_1$ and next let $\varrho_0 \to 0$. Under the polar coordinate parameterization of $\boldsymbol{\theta} = [\varrho \cos(\omega), \varrho \sin(\omega)]^T$, with the covariance function expressed as

$\mathbb{C}(\boldsymbol{\theta}, \boldsymbol{\theta}^\dagger) = \exp[\varrho\, \varrho^\dagger \cos(\omega - \omega^\dagger)] - 1$, it follows that

$$\begin{aligned}
\kappa_0 &= \int_{\varrho_0}^{\varrho_1} \int_0^{2\pi} [\mathbb{C}(\boldsymbol{\theta}, \boldsymbol{\theta})]^{-3/2} \det \begin{pmatrix} \exp(\varrho^2) - 1 & \varrho \exp(\varrho^2) & 0 \\ \varrho \exp(\varrho^2) & (1 + \varrho^2) \exp(\varrho^2) & 0 \\ 0 & 0 & \varrho^2 \exp(\varrho^2) \end{pmatrix}^{1/2} d\,\omega\, d\varrho \\
&= 2\pi \int_{\varrho_0}^{\varrho_1} \left[\frac{\varrho^2 \exp(3\varrho^2) - \varrho^2 (1 + \varrho^2) \exp(2\varrho^2)}{\{\exp(\varrho^2) - 1\}^3} \right]^{1/2} d\varrho.
\end{aligned}$$

The integrand has a finite limit as $\varrho \to 0$; therefore, the integral is still valid when $\varrho_0 = 0$.

Next, we consider the boundaries at $\varrho = \varrho_0$ and $\varrho = \varrho_1$. For an arbitrary ϱ, the length of the boundary is

$$\ell_0 = \int_0^{2\pi} [\mathbb{C}(\boldsymbol{\theta}, \boldsymbol{\theta})]^{-1} \det \begin{pmatrix} \exp(\varrho^2) - 1 & 0 \\ 0 & \varrho^2 \exp(\varrho^2) \end{pmatrix}^{1/2} d\omega = 2\pi \sqrt{\frac{\varrho^2 \exp(\varrho^2)}{[\exp(\varrho^2) - 1]}}.$$

Therefore,

$$\ell_0 = 2\pi \left[\sqrt{\frac{\varrho_0^2 \exp(\varrho_0^2)}{[\exp(\varrho_0^2) - 1]}} + \sqrt{\frac{\varrho_1^2 \exp(\varrho_1^2)}{[\exp(\varrho_1^2) - 1]}} \right] \longrightarrow 2\pi \left[1 + \sqrt{\frac{\varrho_1^2 \exp(\varrho_1^2)}{\{\exp(\varrho_1^2) - 1\}}} \right]$$

as $\varrho_0 \to 0$. The contribution from the inner boundary does not disappear as $\varrho_0 \to 0$, instead it converges to 2π. This implies that the manifold $\mathcal{M}$ corresponding to this process has a hole and $\mathcal{M}$ has an Euler-Poincare characteristic of $\mathcal{E} = 0$. The tail-probability approximation of Theorem 4 simplifies to

$$\begin{aligned}
\mathbb{P}\left(\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} Z(\boldsymbol{\theta}) \geq c \right) &\approx \frac{\kappa_0}{4\pi} \mathbb{P}\left(\chi_3^2 \geq c^2\right) + \frac{\ell_0}{4\pi} \mathbb{P}\left(\chi_2^2 \geq c^2\right) - \frac{\kappa_0}{4\pi} \mathbb{P}\left(\chi_1^2 \geq c^2\right) \\
&= \frac{\kappa_0}{2\sqrt{2\pi}}\, c \exp(-c^2/2) + \frac{\ell_0}{4\pi} \exp(-c^2/2) \quad \text{as} \quad n \to \infty.
\end{aligned}$$

The interior hole occurs in any two-parameter problem, as the next lemma demonstrates.

Lemma 1 *Suppose $\boldsymbol{\theta}$ is of dimension $d = 2$ and there exists a $\boldsymbol{\lambda}$ such that $f(\cdot; \boldsymbol{\lambda}) = \psi(\cdot; \boldsymbol{\theta})$. The normalized score process $S^\star(\boldsymbol{\theta})$ has a singularity at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ and correspondingly, the manifold $\mathcal{M}$ has a hole. The length of the interior boundary of $\mathcal{M}$ is 2π.*

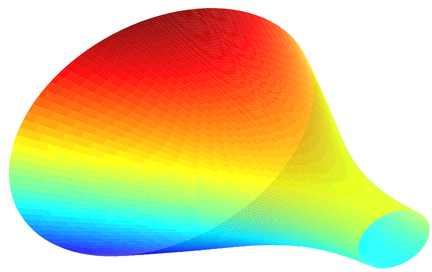

Figure 3: Manifold for the bivariate normal mixture testing problem. The cylindrical manifold has two boundaries: a circle with circumference 2π, corresponding to $\varrho = 0$, and a larger (high dimensional) ring corresponding to $\varrho = \varrho_1$.

Proof. A Taylor series expansion yields

$$S(\boldsymbol{\theta}) = \langle \boldsymbol{\theta} - \boldsymbol{\theta}_0,\, S'(\boldsymbol{\theta}_0)\rangle + o(\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|) \quad \text{as} \quad \boldsymbol{\theta} \to \boldsymbol{\theta}_0.$$

Let $\mathbf{R}$ be a matrix such that $\mathrm{cov}[S'(\boldsymbol{\theta}_0)] = n\,\mathbf{R}^T\mathbf{R}$. Then the normalized score process becomes

$$S^\star(\boldsymbol{\theta}) = \frac{S(\boldsymbol{\theta})}{\sqrt{n\,\mathbb{C}(\boldsymbol{\theta},\boldsymbol{\theta})}} = \left\langle \frac{\mathbf{R}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)}{\|\mathbf{R}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)\|}, \frac{\mathbf{R}^{-1}S'(\boldsymbol{\theta}_0)}{\sqrt{n}} \right\rangle + o(\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|).$$

As $\boldsymbol{\theta}$ varies in a small circle around $\boldsymbol{\theta}_0$, the boundary of the manifold $\mathcal{M}$, $\mathbf{R}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)/\|\mathbf{R}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)\|$, becomes the unit circle in $\mathcal{R}^2$ which has length 2π. ■

For $d = 1$, the manifold $\mathcal{M}$ has $(m+1)$ segments so that $\ell_0 = 2(m+1)$. Approximate critical values are obtained based on Theorem 4 and κ_0 is evaluated using numerical integration. For $d = 2$, the manifold $\mathcal{M}$ has m holes with each hole contributing 2π to the total length of the boundary ℓ_0. The Euler-Poincare characteristic of $\mathcal{M}$ is therefore $(1 - m)$. For the result in Theorem 4, the constant κ_0 and the length of the outer boundary are found using a bivariate and univariate numerical integrations, respectively.

