ISSN 1342-2804

Research Reports on Mathematical and Computing Sciences

Testing Regions with Nonsmooth Boundaries via Multiscale Bootstrap

Hidetoshi Shimodaira

November 2006, B–436

Department of
Mathematical and
Computing Sciences
Tokyo Institute of Technology

SERIES B: Operations Research

Testing Regions with Nonsmooth Boundaries via Multiscale Bootstrap

Hidetoshi Shimodaira

Department of Mathematical and Computing Sciences
Tokyo Institute of Technology
2-12-1 Ookayama, Meguro-ku, Tokyo 152-8552, Japan
shimo@is.titech.ac.jp

RUNNING HEAD: MULTISCALE BOOTSTRAP

Footnotes

- Supported by Grant KAKENHI-17700276 from MEXT of Japan.
- AMS 2000 subject classifications. Primary 62G10; secondary 62G09.
- *Key words and phrases.* problem of regions, bootstrap probability, bias correction, Fourier transform, nearly flat surfaces, scaling-law, multiple comparisons.

Summary

Consider a binary response function of data, for example, whether hierarchical clustering produces a particular dendrogram of interest. An arbitrary-shaped region of the parameter space may represent the null hypothesis defined by the binary response to the population. The bootstrap probability is a widely used p-value, and its calibration has been attempted in the literature; the test bias is estimated as curvature of smooth boundary surfaces of the region. However boundaries are nonsmooth for regions of practical importance such as cones. To treat such singularities, the Fourier transforms of surfaces are employed in this paper. Computation requires only the binary responses to bootstrap samples of size n' generated from data of size n. Our method first computes bootstrap probabilities for several values of n' around n, and then extrapolates them, after some transformation, back to $n' = -n$. This gives corrected p-values related to the bootstrap iteration.

1. Introduction

Let Y be a random vector of dimension $m+1$ for some integer $m \geq 1$, and $y \in \mathbb{R}^{m+1}$ be its observed value in $m+1$ dimensional Euclidean space. Our argument is based on the multivariate normal model with unknown mean vector $\mu \in \mathbb{R}^{m+1}$ and covariance identity I_{m+1},

$$Y \sim N_{m+1}(\mu, I_{m+1}).$$

This is a simplification of reality. Let $\mathcal{X} = (x_1, \ldots, x_n)$ be a sample of size n. We assume there is an approximate transformation, depending on n, from $\mathcal{X}$ to y so that Y is appropriately normalized. Let $\mathcal{H} \subset \mathbb{R}^{m+1}$ be an arbitrary-shaped region of positive volume. We would like to calculate a p-value $p(y)$ for testing the null hypothesis $\mu \in \mathcal{H}$. This is the "problem of regions" discussed previously in Efron and Tibshirani (1998), where smoothness of the boundary surface $\partial\mathcal{H}$ has been assumed. Our purpose here is to extend the argument to nonsmooth $\partial\mathcal{H}$.

The most common practice for calculating $p(y)$ can be described as follows. Let $\theta \in \mathbb{R}^m$ be a nuisance parameter vector with domain Θ, and $\mu(\theta)$ be a function to express the boundary surface as $\partial\mathcal{H} = \{\mu(\theta) \mid \theta \in \Theta\}$. By assuming $p(y)$ is decreasing as y moves away from $\mathcal{H}$, we consider the test bias only on the boundary surface as

$$P(p(Y) < \alpha \mid \mu(\theta)) = \alpha + \mathrm{bias}(\theta),$$

where α is a significance level and $P(\cdot \mid \mu)$ denotes probability with respect to Y. Let $T(y)$ be a test statistic, which is increasing as y moves away from $\mathcal{H}$, and define a p-value $p(y|\theta) = P(T(Y) \geq T(y) \mid \mu(\theta))$ for testing the simple hypothesis $\mu = \mu(\theta)$. Then, $p_{\sup}(y) = \sup_{\theta\in\Theta} p(y|\theta)$ controls the type-I error, i.e., $\mathrm{bias}(\theta) \leq 0$ for all $\theta \in \Theta$.

Let $\hat{\theta}(y)$ be the maximum likelihood estimate of θ, i.e., the value which minimizes $\|y - \mu(\theta)\|$. Berger and Boos (1994) argued that "it seems a waste of information in the data to take the sup over all values of θ." $p_{\sup}(y)$ can be improved if the point achieving the sup is very far from $\hat{\theta}(y)$ so that $p_{\sup}(y)$ is much larger than $p(y|\hat{\theta}(y))$. They gave $p_\beta(y) = \sup_{\theta\in C_\beta(y)} p(y|\theta) + \beta$ for a very small β such as 0.001, where $C_\beta(y)$ is a $1-\beta$ confidence set for θ. $p_\beta(y)$ controls the type-I error and it is less conservative than $p_{\sup}(y)$, although not yet unbiased in general.

Our mathematical formulation focuses on the unbiasedness, i.e., $\mathrm{bias}(\theta) = 0$ for all $\theta \in \Theta$. Since $p_\beta(y)$ is closer to being unbiased than $p_{\sup}(y)$, $p_\beta(y)$ is preferable to $p_{\sup}(y)$. More generally, we take the position that unbiasedness is an ideal property, which automatically leads to the control of type-I error. Another position is that the unbiasedness is a secondary criterion and it can be even inappropriate since it does not directly address the evidence in the data against the null hypothesis (Perlman and Wu, 1999). Both positions are compromised in our search for *approximately* unbiased tests, not the exact one.

The paper is organized as follows. In Section 2, we describe our method and the main result after some background. In Section 3, we introduce the notion of *nearly flat surfaces* and develop a nonstandard asymptotic theory utilizing Fourier transforms of surfaces. This theory is quite simple mathematically, yet it provides the basis of our method. In Section 4, numerical examples are given for demonstrating the importance of approximately unbiased tests for nonsmooth surfaces. In Section 5, practical issues are discussed. All mathematical proofs are given in the Appendix.

2. Approximately unbiased tests

2.1. **Background.** An example of an approximately unbiased test is given simply by $p(y) = p(y|\hat{\theta}(y))$. The test bias may reduce asymptotically as $n \to \infty$. In fact, this $p(y)$ has $\text{bias}(\theta) = O(n^{-3/2})$ for smooth $\partial\mathcal{H}$ when $T(y)$ is the signed distance, i.e., $\pm\|y - \mu(\hat{\theta}(y))\|$ with positive sign for $y \notin \mathcal{H}$ and negative sign for $y \in \mathcal{H}$. This is a one-sided analogue of the Bartlett correction applied to $\|y - \mu(\hat{\theta}(y))\|^2$.

Another example is the bootstrap probability introduced by Felsenstein (1985) for phylogenetic inference. For a scale parameter $\sigma > 0$, we define

$$\alpha_{\sigma^2}(y) = P_{\sigma^2}(Y^* \in \mathcal{H} \mid y),$$

where the probability is with respect to

$$Y^* \sim N_{m+1}(y, \sigma^2 I_{m+1}).$$

The p-value $p(y) = \alpha_1(y)$ has been widely used. Given a high-performance computing environment, we can calculate $\alpha_{\sigma^2}(y)$ very easily as an observed frequency of $Y^* \in \mathcal{H}$ even for the situation where complicated data analysis is employed almost as a black-box. The following two properties (a) and (b) of $\alpha_{\sigma^2}(y)$ are important for facilitating the implementation. (a) Only used is a binary response whether $Y^* \in \mathcal{H}$. For example, whether a particular dendrogram or cluster of interest is observed in hierarchical clustering. (b) Only resampling from the data y is performed. We assume that the above normal model is obtained at least approximately from bootstrap sample $\mathcal{X}^* = (x_1^*, \ldots, x_{n'}^*)$ for $n' > 0$ by applying the same transformation used for $\mathcal{X}$; thus $\mathcal{H}$ is unchanged for Y^*. The normalizing constant is $\sqrt{n}$ for many situations (Bickel and Freedman, 1981), and the scale is $\sigma^2 = n/n'$. This is because we used $\sqrt{n}$ for $\mathcal{X}^*$, for which the normalizing constant should have been $\sqrt{n'}$.

