

Research Reports on Mathematical and Computing Sciences

Approximately Unbiased Tests for Singular Surfaces
via Multiscale Bootstrap Resampling

Hidetoshi Shimodaira

May 2006, B-430

Department of
Mathematical and
Computing Sciences
Tokyo Institute of Technology

SERIES **B:** *Operations Research*

Approximately Unbiased Tests for Singular Surfaces via Multiscale Bootstrap Resampling

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

May, 2006

RUNNING HEAD: MULTISCALE BOOTSTRAP

Summary

A class of approximately unbiased tests based on bootstrap probabilities is considered for the normal model with unknown expectation parameter vector, where the null hypothesis is represented as an arbitrary-shaped region with possibly singular boundary surfaces. We alter the sample size n' of replicated datasets from the sample size n of the observed dataset, and calculate bootstrap probabilities at several n' values. As shown in a previous paper, this multiscale bootstrap gives a bias-corrected p -value with third-order accuracy by differentiating bootstrap z -value with respect to $\sqrt{n'}$. However, the asymptotic theory is justified only for smooth boundary surfaces, and a breakdown of the theory is observed for cone shaped regions derived from inequality constraints. In this paper, we develop a linear theory, where the Fourier transformation of the boundary surface, instead of the Taylor series, plays an important role. A low-pass filter and its inverse filter represent calculation of bootstrap probability and an unbiased p -value, respectively. It turns out that the unbiased p -value is expressed as the bootstrap probability with $n' = -n$. The obtained class of p -values includes, as special cases, the bootstrap probability and the third-order accurate p -value, and is equivalent in a certain sense to the bootstrap iteration. A new geometrical insight into the controversy over unbiasedness and monotonicity is given by showing that monotone tests, not the approximately unbiased tests, can be counterintuitive. The proposed procedure is illustrated in examples of phylogenetic inference and multiple comparisons.

Footnotes

- Supported by Grant KAKENHI-17700276 from MEXT of Japan.
- AMS 2000 subject classifications. Primary 62G10; secondary 62G09.
- *Key words and phrases.* Approximately unbiased tests, bootstrap probability, bias correction, Fourier transformation, monotonicity, multiple testing, problem of regions, scaling-law.

1 Introduction

The problem we are going to discuss is expressed simply as follows. Let Y be a $m + 1$ -dimensional multivariate normal vector with unknown mean vector μ and covariance identity I_{m+1} ,

$$(1.1) \quad Y \sim N_{m+1}(\mu, I_{m+1}).$$

A null hypothesis of our interest is represented as $\mu \in \mathcal{H}$, where \mathcal{H} is an arbitrarily-shaped region in $m + 1$ -dimensional space. Observing y , we calculate $\hat{\alpha}(y)$, a probability value (p -value) for testing the null hypothesis. We would like to devise $\hat{\alpha}(y)$ so that the bias of hypothesis testing is small while keeping the computational cost manageable and also taking account of other properties of testing called locality and monotonicity. The bias of testing is the difference between rejection probability and significance level $0 < \alpha < 1$ for μ being on the boundary of region $\partial\mathcal{H}$;

$$(1.2) \quad P(\hat{\alpha}(Y) < \alpha \mid \mu) = \alpha + \text{bias}, \quad \mu \in \partial\mathcal{H},$$

where $P(\cdot)$ denotes probability with respect to (1.1). This is also said the bias of the p -value interchangeably.

We do not assume our knowledge on the value of m nor the shape of \mathcal{H} in the calculation of $\hat{\alpha}(y)$, but only assume that a mechanism is available to us for generating bootstrap replicates and identifying whether outcomes are in the region or not. Let y_1^*, \dots, y_B^* be B (say, $B = 10,000$) bootstrap replicates generated from

$$(1.3) \quad Y^* \sim N_{m+1}(y, \sigma^2 I_{m+1}).$$

Typically, the scale parameter is $\sigma = 1$, but it can be altered to any positive value by us. We count how many times the replicates are in the region; $C = \#\{y_1^*, \dots, y_B^* \in \mathcal{H}\}$. The bootstrap probability is calculated as C/B , or formally defined as

$$\tilde{\alpha}_\sigma(y) = P_\sigma(Y^* \in \mathcal{H} \mid y)$$

by considering the limit $B \rightarrow \infty$. Here $P_\sigma(\cdot)$ denotes probability with respect to (1.3).

Our calculation of $\hat{\alpha}(y)$ is based only on $\tilde{\alpha}_\sigma(y)$ values, so that the method can be used for complicated practical applications. Let $\mathcal{D} = \{d_1, \dots, d_n\}$ be a observed sample of size n , and $\mathcal{D}^* = \{d_1^*, \dots, d_{n'}^*\}$ be a bootstrap sample of size n' . We are interested in how much confidence we have in the outcome of a 0-1 valued function $g(\mathcal{D})$. For example, \mathcal{D} is an array of DNA sequences in the hierarchical clustering problem (phylogenetic inference) of Section 2. A particular dendrogram of interest corresponds to $g(\mathcal{D}) = 1$.

The bootstrap probability is easily calculated, although computationally demanding, by repeatedly applying a computer software calculating $g(\mathcal{D})$ to \mathcal{D}^* 's. We assume that there is a transformation f from \mathcal{D} to y such that $y = f(\mathcal{D})$ and $y^* = f(\mathcal{D}^*)$ satisfy (1.1) and (1.3), respectively, at least approximately with scale parameter

$$\sigma^2 = \frac{n}{n'},$$

and that $g(\mathcal{D}) = 1 \Leftrightarrow y \in \mathcal{H}$. For the case of d_1, \dots, d_n being real vectors with sufficiently large n , such a transformation may be given by $y = \sqrt{n} \times \frac{1}{n}(d_1 + \dots + d_n)$ and $y^* = \sqrt{n} \times \frac{1}{n'}(d_1^* + \dots + d_{n'}^*)$. Since the bootstrap probability is transformation invariant, i.e., it does not change by the transformation f , we only assume the existence of f without knowing the functional form of f .

Several versions of $\hat{\alpha}(y)$, denoted $\hat{\alpha}_k(y)$, $k = 1, 2, \dots$, are defined from $\tilde{\alpha}_\sigma(y)$. The simplest one is the bootstrap probability with $\sigma = 1$, denoted $\hat{\alpha}_1(y) = \tilde{\alpha}_1(y)$. This has been used widely since its application to phylogenetic inference in Felsenstein (1985), and there have been attempts to reduce bias of $\hat{\alpha}_1(y)$. The two-level bootstrap of Efron et al. (1996) and Efron and Tibshirani (1998) calculates a bias-corrected p -value based on the ABC bias correction of Efron (1987). Let $\hat{\mu}(y)$ be the maximum likelihood estimate of μ under the restriction that $\mu \in \partial\mathcal{H}$; i.e., the point on the boundary which minimizes the distance to y . The bias of $\hat{\alpha}_1(y)$ mainly comes from the curvature of $\partial\mathcal{H}$ at $\hat{\mu}(y)$. For bias-correction, the two-level bootstrap estimates the curvature as the deviation of $\hat{\alpha}_1(\hat{\mu}(y))$ from 0.5. Computation of $\hat{\mu}(y)$, however, can be an implementational difficulty in complex problems. By introducing a scaling-law to their asymptotic theory, the multiscale bootstrap method of Shimodaira (2002) calculates a even less biased p -value. It estimates the geometric quantities such as distance and curvature from $\tilde{\alpha}_{\sigma_i}(y)$ at arbitrary σ_i , $i = 1, \dots, M$ ($M \geq 2$) without calculating $\hat{\mu}(y)$. This is equivalent to

$$(1.4) \quad \hat{\alpha}_2(y) = 1 - \Phi \left(\left. \frac{d\tilde{z}_\sigma(y)}{d(1/\sigma)} \right|_1 \right),$$

where the bootstrap z -value is defined as

$$(1.5) \quad \tilde{z}_\sigma(y) = \Phi^{-1}(1 - \tilde{\alpha}_\sigma(y)).$$

Here $\Phi(\cdot)$ is the standard normal distribution function, and $\Phi^{-1}(\cdot)$ is its inverse. When $\partial\mathcal{H}$ is a smooth surface, these three p -values are approximately unbiased in the sense that the bias in (1.2) reduces to zero asymptotically as $n \rightarrow \infty$. $\hat{\alpha}(y)$ is said to be k -th order accurate if the bias is $O(n^{-k/2})$; $\hat{\alpha}_1(y)$ is first-order accurate, the two-level bootstrap is second-order accurate, and $\hat{\alpha}_2(y)$ is third-order accurate. This argument is generalized to exponential family of distributions in Shimodaira (2004) by introducing a multistep-multiscale bootstrap method.

Smoothness of the boundary surface $\partial\mathcal{H}$ is essential for justification of the above result. For convenience, we write $y = (u, v)$, $u = (u_1, \dots, u_m)$ and $\mu = (\theta, \lambda)$, $\theta = (\theta_1, \dots, \theta_m)$. (1.1) and (1.3) are now

$$(1.6) \quad U \sim N_m(\theta, I_m), \quad V \sim N(\lambda, 1), \quad U^* \sim N_m(u, \sigma^2 I_m), \quad V^* \sim N(v, \sigma^2).$$

By taking the coordinates properly, a hypothesis region with smooth boundary can be expressed asymptotically as

$$\mathcal{H} = \{(\theta, \lambda) \mid \lambda \leq -h(\theta)\}$$

in a neighborhood of u . $\partial\mathcal{H}$ is specified by $\lambda = -h(\theta)$. The Taylor series of $h(u^*)$ around u is written as

$$\begin{aligned} h(u^*) &= h(u) + \frac{1}{2} \sum_{a,b=1}^m \frac{\partial^2 h}{\partial u_a \partial u_b} \Big|_u (u_a^* - u_a)(u_b^* - u_b) \\ &\quad + \frac{1}{6} \sum_{a,b,c=1}^m \frac{\partial^3 h}{\partial u_a \partial u_b \partial u_c} \Big|_u (u_a^* - u_a)(u_b^* - u_b)(u_c^* - u_c) + \dots \end{aligned}$$

The asymptotic order of coefficients is $\partial^k h / \partial u^k = O(n^{-(k-1)/2})$, $k \geq 2$, because \mathcal{H} is magnified by the rescaling factor \sqrt{n} to keep the variance in (1.1) constant; see the argument below eq. (2.12) of Efron and Tibshirani (1998). The surface becomes flat at the limit $n \rightarrow \infty$, in which case $\tilde{\alpha}_\sigma(y) = 1 - \Phi((v + h(u))/\sigma)$ and an exactly unbiased p -value is given by $\hat{\alpha}(y) = 1 - \Phi(v + h(u)) = \tilde{\alpha}_1(y)$. When the surface is curved, the two-level bootstrap adjusts the curvature term $\partial^2 h / \partial u^2 = O(n^{-1/2})$, and the multiscale bootstrap adjusts up to $\partial^3 h / \partial u^3 = O(n^{-1})$ term.