5.5 Simulation Experiments

In order to demonstrate the power of the proposed methods, we present two simulation studies and illustrate the process of building mixture models.

We consider the simulated dataset shown in Fig. 5.5(a), consisting of a sample of size $n = 100$ drawn from the two-component normal mixture model $0.5N(-2, 1)+0.5N(2, 1)$.

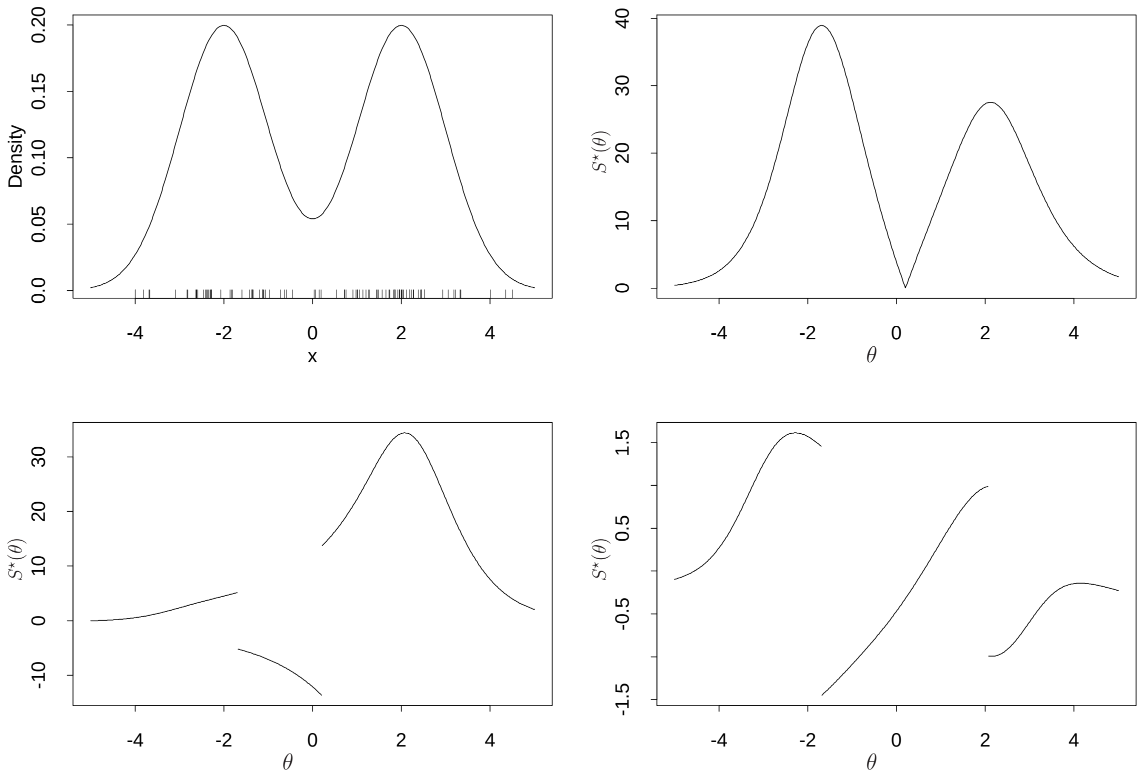

Figure 4: (a) Simulated data from the two-component normal mixture: $0.5N(-2,1) + 0.5N(2,1)$; (b) $S^\star(\theta)$ for the one-component mixture model; (c) $S^\star(\theta)$ for the two-component mixture model; (d) $S^\star(\theta)$ for the model with a third component included and the first component removed.

The model building process starts with the first component at the sample mean $\widehat{\theta}_1 = \overline{X} = 0.20322$. The starting model is obviously a poor fit for the dataset. Fig. 5.5(b) presents the fitted normalized score process $S^\star(\theta)$, showing two peaks in the vicinity of the true mixture components. An application of the volume-of-tube formula in (3.10) to this model yields $\kappa_0 = 5.72$ and $\ell_0 = 4$ with the critical value of $c = 2.518$ at the 5% level. Clearly the peaks are highly significant. A second component at $\theta_2 = -1.68929$ [the location of the larger left peak in Fig. 5.5(b)] is included in the model and the vector of estimated mixing weights is $\widehat{\boldsymbol{\beta}} = (0.67315, 0.32685)^T$. The incorrect first component θ_1 still dominates the fitted mixture model.

The normalized score process relative to the two-component mixture is shown in

Fig. 5.5(c). The striking feature of this plot is the two discontinuities at the fitted components $\theta_1 = 0.203$ and $\theta_2 = -1.689$. These discontinuities occur due to the zeroes of the covariance function since $\mathbb{C}^\dagger(\theta_1, \theta_1) = \mathbb{C}^\dagger(\theta_2, \theta_2) = 0$ which in turn corresponds to the singularities in $S^\star(\theta)$. The manifold $\mathcal{M}$ for this process has three pieces so that $\kappa_0 = 5.082$ and $\ell_0 = 6$. The critical value $c = 2.571$ and the right peak is still highly significant. The maximum occurs at $\widehat{\theta}_3 = 2.07328$ which is included as a third component in the model. Since $\widehat{\boldsymbol{\beta}} = (0, 0.45616, 0.54384)^T$, the first component is removed from the model. For the two-component mixture model with $\widehat{\theta}_2$ and $\widehat{\theta}_3$, the constants $\kappa_0 = 5.082$ and $\ell_0 = 6$ yielding $c = 2.571$. Fig. 5.5(d) presents the process $S^\star(\theta)$ and it is entirely below the critical value c; therefore, the two-component mixture model with $\widehat{\theta}_2$ and $\widehat{\theta}_3$ is the final fitted model.

The true density is chosen as $p(x; \eta, \theta) = 0.5(1 - \eta)\,\psi(x; -2) + \eta\,\psi(x; 0) + 0.5(1 - \eta)\,\psi(x; 2)$ for $\eta \in \{0, 0.1, 0.2\}$ and $\psi(\cdot; \theta)$ is the normal density with mean θ and unit variance. This density has two large with well separated components and our goal is to test for the presence of the poorly separated third component. We present simulation studies using 1000 data sets under the following three different scenarios:

Model 1: $f(x; \boldsymbol{\lambda}) \equiv \mathrm{g}(x; \mathcal{Q}) = [0.5\,\psi(x; -2) + 0.5\,\psi(x; 2)]$ is completely specified.