Although easy to implement, the bias of $\alpha_1(y)$ is rather large; $\text{bias}(\theta) = O(n^{-1/2})$ even for smooth $\partial\mathcal{H}$. There have been several attempts to improve $\hat{\alpha}_1(y)$. Let $\Phi(\cdot)$ and $\Phi^{-1}(\cdot)$

be the distribution function of $N(0,1)$ and its inverse, respectively. For convenience, we often work on the bootstrap z-value and the z-value of $p(y)$ defined respectively by

$$z_{\sigma^2}(y) = \Phi^{-1}(1 - \alpha_{\sigma^2}(y)), \quad q(y) = \Phi^{-1}(1 - p(y)).$$

Efron et al. (1996) gave a corrected p-value with bias $O(n^{-1})$ by $q(y) = z_1(y) - 2c(y)$, where $c(y) = z_1(\mu(\hat{\theta}(y)))$ term adjusts the bias due to the curvature of $\partial\mathcal{H}$. Efron and Tibshirani (1998) mentioned another p-value with bias $O(n^{-3/2})$ defined by $p(y) = p(y|\hat{\theta}(y))$ using $T(y) = z_1(y)$; this is an instance of the double bootstrap of Hall (1992). These two methods require resampling from $\mu(\hat{\theta}(y))$, and thus they do not satisfy (b) but only (a).

For easier implementation, although based on the theory of Efron et al. (1996), a corrected p-value with bias $O(n^{-3/2})$ satisfying both (a) and (b) was given by the multiscale bootstrap method of Shimodaira (2002). Considering $z_{\sigma^2}(y)$ as a function of $1/\sigma$, $q(y)$ is defined as the slope of this function at $1/\sigma = 1$;

$$q(y) = \frac{\partial z_{\sigma^2}(y)}{\partial(1/\sigma)}\bigg|_1 . \tag{2.1}$$

This $p(y)$ as well as its extension to the exponential family of distributions (Shimodaira, 2004) is shown to be equivalent to other p-values of bias $O(n^{-3/2})$ such as $p(y) = p(y|\hat{\theta}(y))$ using $T(y) = z_1(y)$ or the signed distance, and $p(y)$ obtained from p^*-formula (Barndorff-Nielsen, 1986).

All the above methods for approximately unbiased tests are derived under the assumption that $\partial\mathcal{H}$ is smooth. In this paper, we develop a method without this assumption. We consider the following property. (c) The boundary surface $\partial\mathcal{H}$ is possibly nonsmooth, including smooth surfaces as special cases. This is not merely a theoretical interest, but important for applications such as phylogenetic inference. An attempt for (c) is found in Liu and Singh (1997, Remark 4.2), but it does not satisfy (a) nor (b).

2.2. **Our method.** Here we propose a generalization of the multiscale bootstrap method. This new method satisfies (a) and (b), and it is justified for (c). We work on $\sigma z_{\sigma^2}(y)$, which may be called the *normalized bootstrap z-value.* Considering $\sigma z_{\sigma^2}(y)$ as a function of σ^2, we will specify parametric models of this function in Section 3.4. The models are denoted as $\psi(\sigma^2|\beta(y))$ with a parameter vector $\beta(y)$. The dependency on y may be suppressed in the notation. The new method proceeds as follows.

Step 1. Calculate $\alpha_{\sigma^2}(y)$ at several $\sigma^2 > 0$ values specified in advance. In reality, each $\alpha_{\sigma^2}(y)$ is estimated by counting the frequency of $Y^* \in \mathcal{H}$.

Step 2. Estimate the parameter $\beta(y)$ by fitting the model

$$\sigma z_{\sigma^2}(y) = \psi(\sigma^2|\beta(y)) \tag{2.2}$$

to the bootstrap probabilities obtained in Step 1. An estimated model with the estimated parameter value is written as $\psi(\sigma^2|\hat\beta(y))$.

Step 3. Calculate $q(y)$ by extrapolating $\psi(\sigma^2|\hat\beta(y))$ back to $\sigma^2=-1$. More specifically, for an integer $k>0$ and a real number $\sigma_0^2>0$, we define

$$q_k(y)=\sum_{j=0}^{k-1}\frac{(-1-\sigma_0^2)^j}{j!}\frac{\partial^j\psi(\sigma^2|\hat\beta(y))}{\partial(\sigma^2)^j}\Big|_{\sigma_0^2} \tag{2.3}$$

by using the first k terms of the Taylor series around σ_0^2. Then $p_k(y)=1-\Phi(q_k(y))$, $k=1,2,\ldots$, constitute a class of corrected p-values. The dependency on σ_0^2 is implicit in this notation, and we use $\sigma_0^2=1$ throughout.

The idea of extrapolation to $\sigma^2=-1$ is intimately analogous to the SIMEX, simulation-extrapolation, method for measurement error models (Cook and Stefanski, 1994). The mysterious $\sigma^2=-1$ will be explained by the fact that σz_{σ^2} is obtained from an unbiased q by applying the Gaussian kernel smoothing filter with variance σ^2+1 in the space of y; we notice that $\sigma z_{\sigma^2}=q$ when the variance of the filter is zero.

The bootstrap probability corresponds to $k=1$, and the corrected p-value of Shimodaira (2002) corresponds to $k=2$. In fact, it is easily verified from (2.1), (2.2) and (2.3) that

$$\psi(1|\hat\beta)=q_1,\quad \frac{\partial(\psi(\sigma^2|\hat\beta)/\sigma)}{\partial(1/\sigma)}\Big|_1=q_2.$$

The main result shown in the following sections is that bias$(\theta)\to 0$, ignoring some asymptotic errors, as $k\to\infty$ even for nonsmooth $\partial\mathcal{H}$. Although it happened that the procedure of Shimodaira (2002) is a special case of the new method, their justifications are based on very different asymptotic frameworks. We will also show similarities between our method and the k-th iterated bootstrap method for interpreting the p-values.

3. Asymptotic theory of nearly flat surfaces

3.1. **Nearly flat surfaces.** We describe conditions imposed on $\partial\mathcal{H}$, and some notation for the Fourier transforms.

First we consider a parametric form of $\partial\mathcal{H}$. Let $y_j=u_j$, $j=1,\ldots,m$, for $u\in\mathbb{R}^m$, and $y_{m+1}=v$ for $v\in\mathbb{R}$. A point is simply written as $y=(u,v)$. Using a continuous function $h(u)$, we write $\mathcal{H}$, at least locally, as

$$\mathcal{H}=\{(u,v)\mid v\le -h(u),u\in\mathbb{R}^m\}.$$

We will use the L^1-norm and L^∞-norm of h defined as

$$\|h\|_1=\int_{\mathbb{R}^m}|h(u)|\,du,\quad \|h\|_\infty=\sup_{u\in\mathbb{R}^m}|h(u)|.$$

In ordinary asymptotic theory as $n \to \infty$, the shape of $\mathcal{H}$ in the space of normalized Y is magnified by the factor $\sqrt{n}$. For asymptotic expansions, we utilize the fact that a smooth $\partial\mathcal{H}$ approaches a flat surface in a neighborhood of any point on $\partial\mathcal{H}$. More specifically, it follows from the argument below eq. (2.12) of Efron and Tibshirani (1998) that $h(0) = \partial h/\partial\theta|_0 = 0$ and $\partial^j h/\partial\theta^j|_0 = O(n^{-\frac{j-1}{2}})$, $j \geq 2$, by choosing coordinates without losing generality. This argument, however, does not apply to nonsmooth $\partial\mathcal{H}$. For example, a cone-shaped $\mathcal{H}$ is invariant under the magnification.

In the asymptotic argument of this paper, we do not let $n \to \infty$. Instead, we introduce an artificial parameter τ and let $\tau \to 0$. The idea is to assume $\|h\|_\infty = O(\tau)$ so that any surfaces approach flat surfaces. For smooth surfaces, this assumption can be interpreted as $\partial^j h/\partial\theta^j = O(\tau)$, $j \geq 0$. We will work on asymptotic expansions in terms of $\tau, \tau^2, \tau^3, \ldots$, and take only finite terms. In reality, $\|h\|_\infty$ is not necessarily small enough, and the theory may be confirmed numerically; we do the same thing when applying the ordinary asymptotic theory to a small or moderate n. Although n is irrelevant to our argument, it is implicitly assumed to be large enough to justify the normal model. Some further notion is discussed at the end of Section 3.4.