In this paper, we remove the restriction of smoothness by allowing singularities of surfaces while keeping the normality assumption. A simple example is a circular cone; $h(u) = a\|u\|$ is singular at the vertex $u = 0$. The asymptotic theory based on smooth surfaces may lead to seriously biased p -values when μ is close to the vertex, while the singularity can be ignored when μ is far from the vertex. In practice, \mathcal{H} is often a cone at least approximately in a neighborhood of singular points. This is exactly the case for the selection problem of Section 3, where \mathcal{H} is a polyhedral convex cone. More generally, a polyhedral \mathcal{H} is derived from linear inequalities on μ ; it can be regarded as a cone asymptotically by magnification centered at a singular point. It is noted that the shape of cone does not change by the magnification, e.g., a of the circular cone is constant as $n \rightarrow \infty$, a very distinctive feature of cones compared with smooth surfaces.

It was the key to success in the asymptotic theory that smooth surfaces become flat as $n \rightarrow \infty$. In reality, n is finite and the deviation from the flat surface is adjusted by

taking account of the first few terms in the Taylor series. We apply this same idea to singular surfaces. For example, the circular cone $h(u) = a\|u\|$ is flat if $a = 0$, and thus we may develop an asymptotic theory in terms of $a \rightarrow 0$ instead of $n \rightarrow \infty$. We consider the power series expansion in terms of a , and the deviation from the flat surface is adjusted by taking account of first few terms. In this paper, we discuss only up to $O(a)$ term ignoring $O(a^2)$. This linear theory for singular surfaces calculates, if compared with the asymptotic theory for smooth surfaces, up to $O(n^{-1/2})$ term ignoring $O(n^{-1})$. For singular surfaces, including smooth surfaces as special cases, we introduce a notion of *nearly flat surfaces* as follows. First we accept the idea that p -value calculation should depend only on the shape of \mathcal{H} in the neighborhood of y , and that the influence of $h(u^*)$ approaches zero as $\|u^* - u\| \rightarrow \infty$. This property of $\hat{\alpha}(y)$, which we call *locality*, is inevitable for p -values based only on $\tilde{\alpha}_\sigma(y)$. Without losing generality, we assume that y and μ are not very far from the origin, and that $h(u^*)$, $\|u^*\| > L$, can be ignored for sufficiently large L . More specifically, we define

$$h_L(u) = h(0) + (h(u) - h(0))e^{-\|u\|^2/2L^2},$$

and ignore the difference $|h_L(u) - h(u)|$. For a technical reason, we assume that $h_L(u) - h_L(0)$ is absolutely integrable;

$$\int |h_L(u) - h_L(0)| du < \infty,$$

where the integration is over m -dimensional space. Then, we say $h(u)$ is nearly flat if

$$\Delta h = \sup_u |h_L(u) - h_L(0)|$$

is sufficiently small. In our linear theory, calculation is made only up to $O(\Delta h)$ term ignoring $O((\Delta h)^2)$, and equality up to $O(\Delta h)$ is indicated by “ \approx ”.

In Section 2, we will define a class of approximately unbiased p -values $\hat{\alpha}_k(y)$, $k \geq 1$, and its generalization. The bootstrap probability and the bias-corrected p -value of Shimodaira (2002) correspond to $k = 1$ and 2, respectively. It will be shown in Section 5 that $\hat{\alpha}_k(y)$ adjusts up to $\partial^{2k-1}h/\partial u^{2k-1}$ term, and the bias approaches zero (in the sense of “ \approx ”) even for singular surfaces as $k \rightarrow \infty$. It will be also shown there that the p -value calculated by k bootstrap iterations of Hall (1992) shares the same property. On the other hand, it follows from Lehmann (1952) that assuming the power function to be analytic, unbiased tests, except for the trivial randomized test, do not exist if $\partial\mathcal{H}$ has cone singularities. This apparent inconsistency is explained by the fact that $\hat{\alpha}_k(y)$ does not converge as $k \rightarrow \infty$ for singular surfaces. In other words, the boundary surface of rejection region oscillates wildly for large k , a symptom of inappropriate, or counterintuitive, “Emperor’s new tests”

criticized by Perlman and Wu (1999); this symptom often implies that the rejection region is not monotone in the sense of Lehmann (1952). There is a controversy (Perlman and Wu, 1999, Comments and Rejoinder) over which of unbiasedness and monotonicity is important. Although we are seeking for unbiased tests, I agree to the importance of monotonicity. A problem is that monotonicity is geometrically a global property which is incompatible with locality. A compromise can be made by relaxing the strict use of monotonicity so that it holds only approximately in a neighborhood of each point. It will be seen that a good balance is made by using a finite k .

This paper proceeds as follows. The multiscale bootstrap method for nearly flat surfaces is presented through an illustrative example of phylogenetic inference in Section 2. Properties of testing such as unbiasedness and monotonicity are discussed from geometrical viewpoints via a numerical study of selection problem in Section 3. A formal argument of the linear theory is given for location-scale family of distributions in Section 4, where the Fourier transformation of the boundary surface, instead of the Taylor series, plays an important role. The theory is applied to the normal model in Section 5 to derive the approximately unbiased tests. Technical details are given in Appendix and Shimodaira (2006).

2 Multiscale bootstrap in an illustrative example

2.1 Phylogenetic inference

Phylogeny is the history of evolution, represented as a tree with labeled leaves corresponding to species. It has become a common practice for biologists to analyze DNA sequences for phylogenetic tree inference. We are interested in the branching order of species, called tree topology. The maximum likelihood method has been used widely since Felsenstein (1981); each tree gives a parametric model, and the tree with the largest likelihood value is selected. For illustration, the dataset of Shimodaira and Hasegawa (1999), consisting of $n = 3414$ amino acids for six mammalian species, was reanalyzed by using the same model of evolution as in that paper. Fifteen tree topologies are considered and numbered 1 to 15 in decreasing order of likelihood values; see Table 1 of Shimodaira and Hasegawa (1999) or Table 3 of Shimodaira (2002). The maximum likelihood estimate is tree-1, represented as (((human, (seal, cow)), rabbit), mouse), opossum) indicating rabbit is closer to human than to mouse, while a traditional hypothesis is tree-7, represented as (((human, (rabbit, mouse)), (seal, cow)), opossum) indicating rabbit is closer to mouse than to human. Our null hypothesis is that tree-7 is the true history.

2.2 Multiscale bootstrap resampling

First we performed bootstrap resampling with $n' = n$. The number of bootstrap samples is $B = 10^5$, which may be too large, but for a clear result of illustration. The number of times that tree-7 became the maximum likelihood estimate is $C = 1510$, and thus the bootstrap probability is estimated as $\tilde{\alpha}_1 = 0.0151$. The null hypothesis is then rejected at significance level $\alpha = 0.05$.

We next performed $M = 13$ sets of bootstrap resampling with $n'_i = 13656, 10838, 8602, 6828, 5419, 4301, 3414, 2709, 2150, 1707, 1354, 1075, 853$ for $i = 1, \dots, M$ ($n' = n$ is included). These values are chosen so that $\sigma_i^2 = n/n_i$ is equally spaced in log-scale over the range of $2^{-2}, \dots, 2^2$. The frequency of selecting tree-7 was counted to get $C_i = 20, 57, 140, 308, 582, 971, 1510, 2068, 2713, 3325, 4149, 4496, 5137$, respectively. $\tilde{\alpha}_{\sigma_i}$ are estimated as C_i/B , and \tilde{z}_{σ_i} are plotted in Fig. 1(a).

— Insert Figure 1 Here —

2.3 Model fitting

We will work on a function $\psi(\sigma^2)$, instead of \tilde{z}_σ , defined simply as

$$(2.1) \quad \psi(\sigma^2) = \sigma \tilde{z}_\sigma, \quad \sigma > 0.$$

We consider parametric models of $\psi(\sigma^2)$ to understand how the bootstrap probability depends on σ . The simplest model would be a polynomial with respect to σ^2 ;

$$(2.2) \quad \psi_k(\sigma^2 | \beta) = \sum_{j=0}^{k-1} \beta_j \sigma^{2j},$$

where $\beta = (\beta_0, \dots, \beta_{k-1})$ is a parameter vector. This model turns out to be correct for smooth surfaces. In fact, the linear theory of Section 5 (and Appendix A.1) shows that $\psi(\sigma^2) \approx \psi_k(\sigma^2 | \beta)$ if $h(u)$ is a polynomial of degree $2k - 1$. For $k = 2$, (2.2) is written as $\tilde{z}_\sigma = \beta_0 \sigma^{-1} + \beta_1 \sigma$, which is exactly the model of Shimodaira (2002) with signed distance β_0 and curvature correction term β_1 .

In practice, we estimate $\psi(\sigma^2)$ from observed bootstrap probabilities. Let $\psi(\sigma^2 | \beta)$ be a model of $\psi(\sigma^2)$. Since C_i is distributed as binomial with number of trials B and success probability $\tilde{\alpha}_{\sigma_i} = 1 - \Phi(\psi(\sigma_i^2 | \beta)/\sigma_i)$, the maximum likelihood estimate $\hat{\beta}$ is obtained numerically by maximizing the log-likelihood function

$$\ell(\beta) = \sum_{i=1}^M (C_i \log(\tilde{\alpha}_{\sigma_i}) + (B - C_i) \log(1 - \tilde{\alpha}_{\sigma_i})).$$

For example, the parameter of $\psi_2(\sigma^2|\beta)$ is estimated as

$$\hat{\beta}_0 = 1.7733, \quad \hat{\beta}_1 = 0.3895.$$

The result of model fitting is shown in Fig. 1 for $k = 2, 3, 4$. The goodness of fit is measured by the difference of AIC values between the specified model and an unconstrained binomial model;

$$\text{AIC} = (-2\ell(\hat{\beta}) + 2k) - (-2\hat{\ell} + 2M),$$

where $\hat{\ell} = \sum_{i=1}^M (C_i \log(C_i/B) + (B - C_i) \log(1 - C_i/B))$.

2.4 Approximately unbiased p -values

A surprising consequence of Section 5 is that an approximately unbiased p -value with bias ≈ 0 is expressed, if exists, as

$$(2.3) \quad \hat{\alpha}_\infty = 1 - \Phi(\psi(-1)).$$

In other words, $\hat{\alpha}_\infty$ is obtained from the bootstrap probability with $\sigma^2 = -1$ or $n' = -n$, by extrapolating $\psi(\sigma^2)$ from those of $\sigma^2 > 0$ to that of $\sigma^2 = -1$. By letting $\psi(\sigma^2) = \psi_k(\sigma^2|\beta)$ in (2.3), we define

$$(2.4) \quad \hat{\alpha}_k = 1 - \Phi\left(\sum_{j=0}^{k-1} (-1)^j \beta_j\right),$$

where β is estimated by $\hat{\beta}$ in practice. For $k = 2$, this gives the approximately unbiased p -value of Shimodaira (2002) as

$$(2.5) \quad \hat{\alpha}_2 = 1 - \Phi(\hat{\beta}_0 - \hat{\beta}_1) = 1 - \Phi(1.7733 - 0.3895) = 0.0832,$$

which does not reject the null hypothesis at $\alpha = 0.05$. In Table 1, $\hat{\alpha}_k$ is shown at row ψ_k of column $\hat{\alpha}_{k,1}$ for $k = 1, \dots, 4$. Note that (2.4) differs from the description of Section 1 for $k = 1, 2$. However $\hat{\alpha}_1(y) = 1 - \Phi(\hat{\beta}_0)$ converges to $\tilde{\alpha}_1(y)$ as all $\sigma_i \rightarrow 1$. Similarly $\hat{\alpha}_2(y)$ of (2.5) converges to (1.4).