Model 2: $f(x; \boldsymbol{\lambda}) \equiv \mathrm{g}(x; \mathcal{Q}) = [\beta_1\,\psi(x; -2) + \beta_2\,\psi(x; 2)]$, where β_1 and β_2 are estimated.

Model 3: $f(x; \boldsymbol{\lambda}) \equiv \mathrm{g}(x; \mathcal{Q}) = [\beta_1\,\psi(x; \theta_1) + \beta_2\,\psi(x; \theta_2)]$, where βs and θs are estimated.

Table 1: Rejection rates for three different null models under three different perturbation sizes based on 1000 simulation studies.

	$n = 200$			$n = 1000$		
Model	$\eta = 0.0$	$\eta = 0.1$	$\eta = 0.2$	$\eta = 0.0$	$\eta = 0.05$	$\eta = 0.1$
1	79	537	975	74	636	990
2	78	583	985	76	673	992
3	74	292	588	61	371	817

Table 1 presents the rejection rates for 1000 simulations under two sample sizes. When $\eta = 0$, $\mathcal{H}_0$ is true and hence we expect the rejection rate to be close to the nominal significance level of 5%. As η increases, the power increases as expected. As the null assumptions are relaxed, the power decreases which again is to be expected. The poor separation between the components makes it difficult for the test to detect the third component which is more prominent for model 3. Naturally, estimating the nuisance parameters under the null model has an effect on the power of the test.

6 Discussion

In this article, we introduced a general class of models, perturbation models, and proposed a test statistic (asymptotically equivalent to the LRT statistic) based on the score process to detect the presence of perturbation. We derived general inferential theory for the asymptotic null distribution of the test statistic for a class of non-regular problems using the Hotelling-Weyl-Naiman volume-of-tube formula. The resulting theory is extended to solve the long-pending fundamental problem of testing for the mixture complexity, including the case when the null model includes a set of nuisance parameters. Our theory is applicable to a general family of mixture models including the multivariate family of mixtures. Other applications to the general theory include spatial scan analysis, latent class models (employed in social research) and Rasch models (employed in educational testing and survey sampling). The inferential theory developed in this article provides a solution to an important class of statistical problems involving loss of identifiability and/or when some of the parameters are on the boundary of the parametric space.

The explicit determination of the geometric constants appearing in the tube formula are carried out using the `Libtube` software (Loader, 2005). Our theory is general enough to be applicable to scalar or vector $\boldsymbol{\lambda}$ and univariate or multivariate data. The advantage of our approach is that the tube formula provides an elegant approximation to the asymptotic null distribution compared to those based on simulations or bootstrap based procedures.

7 Proofs

In this section we provide proofs of the main theorems. As before, notation $'$ is used to denote derivative with respect to the appropriate term.

Proof of Theorem 2. Let

$$K(\eta, \boldsymbol{\theta}) = \sum_{i=1}^{n} \log \left[1 + \frac{\eta \left\{ \psi(x_i; \boldsymbol{\theta}) - f(x_i; \boldsymbol{\lambda}) \right\}}{f(x_i; \boldsymbol{\lambda})} \right].$$

The LRT statistic becomes $\sup_{\boldsymbol{\theta}, \eta > 0} K(\eta, \boldsymbol{\theta})$. For any $\eta > 0$, a Taylor series expansion yields

$$\begin{aligned} K(\eta/\sqrt{n}, \boldsymbol{\theta}) &= K(0, \boldsymbol{\theta}) + \frac{\eta}{\sqrt{n}} K'(0, \boldsymbol{\theta}) + \frac{\eta^2}{2n} K''(\eta^\star, \boldsymbol{\theta}) \quad \text{for} \quad 0 \leq \eta^\star \leq \frac{\eta}{\sqrt{n}} \\ &= \frac{\eta}{\sqrt{n}} \sum_{i=1}^{n} \left[\frac{\psi(x_i; \boldsymbol{\theta})}{f(x_i; \boldsymbol{\lambda})} - 1 \right] \\ &\quad - \frac{\eta^2}{2n} \sum_{i=1}^{n} \left[\frac{\left\{ \psi(x_i; \boldsymbol{\theta}) - f(x_i; \boldsymbol{\lambda}) \right\}^2 / \{ f(x_i; \boldsymbol{\lambda}) \}^2}{1 + \eta^\star \{ \psi(x_i; \boldsymbol{\theta}) / f(x_i; \boldsymbol{\lambda}) - 1 \}} \right]. \end{aligned}$$

Under an implicit assumption that convergence statements are uniform in $\boldsymbol{\theta}$ for bounded sets and from the results in Rubin (1956), it follows that

$$K''(\eta^\star, \boldsymbol{\theta}) = \sum_{i=1}^{n} \left[\frac{\left\{ \psi(x_i; \boldsymbol{\theta}) - f(x_i; \boldsymbol{\lambda}) \right\}^2 / \{ f(x_i; \boldsymbol{\lambda}) \}^2}{1 + \eta^\star \{ \psi(x_i; \boldsymbol{\theta}) / f(x_i; \boldsymbol{\lambda}) - 1 \}} \right]$$

is uniformly converging to $\mathbb{C}(\boldsymbol{\theta}, \boldsymbol{\theta})$. Therefore,

$$K(\eta/\sqrt{n}, \boldsymbol{\theta}) \quad = \quad \frac{\eta}{\sqrt{n}} S(\boldsymbol{\theta}) - \frac{\eta^2}{2} \mathbb{C}(\boldsymbol{\theta}, \boldsymbol{\theta}) + o_p(1),$$

where the $o_p(1)$ term is uniform in η and $\boldsymbol{\theta}$ on compact sets. In effect, $\sup_{\eta \geq 0} K(\eta/\sqrt{n}, \boldsymbol{\theta}) = (1/2) \max\{0, S^\star(\boldsymbol{\theta})\}^2 + o_p(1)$. ∎

On the way to proving Theorem 3, we derive a series of technical results.

Lemma 2 *Let $a(\theta)$ be a continuously differentiable function on an interval Θ. Let $a_\star = [a(\theta_1) - a(\theta_0)]$. Then*

$$\int_{\theta_0}^{\theta_1} [a'(\theta)]^2 \, d\theta \geq \frac{a_\star^2}{|\theta_1 - \theta_0|},$$

where $a'(\theta) = da(\theta)/d\theta$.