For each τ, we impose two extra conditions on h as follows. They are for technical reasons related to the Fourier transforms.

We assume that h is absolutely integrable, i.e., $\|h\|_1 < \infty$ for each τ. Then the Fourier transform of h is defined as

$$\tilde{h}(\omega) = \mathcal{F}h(\omega) = \int_{\mathbb{R}^m} e^{-i\omega\cdot u}\, h(u)\, du,$$

where $\omega \in \mathbb{R}^m$ is a spatial angular frequency vector, $\omega \cdot u = \sum_{j=1}^m \omega_j u_j$ is inner product, and $i = \sqrt{-1}$ is the imaginary unit. The condition $\|h\|_1 < \infty$ implies $\|\tilde{h}\|_\infty < \infty$. It follows from the Riemann-Lebesgue lemma that $\lim_{\|\omega\|\to\infty} \tilde{h}(\omega) = 0$, and $\tilde{h}$ is uniformly continuous.

For a given $\tilde{h}(\omega)$, we next assume that $\|\tilde{h}\|_1 < \infty$ for each τ. Then, the inverse Fourier transform is defined as

$$h(u) = \mathcal{F}^{-1}\tilde{h}(u) = \frac{1}{(2\pi)^m} \int_{\mathbb{R}^m} e^{i\omega\cdot u}\, \tilde{h}(\omega)\, d\omega,$$

where the above integral can be interpreted in the same way as $\mathcal{F}$; it is a Lebesgue integral instead of an improper Riemann integral. We also have $\|h\|_\infty < \infty$, $\lim_{\|u\|\to\infty} h(u) = 0$, and h is uniformly continuous. The convolution of g and h for $\|g\|_1 < \infty$ is written as

$$(g * h)(u) = \int_{\mathbb{R}^m} g(u')h(u - h')\, du' = \mathcal{F}^{-1}[\tilde{g}(\omega)\tilde{h}(\omega)](u).$$

Note that $g * h$ also satisfies the same conditions as h; $\|g * h\|_1 \leq \|g\|_1\|h\|_1 < \infty$, and $\|\mathcal{F}(g * h)\|_1 = \|\tilde{g}(\omega)\tilde{h}(\omega)\|_1 \leq \|\tilde{g}\|_\infty\|\tilde{h}\|_1 < \infty$.

The conditions are summarized as $\|h\|_1 < \infty$, $\|\tilde{h}\|_1 < \infty$, and $\|h\|_\infty = O(\tau)$. In this paper, an h satisfying these conditions is called *nearly flat*.

3.2. **Expectation operator and its inverse.** Consider the expected value of $h(U^*)$ with respect to $U^* \sim N_m(u, \sigma^2 I_m)$, and denote it as

$$E_{\sigma^2} h(u) = E_{\sigma^2}(h(U^*)|u).$$

The left-hand side is meant to be an operator applied to the function h. This notation will be used repeatedly in the following sections. Here we give expressions for $E_{\sigma^2} h(u)$ and its inverse using the Fourier transforms.

Let $f_{\sigma^2}(u) = (2\pi\sigma^2)^{-\frac{m}{2}} e^{-\frac{\|u\|^2}{2\sigma^2}}$ be the density of $N_m(0, \sigma^2 I_m)$, and $\tilde{f}_{\sigma^2}(\omega) = e^{-\sigma^2 \frac{\|\omega\|^2}{2}}$ be its Fourier transform. Then $E_{\sigma^2} h(u)$ can be expressed as a convolution,

$$\int_{\mathbb{R}^m} f_{\sigma^2}(u^* - u) h(u^*)\, du^* = f_{\sigma^2}(-u) * h(u),$$

and its Fourier transform becomes $\mathcal{F}[E_{\sigma^2} h](\omega) = \tilde{f}_{\sigma^2}(\omega)\, \tilde{h}(\omega)$. By applying the inverse Fourier transform, we obtain

$$E_{\sigma^2} h(u) = \mathcal{F}^{-1}\Big[e^{-\sigma^2 \frac{\|\omega\|^2}{2}}\, \tilde{h}(\omega)\Big](u).$$

This is a Gaussian kernel smoothing of scale σ, and can interpreted as an application of the Gaussian low-pass filter $\tilde{f}_{\sigma^2}$ to $\tilde{h}$.

The inverse filter of $\tilde{f}_{\sigma^2}$ defines $\mathcal{F}[E_{\sigma^2}^{-1} h](\omega) = (1/\tilde{f}_{\sigma^2}(\omega))\, \tilde{h}(\omega)$. Its inverse Fourier transform gives

$$\begin{aligned} E_{\sigma^2}^{-1} h(u) &= \mathcal{F}^{-1}\Big[e^{\sigma^2 \frac{\|\omega\|^2}{2}}\, \tilde{h}(\omega)\Big](u) \\ &= E_{-\sigma^2} h(u). \end{aligned}$$

Although $E_{\sigma^2} h$ is nearly flat for $\sigma^2 > 0$, $E_{\sigma^2}^{-1} h$ may not be defined in general unless $\|e^{\sigma^2 \frac{\|\omega\|^2}{2}}\, \tilde{h}(\omega)\|_1 < \infty$.

3.3. **Bootstrap probability.** We give an expression for α_{σ^2} using the expectation operator. A simple linear theory will be discussed by taking only up to $O(\tau)$ terms.

For $y = (u, v)$, we write $\alpha_{\sigma^2}(y)$ as

$$\begin{aligned} \alpha_{\sigma^2}(u, v) &= P_{\sigma^2}(V^* \le -h(U^*) \mid u, v) \\ &= E_{\sigma^2}\Big[\Phi\Big(\frac{-h(U^*) - v}{\sigma}\Big) \mid u\Big]. \end{aligned}$$

By letting $x = (-E_{\sigma^2} h(u) - v)/\sigma$ and $\varepsilon = (-h(U^*) + E_{\sigma^2} h(u))/\sigma$, it becomes $\alpha_{\sigma^2}(u, v) = E_{\sigma^2}(\Phi(x+\varepsilon)|u)$. Considering $E_{\sigma^2}(\varepsilon|u) = 0$ and $E_{\sigma^2}(\varepsilon^k|u) = O(\tau^k)$, $k = 2, 3, \ldots$, we obtain the asymptotic expansion as

$$
\begin{aligned}
\alpha_{\sigma^2}(u,v) &= E_{\sigma^2}\big(\Phi(x)+\phi(x)\varepsilon-\phi(x)x\varepsilon^2/2 \mid u\big)+O(\tau^3)\\
&= \Phi(x)-\phi(x)xE_{\sigma^2}(\varepsilon^2|u)/2+O(\tau^3)\\
&= \Phi\big(x-xE_{\sigma^2}(\varepsilon^2|u)/2+O(\tau^3)\big). \qquad (3.1)
\end{aligned}
$$

If we take only $x=O(\tau)$ above, a scaling-law for α_{σ^2} is expressed as

$$
\sigma z_{\sigma^2}(u,v)=v+E_{\sigma^2}h(u)+O(\tau^2). \qquad (3.2)
$$

3.4. **Models.** To specify appropriate parametric models $\psi(\sigma^2|\beta)$, we would like to give expressions for $E_{\sigma^2}h(u)$ for concrete cases of h such as polynomials and cones. For this purpose, h is allowed to be unbounded, and a justification of this argument is given at the end of this section.

For a smooth h, it is shown in Appendix A.1 that

$$
E_{\sigma^2}h(u)=\sum_{j=0}^{\infty}\sigma^{2j}\beta_j(u), \qquad (3.3)
$$

where the Taylor series of $h(u^*)$ around u gives

$$
\beta_j(u)=\frac{1}{2^j j!}\sum_{j_1+\cdots+j_m=j}\frac{j!}{j_1!\cdots j_m!}\frac{\partial^{2j}h}{\partial u_1^{2j_1}\cdots\partial u_m^{2j_m}}, \quad j\ge 0.
$$

The summation ranges over all combinations of nonnegative integers with the sum fixed at j. If h is a polynomial of degree $2k-1$, we may redefine $\beta_0 := v+\beta_0(u)$ to get

$$
\psi_{\mathrm{poly},k}(\sigma^2|\beta)=\sum_{j=0}^{k-1}\beta_j\sigma^{2j}, \quad k\ge 1.
$$

This model specifies σz_{σ^2} correctly by ignoring $O(\tau^2)$ terms.