— Insert Table 1 Here —

2.5 Singular models

We next performed $M = 13$ sets of bootstrap resampling again but with $n'_i = 873984, 346840, 137643, 54624, 21677, 8602, 3414, 1354, 537, 213, 84, 33, 13$ for $i = 1, \dots, M$ so that $\sigma_i^2 = n/n_i$ is equally spaced in log-scale over the range of $2^{-8}, \dots, 2^8$. The frequency

of selecting tree-7 was counted to get $C_i = 0, 0, 0, 0, 1, 146, 1499, 4017, 5942, 6703, 6794, 5827, 5395$, respectively. Except for those of $C_i = 0$, \tilde{z}_{σ_i} are plotted in Fig. 1(b).

The corrected p -value is calculated for $k = 2$, say, as

$$(2.6) \quad \hat{\alpha}_2 = 1 - \Phi(2.6743 - 0.1182) = 0.0053,$$

which rejects the null hypothesis at $\alpha = 0.05$. We observe a large difference in the p -value between (2.5) and (2.6), and also in Table 1 between the narrow σ range ($\sigma_i^2 = 2^{-2}, \dots, 2^2$) and the wide σ range ($\sigma_i^2 = 2^{-8}, \dots, 2^8$). Looking at the curves in Fig. 1, we notice a quite bad fitting of ψ_k , $k = 2, 3, 4$ for the wide σ range, although the fitting improves as k increases. It was mentioned in Shimodaira (2002) that the misspecification of models indicates a ‘‘breakdown of the asymptotic theory’’ and that the corrected p -value should not be used then.

The bad fitting of ψ_k models is explained if $h(u)$ is a singular surface. $\psi(\sigma^2)$ is expanded in the form of

$$(2.7) \quad \psi(\sigma^2) = \sum_{j=0}^{\infty} \beta_j \sigma^{1-j} = \beta_0 \sigma + \beta_1 + \beta_2 \sigma^{-1} + \dots$$

in a neighborhood of the vertex of cones as shown in Appendix A.3. We call $\psi(\sigma^2|\beta)$ a *singular model* if it takes account of such singularities. For example, a simple singular model is defined by

$$(2.8) \quad \psi_s(\sigma^2|\beta_0, \beta_1, \beta_2) = \beta_0 + \beta_1 \sigma^2 (1 + \beta_2 \sigma)^{-1},$$

which includes the first two terms in (2.7) by letting $\beta_2 \rightarrow \infty$ with β_1/β_2 fixed, while it becomes ψ_2 by letting $\beta_2 \rightarrow 0$. The fitting of this model is extremely good in Fig. 1.

Substituting $\sigma^2 = -1$ in (2.8) does not make sense, because $\sigma = \sqrt{-1}$ is an imaginary number. In general, let $\psi(\sigma^2)$ be a smooth function for $\sigma > 0$. We define a class of approximately unbiased p -values from $\psi(\sigma^2)$ by

$$(2.9) \quad \hat{\alpha}_{k,\sigma_0} = 1 - \Phi \left(\sum_{j=0}^{k-1} \frac{(-1 - \sigma_0^2)^j}{j!} \frac{d^j \psi(x)}{dx^j} \Big|_{\sigma_0^2} \right)$$

for $k \geq 1$ and $\sigma_0 > 0$. This is simply a Taylor series of $\psi(\sigma^2)$ around σ_0^2 terminated at the k -th term for calculating $\psi(-1)$. $\hat{\alpha}_{1,\sigma_0} = \tilde{\alpha}_{\sigma_0}$ for $k = 1$; this is the reason why $\tilde{\alpha}_1$ is denoted as $\hat{\alpha}_1$ in Section 1 for notational simplicity. $\hat{\alpha}_{k,\sigma_0}$ reduces to $\hat{\alpha}_{k'}$ if $\psi(\sigma^2) = \psi_{k'}(\sigma^2|\beta)$, $k' \leq k$.

For models $\psi_{k'}$, $k' = 1, \dots, 4$ and ψ_s , explicit formulas of the corrected z -value $\hat{z}_{k,\sigma_0} = \Phi^{-1}(1 - \hat{\alpha}_{k,\sigma_0})$ are given for $k = 1, \dots, k'$ ($k' = 4$ for ψ_s) as follows. Model $\psi_1(\sigma^2|\beta)$;

$\hat{z}_{1,\sigma_0} = \beta_0$. Model $\psi_2(\sigma^2|\beta)$; $\hat{z}_{1,\sigma_0} = \beta_0 + \sigma_0^2\beta_1$, $\hat{z}_{2,\sigma_0} = \beta_0 - \beta_1$. Model $\psi_3(\sigma^2|\beta)$; $\hat{z}_{1,\sigma_0} = \beta_0 + \sigma_0^2\beta_1 + \sigma_0^4\beta_2$, $\hat{z}_{2,\sigma_0} = \beta_0 - \beta_1 - \sigma_0^2(2 + \sigma_0^2)\beta_2$, $\hat{z}_{3,\sigma_0} = \beta_0 - \beta_1 + \beta_2$. Model $\psi_4(\sigma^2|\beta)$; $\hat{z}_{1,\sigma_0} = \beta_0 + \sigma_0^2\beta_1 + \sigma_0^4\beta_2 + \sigma_0^6\beta_3$, $\hat{z}_{2,\sigma_0} = \beta_0 - \beta_1 - \sigma_0^2(2 + \sigma_0^2)\beta_2 - \sigma_0^4(3 + 2\sigma_0^2)\beta_3$, $\hat{z}_{3,\sigma_0} = \beta_0 - \beta_1 + \beta_2 + \sigma_0^2(3 + 3\sigma_0^2 + \sigma_0^4)\beta_3$, $\hat{z}_{4,\sigma_0} = \beta_0 - \beta_1 + \beta_2 - \beta_3$. Model $\psi_s(\sigma^2|\beta)$; $\hat{z}_{1,\sigma_0} = \beta_0 + \beta_1\sigma_0^2(1 + \sigma_0\beta_2)^{-1}$, $\hat{z}_{2,\sigma_0} = \beta_0 - \beta_1(1 + \frac{1}{2}\sigma_0(1 - \sigma_0^2)\beta_2)(1 + \sigma_0\beta_2)^{-2}$, $\hat{z}_{3,\sigma_0} = \beta_0 - \beta_1(1 + \frac{3+18\sigma_0^2-\sigma_0^4}{8\sigma_0}\beta_2 + \frac{1+6\sigma_0^2-3\sigma_0^4}{8}\beta_2^2)(1 + \sigma_0\beta_2)^{-3}$, $\hat{z}_{4,\sigma_0} = \beta_0 - \beta_1(1 + \frac{1+9\sigma_0^2+55\sigma_0^4-\sigma_0^6}{16\sigma_0^3}\beta_2 + \frac{1+5\sigma_0^2+15\sigma_0^4-\sigma_0^6}{4\sigma_0^2}\beta_2^2 + \frac{1+5\sigma_0^2+15\sigma_0^4-5\sigma_0^6}{16\sigma_0}\beta_2^3)(1 + \sigma_0\beta_2)^{-4}$.

For good fitting models, the corrected p -value should not be influenced very much by the choice of σ_i values. This is the case for ψ_s model in Fig. 1. Let us apply (2.9) to $\psi(\sigma^2) = \psi_s(\sigma^2|\hat{\beta})$ with $k = 2, \sigma_0 = 1$. The corrected p -value is written as

$$\hat{\alpha}_{2,1} = 1 - \Phi\left(\hat{\beta}_0 - \frac{\hat{\beta}_1}{(1 + \hat{\beta}_2)^2}\right).$$

The maximum likelihood estimate $(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2) = (1.6374, 0.7858, 0.4643)$ for the narrow σ range gives $\hat{\alpha}_{2,1} = 0.1019$, and $(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2) = (1.6606, 0.7361, 0.4293)$ for the wide σ range gives $\hat{\alpha}_{2,1} = 0.0968$. The difference is negligible here for $k = 2$, and so for $k = 1, \dots, 4$ in Table 1. In addition, the corrected p -value should not be influenced very much by the choice of model itself as far as good fitting models are used. This is the case for $\psi_2, \psi_3, \psi_4, \psi_s$ in the narrow σ range, and for ψ_s in the wide σ range.

2.6 Sampling errors in bootstrap

There are sampling errors caused by the bootstrap resampling in the maximum likelihood estimates as well as the corrected p -values. This error of order $O(\sqrt{MB})$ approaches zero as the total number of bootstrap replicates MB increases, but it is limited by computing resources. For making the error as small as possible, we carefully choose σ_i values, models, and (k, σ_0) of $\hat{\alpha}_{k,\sigma_0}$.

The error will be smaller for wider σ range and also for simpler models (with smaller number of parameters), but the model fitting will become worse then. Considering this trade-off, we may choose the wide σ range and ψ_s model in the phylogeny example. But too wide σ range again inefficient, because C_i or $B - C_i$ can be very small or even zero for some σ_i .

As seen in Table 1, the standard error becomes larger as k increases. Typically, we use $B = 10^4$ in practice, meaning the standard errors are three times larger than those in Table 1, and then $k \leq 4$ is advised. The error will also be large if σ_0 is not covered well in the range of σ_i 's.

Although the above issues are important in practice, the problem vanishes ultimately

by increasing MB . In the next section, we discuss the choice of (k, σ_0) by assuming the correct model and $B = \infty$.

3 Geometry of rejection regions

3.1 Subset selection of three normal populations

Let us consider multiple comparisons of independent normal populations

$$X_i \sim N(\eta_i, 1), \quad i = 1, \dots, m + 2,$$

where we would like to find i of the largest η_i value by selecting a subset of $1, \dots, m + 2$. This problem is equivalent to consider, for each i , the null hypothesis that $\eta_i \geq \max_{j \neq i} \eta_j$; i is included in the subset if the hypothesis is not rejected.

We discuss the case of $i = 1$, $m = 1$; the null hypothesis is

$$(3.1) \quad \eta_2 - \eta_1 \leq 0, \quad \eta_3 - \eta_1 \leq 0.$$

Geometry is discussed easily if we work on transformed variables

$$y_1 = \frac{-x_2 + x_3}{\sqrt{2}}, \quad y_2 = \frac{-2x_1 + x_2 + x_3}{\sqrt{6}}, \quad y_3 = \frac{x_1 + x_2 + x_3}{\sqrt{3}}.$$

Since y_3 is not relevant to (3.1), we consider mean parameters for y_1 and y_2 as

$$\theta = \frac{-\eta_2 + \eta_3}{\sqrt{2}}, \quad \lambda = \frac{-2\eta_1 + \eta_2 + \eta_3}{\sqrt{6}}.$$

Now we consider two independent normal variables

$$Y_1 \sim N(\theta, 1), \quad Y_2 \sim N(\lambda, 1)$$

with the null hypothesis $\lambda \leq -|\theta|/\sqrt{3}$ as shown in Fig. 2(a).

In the following, we will consider several tests to explain properties of the approximately unbiased tests. A similar argument with reviews on subset selection procedures is found in Somerville (1986).