Proof. Let $\theta_\star = (\theta_1 - \theta_0)$ so that

$$\begin{aligned} \int_{\theta_0}^{\theta_1} [a'(\theta)]^2 \, d\theta &= \int_{\theta_0}^{\theta_1} \left(a'(\theta) - \frac{a_\star}{\theta_\star} + \frac{a_\star}{\theta_\star} \right)^2 d\theta \\ &= \int_{\theta_0}^{\theta_1} \left(a'(\theta) - \frac{a_\star}{\theta_\star} \right)^2 d\theta + \int_{\theta_0}^{\theta_1} \frac{a_\star^2}{\theta_\star^2} \, d\theta + \frac{2a_\star}{\theta_\star} \int_{\theta_0}^{\theta_1} \left(a'(\theta) - \frac{a_\star}{\theta_\star} \right) d\theta. \end{aligned}$$

Note that the first integral is non-negative and the third one is zero. ∎

Lemma 3 *Suppose* $\theta_0 < \theta_2$ *and* $a(\theta_0) = a(\theta_2) = 0$, *then*

$$\int_\Theta [a'(\theta)]^2 \, d\theta \geq \frac{4}{|\theta_2 - \theta_0|} \left(\sup_{\theta_0 \leq \theta \leq \theta_2} |a(\theta)| \right)^2 .$$

Proof. Suppose the supremum occurs at $(\theta_1, a_\star)$ with $\theta_0 < \theta_1 < \theta_2$. An application of Lemma 2 separately over $[\theta_0, \theta_1]$ and $[\theta_1, \theta_2]$ yields

$$\int_\Theta [a'(\theta)]^2 \, d\theta \geq \int_{\theta_0}^{\theta_2} [a'(\theta)]^2 \, d\theta \geq a_\star^2 \left[\frac{1}{(\theta_1 - \theta_0)} + \frac{1}{(\theta_2 - \theta_1)} \right] \geq \frac{4a_\star^2}{(\theta_2 - \theta_0)}.$$

∎

Lemma 4 *Suppose* $b(\theta)$ *is continuously differentiable. For* $\delta > 0$, *let* $b_\delta(\theta)$ *be the linear interpolant between the points* $0, \pm\delta, \pm 2\delta, \ldots$. *Then*

$$\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left| b_\delta(\theta) - b(\theta) \right|^2 \leq \delta \int_\Theta [b'(\theta)]^2 \, d\theta.$$

Proof. Once again, let $a_\star$ be the supremum. An application of Lemma 3 to $a(\theta) = [b_\delta(\theta) - b(\theta)]$ yields

$$\begin{aligned} a_\star^2 &\leq \frac{\delta}{4} \int_\Theta \left[b_\delta'(\theta) - b'(\theta) \right]^2 \, d\theta \leq \frac{\delta}{2} \int_\Theta \left[\{b_\delta'(\theta)\}^2 + \{b'(\theta)\}^2 \right] \, d\theta \\ &\leq \delta \int_\Theta [b'(\theta)]^2 \, d\theta. \end{aligned}$$

The final inequality holds since $\int_\Theta [b_\delta'(\theta)]^2 \, d\theta \leq \int_\Theta [b'(\theta)]^2$; this follows from the application of Lemma 2 between each pair of knots of $b_\delta(\cdot)$. ∎

Lemma 5 *Let* $Y(\boldsymbol{\theta})$ *be a stochastic process with continuously differentiable sample paths and let* $Y_\delta(\boldsymbol{\theta})$ *be its linear interpolant between points* $0, \pm\delta, \ldots$. *Then*

$$\mathbb{P} \left(\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} |Y_\delta(\boldsymbol{\theta}) - Y(\boldsymbol{\theta})| \geq \epsilon \right) \leq \frac{\delta}{\epsilon^2} \mathbb{E} \int_{\boldsymbol{\Theta}} [Y'(\boldsymbol{\theta})]^2 \, d\boldsymbol{\theta}.$$

Uniform convergence holds if the expectation is finite:

$$\lim_{\delta \to 0} \mathbb{P} \left(\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} |Y_\delta(\boldsymbol{\theta}) - Y(\boldsymbol{\theta})| \geq \epsilon \right) = 0 \quad \textit{for all} \quad \epsilon > 0. \tag{7.1}$$

Proof. From Lemma 4, it follows that

$$\begin{aligned}\mathbb{P}\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |Y_\delta(\boldsymbol{\theta}) - Y(\boldsymbol{\theta})| \geq \epsilon\right) &\leq \mathbb{P}\left(\delta \int_{\boldsymbol{\Theta}} [Y'(\boldsymbol{\theta})]^2 \, d\boldsymbol{\theta} \geq \epsilon^2\right) \\ &\leq \frac{\delta}{\epsilon^2}\, \mathbb{E} \int_{\boldsymbol{\Theta}} [Y'(\boldsymbol{\theta})]^2 \, d\boldsymbol{\theta},\end{aligned}$$

where the last line follows from the Markov's inequality for any non-negative random variable. ∎

Lemma 6 *If $Y_\delta(\boldsymbol{\theta})$ converges uniformly to $Y(\boldsymbol{\theta})$, as defined in (7.1), then*

$$\lim_{\delta\to 0} \mathbb{P}\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} Y_\delta(\boldsymbol{\theta}) \geq c\right) = \mathbb{P}\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} Y(\boldsymbol{\theta}) \geq c\right) \quad \textit{for any} \quad c,$$

where the right hand side is continuous.

Proof. For any $\epsilon > 0$,

$$\mathbb{P}\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} Y_\delta(\boldsymbol{\theta}) \geq c\right) \geq \mathbb{P}\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} Y(\boldsymbol{\theta}) \geq c+\epsilon\right) - \mathbb{P}\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |Y_\delta(\boldsymbol{\theta}) - Y(\boldsymbol{\theta})| > \epsilon\right).$$

Consequently, $\liminf_{\delta\to 0} \mathbb{P}\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} Y_\delta(\boldsymbol{\theta}) \geq c\right) \geq \mathbb{P}\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} Y(\boldsymbol{\theta}) \geq c+\epsilon\right)$. However, since ϵ is arbitrary,

$$\liminf_{\delta\to 0} \mathbb{P}\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} Y_\delta(\boldsymbol{\theta}) \geq c\right) \geq \mathbb{P}\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} Y(\boldsymbol{\theta}) \geq c\right).$$