Since $\psi_{\mathrm{poly},k}(\sigma^2|\beta)$ is nothing but a polynomial of σ^2 of degree $k-1$, it approximates arbitrarily well any continuous function within a finite interval of $\sigma^2>0$ by increasing k, so that it might suffice even for nonsmooth h. However, it would most certainly be better to have a concise model, because the parameters must be estimated from bootstrap probabilities. For a cone-shaped $\mathcal{H}$ with vertex at the origin, it is shown in Appendix A.2 that

$$
E_{\sigma^2}h(u)=\sum_{j=0}^{\infty}\sigma^{1-j}\beta_j(u) \qquad (3.4)
$$

in a neighborhood of the vertex, where $\beta_j(u)=O(\|u\|^j)$ as $\|u\|\to 0$; these $\beta_j(u)$ are not relevant to those in (3.3). A model that takes conical singularity into account may be defined as

$$\psi_{\mathrm{sing},k}(\sigma^2|\beta) = \beta_0 + \sum_{j=1}^{k-2} \frac{\beta_j \sigma^{2j}}{1+\beta_{k-1}(\sigma-1)}, \quad k \geq 3,$$

where $0 \leq \beta_{k-1} \leq 1$. If $\beta_{k-1} = 0$, $\psi_{\mathrm{sing},k}(\sigma^2|\beta)$ reduces to $\psi_{\mathrm{poly},k-1}(\sigma^2|\beta)$. If $\beta_{k-1} = 1$, then $\psi_{\mathrm{sing},k}(\sigma^2|\beta)$ includes the first two terms of (3.4).

Polynomials and cones are unbounded, and they are obviously not nearly flat. When an unbounded h is used in this paper, it is in fact meant to be g as defined below. Let $g(u)$ be a continuous function of slow growth, i.e., for some k, $|g(u)| = O(\|u\|^k)$, as $\|u\| \to \infty$. We assume $g(u) = O(\tau)$ for each u. Let $h(u) = f_{\delta^2}(u) * (w_{L^2}(u) g(u))$, where $w_{L^2}(u) = \tilde{f}_{L^{-2}}(u)$ is a window function and δ and L are constants. It is shown in Appendix A.3 that $h(u)$ is nearly flat and the difference between $g(u)$ and $h(u)$ can be made arbitrarily small by letting $\delta \to 0$ and $L \to \infty$. We ignore this difference by assuming δ is sufficiently small and L is sufficiently large.

3.5. **Unbiased surfaces.** Let $\mathcal{R}$ and $\mathcal{S}$ be regions in $\mathbb{R}^{m+1}$. If $\mathcal{R}$ is the rejection region of an unbiased test of the null hypothesis $\mathcal{S}$, then $\partial\mathcal{S}$ is called an *unbiased surface* of $\mathcal{R}$ in this paper. This test is unbiased for $\mathcal{H}$ if $\mathcal{H} = \mathcal{S}$.

We first derive an expression of $\mathcal{S}$ for a given $\mathcal{R}$. Let r and s be nearly flat continuous functions, and $z = \Phi^{-1}(1-\alpha)$ for significance level α. In a similar way to $\mathcal{H}$, we write

$$\mathcal{R} = \{(u,v) \mid v > z - r(u)\}, \quad \mathcal{S} = \{(u,v) \mid v \leq -s(u)\}.$$

The rejection probability at $\mu = (\theta, -s(\theta))$ for $\theta \in \mathbb{R}^m$ is

$$P_1(V > z - r(U) \mid \theta, -s(\theta)) = 1 - E_1(\Phi(z - r(U) + s(\theta)) \mid \theta),$$

where P_{σ^2} and E_{σ^2}, previously defined for $Y^* = (U^*, V^*)$, are used for $Y = (U,V)$. By letting $x = z - E_1 r(\theta) + s(\theta)$, $\varepsilon = -r(U) + E_1 r(\theta)$ in (3.1), we have

$$\Phi^{-1}(1 - P_1(\text{reject } \mathcal{S} \mid \theta)) = z - E_1 r(\theta) + s(\theta) + O(\tau^2). \tag{3.5}$$

Since the left-hand side is equal to z, we obtain

$$s(\theta) = E_1 r(\theta) + O(\tau^2). \tag{3.6}$$

We next derive an expression of the unbiased z-value $q(u,v)$ for the null hypothesis $\mathcal{S}$. From (3.6), we have

$$r(u) = E_{-1} s(u) + O(\tau^2). \tag{3.7}$$

Since $q(u,v) = z$ for $y \in \partial\mathcal{R}$, we substitute $r(u) = q(u,v) - v$ for the left-hand side of (3.7), and we obtain

$$q(u,v) = v + E_{-1} s(u) + O(\tau^2), \tag{3.8}$$

from which the unbiased p-value is calculated as $p(u,v) = 1 - \Phi(q(u,v))$.

Assume $s(u) = h(u) + O(\tau^2)$ and $\sigma z_{\sigma^2}(u,v) = \psi(\sigma^2|\beta(u)) + O(\tau^2)$. By comparing (3.2) and (3.8), we obtain

$$q(u,v) = \psi(-1|\beta(u)) + O(\tau^2) \tag{3.9}$$

for an approximately unbiased test of $\mathcal{H}$ with bias $O(\tau^2)$, if the right-hand side exists. A refinement of this result is given below.

Theorem 1. *For a nearly flat h, assume that $E_{-1}h$ and $E_{-1}h^2$ exist. Define $q(u,v)$ by extrapolating the value of $\sigma z_{\sigma^2}(u,v)$ to $\sigma^2 = -1$. This gives an approximately unbiased test of $\mathcal{H}$ with bias only $O(\tau^3)$.*

For a given $\mathcal{S}$, let us define $q(u,v)$ by (3.8), and use it for testing $\mathcal{H}$. The rejection probability at $\mu \in \partial\mathcal{H}$ is obtained by evaluating (3.5) for $\mu = (\theta, -h(\theta))$,

$$\Phi^{-1}(1 - P_1(\text{reject } \mathcal{H} \mid \theta)) = z - s(\theta) + h(\theta) + O(\tau^2). \tag{3.10}$$

Therefore the test bias in terms of z-value is $h(\theta) - s(\theta) + O(\tau^2)$ at each θ. In the next section, we examine the test bias for choices of $s(\theta)$.

3.6. **A class of approximately unbiased tests.** We evaluate the bias of our method proposed in Section 2, and consider its generalization. The argument is made within the linear theory by ignoring $O(\tau^2)$ terms.

If h is a polynomial of degree $2k-1$, $\psi_{\text{poly},k}(\sigma^2|\beta)$ is a correct model for σz_{σ^2}. Thus $q_k(y)$ of (2.3) using the correct model becomes simply $q_k(u,v) = \psi_{\text{poly},k}(-1|\beta(u))$. This is an example of (3.9), where the right-hand side exists. For any continuous h, if considered only within a bounded domain $\|u\| < L$, say, it can be approximated arbitrarily well by a polynomial of large k. Therefore the bias of q_k, if using the correct model, approaches zero as $k \to \infty$, by ignoring $O(\tau^2)$ terms and the effect from the outside of the bounded domain. Our q_k is approximately unbiased for any continuous h in this sense.