— Insert Figure 2 Here —

3.2 Normal test and multiple comparisons

First we consider Point 1 in Fig. 2, corresponding to

$$x_2 - x_1 = 1.5, \quad x_3 - x_1 = 2.5.$$

A one-sided normal test does not reject $\eta_2 - \eta_1 \leq 0$ at $\alpha = 0.05$ because $x_2 - x_1 \leq 1.645 \times \sqrt{2} = 2.326$, but it does reject $\eta_3 - \eta_1 \leq 0$ because $x_3 - x_1 > 2.326$. A simple-minded use of the normal test, therefore, rejects (3.1) because $\max(x_2 - x_1, x_3 - x_1) > 2.326$. The p -value is calculated as $\hat{\alpha}_{\text{nt}} = 1 - \Phi(2.5/\sqrt{2}) = 0.0385$; nt stands for normal test.

The subset selection procedure of Gupta (1965) is designed to control a type-I error by considering the multiple comparisons problem. The critical constant $1.916 \times \sqrt{2} = 2.710$ is larger than that of the normal test, and (3.1) is not rejected because $\max(x_2 - x_1, x_3 - x_1) \leq 2.710$. The p -value is calculated as $\hat{\alpha}_{\text{mc}} = P(\max(X_2 - X_1, X_3 - X_1) > 2.5 | \eta_1 = \eta_2 = \eta_3) = 0.0686$; mc stands for multiple comparisons.

In phylogenetic inference, versions of these two tests are widely used for comparing log-likelihood values. The test of Kishino and Hasegawa (1989) is a normal test, and that of Shimodaira and Hasegawa (1999) is a multiple comparisons test.

The multiple comparisons test is valid, but often said conservative. For a fixed value of the test statistic, the p -value approaches one as $m \rightarrow \infty$. The p -value is $P(\max_{i=2}^{m+2} X_i - X_1 > 2.5 | \eta_1 = \dots = \eta_{m+2}) = 0.191$ for $m = 8$, and it becomes 0.498 for $m = 98$.

Imagine a situation that $x_2 - x_1 = 2.5$, $x_i - x_1 = -10^{10}$, $i = 3, \dots, 100$. The normal test gives p -value 0.0385 by ignoring extremely small x_i 's. On the other hand, the multiple comparisons test gives 0.498 by considering the least favorable configuration $\eta_1 = \dots = \eta_{100}$, i.e., the vertex of cone, although it is unrealistic for the observed x . It is the multiple comparisons test that is counterintuitive in this case, not the approximately unbiased tests.

3.3 Multiple range subset selection

The multiple range subset selection procedure of Somerville (1984) introduces a series of critical constants in the multiple comparisons. As a consequence, it has the locality property and remedies the counterintuitive behavior. When applied to the three populations case, it first uses critical constant $1.969 \times \sqrt{2} = 2.784$, slightly larger than that of Gupta (1965). It is shown in DuPreez et al. (1985) that the multiple range test controls the type-I error for $m = 1$, but only numerical studies seem available for $m \geq 2$ (Somerville, 1986).

We consider Point 2 in Fig. 2, corresponding to

$$x_2 - x_1 = -2, \quad x_3 - x_1 = 2.5.$$

At the first stage, the multiple range test does not reject (3.1) because $\max(x_2 - x_1, x_3 - x_1) \leq 2.784$. Next we see if any of x_2 or x_3 is small enough to be excluded for further consideration; we ignore x_2 because $|x_2 - x_3| > 2.784$ using the same critical constant as the first stage. At the last stage, (3.1) is rejected because $\max(x_2 - x_1, x_3 - x_1) > 2.326$ by applying the normal test. The p -value is $\hat{\alpha}_{\text{mr}} = 0.0385$; mr stands for multiple range.

Table 2 compares the p -values for the four points of Fig. 2, including those for a concave case discussed later. $\hat{\alpha}_{\text{mr}}$ behaves similarly as $\hat{\alpha}_{\text{mc}}$ for points close to the vertex (Points 1, 3) and $\hat{\alpha}_{\text{mr}}$ behaves similarly as $\hat{\alpha}_{\text{nt}}$ for points far from the vertex (Points 2, 4). This same tendency is observed for the shapes of the rejection regions in Fig. 2, and also for the rejection probabilities in Table 3. In fact, the multiple range test is approximately unbiased in the sense that the rejection probability is equal to the significance level at η with $\eta_1 = \eta_i$ for $i \in I \subset \{2, \dots, m + 2\}$ and $\eta_j = -\infty$ for the rest of j ; these points are regarded locally as a vertex of cone in $|I|$ dimensional space.

Recently, Lehmann et al. (2005) discussed some optimality of monotone step-down procedures; they are to identify which of the inequalities hold. Although the appearance of the multiple range test, which considers only the overall hypothesis, is similar to the step-down procedures, they differ.

— Insert Table 2 Here —

— Insert Table 3 Here —

3.4 Concave cones

Sometimes we are interested in the null hypothesis; $\eta_i \leq \max_{j \neq i} \eta_j$. Its rejection indicates a significant evidence that η_i is the largest. For $i = 1$, $m = 1$, the null hypothesis is

$$(3.2) \quad \eta_2 - \eta_1 \geq 0 \quad \text{or} \quad \eta_3 - \eta_1 \geq 0.$$

The signs of y_2 and λ are reversed so that (3.2) is a lower part in Fig. 2(b). We say a “concave cone” for the complement of a convex cone. The p -value are given in Table 2 for the two points of Fig. 2(b); Point 3 is $x_2 - x_1 = -3$, $x_3 - x_1 = -2$, and Point 4 is $x_2 - x_1 = -6.5$, $x_3 - x_1 = -2$.

We define $\hat{\alpha}_{\text{nt}}$, $\hat{\alpha}_{\text{mc}}$, and $\hat{\alpha}_{\text{mr}}$ in the same way as before, but using tests statistics with their signs reversed. Note, however, that $\hat{\alpha}_{\text{mc}}$ and $\hat{\alpha}_{\text{mr}}$ are nonstandard and do not control the type-I error for the concave case, whereas $\hat{\alpha}_{\text{nt}}$ does so. The approximately unbiased test of Liu and Berger (1995), which controls the type-I error, is constructed from the normal test by carefully enlarging the rejection region.

Note that $\hat{\alpha}_{k,\sigma_0}$ for (3.2) is given by $1 - \hat{\alpha}_{k,\sigma_0}$ using $\hat{\alpha}_{k,\sigma_0}$ for (3.1), because the normal density is symmetric and $1 - \Phi(x) = \Phi(-x)$ holds. This applies for any hypothesis region obtained by the complement of another region.

3.5 Approximately unbiased tests

We calculated p -values $\hat{\alpha}_{k,1}$, $k = 1, \dots, 4$ from (2.9). Bootstrap resampling is not used, and $\psi(\sigma^2)$ is obtained directly by numerical integration from (2.1). As expected, the rejection probability approaches α as k increases, but the convergence is slow at the vertex ($\delta = 0$), particularly for the concave case. Overall, $\hat{\alpha}_{k,1}$ with $k = 2, 3, 4$ behaves similarly as $\hat{\alpha}_{\text{mr}}$.

Their rejection regions have a “dent” near the vertex of the convex cone, and a “bump” for the concave case. This common feature for tests attempting unbiasedness violates the monotonicity of rejection regions. A rejection region \mathcal{R} is said monotone for testing (3.1), if $x' \in \mathcal{R}$ is implied by $x \in \mathcal{R}$ and $x'_i - x'_1 \geq x_i - x_1$. For testing (3.2) the inequalities are reversed. In terms of p -values, $\hat{\alpha}$ is monotone if $\hat{\alpha}(x) \geq \hat{\alpha}(x')$. In testing (3.1), for example, Point 1 is a stronger evidence than Point 2. However $\hat{\alpha}_{\text{mr}}$, $\hat{\alpha}_{k,1}$, $k = 2, 3, 4$ claim the opposite, and thus they are not monotone, whereas $\hat{\alpha}_{\text{nt}}$, $\hat{\alpha}_{\text{mc}}$, and $\hat{\alpha}_{1,1}$ are monotone.

A measure of the amount of nonmonotonicity may be defined by

$$\gamma_\epsilon(x) = \sup_{x' \succeq x, \|x' - x\| < \epsilon} \hat{\alpha}(x') - \hat{\alpha}(x),$$

where \succeq denotes the partial ordering of monotonicity. For monotone $\hat{\alpha}$, $\gamma_\epsilon(x) \leq 0$ for all $\epsilon > 0$ at any point x . A large positive $\gamma_\epsilon(x)$, or $\gamma(x) = \lim_{\epsilon \rightarrow 0} \gamma_\epsilon(x)/\epsilon$, results in $\partial\mathcal{R}$ oscillating wildly, whereas $\partial\mathcal{R}$ extends nearly parallel to $\partial\mathcal{H}$ for monotone tests. As seen in Fig. 2, the violation of monotonicity is minor for $\hat{\alpha}_{k,1}$, $k = 2, 3$, although it becomes serious as k increases.

We see that unbiasedness improves as k increases but monotonicity becomes worse then. In Section 5, we see that unbiasedness improves slightly as σ_0 approaches zero but monotonicity becomes much worse then. It seems that a sensible choice is between $2 \leq k \leq 4$ with $\sigma_0 = 1$. Further experience is needed for a proper choice of (k, σ_0) in applications.

Note that the nonmonotonicity criticism in literature has been made mostly to the concave case; Perlman and Wu (2003, Fig. 11) proposed a likelihood ratio test for the convex case qualitatively similar to ours, and their Remark 6.1 says “monotonicity is not necessarily relevant for testing problems with multivariate one-sided hypotheses.”

4 Linear theory of location-scale family

4.1 Bootstrap probability

Let us consider a location-scale family

$$(4.1) \quad U \sim f(u - \theta), \quad V \sim g(v - \lambda), \quad U^* \sim f_\sigma(u^* - u), \quad V^* \sim g_\sigma(v^* - v)$$

as a generalization of (1.6). $f(\cdot)$ and $g(\cdot)$ are arbitrary densities, and $f_\sigma(u) = f(u/\sigma)/\sigma^m$ and $g_\sigma(v) = g(v/\sigma)/\sigma$. The bootstrap probability is written as

$$\tilde{\alpha}_\sigma(u, v) = P_\sigma(V^* \leq -h(U^*) \mid u, v) = E_\sigma \left(G \left(\frac{-v - h(U^*)}{\sigma} \right) \mid u \right),$$

where $G(x) = \int_{-\infty}^x g(v) dv$ and $E_\sigma(\cdot)$ denotes the expectation at scale σ . Applying the Taylor series around $a = -v - h(u)$ twice, we obtain $\tilde{\alpha}_\sigma(u, v) \approx G\left(\frac{a}{\sigma}\right) - g_\sigma(a)E_\sigma(h(U^*) - h(u)|u) \approx G\left(\frac{a}{\sigma} - \frac{1}{\sigma}E_\sigma(h(U^*) - h(u)|u)\right)$ by ignoring $O((\Delta h)^2)$ terms. As a result, we have exchanged the order of the two operators $G(\cdot)$ and $E_\sigma(\cdot|u)$ to get

$$(4.2) \quad \tilde{\alpha}_\sigma(u, v) \approx G \left(-\frac{v + E_\sigma(h(U^*)|u)}{\sigma} \right).$$

We define a modification of bootstrap z -value as $\tilde{z}_\sigma = -G^{-1}(\tilde{\alpha}_\sigma)$ via $G(\cdot)$ instead of $\Phi(\cdot)$. Then (4.2) is rewritten as

$$(4.3) \quad \sigma \tilde{z}_\sigma(u, v) \approx v + E_\sigma(h(U^*)|u).$$

4.2 Unbiased surfaces

We consider a rejection region of the form

$$\mathcal{R} = \{(u, v) \mid v > -r(u)\}.$$

The dependence on α is implicit in the notation. We assume that $\partial\mathcal{R}$ is nearly flat with $\Delta r = O(\Delta h)$ and $r(u)$ satisfies the same conditions as $h(u)$ does. Then it follows from (4.2) that the acceptance probability is written as

$$(4.4) \quad P(V \leq -r(U) \mid \theta, \lambda) \approx G(-\lambda - E(r(U) \mid \theta))$$

by replacing $u, v, h, U^*, V^*, \sigma$ in $P_\sigma(V^* \leq -h(U^*) \mid u, v)$ with $\theta, \lambda, r, U, V, 1$, respectively.