By a similar argument, it follows that

$$\limsup_{\delta\to 0} \mathbb{P}\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} Y_\delta(\boldsymbol{\theta})\right) \geq \mathbb{P}\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} Y(\boldsymbol{\theta}) \geq c\right)$$

which completes the proof. ∎

Proof of Theorem 3. First, convergence of finite-dimensional distributions is a consequence of the multivariate central limit theorem. Since a linear interpolant is always maximized at one of the knots, this implies that the theorem holds for a linear interpolant:

$$\lim_{n\to\infty} \mathbb{P}\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} S^\star_\delta(\boldsymbol{\theta}) \geq c\right) = \mathbb{P}\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} Z_\delta(\boldsymbol{\theta}) \geq c\right)$$

for any $\delta > 0$. For any $\epsilon > 0$, Lemma 5 implies that

$$\begin{aligned}
\mathbb{P}\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} S^{\star}(\boldsymbol{\theta}) \geq c\right) &\leq \mathbb{P}\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} S^{\star}_{\delta}(\boldsymbol{\theta}) \geq c-\epsilon\right) + \mathbb{P}\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |S^{\star}(\boldsymbol{\theta}) - S^{\star}_{\delta}(\boldsymbol{\theta})| \geq \epsilon\right) \\
&\leq \mathbb{P}\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} S^{\star}_{\delta}(\boldsymbol{\theta}) \geq c-\epsilon\right) + \frac{\delta}{\epsilon^2}\,\mathbb{E}\int_{\boldsymbol{\Theta}}\left(\frac{\partial}{\partial\boldsymbol{\theta}} S^{\star}_{\delta}(\boldsymbol{\theta})\right)^2 d\boldsymbol{\theta} \\
&= \mathbb{P}\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} S^{\star}_{\delta}(\boldsymbol{\theta}) \geq c-\epsilon\right) + \frac{\delta}{\epsilon^2}\,\mathbb{E}\int_{\boldsymbol{\Theta}}\left[Z'_{\delta}(\boldsymbol{\theta})\right]^2 d\boldsymbol{\theta},
\end{aligned}$$

where $Z'_{\delta}(\boldsymbol{\theta}) = \partial Z_{\delta}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$. The last equality follows from the fact that Z_{δ} and $S^{\star}_{\delta}$ have the same covariance function. Assumption A4 implies that the expectation is finite. From the convergence of finite-dimensional distributions, it follows that

$$\limsup_{n\to\infty} \mathbb{P}\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} S^{\star}(\boldsymbol{\theta}) \geq c\right) \leq \mathbb{P}\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} Z_{\delta}(\boldsymbol{\theta}) \geq c-\epsilon\right) + \frac{\delta}{\epsilon^2}\,\mathbb{E}\int_{\boldsymbol{\Theta}}\left[Z'_{\delta}(\boldsymbol{\theta})\right]^2 d\boldsymbol{\theta}.$$

First, let $\delta \to 0$ and apply Lemma 6 to Z_{δ}. Next, let $\epsilon \to 0$ to obtain

$$\limsup_{n\to\infty} \mathbb{P}\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} S^{\star}(\boldsymbol{\theta}) \geq c\right) \leq \mathbb{P}\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} Z(\boldsymbol{\theta}) \geq c\right).$$

A similar argument shows that

$$\liminf_{n\to\infty} \mathbb{P}\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} S^{\star}(\boldsymbol{\theta}) \geq c\right) \geq \mathbb{P}\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} Z(\boldsymbol{\theta}) \geq c\right)$$

which completes the proof. ∎

Proof of Theorem 4. We assume the regularity conditions 1 to 4 in Adler (2000). The integral in (3.8) can be expressed as

$$\begin{aligned}
\int_{c^2}^{\infty} \mathbb{P}\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \langle \mathbf{U}_J, \boldsymbol{\xi}_J(\boldsymbol{\theta})\rangle \geq \frac{c}{\sqrt{y}}\right) h_J(y)\,dy &= \int_{c^2}^{\frac{c^2}{w_0}} \mathbb{P}\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \langle \mathbf{U}_J, \boldsymbol{\xi}_J(\boldsymbol{\theta})\rangle \geq \frac{c}{\sqrt{y}}\right) h_J(y)\,dy \\
&\quad + \int_{\frac{c^2}{w_0}}^{\infty} \mathbb{P}\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \langle \mathbf{U}_J, \boldsymbol{\xi}_J(\boldsymbol{\theta})\rangle \geq \frac{c}{\sqrt{y}}\right) h_J(y)\,dy,
\end{aligned} \tag{7.2}$$

where $w_0 = (1 - r_0^2/2)$ and r_0 is the critical radius of the tube. The volume-of-tube formula given in (A.4) is exact when $y \in [c^2, c^2/w_0]$ and it is only approximate when $y \in [c^2/w_0, \infty)$. In the former case, from (3.9)

$$\mathbb{P}\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \langle \mathbf{U}_J, \boldsymbol{\xi}_J(\boldsymbol{\theta})\rangle \geq \frac{c}{\sqrt{y}}\right) = \sum_{t=0}^{d} \frac{\zeta_t^J}{A_{d+1-t}} \mathbb{P}\left[B_{(d+1-t)/2,(J-d-1+t)/2} \geq w^2\right].$$

We express the first integral in (7.2) as $F(c^2) - F(c^2/w_0)$, where

$$F(x) = \sum_{t=0}^{d} \frac{\zeta_t^J}{A_{d+1-t}} \int_x^\infty \mathbb{P}\left[B_{(d+1-t)/2,(J-d-1+t)/2} \geq w^2\right] h_J(y)\, dy.$$

Note that the second integral in (7.2) is ≥ 0 providing a lower bound. Furthermore,

$$\int_{c^2/w_0}^{\infty} \mathbb{P}\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \langle \mathbf{U}_J, \boldsymbol{\xi}_J(\boldsymbol{\theta})\rangle \geq \frac{c}{\sqrt{y}}\right) h_J(y)\, dy \leq \int_{c^2/w_0}^{\infty} h_J(y) dy = \mathbb{P}\left(\chi_J^2 \geq \frac{c^2}{w_0}\right).$$

Therefore, $F(c^2)-F(c^2/w_0) \leq \mathbb{P}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} Z_J(\boldsymbol{\theta}) \geq c\right] \leq F(c^2)-F(c^2/w_0)+\mathbb{P}\left(\chi_J^2 \geq c^2/w_0\right)$. As $c \to \infty$, $F(c^2) - F(c^2/w_0) \approx F(c^2)$. Therefore, $\mathbb{P}\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} Z_J(\boldsymbol{\theta}) \geq c\right) \approx F(c^2)$ as $c \to \infty$. By performing the integration in $F(c^2)$, it follows that

$$\mathbb{P}\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} Z_J(\boldsymbol{\theta}) \geq c\right) = \sum_{t=0}^{d} \frac{\zeta_t^J}{A_{d+1-t}} \mathbb{P}\left(\chi_{d+1-t}^2 \geq c^2\right) + o[c^{-1}\exp(-c^2/2)] \quad \text{as} \quad c \to \infty.$$