Is there any other q_k with this property? To answer this question, we first rewrite our q_k, $k \geq 1$, as follows.

$$\begin{aligned} q_k(u,v) &= v + \sum_{j=0}^{k-1} \frac{(-1-\sigma_0^2)^j}{j!} \frac{\partial^j}{\partial(\sigma^2)^j}\bigg|_{\sigma_0^2} \mathcal{F}^{-1}\left[\tilde{h}(\omega) e^{-\sigma^2 \frac{\|\omega\|^2}{2}}\right](u) + O(\tau^2) \\ &= v + \mathcal{F}^{-1}\left[\tilde{h}(\omega) e^{\frac{\|\omega\|^2}{2}} (1 - J_k(\omega))\right](u) + O(\tau^2), \end{aligned} \tag{3.11}$$

where $J_k(\omega)$ for our q_k is given by straightforward calculation as

$$\begin{aligned} J_k(\omega) &= 1 - e^{-(1+\sigma_0^2)\frac{\|\omega\|^2}{2}} \sum_{j=0}^{k-1} \frac{(1+\sigma_0^2)^j}{j!} \Big(\frac{\|\omega\|^2}{2}\Big)^j \\ &= \frac{\gamma(k, (1+\sigma_0^2)\frac{\|\omega\|^2}{2})}{\Gamma(k)} = \sum_{j=k}^{\infty} \frac{(-1)^{j-k}(1+\sigma_0^2)^j \|\omega\|^{2j}}{(k-1)!(j-k)!j2^j}. \end{aligned}$$

By comparing (3.8) and (3.11), we may define s for q_k and denote it as s_k. The test bias in terms of z-value is then

$$h(\theta) - s_k(\theta) = \mathcal{F}^{-1}[\tilde{h}(\omega) J_k(\omega)](\theta) + O(\tau^2).$$

Therefore, one may think that any q_k can be approximately unbiased in the same sense as our q_k, if $J_k(\omega) \to 0$ as $k \to \infty$. This idea is stated more formally in the below.

Theorem 2. *Define a class of corrected z-values $q_k(u, v)$, $k = 1, 2, \ldots$, by (3.11) for a given $J_k(\omega)$. We assume that h is nearly flat and $J_k(\omega)$ satisfies the following three conditions. (i) $J_k(\omega) \to 0$ as $k \to \infty$ at each ω. (ii) For some $C > 0$, $\|J_k\|_\infty < C$ holds for all k. (iii) $\|e^{\frac{\|\omega\|^2}{2}}(1 - J_k(\omega))\|_\infty < \infty$ for each k. Then, we have*

$$P_1(q_k(U, V) > z \mid \theta, -h(\theta)) \to \alpha + O(\tau^2)$$

as $k \to \infty$ at each θ.

In addition to the above three conditions, let us assume that (iv) $J_k(\omega)$ is expressed as $J_k(\omega) = \sum_{j=k}^{\infty} a_{k,j} \|\omega\|^{2j}$, where $a_{k,j}$ are coefficients. It is shown in Appendix A.6 that the bias of q_k is $O(\tau^2)$ if h is a polynomial of degree less than or equal to $2k - 1$.

It is easy to verify that the $J_k(\omega)$ of our q_k satisfies conditions (i)-(iv). Another example of a $J_k(\omega)$ satisfying the four conditions is defined below. Let q_k be the corrected z-value of the k-th iterated bootstrap applied to the bootstrap probability. For $k = 1$, $q_1(u, v) = z_1(u, v)$, and for $k \geq 1$,

$$q_{k+1}(u, v) = \Phi^{-1}\Big\{P_1\Big(q_k(U^*, V^*) \leq q_k(u, v) \mid \hat{\theta}(u, v), -h(\hat{\theta}(u, v))\Big)\Big\}.$$

It is shown in Appendix A.7 that the $J_k(\omega)$ of this q_k is given by

$$\begin{aligned} J_k(\omega) &= (1 + e^{-\frac{\|\omega\|^2}{2}})(1 - e^{-\frac{\|\omega\|^2}{2}})^k \\ &= (-1)^k k! \sum_{j=k}^{\infty} (S2(j, k) + S2(j+1, k+1)) \frac{(-1)^j \|\omega\|^{2j}}{2^j j!}, \end{aligned}$$

where $S2(j, k) = \sum_{i=0}^{k} (-1)^{k-i} i^j / i!(k-i)!$ are the Stirling numbers of the second kind.

4. Numerical examples

4.1. **Multiple comparisons.** We consider a polyhedral convex cone expressed as

$$\mathcal{H} = \{\mu \mid \mu_1 \geq \max_{i=2}^{M} \mu_i\},$$

for $M = m+1$; $\mathcal{H}$ can be expressed in terms of $h(u)$ after rotation of the coordinates. We consider y of the form $y_2 = y_1 + 1$, $y_3 = \cdots = y_M = y_1 - d$ below. Table 1 shows p-values for $d = -1$, $M = 10$ and $d = 5$, $M = 10$.

This $\mathcal{H}$ appears in the multiple comparisons with the "best" (Hsu, 1981), or equivalently the ranking and selection of Gupta (1965). The test statistic $T(y) = \max_{i=2}^{M}(y_i - y_1)/\sqrt{2}$ is used often for the multiple comparisons (denoted MC for short). The p-value of MC is $p_{\sup}(y)$, where the sup is attained at the least favorable configuration $\mu_1 = \cdots = \mu_M$. Since $T(y) = 1/\sqrt{2}$ is the same for $d = -1$ and $d = 5$, $p_{\sup}(y)$ gives the same value for the two cases.

For $d = 5$, a statistician may decide by using common sense that $y_3, \ldots, y_M$ should be ignored for calculating the p-value because they are too small. Then, the z-test gives $p(y) = 1 - \Phi(T(y))$. In fact, this is valid for $d \to \infty$, where $p_\beta(y)$ can be made arbitrarily close to the z-test.

This situation is often encountered when using maximum likelihood inference for phylogenetic trees. It is an example of non-nested model selection, where y_i is the log-likelihood of i-th tree, and a large sample size, say, $n > 3000$, justifies the central limit theorem for y. A resampling-based MC (Shimodaira and Hasegawa, 1999) as well as the z-test (Kishino and Hasegawa, 1989) has been widely used. Practitioners complain that MC is conservative and we discover nothing, whereas the z-test tends to give false discoveries.

Our method seems to have the advantages of both MC and z-test; it behaves in a way similar to MC near $d = 0$ and also to the z-test as $d \to \infty$. Our p_k are calculated from thirteen α_{σ^2} values of the multiscale bootstrap (Figure 1). For $d = 5$, p_k, $k = 1, \ldots, 4$, give almost the same values as the z-test. For $d = 0$, p_1 is too small, but it is calibrated by our method so that p_4 is almost the same as MC.

— Insert Table 1 Here —

— Insert Figure 1 Here —

4.2. **Rejection probabilities.** We consider a region in $\mathbb{R}^2$ defined as

$$\mathcal{H} = \Big\{(\mu_1, \mu_2) \mid \mu_2 \leq -(a + \mu_1^2/3)^{1/2}\Big\},$$

where $a = 1$ or $a = 0$. Both cases are shown in Figure 2. Table 2 gives rejection probabilities at significance level $\alpha = 0.05$ for $\mu \in \partial\mathcal{H}$; we chose $\mu_1 = d/\sqrt{2}$, $d = 0, 1, 2, 4, 8$. By applying a linear transformation to $\mathcal{H}$, it can be shown that the case when $a = 0$

corresponds to the case when $M = 3$ for MC with $\mu_2 = \mu_1$, $\mu_3 = \cdots = \mu_M = \mu_1 - d$. The results for $M = 10$ are also given.

MC is unbiased at $d = 0$, and so is the z-test as $d \to \infty$. However, MC performs very poorly for $d \to \infty$ and so does the z-test at $d = 0$. On the other hand, the corrected p-values, especially p_4, are much closer to being unbiased overall. More detailed examination of the results is given below.

The rejection region $\mathcal{R}$ is defined by $p_k < \alpha$ for each k, and it corresponds to the region above the curve $p_k = \alpha$. As expected, $P(p_k(Y) < \alpha)$ approaches α as d or k increases. However, the conical singularity made the convergence slower for $a = 0$ than for $a = 1$, and an increase of dimensionality made it even slower for $M = 10$ than for $M = 3$.

Let $\mathcal{H}'$ be the complement of $\mathcal{H}$ in $\mathbb{R}^{m+1}$, and denote α_{σ^2} and p_k by α'_{σ^2} and p'_k respectively when $\mathcal{H}'$ is the null hypothesis. Bayesian-like formulas, $\alpha_{\sigma^2} + \alpha'_{\sigma^2} = 1$ and $p_k + p'_k = 1$, hold for our method. The rejection probability $P(p_k(Y) > 1 - \alpha)$ for $\mathcal{H}'$ behaves similarly to that for $\mathcal{H}$, but the sign of test bias is reversed.