Let us consider a region of the form

$$\mathcal{S} = \{(\theta, \lambda) \mid \lambda \leq -s(\theta)\}.$$

Given \mathcal{R} , we define the function $s(\theta)$ so that the rejection probability is equal to α for any $(\theta, \lambda) \in \partial\mathcal{S}$. We call $\partial\mathcal{S}$ an *unbiased surface*, because $\mathcal{S} = \mathcal{H}$ implies unbiasedness of the test. By substituting $1 - \alpha$ for the left side of (4.4), and letting $\lambda = -s(\theta)$ on the right side, we get

$$(4.5) \quad s(\theta) \approx E(r(U)|\theta) + G^{-1}(1 - \alpha).$$

We formally denote the inverse operator of $E(\cdot|\theta)$ as $E^{-1}(\cdot|u)$. Then (4.5) is rewritten as

$$(4.6) \quad r(u) \approx E^{-1}(s(\theta)|u) - G^{-1}(1 - \alpha).$$

Let $\hat{\alpha}(u, v)$ be the p -value associated with \mathcal{R} , meaning that $\hat{\alpha}(u, v) < \alpha$ is equivalent to $v > -r(u)$. By letting $\alpha = \hat{\alpha}(u, v)$, $r(u) = -v$ in (4.6) for $(u, v) \in \partial\mathcal{R}$, we obtain

$$(4.7) \quad \hat{\alpha}(u, v) \approx 1 - G(v + E^{-1}(s(\theta)|u)).$$

We define a modification of corrected z -value as $\hat{z} = G^{-1}(1 - \hat{\alpha})$ via $G(\cdot)$ instead of $\Phi(\cdot)$. Then, (4.7) is rewritten as

$$(4.8) \quad \hat{z}(u, v) \approx v + E^{-1}(s(\theta)|u).$$

4.3 Fourier transformation

Let \mathcal{F} be the Fourier transformation operator

$$\mathcal{F}h(\omega) = \int e^{-i\omega \cdot u} h(u) du,$$

where $\omega = (\omega_1, \dots, \omega_m)$ is an angular frequency vector, $\omega \cdot u = \sum_{j=1}^m \omega_j u_j$ is the inner product, $i = \sqrt{-1}$ is the imaginary unit, and integrals are over m dimensional space otherwise stated. The inverse operator \mathcal{F}^{-1} is given as

$$h(u) = \mathcal{F}^{-1}(\mathcal{F}h(\omega))(u) = \frac{1}{(2\pi)^m} \int e^{i\omega \cdot u} \mathcal{F}h(\omega) d\omega.$$

Since $h(u)$ must be absolutely integrable for justifying the transformation, we in fact work on $h_L(u) - h_L(0)$, and formally denote $\mathcal{F}h$ for a nearly flat $h(u)$.

Here we overview properties of the Fourier transformation. Let $(Eh)(\theta)$ denote the expectation operator defined by

$$E(h(U)|\theta) = \int h(u) f(u - \theta) du.$$

Since it is a convolution, we can write

$$(4.9) \quad (\mathcal{F}Eh)(\omega) = (\mathcal{F}h)(\omega) (\mathcal{F}f)(-\omega),$$

and its inverse becomes

$$(4.10) \quad (\mathcal{F}E^{-1}h)(\omega) = (\mathcal{F}h)(\omega)/(\mathcal{F}f)(-\omega).$$

The scaling-law of density is

$$(4.11) \quad (\mathcal{F}f_\sigma)(\omega) = (\mathcal{F}f)(\sigma\omega),$$

and it follows from (4.9) and (4.11) that

$$(4.12) \quad (\mathcal{F}E_\sigma h)(\omega) = (\mathcal{F}h)(\omega) (\mathcal{F}f)(-\sigma\omega).$$

Results of Sections 4.1 and 4.2 can be written down explicitly via the Fourier transformation. For notational brevity, we write

$$H(\omega) = \frac{1}{(2\pi)^m} \mathcal{F}h(\omega), \quad S(\omega) = \frac{1}{(2\pi)^m} \mathcal{F}s(\omega), \quad F(\omega) = \mathcal{F}f(-\omega).$$

Then applying (4.12) to (4.3) gives

$$(4.13) \quad \sigma \tilde{z}_\sigma(u, v) \approx v + \int e^{i\omega \cdot u} H(\omega) F(\sigma\omega) d\omega,$$

and applying (4.10) to (4.8) gives

$$(4.14) \quad \hat{z}(u, v) \approx v + \int e^{i\omega \cdot u} \frac{S(\omega)}{F(\omega)} d\omega.$$

An approximately unbiased z -value with bias ≈ 0 is expressed, if exists, as (4.14) by letting $S(\omega) = H(\omega)$; we denote it as $\hat{z}_\infty(u, v)$, and the corresponding p -value as $\hat{\alpha}_\infty(u, v) = 1 - G(\hat{z}_\infty(u, v))$.

These two formulas provide intuitive interpretations. Calculation of $\tilde{\alpha}_\sigma$ is equivalent to applying the linear filter $F(\sigma\omega)$ to $\partial\mathcal{H}$, which is often a low-pass filter. Calculation of $\hat{\alpha}_\infty$ is equivalent to applying the inverse filter $1/F(\omega)$ to $\partial\mathcal{H}$, which is often an unstable high-pass filter. This explains the reason why $\partial\mathcal{R}$ oscillates wildly for ‘‘Emperor’s new tests’’; singularities imply high-frequency components in $H(\omega)$, and they are much emphasized in $H(\omega)/F(\omega)$. In fact, we will see $H(\omega)/F(\omega) \rightarrow \infty$ as $\|\omega\| \rightarrow \infty$ for singular surfaces in Section 5.

4.4 Bootstrap iteration

We consider the bootstrap iteration of Hall (1992) for calculation of corrected p -values. Let $\hat{\alpha}_{k+1, \text{bi}}$ be the p -value obtained by applying the bootstrap iteration k times to $\tilde{\alpha}_1$, in

the sense of Efron and Tibshirani (1998); bi stands for bootstrap iteration. In terms of $\hat{z}_{k,\text{bi}} = G^{-1}(1 - \hat{\alpha}_{k,\text{bi}})$, the k -th iteration is written as

$$\hat{z}_{k+1,\text{bi}}(u, v) = G^{-1} \left(P \left(\hat{z}_{k,\text{bi}}(U^*, V^*) \leq \hat{z}_{k,\text{bi}}(u, v) \mid \hat{\theta}(u, v), -h(\hat{\theta}(u, v)) \right) \right)$$

for $k \geq 1$ and $\hat{z}_{1,\text{bi}}(u, v) = \tilde{z}_1(u, v)$, where $\hat{\theta}(u, v)$ is the maximum likelihood estimate of θ under the restriction that $(\theta, \lambda) \in \partial\mathcal{H}$. For the normal case, $(\hat{\theta}, -h(\hat{\theta}))$ is the point on $\partial\mathcal{H}$ closest to (u, v) .

It is shown in Appendix A.4 that, using a function $d_k(u) = O(\Delta h)$, the z -value is in the form of

$$(4.15) \quad \hat{z}_{k,\text{bi}}(u, v) \approx v + d_k(u),$$

and the iteration is

$$(4.16) \quad d_{k+1}(u) \approx d_k(u) - E(d_k(U^*)|u) + h(u)$$

for $k \geq 1$ and $d_1(u) \approx E(h(U^*)|u)$. This result can be written down explicitly via the Fourier transformation. The iteration of $D_k(\omega) = \mathcal{F}d_k(\omega)/(2\pi)^m$ is $D_{k+1}(\omega) \approx (1 - F(\omega))D_k(\omega) + H(\omega)$ for $k \geq 1$ and $D_1(\omega) \approx H(\omega)F(\omega)$. Thus, by induction,

$$(4.17) \quad \begin{aligned} D_k(\omega) &= H(\omega) \frac{1 - (1 + F(\omega))(1 - F(\omega))^k}{F(\omega)} \\ &= H(\omega) \sum_{j=0}^k \frac{k!(k-1-2j)}{(j+1)!(k-j)!} (-F(\omega))^j. \end{aligned}$$

By noting $d_k(u) = \int e^{i\omega \cdot u} D_k(\omega) d\omega$, (4.17) is compared with (4.14). Bootstrap iteration modifies the inverse filter of unbiased test by multiplying the factor $1 - (1 - F(\omega))^k(1 + F(\omega))$. The modified filter is bounded and thus stable, because it includes only $F(\omega)^j$, $j \geq 0$.

5 Normal model

5.1 An approximately unbiased test

The argument of Section 4 is simplified for the normal model of (1.6);

$$G(v) = \Phi(v), \quad F(\omega) = e^{-\frac{\|\omega\|^2}{2}}.$$

The two z -value definitions are $G^{-1}(1 - \alpha) = -G^{-1}(\alpha)$, and the scaling-law of the density is $F(\sigma\omega) = F(\omega)\sigma^2$. The latter implies the main result (2.3), because the inverse filter

$1/F(\omega)$ is now given by $F(\sigma\omega)$ with $\sigma = \sqrt{-1}$. More specifically, (4.13) and (4.14) are now

$$(5.1) \quad \sigma \tilde{z}_\sigma(u, v) \approx v + \int e^{i\omega \cdot u - \sigma^2 \frac{\|\omega\|^2}{2}} H(\omega) d\omega,$$

$$(5.2) \quad \hat{z}(u, v) \approx v + \int e^{i\omega \cdot u + \frac{\|\omega\|^2}{2}} S(\omega) d\omega.$$

Therefore (2.3) follows by noting (5.1) \approx (5.2) if we let $\sigma^2 = -1$ and $S(\omega) = H(\omega)$.

5.2 Smooth surfaces

Assume $h(u)$ is a smooth surface, and define $\psi_\infty(\sigma^2|\beta) = \sum_{j=0}^{\infty} \beta_j \sigma^{2j}$ with coefficients

$$(5.3) \quad \beta_j = \frac{1}{2^j j!} \sum_{j_1 + \dots + j_m = j} \frac{j!}{j_1! \dots j_m!} \frac{\partial^{2j} h}{\partial u_1^{2j_1} \dots \partial u_m^{2j_m}} \Big|_u$$

for $j \geq 1$ and $\beta_0 = v + h(u)$. Considering (4.3) and the series expansion of $E_\sigma(h(U^*)|u)$ in Appendix A.1, we obtain $\psi(\sigma^2) \approx \psi_\infty(\sigma^2|\beta)$, and thus ψ_k is correct up to $O(\Delta h)$ if $h(u)$ is a polynomial of degree $2k - 1$.