When the Karhunen-Loève expansion is infinite, the above result for the truncated Gaussian random field $Z_J(\boldsymbol{\theta})$ is extended by letting $J \to \infty$ as follows. Uniform convergence of the Karhunen-Loève expansion implies that $Z_J(\boldsymbol{\theta}) \longrightarrow Z(\boldsymbol{\theta})$ uniformly and hence

$$\mathbb{P}\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} Z_J(\boldsymbol{\theta}) \geq c\right) \longrightarrow \mathbb{P}\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} Z(\boldsymbol{\theta}) \geq c\right) \quad \text{as} \quad J \to \infty. \tag{7.3}$$

The volume-of-tube formula given in (A.4) is in terms of ζ_t^J; however, as $J \to \infty$ and for $t = 0, \ldots, d$, $\zeta_t^J \to \zeta_t$, the corresponding geometric term found via $\rho(\boldsymbol{\theta}, \boldsymbol{\theta})$, Therefore the result (3.10) holds. For example, the expression for $\kappa_0 \equiv \zeta_0$ is derived by approximating $\mathcal{M}$ by a series of short line segments to obtain

$$\kappa_0 = \int_{\boldsymbol{\theta}} \left| \det\left[\nabla_1 \nabla_2^T \rho(\boldsymbol{\theta}, \boldsymbol{\theta})\right] \right|^{1/2} d\boldsymbol{\theta}. \qquad \blacksquare$$

Remark 3: We take sufficiently large J so that the relation (7.3) holds. In practice, it is not necessary to employ a truncated covariance function (3.7) that requires specification of J and the manifold $\mathcal{M}$. Our calculations are carried out in terms of the covariance function $\mathbb{C}(\boldsymbol{\theta}, \boldsymbol{\theta}^\dagger)$. In effect, knowledge of J and the specification of $\mathcal{M}$ does not arise in practice.

Appendix A: Explicit Expressions for Geometric Constants in (3.10)

We consider finite Karhunen-Loève expansion with J terms in deriving the geometric constants. As a first step, we partition the manifold $\mathcal{M}$, correspondingly the tube $\mathfrak{T}(r, \mathcal{M})$ and the parameter space $\boldsymbol{\Theta}$ into various boundary regions. First, each point in $\mathfrak{T}(r, \mathcal{M})$ is linked to a point in $\mathcal{M}$ by a perpendicular projection. Correspondingly, each point in $\mathcal{M}$ is linked to a set of points in $\mathfrak{T}(r, \mathcal{M})$. Second, partition $\mathcal{M}$ into regions $\mathcal{M}_0, \ldots, \mathcal{M}_d$ based on the dimension of the linked sets, where $\mathcal{M}_0$ represents the main part of the manifold and $\mathcal{M}_1, \ldots, \mathcal{M}_d$ represent boundary regions. For example, when $d = 1$, $\mathcal{M}_1$ corresponds to the two end-points and $\mathcal{M}_0$ corresponds to the rest of the tube (see Fig. 1). If $d = 2$, manifold $\mathcal{M}$ is a polygon so that $\mathcal{M}_2$ represents the corners, $\mathcal{M}_1$ the edges and $\mathcal{M}_0$ the interior. In effect, for a d-dimensional manifold $\mathcal{M}$, we can partition both $\mathfrak{T}(r, \mathcal{M})$ and the space $\boldsymbol{\Theta}$ into $(d+1)$ regions to express $\vartheta(r, \mathcal{M}) = V_0 + V_1 + \cdots + V_d$. The main part of the tube can be represented as

$$\left[(1 + \|\boldsymbol{\tau}\|^2)^{-1/2}\ (\boldsymbol{\xi}(\boldsymbol{\theta}) + \mathbf{Q}(\boldsymbol{\theta})\,\boldsymbol{\tau})\colon \boldsymbol{\theta} \in \boldsymbol{\Theta}, \|\boldsymbol{\tau}\| \leq \tau_0\right], \tag{A.1}$$

where $\tau_0 = \sqrt{1 - w^2}/w$, $\mathbf{Q}(\boldsymbol{\theta})$ is an orthonormal basis matrix for the normal space at $\boldsymbol{\xi}(\boldsymbol{\theta})$. Provided that this transformation is one-to-one, the volume V_0 can be expressed as $V_0 = \int_{\boldsymbol{\theta}} \int_{\boldsymbol{\tau}} \left| \det[\mathbf{J}(\boldsymbol{\theta}, \boldsymbol{\tau})] \right| d\boldsymbol{\theta}\, d\boldsymbol{\tau}$, where $\mathbf{J}(\boldsymbol{\theta}, \boldsymbol{\tau})$ is the Jacobian of the representation (A.1). The determinant of the Jacobian can be expressed as $\det[\mathbf{J}(\boldsymbol{\theta}, \boldsymbol{\tau})] = P_{\boldsymbol{\theta}}(\boldsymbol{\tau})(1+\|\boldsymbol{\tau}\|^2)^{-n/2}$, where $P_{\boldsymbol{\theta}}(\boldsymbol{\tau})$ is a dth degree polynomial in $\boldsymbol{\tau}$ with coefficients depending on $\boldsymbol{\theta}$. This representation allows the integral defining V_0 to be split into its $\boldsymbol{\theta}$ and $\boldsymbol{\tau}$ components, leading to a finite series expansion, for a truncated $Z_J(\boldsymbol{\theta})$,

$$V_0 = \sum_{t=0}^{d} \kappa_t \frac{2A_J}{A_{t+1} A_{d+1-t}} \mathbb{P}\left[B_{(d+1-t)/2,(J-d-1+t)/2} \geq w^2\right],$$

where, κ_t are the polynomial coefficients integrated over $\mathcal{M}$ for even-order t and the partial beta terms arise from integrating the τ parts. Odd-order terms integrate to 0 by symmetry; therefore, we set $\kappa_t = 0$ when t is odd. Recall that $A_t = 2\pi^{t/2}/\Gamma(t/2)$ is the $(t-1)$-dimensional volume of the unit sphere $\mathcal{S}^{(t-1)}$ in $\mathcal{R}^t$. The first constant κ_0 is the d-dimensional volume of the manifold $\mathcal{M}$, represented in terms of the covariance function, expressed as