For $\mathcal{H}$ and $\mathcal{H}'$, the test bias of p_k tends to be larger as the singularity of the surface becomes more evident by increasing M or decreasing d. The test bias of p_k reduces even for such cases, although very slowly, as k increases.

— Insert Figure 2 Here —

— Insert Table 2 Here —

5. Discussion

Theorem 2 states that the test bias of our p_k approaches zero, ignoring the asymptotic error of $O(\tau^2)$, as $k \to \infty$ even for nonsmooth $\partial\mathcal{H}$. This is confirmed numerically in Section 4.2. However, the argument of Lehmann (1952) implies that an unbiased test does not exist for cone-shaped $\mathcal{H}$. This nonexistence does not exclude the possibility of approximately unbiased tests but it *does* lead to some difficulty as explained below.

In Figure 2, $\partial\mathcal{R}$ oscillates more wildly for $a = 0$ than for $a = 1$. This is a consequence of our attempt to reduce the test bias, although a larger bias still remains in p_k for $a = 0$ than for $a = 1$ (Table 2). Shapes of $\mathcal{R}$ similar to those in the panel (b) of Figure 2 are found in DuPreez et al. (1985) and Perlman and Wu (2003, Fig. 11) for regions $\mathcal{H}$ of convex cones, and also Liu and Berger (1995) for regions $\mathcal{H}$ of concave cones, i.e., the complements of convex cones.

These $\mathcal{R}$ violate monotonicity in the sense of Lehmann (1952), and they, particularly in the concave cases, are criticized by Perlman and Wu (1999). I *do* agree with this criticism if the oscillation of $\partial\mathcal{R}$ is large for interpreting the p-value as the evidence against $\mathcal{H}$. Our numerical examples suggest that a compromise between unbiasedness and monotonicity is made by taking around $k = 3$. For example, the curve of $p_4 = 0.95$ for $a = 0$ in

Figure 2 oscillates too much, while p_4 improves the bias of p_3 only slightly in Table 2. For determining an optimal k, the balance should be formulated as mathematical criteria in future work.

The above problem with our p_k, which also applies to the bootstrap iteration, is explained technically by the fact that $\|e^{\frac{\|\omega\|^2}{2}}(1 - J_k(\omega))\|_\infty$ is unbounded for $k \to \infty$ in the condition (iii) of Theorem 2. Let r_k be the r for p_k, and note that $\tilde{r}_k(\omega) = \tilde{h}(\omega)e^{\frac{\|\omega\|^2}{2}}(1 - J_k(\omega))$. Since high frequency components in a nonsmooth h are large, $|\tilde{h}(\omega)|$ reduces only slowly as $\|\omega\| \to \infty$ so that r_k diverges as $k \to \infty$.

In reality, several factors other than k must also be determined; they include σ_0^2, $\psi(\sigma^2|\beta)$, the number of resamples, the number of σ^2 and their values. The scaleboot software (R package is available from CRAN), which comes with examples of real applications, implements maximum likelihood estimation of β by numerical optimization of the binomial likelihood function and model selection by AIC from $\psi_{\mathrm{poly},k}$ and $\psi_{\mathrm{sing},k}$ models. A narrow range of σ^2, say, $\sigma^{-2} = 0.5, 0.6, \ldots, 1.4$, suffices for p_2. For p_k, $k \geq 3$, a wider range of σ^2 as those in Section 4 is needed for estimating the higher order derivatives of σz_{σ^2}. The standard errors of p_k due to the resampling increases as k becomes larger; this also restricts the use of a large k.

Acknowledgments

I wish to thank Michael D. Perlman and the anonymous referees for their comments on the earlier versions of this paper, and Takafumi Kanamori for pointing out the similarity between our method and the SIMEX algorithm. I thank Paul A. Sheridan for his comments to improve the manuscript.

Appendix A. Proofs

A.1. **Expected value of smooth surfaces.** (3.3) follows by applying below to each term of the Taylor series.

$$E_{\sigma^2}\{(U_1^* - u_1)^{k_1} \cdots (U_m^* - u_m)^{k_m} \mid u\} = \sigma^{k_1+\cdots+k_m}(k_1 - 1)!! \cdots (k_m - 1)!!$$

if all $k_1, \ldots, k_m$ are even, and it becomes zero otherwise.

A.2. **Expected value of cone surfaces.** Consider a kind of polar coordinates $(\|u\|, t) \leftrightarrow u$, where $t = u/\|u\|$. Let $c(t)$ be a continuous function on the sphere. Then, cones are expressed as $h(u) = c(t)\,\|u\|$. By considering $f_{\sigma^2}(u^* - u)/f_{\sigma^2}(u^*)$, we have $E_{\sigma^2}h(u) = e^{-\|u\|^2/2\sigma^2}E_{\sigma^2}(h(U^*)e^{U^*\cdot u/\sigma^2}|0)$, where $U^* \sim N_m(0, \sigma^2 I_m)$. In this expression, $\|U^*\|$ and T^* are independent, where $\|U^*\|^2/\sigma^2 \sim \chi_m^2$ and T^* is distributed uniformly on the sphere.

By substituting $h(U^*) = c(T^*)\|U^*\|$ and $U^* \cdot u = \|U^*\| (T^* \cdot u)$ for those in above, we obtain

$$E_{\sigma^2} h(u) = \sigma e^{-\frac{\|u\|^2}{2\sigma^2}} \sum_{j=0}^{\infty} (j!\sigma^j)^{-1} E(\chi_m^{j+1}) E(c(T^*)\,(T^* \cdot u)^j),$$

from which (3.4) immediately follows.

A.3. **Functions of slow growth.** Let $t(u) = w_{L^2}(u)g(u)$. Since $\|t\|_1 < \infty$, we have $\|h\|_1 \le \|f_{\delta^2}\|_1 \, \|t\|_1 < \infty$ and $\|\tilde{h}\|_1 \le \|\tilde{f}_{\delta^2}\|_1 \, \|\tilde{t}\|_\infty < \infty$. Therefore h is nearly flat. At each u, $h(u) \to t(u)$ as $\delta \to 0$ (Lukacs, 1964, p. 781), and $t(u) \to g(u)$ as $L \to \infty$, since $w_{L^2}(u) \to 1$. Then, $h(u) \to g(u)$ as $\delta \to 0$ and $L \to \infty$ at each u.

A.4. **Proof of Theorem 1.** By calculating $E_{\sigma^2}(\varepsilon^2|u)$ in (3.1), the scaling-law of (3.2) becomes

$$\text{(A.1)} \qquad \sigma z_{\sigma^2}(u,v) = v + E_{\sigma^2} h(u) - \frac{v}{2\sigma^2} \Big\{ E_{\sigma^2} h^2(u) - (E_{\sigma^2} h(u))^2 \Big\} + O(\tau^3).$$

We will show that (3.8), if calculated up to $O(\tau^2)$ terms, becomes

$$\text{(A.2)} \qquad q(u,v) = v + E_{-1} s(u) + \frac{v}{2} \Big\{ E_{-1} s^2(u) - (E_{-1} s(u))^2 \Big\} + O(\tau^3).$$

By letting $\sigma^2 = -1$ in (A.1) and $s(u) = h(u) + O(\tau^3)$ in (A.2), we observe that these two formulas are equivalent up to $O(\tau^2)$ terms. This completes the proof.

We first define $\mathcal{R}$ from (A.2) by $q(u,v) > z$, and then we will show this $\mathcal{R}$ is the rejection region of an unbiased test of $\mathcal{S}$, ignoring $O(\tau^3)$ terms. Solve $q(u,v) = z$ with respect to v, and substitute it for v on the right-hand side of $r(u) = z - v$. Then we have

$$\text{(A.3)} \qquad r(u) = E_{-1} s(u) + \frac{z}{2} \Big\{ E_{-1} s^2(u) - (E_{-1} s(u))^2 \Big\} + O(\tau^3).$$

On the other hand, (3.5) becomes, by calculating $E_{\sigma^2}(\varepsilon^2|u)$ in (3.1),

$$\Phi^{-1}(1 - P_1(\text{reject } \mathcal{S} \mid \theta)) = z - E_1 r(\theta) + s(\theta) - \frac{z}{2} \Big\{ E_1 r^2(\theta) - (E_1 r(\theta))^2 \Big\} + O(\tau^3).$$

The right-hand side becomes $z + O(\tau^3)$, by substituting (A.3) for $r(u)$.