On the other hand, we also have

$$(5.4) \quad \beta_j = \frac{1}{2^j j!} \int e^{i\omega \cdot u} (-1)^j \|\omega\|^{2j} H(\omega) d\omega$$

for $j \geq 1$ by substituting $e^{-\sigma^2 \frac{\|\omega\|^2}{2}} = \sum_{j=0}^{\infty} (-\sigma^2 \|\omega\|^2 / 2)^j / j!$ in (5.1) and exchanging the order of summation and integration. Therefore, $(-1)^j \|\omega\|^{2j}$ component in a linear filter, if applied to smooth $\partial\mathcal{H}$, gives $\partial^{2j} h / \partial u^{2j}$.

5.3 Singular surfaces

We consider a simple example of singular surface of the form

$$h(u) = \|u\|^q$$

for a real number $q > 0$. This is singular except for even numbers $q = 2, 4, \dots$, and it is a cone for $q = 1$. First note that, for χ_m^2 , a chisquare random variable of m degrees of freedom, the expected value of $(\chi_m^2)^{q/2}$ is $E(\chi_m^q) = 2^{q/2} \Gamma(\frac{m+q}{2}) / \Gamma(\frac{m}{2})$. Then it is shown in Appendix A.2 that

$$(5.5) \quad E_\sigma(\|U^*\|^q | u) = \sigma^q E(\chi_m^q) {}_1F_1 \left(-\frac{q}{2}, \frac{m}{2}, -\frac{\|u\|^2}{2\sigma^2} \right),$$

where the confluent hypergeometric function is

$${}_1F_1(a, b, z) = 1 + \frac{az}{b} + \frac{a(a+1)z^2}{2b(b+1)} + \cdots = \sum_{j=0}^{\infty} \frac{(a)_j z^j}{j!(b)_j}$$

using the Pochhammer symbol $(a)_j = a(a+1)\cdots(a+j-1)$. Thus, from (4.3), we see that $\psi(\sigma^2) \approx v + \sum_{j=0}^{\infty} \beta_j \|u\|^{2j} \sigma^{q-2j}$ with β_j 's defined from m and q ; this reduces to ψ_k with $k = \frac{q}{2} + 1$ for smooth surfaces, because the summation terminates at $j = q/2$ for even q .

The Fourier transformation of $h_L(u)$ is

$$(5.6) \quad H(\omega) = (2\pi)^{-\frac{m}{2}} L^{m+q} E(\chi_m^q) {}_1F_1\left(\frac{m+q}{2}, \frac{m}{2}, -\frac{L^2\|\omega\|^2}{2}\right).$$

For sufficiently large $x = L^2\|\omega\|^2/2$,

$${}_1F_1\left(\frac{m+q}{2}, \frac{m}{2}, -x\right) = \begin{cases} (-1)^{q/2} \frac{\Gamma(\frac{m}{2})}{\Gamma(\frac{m+q}{2})} e^{-x} x^{q/2} (1 + O(x^{-1})) & q \text{ is even} \\ \frac{\Gamma(\frac{m}{2})}{\Gamma(-\frac{q}{2})} x^{-\frac{m+q}{2}} (1 + O(x^{-1})) & \text{otherwise.} \end{cases}$$

This result is consistent with Section 5.2 for smooth surfaces; the e^{-x} term above makes $|\beta_j| < \infty$ in (5.4). For singular surfaces, however, the integration in (5.4) does not converge for $j \geq (m+q-1)/2$, indicating that \hat{z}_∞ does not exist. We can even say that $H(\omega)/F(\omega) \rightarrow \infty$ as $\|\omega\| \rightarrow \infty$.

— Insert Figure 3 Here —

5.4 Filter representations

We consider a corrected z -value, using a function $J(\omega)$, of the form

$$(5.7) \quad \hat{z}(u, v) \approx v + \int e^{i\omega \cdot u + \frac{\|\omega\|^2}{2}} (1 - J(\omega)) H(\omega) d\omega,$$

which reduces to $\hat{z}_\infty(u, v)$ if $J(\omega) = 0$. The Fourier transformation of the unbiased surface is $S(\omega) = (1 - J(\omega))H(\omega)$; we call $1 - J(\omega)$ as an *unbiased filter*. The linear filter to be applied to $\partial\mathcal{H}$ in (5.7) is $e^{\frac{\|\omega\|^2}{2}}(1 - J(\omega))$, and we call it as a *rejection filter*. The rejection filter calculates not only $\hat{z}(u, v)$ but also $\partial\mathcal{R}$ by $r(u) \approx \hat{z}(u, v) - v - z$. In the following, we call $r(u) + z$ as a *rejection surface*. This is the boundary surface of rejection region at $\alpha = 0.5$, so that it is compared with $h(u)$ easily.

Consider a series of $J(\omega)$, denoted as $J_k(\omega)$, $k \geq 1$. We will see both the multiscale bootstrap and the bootstrap iteration lead to versions of $J_k(\omega)$ with the following two

properties. (i) $J_k(\omega)$ is a polynomial of $\|\omega\|^2$ in which coefficients are zero for $\|\omega\|^{2j}$, $j = 0, \dots, k-1$. Thus $e^{\frac{\|\omega\|^2}{2}}(1 - J_k(\omega))$ is equivalent to $e^{\frac{\|\omega\|^2}{2}}$ up to $\|\omega\|^{2k-2}$ term. Considering (5.4), $J_k(\omega)$ calculates $\hat{z} \approx \hat{z}_\infty$ if $h(u)$ is a polynomial of degree $2k-1$. (ii) $J_k(\omega)$ is bounded for all k and ω , and $J_k(\omega) \rightarrow 0$ as $k \rightarrow \infty$ at each ω . Since $H(\omega) - S(\omega) = J(\omega)H(\omega)$, this implies $s(u) \rightarrow h(u)$, meaning that the test bias reduces zero up to $O(\Delta h)$, although the corrected p -value does not converge for singular surfaces.

5.5 Multiscale bootstrap

Let $J_{k,\sigma_0}(\omega)$ be $J(\omega)$ in (5.7) for $\hat{\alpha}_{k,\sigma_0}$ defined in (2.9). By considering the j -th derivative of (5.1) with respect to σ^2 , we have

$$(5.8) \quad \frac{d^j \psi(\sigma^2)}{d(\sigma^2)^j} \Big|_{\sigma_0^2} \approx \int e^{i\omega \cdot u - \sigma_0^2 \frac{\|\omega\|^2}{2}} \left(-\frac{\|\omega\|^2}{2} \right)^j H(\omega) d\omega.$$

Therefore, the rejection filter is

$$e^{\frac{\|\omega\|^2}{2}} (1 - J_{k,\sigma_0}(\omega)) = \sum_{j=0}^{k-1} \frac{(1 + \sigma_0^2)^j \|\omega\|^{2j}}{2^j j!} e^{-\sigma_0^2 \frac{\|\omega\|^2}{2}}.$$

Using the incomplete gamma function $\gamma(k, z) = \int_0^z t^{k-1} e^{-t} dt$, we have

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

Note that a smaller σ_0 leads to less biased p -value, but $J_{k,\sigma_0}(\omega)$ does not vanish even at $\sigma_0 = 0$.

If we go back to (4.3), the rejection surface is written as

$$(5.9) \quad r(u) + z \approx \sum_{j=0}^{k-1} \frac{(-1 - \sigma_0^2)^j}{j!} \frac{\partial^j E_\sigma(h(U^*) | u)}{\partial(\sigma^2)^j} \Big|_{\sigma_0^2}.$$

Considering (5.8), the unbiased surface $s(u)$ is given by (5.9) but the derivative is evaluated at $\sigma_0^2 + 1$ instead of σ_0^2 . These surfaces are drawn for $h(u) = \|u\|^q$ with $m = 1$, $q = 1$ in Fig. 3 (a, b) using eq. (A.1) in Appendix.

5.6 Bootstrap iteration

Let $J_{k,\text{bi}}(\omega)$ be $J(\omega)$ in (5.7) for $\hat{\alpha}_{k,\text{bi}}$ defined in Section 4.4. We immediately obtain from (4.17) that

$$(5.10) \quad \begin{aligned} J_{k,\text{bi}}(\omega) &= (1 + e^{-\frac{\|\omega\|^2}{2}})(1 - e^{-\frac{\|\omega\|^2}{2}})^k \\ &= (-1)^k k! \sum_{j=k}^{\infty} (S2(j, k) + S2(j+1, k+1)) \frac{(-1)^j \|\omega\|^{2j}}{2^j j!}, \end{aligned}$$

where $S2(j, k) = \sum_{i=0}^k (-1)^{k-i} j^i / i! (k-i)!$ are the Stirling numbers of the second kind. Note that the rejection filter converges to $k-1$ as $\|\omega\| \rightarrow \infty$, while it converges to zero for the multiscale bootstrap. This difference leads to the pointy shape of the rejection surface at $\theta = 0$ observed in Fig. 3(c).

The rejection surface is obtained from (4.17), by noting $F(\omega)^j = F(\sqrt{j}\omega)$, as

$$(5.11) \quad r(u) + z \approx \sum_{j=0}^k \frac{(-1)^j k! (k-1-2j)}{(j+1)! (k-j)!} E_{\sqrt{j}}(h(U^*) | u),$$

where $E_0(h(U^*) | u) = h(u)$ for $j = 0$. The unbiased surface $s(u)$ is given by (5.11) but $E_{\sqrt{j}}(h(U^*) | u)$ is replaced by $E_{\sqrt{j+1}}(h(U^*) | u)$.

A Appendix

A.1 Expected value of smooth surfaces

We consider the expected value of smooth $h(U^*)$ under the normal model. Note that $E_{\sigma}(U_j^{*q} | 0) = \sigma^q (q-1)!! = \sigma^q 2^{q/2} \Gamma(\frac{1+q}{2}) / \Gamma(\frac{1}{2})$ for even q , and $E_{\sigma}(U_j^{*q} | 0) = 0$ for odd q . Thus, if all q_1, \dots, q_m are even, $E_{\sigma}(\prod_{j=1}^m (U_j^* - u_j)^{q_j} | u) = \sigma^q \prod_{j=1}^m (q_j - 1)!!$ with $q = \sum_{j=1}^m q_j$. Otherwise, the expectation is zero. Applying this result to the Taylor series of $h(U^*)$ around u gives $E_{\sigma}(h(U^*) | u) = h(u) + \sum_{j=1}^{\infty} \beta_j \sigma^{2j}$ with β_j defined in (5.3).