$$\kappa_0 = \int_{\boldsymbol{\Theta}} \mathbb{C}(\boldsymbol{\theta}, \boldsymbol{\theta})^{-(d+1)/2} \left| \det \begin{bmatrix} \mathbb{C}(\boldsymbol{\theta}, \boldsymbol{\theta}^{\dagger}) & \nabla_2^T\, \mathbb{C}(\boldsymbol{\theta}, \boldsymbol{\theta}^{\dagger}) \\ \nabla_1\, \mathbb{C}(\boldsymbol{\theta}, \boldsymbol{\theta}^{\dagger}) & \nabla_1 \nabla_2^T\, \mathbb{C}(\boldsymbol{\theta}, \boldsymbol{\theta}^{\dagger}) \end{bmatrix} \right|_{\boldsymbol{\theta}^{\dagger} = \boldsymbol{\theta}}^{1/2} d\boldsymbol{\theta}, \tag{A.2}$$

where ∇_1 and ∇_2 denote vectors of partial derivative operators with respect to the components of $\boldsymbol{\theta}$ and $\boldsymbol{\theta}^{\dagger}$ respectively. The geometric constant κ_2 is the measure of curvature of $\mathcal{M}$.

The process for handling boundary corrections is similar. To compute the main boundary corrections, represent the half-tubes around boundaries in a form similar to (A.1), with $\mathbf{Q}(\boldsymbol{\theta})$ supplemented by a vector tangent to $\mathcal{M}$ but normal to $\partial\mathcal{M}$, the boundary of $\mathcal{M}$. The vector $\boldsymbol{\tau}$ is then restricted to a half-sphere. Following the derivation of Weyl (1939), we obtain a series of the form, for truncated $Z_J(\boldsymbol{\theta})$,

$$V_1 = \sum_{t=0}^{d-1} \ell_t \frac{A_J}{A_{t+1}A_{d-t}} \mathbb{P}\left[B_{(d-t)/2,(J-d+t)/2} \geq w^2\right],$$

where ℓ_t terms are the integrals of polynomial coefficients. The first term, ℓ_0 is the $(d-1)$-dimensional volume of $\partial\mathcal{M}$ which has a form similar to (A.2), summed over each of the boundary faces. It is important to note that odd order terms no longer disappear; ℓ_1 is a measure of rotation of $\partial\mathcal{M}$ and ℓ_2 is a measure of curvature similar to κ_2. Similarly, at corners where two boundary faces meet, we can represent

$$V_2 = \sum_{t=0}^{d-2} \nu_t \frac{A_J}{A_{t+1}A_{d-1-t}} \mathbb{P}\left[B_{(d-1-t)/2,(J-d+1+t)/2} \geq w^2\right],$$

where ν_0 measures the rotation angles in the regions of $\partial^2\mathcal{M}$ (the boundary of $\partial\mathcal{M}$) where two boundary faces meet and ν_1 is a combination of rotation angles and rotation of the edges. Currently, our software library enables computing all the terms given in (3.10); effectively yielding a complete implementation of the tube formula up to $d = 3$. To the best of our knowledge, there exist no method for general implementation of higher-order terms with boundary corrections.

Remark 4: When $d = 2$, the fourth order coefficients are $\ell_2 = \nu_1 = m_0 = 0$. Additionally, the Euler-Poincare characteristic (Knowles and Siegmund, 1989) satisfies $\kappa_2 + \ell_1 + \nu_0 = 2\pi\mathcal{E} - \kappa_0$ eliminating the need to compute κ_2, ℓ_1 and ν_0 directly. The Euler-Poincare characteristic is the number of pieces making up the manifold, minus the number of holes. When $\boldsymbol{\Theta}$ is a compact as well as a convex set and $\mathbb{C}(\boldsymbol{\theta}, \boldsymbol{\theta}) > 0$ for all $\boldsymbol{\theta}$ then $\mathcal{E} = 1$.

Combining the above results together, the tube formula, up to fourth order terms, can be expressed as

$$
\begin{aligned}
\vartheta(r,\mathcal{M}) &\approx V_0+V_1+V_2+V_3 \\
&= \kappa_0 \frac{A_J}{A_{d+1}} \mathbb{P}\left[B_{(d+1)/2,(J-d-1)/2} \geq w^2\right] + \ell_0 \frac{A_J}{2A_d} \mathbb{P}\left[B_{d/2,(J-d)/2} \geq w^2\right] \\
&\quad + (\kappa_2+\ell_1+\nu_0) \frac{A_J}{2\pi A_{d-1}} \mathbb{P}\left[B_{(d-1)/2,(J-d-1)/2} \geq w^2\right] \\
&\quad + (\ell_2+\nu_1+m_0) \frac{A_J}{4\pi A_{d-2}} \mathbb{P}\left[B_{(d-2)/2,(J-d-2)/2} \geq w^2\right], \qquad \text{(A.3)}
\end{aligned}
$$

where m_0 measures the size of wedges at corners where three boundary faces of $\mathcal{M}$ meet. After completing evaluation of all terms leads to a series,

$$
\vartheta(r,\mathcal{M}) \approx \sum_{t=0}^{d} \zeta_t^J \frac{A_J}{A_{d+1-t}} \mathbb{P}\left[B_{(d+1-t)/2,(J-d-1+t)/2} \geq w^2\right]. \qquad \text{(A.4)}
$$

The dominant term ζ_0^J can be expressed as

$$
\zeta_0^J = \int_{\boldsymbol{\Theta}} \left\| \frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{\xi}(\boldsymbol{\theta}) \right\| d\boldsymbol{\theta} = \int_{\boldsymbol{\theta}} \left| \det\left[\nabla_1 \nabla_2^T \rho_J(\boldsymbol{\theta},\boldsymbol{\theta})\right] \right|^{1/2} d\boldsymbol{\theta}, \qquad \text{(A.5)}
$$

where ∇_1 and ∇_2 are partial derivative operators with respect to the first and second arguments of $\rho_J(\cdot,\cdot)$, respectively.

The following correspondence (up to $t=3$) holds: $\zeta_0=\kappa_0, \zeta_1=\ell_0/2, \zeta_2=(\kappa_2+\ell_1+\nu_0)/(2\pi)$ and $\zeta_3=(\ell_2+\nu_1+m_0)/(4\pi)$. The tube formula is exact for tubes with radius $r \leq r_0$, the critical radius.

References

Adler, R. J. (1990). *An Introduction to Continuity, Extrema, and Related Topics for General Gaussian Processes*. Institute of Mathematical Statistics, Hayward, CA.