A.5. **Proof of Theorem 2.** From (iii), we have $\|\tilde{h}(\omega) e^{\frac{\|\omega\|^2}{2}} (1 - J_k(\omega))\|_1 \le \|\tilde{h}\|_1 \, \|e^{\frac{\|\omega\|^2}{2}} (1 - J_k(\omega))\|_\infty < \infty$. Thus $\mathcal{F}^{-1}$ in (3.11) exists for each k, and we can write $q_k(u,v) = v + E_{-1} s_k(u) + O(\tau^2)$, where $s_k(u)$ is defined by $\tilde{s}_k(\omega) = \tilde{h}(\omega)(1 - J_k(\omega))$. The proof completes by showing that $\lim_{k\to\infty}(h(\theta) - s_k(\theta)) = 0$ at each θ in the below.

Considering $\tilde{h}(\omega) - \tilde{s}_k(\omega) = \tilde{h}(\omega) J_k(\omega)$ and (ii), we have $|e^{i\omega\cdot u}(\tilde{h}(\omega) - \tilde{s}_k(\omega))| \le C\,|\tilde{h}(\omega)|$. Since $C\,\|\tilde{h}\|_1 < \infty$, the dominated convergence theorem gives

$$\lim_{k\to\infty} \mathcal{F}^{-1}[\tilde{h}(\omega) - \tilde{s}_k(\omega)](\theta) = \mathcal{F}^{-1}[\tilde{h}(\omega) \lim_{k\to\infty} J_k(\omega)](\theta).$$

From (i), the right-hand side becomes $\mathcal{F}^{-1}[\tilde{h}(\omega) \times 0] = 0$.

A.6. **Condition (iv).** Here we employ the Fourier transforms in terms of generalized functions (Zemanian, 1965) so that $\mathcal{F}h$ exists without the justification argument at the end of Section 3.4 for functions of slow growth. Consider a monomial $h(u) = u_1^{b_1} \cdots u_m^{b_m}$, $b = b_1 + \cdots + b_m \leq 2k - 1$. Let $\delta(x)$ be the delta function and $\delta^{(b)}(x)$ be its b-th derivative. Then, $\tilde{h}(\omega) = i^b (2\pi)^m \delta^{(b_1)}(\omega_1) \cdots \delta^{(b_m)}(\omega_m)$. On the other hand, $e^{i\omega \cdot u} J_k(\omega) = e^{i\omega\cdot u} \sum_{j=k}^{\infty} a_{k,j} (\omega_1^2 + \cdots + \omega_m^2)^j$ includes only monomials of form $\omega_1^{c_1} \cdots \omega_m^{c_m}$, $c_1 + \cdots + c_m \geq 2k$. Therefore, $\mathcal{F}^{-1}[\tilde{h}(\omega) J_k(\omega)](\theta)$ is a linear combination of the following terms

$$\begin{aligned} &\int_{\mathbb{R}^m} \delta^{(b_1)}(\omega_1) \cdots \delta^{(b_m)}(\omega_m) \omega_1^{c_1} \cdots \omega_m^{c_m} \, d\omega \\ =&(-1)^b \int_{\mathbb{R}^m} \delta(\omega_1) \cdots \delta(\omega_m) \frac{d^{b_1} \omega_1^{c_1}}{d\omega_1^{b_1}} \cdots \frac{d^{b_m} \omega_m^{c_m}}{d\omega_m^{b_m}} \, d\omega = 0, \end{aligned}$$

and thus $h(\theta) - s_k(\theta) = 0$. This argument is linear with respect to h, so it applies to any polynomial of degree less than or equal to $2k - 1$.

A.7. **Bootstrap iteration.** Assume that $q_k(u, v)$ is expressed using a nearly flat r_k as

$$q_k(u, v) = v + r_k(u) + O(\tau^2). \tag{A.4}$$

For $k = 1$, $q_1(u, v) = z_1(u, v) = v + E_1 h(u) + O(\tau^2)$, and (A.4) holds by letting $r_1(u) = E_1 h(u)$. For brevity, write $u' = \hat{\theta}(u, v) = u + O(\tau)$. Then, $\Phi(q_{k+1}(u, v)) = P_1(V^* + r_k(U^*) \leq v + r_k(u) | u', -h(u')) + O(\tau^2) = E_1\{\Phi(v + r_k(u) - r_k(U^*) + h(u')) | u'\} + O(\tau^2)$. The argument of (3.1) gives $q_{k+1}(u, v) = v + r_k(u) - E_1 r_k(u') + h(u') + O(\tau^2)$, and we obtain $r_{k+1}(u) = r_k(u) - E_1 r_k(u) + h(u) + O(\tau^2)$. This implies that (A.4) holds for $k \geq 1$.

Next we will give an expression of $r_k(u)$. The Fourier transform of the recurrence formula is $\tilde{r}_{k+1}(\omega) = (1 - e^{-\frac{\|\omega\|^2}{2}}) \tilde{r}_k(\omega) + \tilde{h}(\omega) + O(\tau^2)$, and for $k = 1$, $\tilde{r}_1(\omega) = \tilde{h}(\omega) e^{-\frac{\|\omega\|^2}{2}}$. By solving the formula, we obtain $\tilde{r}_k(\omega) = \tilde{h}(\omega) e^{\frac{\|\omega\|^2}{2}} \{1 - (1 + e^{-\frac{\|\omega\|^2}{2}})(1 - e^{-\frac{\|\omega\|^2}{2}})^k\} + O(\tau^2)$. The proof completes by solving $\tilde{r}_k(\omega) = \tilde{h}(\omega) e^{\frac{\|\omega\|^2}{2}} (1 - J_k(\omega)) + O(\tau^2)$ for $J_k(\omega)$.

References

Barndorff-Nielsen, O. E. (1986). Inference on full or partial parameters based on the standardized signed log likelihood ratio. *Biometrika* **73** 307–322.

Berger, R. L. and Boos, D. D. (1994). P-values maximized over a confidence set for the nuisance parameter. *J. Amer. Statist. Assoc.* **89** 1012–1016.

Bickel, P. J. and Freedman, D. A. (1981). Some asymptotic theory for the bootstrap. *Ann. Statist.* **9** 1196–1217.

Cook, J. R. and Stefanski, L. A. (1994). Simulation-extrapolation estimation in parametric measurement error models. *J. Amer. Statist. Assoc.* **89** 1314–1328.

DuPreez, J. P., Swanepoel, J. W. H., Venter, J. H. and Somerville, P. N. (1985). Some properties of Somerville's multiple range subset selection procedure for three populations. *South African Statist. J.* **19** 45–72.

Efron, B. and Tibshirani, R. (1998). The problem of regions. *Ann. Statist.* **26** 1687–1718.

Efron, B., Halloran, E. and Holmes, S. (1996). Bootstrap confidence levels for phylogenetic trees. *Proc. Natl. Acad. Sci. USA* **93** 13429–13434.

Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39** 783–791.

Gupta, S. S. (1965). On some multiple decision (selection and ranking) rules. *Technometrics* **7** 225–245.

Hall, P. (1992). *The bootstrap and Edgeworth expansion.* Springer, New York.

Hsu, J. C. (1981). Simultaneous confidence intervals for all distances from the "best". *Ann. Statist.* **9** 1026–1034.

Kishino, H. and Hasegawa, M. (1989). Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. *J. Mol. Evol.* **29** 170–179.

Lehmann, E. L. (1952). Testing multiparameter hypotheses. *Ann. Math. Statistics* **23** 541–552.

Liu, H. and Berger, R. L. (1995). Uniformly more powerful, one-sided tests for hypotheses about linear inequalities. *Ann. Statist.* **23** 55–72.

Liu, R. Y. and Singh, K. (1997). Notions of limiting p values based on data depth and bootstrap. *J. Amer. Statist. Assoc.* **92** 266–277.

Lukacs, E. (1964). Inversion formulae for characteristic functions of absolutely continuous distributions. *Amer. Math. Monthly* **71** 44–47 (Correction: p. 781).