A.2 Expected value of surfaces of revolution

We consider $h(u) = \|u\|^q$ for a real $q \geq 0$. Note that $E_{\sigma}(\|U^*\|^q | u) = \sigma^q E(\chi_m^q(\delta^2))$, where $\chi_m^2(\delta^2)$ is a noncentral chisquare random variable with noncentrality $\delta^2 = \|u\|^2 / \sigma^2$. Since $E(\chi_m^q(\delta^2)) = \sum_{j=0}^{\infty} \frac{(\delta^2/2)^j}{j!} e^{-\delta^2/2} E(\chi_{m+2j}^q) = E(\chi_m^q) e^{-\delta^2/2} {}_1F_1\left(\frac{m+q}{2}, \frac{m}{2}, \frac{\delta^2}{2}\right)$, eq. (5.5) follows from the identity $e^z {}_1F_1(a, b, -z) = {}_1F_1(b-a, b, z)$. By differentiating (5.5) with respect to σ^2 , we have

$$(A.1) \quad \frac{\partial^j E_{\sigma}(\|U^*\|^q | u)}{\partial(\sigma^2)^j} = \frac{\sigma^{q-2j} \Gamma(1 + \frac{q}{2})}{\Gamma(1 - j + \frac{q}{2})} E(\chi_m^q) {}_1F_1\left(j - \frac{q}{2}, \frac{m}{2}, -\frac{\|u\|^2}{2\sigma^2}\right).$$

The Fourier transformation of $h_L(u)$ is obtained from

$$\int e^{-i\omega \cdot u} e^{-\frac{\|u\|^2}{2L^2}} \|u\|^q du = e^{-\frac{L^2\omega^2}{2}} (2\pi L^2)^{\frac{m}{2}} E_L(\|U^*\|^q \mid L^2 i\omega).$$

Applying (5.5) to it gives (5.6).

A.3 Expected value of cone surfaces

Let us consider a polar coordinate system $(\|u\|, t) \leftrightarrow u$, where $t = u/\|u\|$ indicates the direction instead of angles. A cone shaped surface is, in general, represented in the form of $h(u) = a(t)\|u\|$, which reduces to the circular cone if $a(t) = a$ is a constant. By noting the independence of $\|U^*\|$ and T^* for $u = 0$, we have

$$\begin{aligned} E_\sigma(h(U^*)|u) &= e^{-\frac{\|u\|^2}{2\sigma^2}} E_\sigma\left(h(U^*)e^{\frac{U^* \cdot u}{\sigma^2}} \mid 0\right) \\ &= e^{-\frac{\|u\|^2}{2\sigma^2}} E_\sigma\left(a(T^*)\|U^*\| \sum_{j=0}^{\infty} (\sigma^{2j} j!)^{-1} \|U^*\|^j (T^* \cdot u)^j \mid 0\right) \\ &= e^{-\frac{\|u\|^2}{2\sigma^2}} \sum_{j=0}^{\infty} (\sigma^{2j} j!)^{-1} E_\sigma(\|U^*\|^{j+1} | 0) E(a(T^*)(T^* \cdot u)^j \mid 0). \end{aligned}$$

Since $E_\sigma(\|U^*\|^{j+1} | 0) = \sigma^{j+1} E(\chi_m^{j+1})$, we find that $E_\sigma(h(U^*)|u)$ is expanded in the series of σ^{1-j} , $j \geq 0$; this proves (2.7). For a cone with $a(-t) = a(t)$, such as the circular cone, the series includes only σ^{1-2j} , $j \geq 0$, since $E(a(T^*)(T^* \cdot u)^j \mid 0) = 0$ for odd j .

A.4 Proof of equations (4.15) and (4.16).

Let $u' = \hat{\theta}(u, v) = u + O(\Delta h)$. Assume (4.15) holds for k . Then $G(\hat{z}_{k+1, \text{bi}}(u, v)) = E\{P(V^* + d_k(U^*) \leq v + d_k(u) \mid U^*, -h(u')) \mid u'\} = E\{G(v + d_k(u) - d_k(U^*) + h(u')) \mid u'\}$. Applying the Taylor series around $a = v + d_k(u) - d_k(u') + h(u')$ twice, it becomes $G(\hat{z}_{k+1, \text{bi}}(u, v)) \approx E\{G(a) + g(a)(-d_k(U^*) + d_k(u')) \mid u'\} = G(a) + g(a)E(-d_k(U^*) + d_k(u') \mid u') \approx G(a + E(-d_k(U^*) + d_k(u') \mid u')) = G(E(v + d_k(u) - d_k(U^*) + h(u') \mid u'))$. Thus $d_{k+1}(u) \approx v + d_k(u) - E(d_k(U^*)|u') + h(u')$. (4.16) is shown by noting $h(u') \approx h(u)$ and $E(d_k(U^*)|u') \approx E(d_k(U^*)|u)$. This also proves (4.15) for $k + 1$.

References

- [1] 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.
- [2] Efron, B. and Tibshirani, R. (1998) The problem of regions. *Ann. Statist.* **26** 1687–1718.
- [3] Efron, B. (1987) Better bootstrap confidence intervals. *J. Amer. Statist. Assoc.* **82** 171–185.
- [4] Efron, B., Halloran, E. and Holmes, S. (1996) Bootstrap confidence levels for phylogenetic trees. *Proc. Natl. Acad. Sci. USA* **93** 13429–13434.
- [5] Felsenstein, J. (1981) Evolutionary trees from DNA sequences: A maximum likelihood approach. *J. Mol. Evol.* **17** 368–376.
- [6] Felsenstein, J. (1985) Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39** 783–791.
- [7] Gupta, S. S. (1965) On some multiple decision (selection and ranking) rules. *Technometrics* **7** 225–245.
- [8] Hall, P. (1992) *The bootstrap and Edgeworth expansion*. Springer, New York.
- [9] 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.
- [10] Lehmann, E. L. (1952) Testing multiparameter hypotheses. *Ann. Math. Statistics* **23** 541–552.
- [11] Lehmann, E. L., Romano, J. P. and Shaffer, J. P. (2005) On optimality of stepdown and stepup multiple test procedures. *Ann. Statist.* **33** 1084–1108.
- [12] Liu, H. and Berger, R. L. (1995) Uniformly more powerful, one-sided tests for hypotheses about linear inequalities. *Ann. Statist.* **23** 55–72.
- [13] Perlman, M. D. and Wu, L. (1999) The emperor's new tests. *Statist. Sci.* **14** 355–381.
- [14] 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.

- [15] Shimodaira, H. (2002) An approximately unbiased test of phylogenetic tree selection. *Systematic Biology* **51** 492–508.
- [16] Shimodaira, H. (2004) Approximately unbiased tests of regions using multistep-multiscale bootstrap resampling. *Ann. Statist.* **32** 2616–2641.
- [17] Shimodaira, H. (2006) Technical details of multiscale bootstrap for singular surfaces. Research Report B-431, Dept. Mathematical and Computing Sciences, Tokyo Institute of Technology, Tokyo.
- [18] Shimodaira, H. and Hasegawa, M. (1999) Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol. Biol. Evol.* **16** 1114–1116.
- [19] Somerville, P. N. (1984) A multiple range subset selection procedure. *J. Statist. Computat. Simul.* **19** 215–226.
- [20] Somerville, P. N. (1986) On the geometry of some subset selection procedures. *Amer. J. Math. Management Sci.* **6** 169–184.

Figures

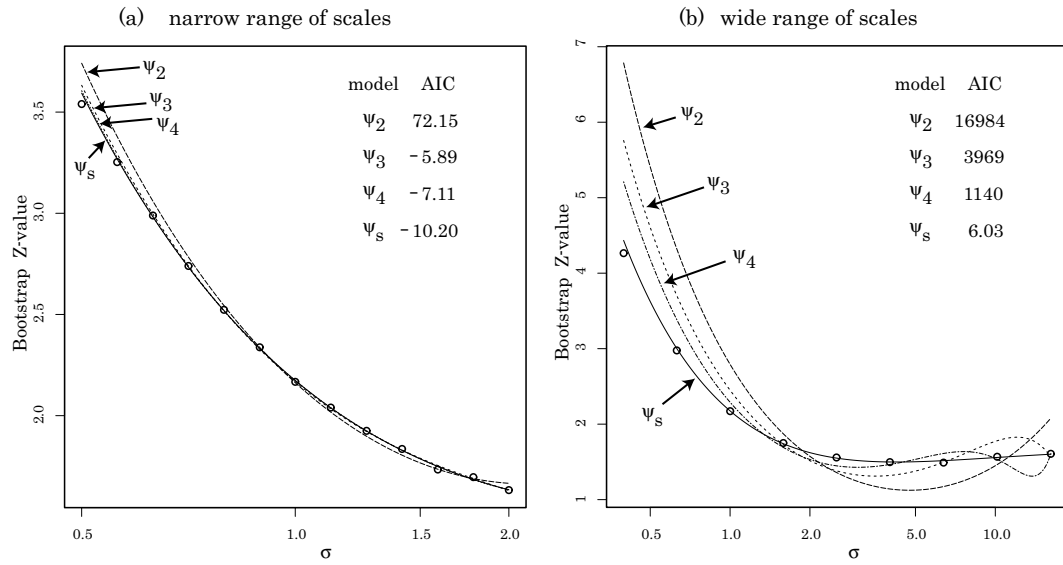


Figure 1: Multiscale bootstrap for the phylogeny example. Bootstrap z -value of eq. (1.5) is plotted against σ in log-scale. Circles are observed z -values of (a) Section 2.2, or (b) Section 2.5. Parametric models ψ_2, ψ_3, ψ_4 of eq. (2.2) and ψ_s of eq. (2.8) are used for curve fitting.

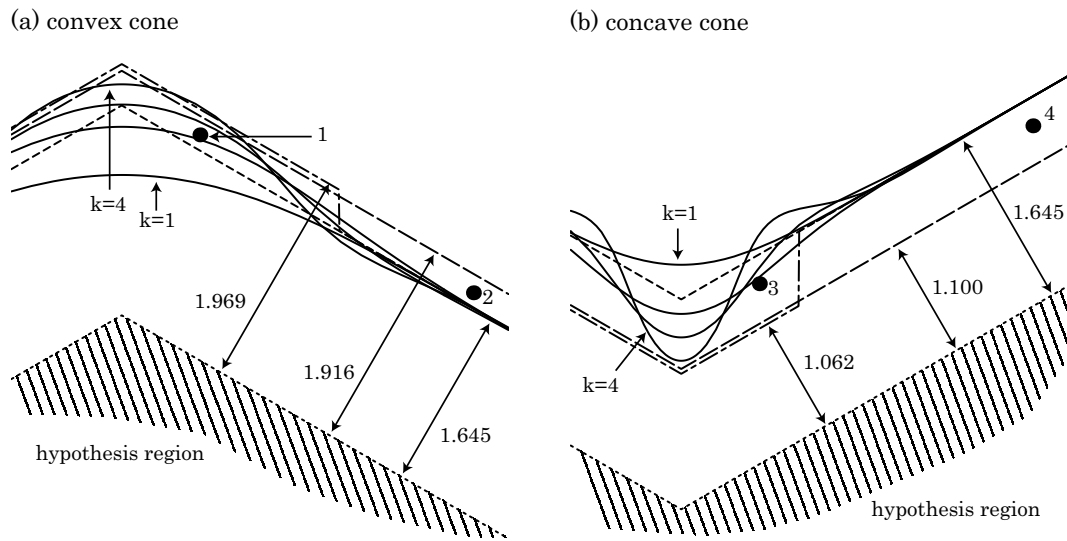


Figure 2: Geometry of the selection example. The axes are (y_1, y_2) or (θ, λ) . The null hypothesis is eq. (3.1) in Panel (a) and eq. (3.2) in Panel (b). The boundary surfaces of rejection regions are shown for the normal test (short dashed lines), the multiple comparisons (long dashed lines), the multiple range procedure (dash-and-dot lines), and $\hat{\alpha}_{k,1}$ for $k = 1, \dots, 4$ (solid lines).