Adler, R. J. (2000). On excursion sets, tube formulas and maxima of random fields. *Ann. Applied Prob.*, 10:1–74.

Azaïs, J.-M. and Wschebor, M. (2005). On the distribution of the maximum of a Gaussian field with d parameters. *Ann. Applied Prob.*, 15:254–278.

Bickel, P. and Chernoff, H. (1993). Asymptotic distribution of the likelihood ratio statistic in a prototypical non regular problem. In *Statistics and Probability: A Raghu Raj Bahadur Fetschrift*, J. K. Ghosh and S. K. Mitra and K. R. Parthasarathy and B. L. S. Prakasa Rao (eds.), 83–96, Wiley Eastern Limited.

Chen, H. and Chen, J. (2001). Large sample distribution of the likelihood ratio test for normal mixtures. *Statistics and Prob. Letters*, 52:125–133.

Chernoff, H. (1954). On the distribution of the likelihood ratio. *Ann. Math. Stat.*, 25:573–578.

Dacunha-Castelle, D. and Gassiat, E. (1999). Testing the order of a model using locally conic parameterization: Population mixtures and stationary arma processes. *Ann. Stat.*, 27:1178–1209.

Davies, R. B. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika*, 64:247–254.

Davies, R. B. (1987). Hypothesis testing when a parameter is present only under the alternative. *Biometrika*, 74:33–43.

Ghosh, J. K. and Sen, P. K. (1985). On the asymptotic performance of the log likelihood ratio statistic for the mixture model and related results. In *Proc. Berkeley Conference in Honor of J. Neyman and J. Kiefer*, Lecam, L. M. and Olshen, R. A. (Eds.), 789–806, Vol. 2, Wadsworth Advanced Books. Belmont, CA.

Glaz, J., Naus, J., and Wallenstein, S. (2001). *Scan Statistics.* Springer, New York.

Hartigan, J. A. (1985). A failure of likelihood asymptotics for normal mixtures. In *Proc. Berkeley Conference in Honor of J. Neyman and J. Kiefer*, Lecam, L. M. and Olshen, R. A. (Eds.), 807–810, Vol. 2, Wadsworth Advanced Books. Belmont, CA.

Hotelling, H. (1939). Tubes and spheres in n-spaces, and a class of statistical problems. *American J. Math.*, 61:440–460.

Johansen, S. and Johnstone, I. (1990). Hotelling's theorem on the volume of tubes: some illustrations in simultaneous inference and data analysis. *Ann. Stat.*, 18:652–684.

Knowles, M. and Siegmund, D. (1989). On Hotelling's geometric approach to testing for a nonlinear parameter in regression. *Intr. Stat. Rev.*, 57:205–220.

Lehmann, E. L. (1999). *Elements of Large-Sample Theory*. Springer-Verlag, New York.

Lin, Y. (1997). *The Likelihood Ratio Test of Mixture Hypothesis and the Tube Volume Problem*. PhD thesis, Department of Statistics, Pennsylvania State University.

Lindsay, B. G. (1995). *Mixture Models: Theory, Geometry and Applications*. IMS NSF-CBMS Regional Conference Series, Hayward.

Liu, X., Pasarica, C., and Shao, Y. (2003). Testing homogeneity in gamma mixture models. *Scandinavian J. Stat.*, 30:227–239.

Loader, C. (1999). *Local Regression and Likelihood*. Springer, New York.

Loader, C. (2005). The volume-of-tubes formula: Computational methods and statistical applications. *E-print Archive: arXiv:math.ST/0511502*.

McLachlan, G. and Peel, D. (2000). *Finite Mixture Models*. Wiley, New York.

Naiman, D. Q. (1987). Simultaneous confidence bounds in multiple regression using predictor variable constraints. *J. American Stat. Assoc.*, 82:214–219.

Naiman, D. Q. (1990). On volumes of tubular neighborhoods of spherical polyhedra and statistical inference. *Ann. Stat.*, 18:685–716.

Pilla, R. S. (2006). Inference under convex cone alternatives for correlated data. *Ann. Stat.*, in press.

Pilla, R. S. and Loader, C. (2005). The volume-of-tube formula: Perturbation tests, mixture models and scan statistics. **E-print Archive:** *math.ST/0511503*, v1.

Pilla, R. S., Loader, C., and Taylor, C. C. (2005). New technique for finding needles in haystacks: Geometric approach to distinguishing between a new source and random fluctuations. *Physical Review Letters*, 95:230202–1–230202–4.

Roeder, K. (1990). Density estimation with confidence sets exemplified by superclusters and voids in the galaxy. *J. American Stat. Assoc.*, 85:617–624.

Rubin, H. (1956). Uniform convergence of random functions with applications to statistics. *Ann. Stat.*, 27:200–203.

Shapiro, A. (1985). Asymptotic distribution for test statistics in the analysis of moment structures under inequality constraints. *Biometrika*, 72:133–144.

Siegmund, D. and Worsley, K. (1995). Testing for a signal with unknown location and scale in a stationary Gaussian random field. *Ann. Stat.*, 23:608–639.

Siegmund, D. and Zhang, H. (1993). The expected number of local maxima of a random field and the volume of tubes. *Ann. Stat.*, 21:1948–1966.

Sun, J. (1993). Tail probabilities of the maxima of Gaussian random fields. *Ann. Prob.*, 21:34–71.

Sun, J. and Loader, C. (1994). Simultaneous confidence bands for linear regression and smoothing. *Ann. Stat.*, 22:1328–1345.

Takemura, A. and Kuriki, S. (2002). On the equivalence of the tube and Euler characteristic methods for the distribution of the maximum of Gaussian fields over piecewise smooth domains. *Ann. Applied Prob.*, 12:768–796.

Uusipaikka, E. (1983). Exact confidence bands for linear regression over intervals. *J. American Stat. Assoc.*, 78:638–644.

Weyl, H. (1939). On the volume of tubes. *American Journal of Mathematics*, 61:461–472.

Wolfe, J. H. (1971). A Monte Carlo study of sampling distribution of the likelihood ratio for mixtures of multinomial distributions. *Technical Bulletin STB 72-2, San Diego: U.S. Naval Personnel and Training Research Laboratory.*

Worsley, K. (1995). Estimating the number of peaks in a random field using the Hadwiger characteristic of excursion sets, with applications to medical images. *Ann. Stat.*, 23:640–669.

Worsley, K. (2001). Testing for signals with unknown location and scale in a χ^2 random field, with an application to fMRI. *Advances in Applied Prob.*, 33:773–793.