Perlman, M. D. and Wu, L. (1999). The emperor's new tests. *Statist. Sci.* **14** 355–381.

Perlman, M. D. and Wu, L. (2003). On the validity of the likelihood ratio and maximum likelihood methods. *J. Statist. Plann. Inference* **117** 59–81.

Shimodaira, H. (2002). An approximately unbiased test of phylogenetic tree selection. *Syst. Biol.* **51** 492–508.

Shimodaira, H. (2004). Approximately unbiased tests of regions using multistep-multiscale bootstrap resampling. *Ann. Statist.* **32** 2616–2641.

Shimodaira, H. and Hasegawa, M. (1999). Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol. Biol. Evol.* **16** 1114–1116.

Zemanian, A. H. (1965). *Distribution Theory and Transform Analysis: An Introduction to Generalized Functions, with Applications.* McGraw-Hill, New York.

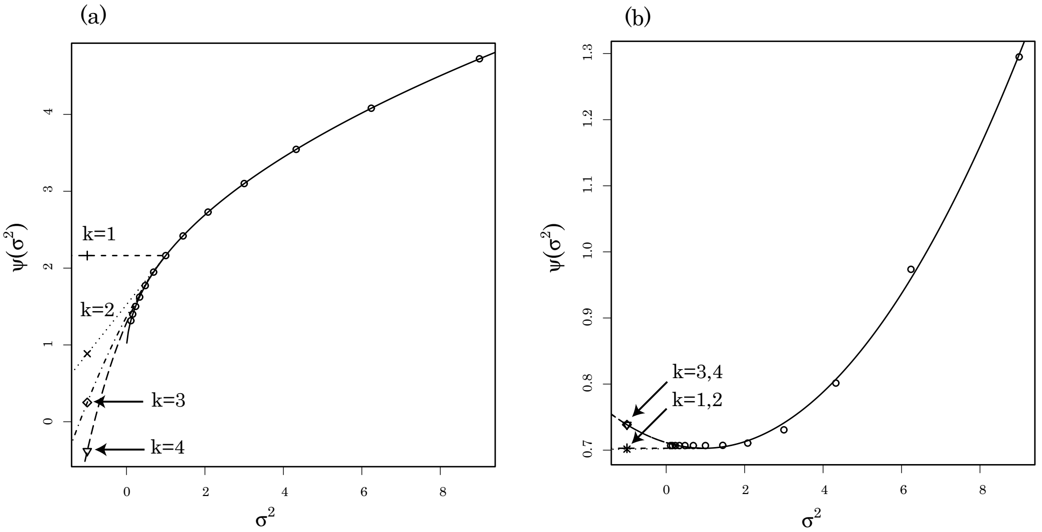

Figure 1. Computation of corrected p-values by the multiscale bootstrap method for (a) $d = -1$, $M = 10$ and (b) $d = 5$, $M = 10$. Observed σz_{σ^2} values are plotted for 13 σ^2-values equally spaced in log-scale between 1/9 and 9. (a) $\psi_{\text{sing},3}$ and (b) $\psi_{\text{poly},3}$ models fit very well. For $d = -1$, $\psi_{\text{sing},3}$ became $\beta_0 + \beta_1\sigma$ because $\hat{\beta}_2 = 1$. Corrected p-values were calculated by (2.3) for $k = 1, \ldots, 4$. Note that α_{σ^2} were in fact calculated from numerical integration here, but the standard errors in Table 1 are given for the case that each α_{σ^2} is estimated from 10,000 resamples.

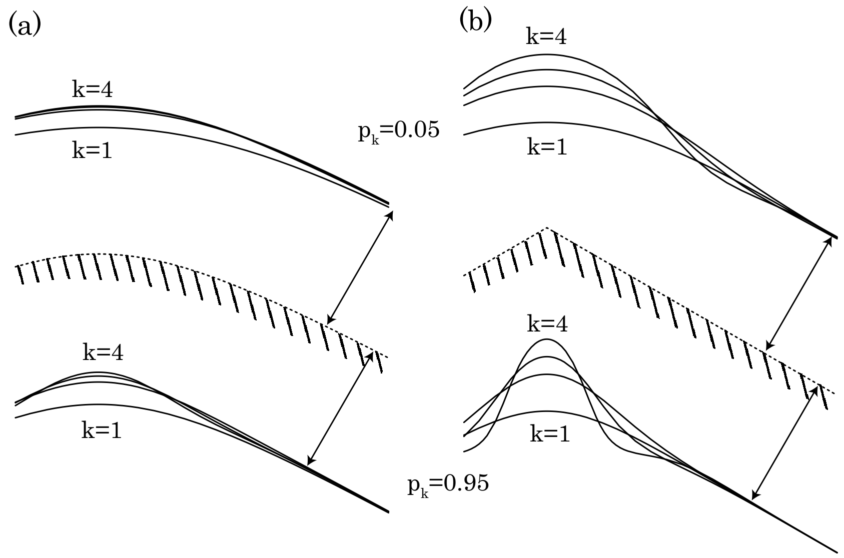

FIGURE 2. Contour lines (solid curves) of $p_k = 0.05$ or 0.95 for (a) $a = 1$ and (b) $a = 0$; lines are drawn for $-1 \leq \mu_1 \leq 3.5$. Hypotheses are regions below the dotted curves. Line segments with arrowheads are of length $\Phi^{-1}(0.95) = 1.64$.

Tables

Table 1. p-values for the two cases of multiple comparisons. The values are in percent, and standard errors are shown in parentheses.

	$d=-1$		$d=5$	
MC	65.91		65.91	
z-test	23.98		23.98	
$k=1$	1.53	(0.02)	24.11	(0.05)
$k=2$	18.79	(0.21)	24.15	(0.09)
$k=3$	40.10	(0.74)	23.01	(0.12)
$k=4$	64.81	(1.66)	23.01	(0.12)

TABLE 2. Rejection probabilities at $\alpha = 0.05$. The values are in percent. They are calculated by numerical integration for $a = 1$ and $a = 0$, and by a Monte-Carlo simulation of 100,000 runs for $M = 10$.

	$P(p_k(Y) < 0.05)$					$P(p_k(Y) > 0.95)$				
d	0	1	2	4	8	0	1	2	4	8
				$a = 1$ (smooth boundary)						
$k = 1$	7.66	7.27	6.51	5.48	5.07	2.65	2.88	3.42	4.45	4.93
$k = 2$	5.33	5.19	4.98	4.89	4.99	4.34	4.56	4.97	5.22	5.01
$k = 3$	5.11	5.04	4.96	4.99	5.00	4.67	4.83	5.08	5.06	5.00
$k = 4$	5.06	5.02	4.98	5.01	5.00	4.79	4.91	5.07	4.99	5.00
				$a = 0$ (nonsmooth boundary) or $M = 3$						
MC	5.00	3.07	2.79	2.77	2.77	5.00	9.18	12.1	13.5	13.6
z-test	8.78	5.59	5.05	5.00	5.00	1.22	2.71	4.09	4.97	5.00
$k = 1$	13.4	7.77	5.79	5.04	5.00	0.85	2.04	3.43	4.85	5.00
$k = 2$	7.66	4.74	4.43	4.88	5.00	1.95	3.99	5.44	5.34	5.00
$k = 3$	6.61	4.51	4.70	5.09	5.00	2.37	4.49	5.52	4.96	5.00
$k = 4$	6.22	4.61	4.99	5.08	5.00	2.67	4.71	5.37	4.89	5.00
				$M = 10$ (multiple comparisons)						
MC	5.1	1.4	0.8	0.8	0.8	5.1	17.3	30.7	38.0	38.7
z-test	23.3	8.9	5.5	4.9	5.0	0.0	0.5	1.9	4.8	5.1
$k = 1$	53.8	25.4	11.1	5.3	5.0	0.0	0.1	0.8	4.0	5.1
$k = 2$	16.9	5.4	3.0	4.1	5.0	0.3	2.0	5.6	7.2	5.1
$k = 3$	10.2	3.4	2.8	4.9	5.0	0.6	3.3	7.3	6.3	5.0
$k = 4$	7.1	2.7	2.7	5.1	5.0	1.3	4.9	8.5	6.2	5.0