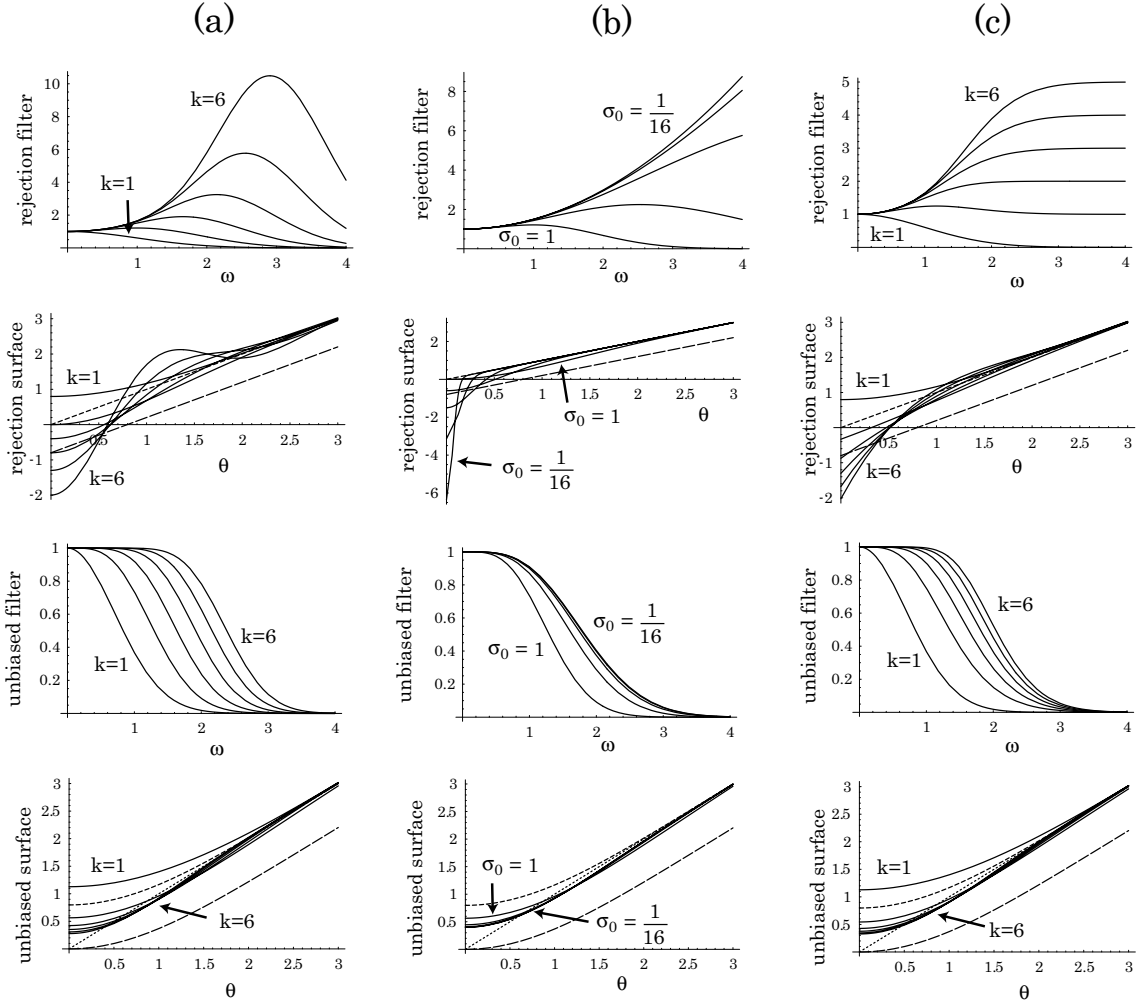


Figure 3: The rejection filter $e^{\frac{\|\varepsilon\|^2}{2}}(1 - J(\omega))$ and the unbiased filter $1 - J(\omega)$ are plotted against $0 \leq \|\omega\| \leq 4$. The rejection surface $r(\theta) + z$ and the unbiased surface $s(\theta)$ are plotted against $0 \leq \theta \leq 3$ for $h(\theta) = |\theta|$ (dotted lines) with $m = 1$. Solid lines are for (a) $\hat{\alpha}_{k,1}$, $k = 1, \dots, 6$, (b) $\hat{\alpha}_{2,2-i}$, $i = 0, \dots, 4$, and (c) $\hat{\alpha}_{k,bi}$, $k = 1, \dots, 6$. Surfaces are also shown for $\hat{\alpha}_{nt}$ (short dashed lines) and $\hat{\alpha}_{mc}$ (long dashed lines). The rejection surface of $\hat{\alpha}_{nt}$ coincides with $h(\theta)$. The unbiased surfaces in (a) and (c) will converge to $h(\theta)$ as $k \rightarrow \infty$, but $s(\theta) > h(\theta)$ near the vertex, indicating that the rejection probability on $\partial\mathcal{H}$ is larger than α for $h(\theta) = a|\theta|$ with $a > 0$ (convex cone), and also it is smaller than α for $a < 0$ (concave cone).

Tables

Table 1: Corrected p -values (in percent) for the phylogeny example. Calculated are $\hat{\alpha}_{k,1}$ of eq. (2.9) for models ψ_1, \dots, ψ_4 , and ψ_s . The bootstrap probability $\tilde{\alpha}_1$ is indicated by *. The values in parentheses are standard errors estimated by the delta method. Only corrected p -values with reasonably small AIC values are justified.

| model | $\hat{\alpha}_{1,1}$ | $\hat{\alpha}_{2,1}$ | $\hat{\alpha}_{3,1}$ | $\hat{\alpha}_{4,1}$ | AIC |
|--|----------------------|----------------------|----------------------|----------------------|--------|
| narrow range ($\sigma^2 = 2^{-2}, \dots, 2^2$) | | | | | |
| | * 1.51 (0.04) | | | | |
| ψ_1 | 0.63 (0.01) | | | | 12708 |
| ψ_2 | 1.53 (0.02) | 8.32 (0.14) | | | 72.15 |
| ψ_3 | 1.51 (0.01) | 10.17 (0.27) | 12.30 (0.54) | | -5.89 |
| ψ_4 | 1.49 (0.02) | 10.34 (0.29) | 14.14 (1.21) | 15.18 (1.82) | -7.11 |
| ψ_s | 1.49 (0.02) | 10.19 (0.26) | 14.24 (0.87) | 17.16 (1.54) | -10.20 |
| wide range ($\sigma^2 = 2^{-8}, \dots, 2^8$) | | | | | |
| | * 1.50 (0.04) | | | | |
| ψ_1 | 0.00 (0.00) | | | | 174279 |
| ψ_2 | 0.26 (0.01) | 0.53 (0.01) | | | 16984 |
| ψ_3 | 0.71 (0.01) | 1.97 (0.04) | 1.98 (0.04) | | 3969 |
| ψ_4 | 1.10 (0.02) | 3.88 (0.07) | 3.94 (0.07) | 3.94 (0.07) | 1140 |
| ψ_s | 1.48 (0.03) | 9.68 (0.25) | 13.25 (0.42) | 15.75 (0.56) | 6.03 |

Table 2: p -values (in percent) for the selection example.

| Point | $\hat{\alpha}_{\text{nt}}$ | $\hat{\alpha}_{\text{mc}}$ | $\hat{\alpha}_{\text{mr}}$ | $\hat{\alpha}_{1,1}$ | $\hat{\alpha}_{2,1}$ | $\hat{\alpha}_{3,1}$ | $\hat{\alpha}_{4,1}$ |
|-------|----------------------------|----------------------------|----------------------------|----------------------|----------------------|----------------------|----------------------|
| 1 | 3.85 | 6.86 | 7.91 | 1.97 | 4.60 | 6.25 | 7.97 |
| 2 | 3.85 | 6.86 | 3.85 | 3.85 | 3.93 | 3.67 | 3.83 |
| 3 | 7.86 | 2.31 | 2.25 | 8.85 | 5.54 | 6.25 | 9.06 |
| 4 | 7.86 | 2.31 | 7.86 | 7.86 | 7.86 | 7.87 | 7.84 |

Table 3: Rejection probabilities (in percent) at significance level $\alpha = 0.05$ for the selection example. These values are obtained by numerical integration. The parameter η is specified by $\eta_1 = \eta_3 = \eta_2 + \delta$, $\delta \geq 0$ so that $\mu \in \partial\mathcal{H}$ with $\theta = \delta/\sqrt{2}$.

| | $\delta = 0$ | $\delta = 1$ | $\delta = 2$ | $\delta = 3$ | $\delta = 4$ | $\delta = 5$ | $\delta = 6$ | $\delta = 7$ | $\delta = 8$ |
|--|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| convex region ($\lambda \leq -\frac{1}{\sqrt{3}} \theta $) | | | | | | | | | |
| $\hat{\alpha}_{\text{nt}}$ | 8.78 | 5.59 | 5.05 | 5.00 | 5.00 | 5.00 | 5.00 | 5.00 | 5.00 |
| $\hat{\alpha}_{\text{mc}}$ | 5.00 | 3.07 | 2.79 | 2.77 | 2.77 | 2.77 | 2.77 | 2.77 | 2.77 |
| $\hat{\alpha}_{\text{mr}}$ | 5.00 | 3.60 | 4.14 | 4.71 | 4.95 | 4.99 | 5.00 | 5.00 | 5.00 |
| $\hat{\alpha}_{1,1}$ | 13.4 | 7.77 | 5.79 | 5.20 | 5.04 | 5.01 | 5.00 | 5.00 | 5.00 |
| $\hat{\alpha}_{2,1}$ | 7.66 | 4.74 | 4.43 | 4.67 | 4.88 | 4.97 | 4.99 | 5.00 | 5.00 |
| $\hat{\alpha}_{3,1}$ | 6.61 | 4.51 | 4.70 | 5.02 | 5.09 | 5.05 | 5.01 | 5.00 | 5.00 |
| $\hat{\alpha}_{4,1}$ | 6.22 | 4.61 | 4.99 | 5.17 | 5.08 | 5.00 | 4.99 | 5.00 | 5.00 |
| concave region ($\lambda \leq \frac{1}{\sqrt{3}} \theta $) | | | | | | | | | |
| $\hat{\alpha}_{\text{nt}}$ | 1.22 | 2.71 | 4.09 | 4.78 | 4.97 | 5.00 | 5.00 | 5.00 | 5.00 |
| $\hat{\alpha}_{\text{mc}}$ | 5.00 | 9.18 | 12.1 | 13.3 | 13.5 | 13.6 | 13.6 | 13.6 | 13.6 |
| $\hat{\alpha}_{\text{mr}}$ | 5.00 | 8.57 | 9.50 | 7.75 | 5.92 | 5.17 | 5.02 | 5.00 | 5.00 |
| $\hat{\alpha}_{1,1}$ | 0.85 | 2.04 | 3.43 | 4.42 | 4.85 | 4.97 | 5.00 | 5.00 | 5.00 |
| $\hat{\alpha}_{2,1}$ | 1.95 | 3.99 | 5.44 | 5.66 | 5.34 | 5.10 | 5.02 | 5.00 | 5.00 |
| $\hat{\alpha}_{3,1}$ | 2.37 | 4.49 | 5.52 | 5.30 | 4.96 | 4.92 | 4.97 | 4.99 | 5.00 |
| $\hat{\alpha}_{4,1}$ | 2.67 | 4.71 | 5.37 | 5.03 | 4.89 | 4.98 | 5.01 | 5.01 | 5.00 |