京都大学
KYOTO UNIVERSITY

RIKEN AIP

DSSV2019 2019/08/13

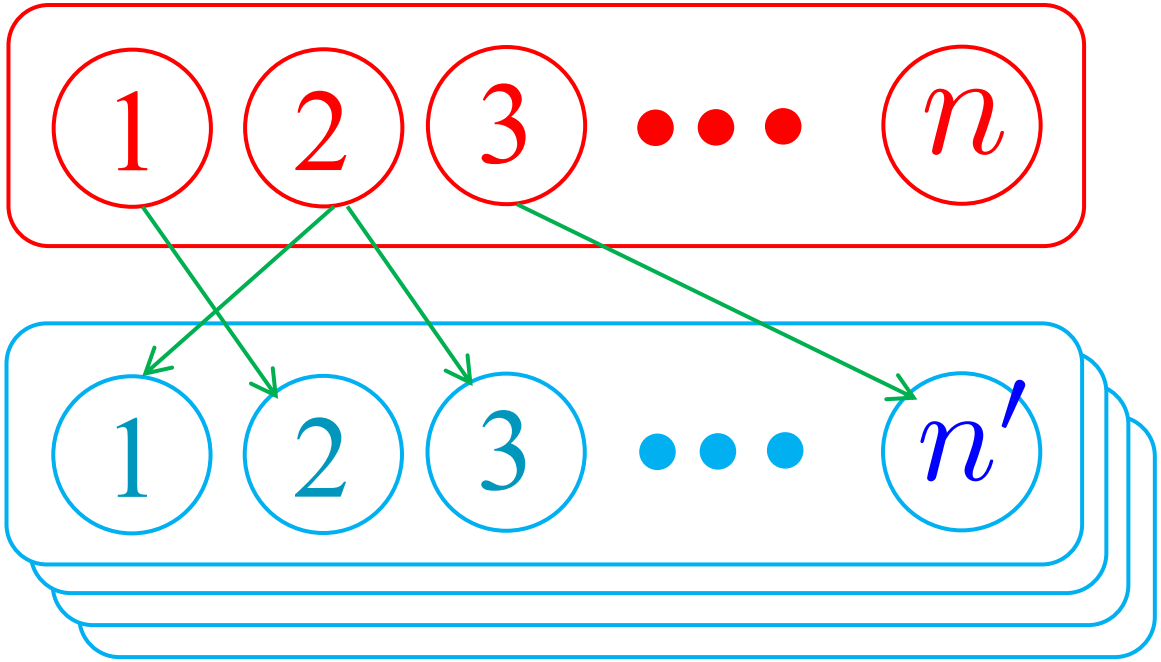
Multiscale bootstrap for selective inference with applications to model selection

Hidetoshi Shimodaira (Kyoto University / RIKEN AIP)

Joint work with

Yoshikazu Terada (Osaka University / RIKEN AIP)

What is Multiscale Bootstrap?



- The bootstrap resampling (Efron 1979) uses n’ = n.
- The ``m-out-of-n’’ bootstrap or subsampling (Politis and Romano 1994, Bickel et al. 1997) typically uses n’<<n for reducing computation. Rescaling the result back to n’ = n.
- Multiscale bootstrap (Shimodaira 2002, 2004, 2008) uses n’ = 0.5n, 1.0n, 1.5n, say, and rescaling the result to n’ = -n (i.e., negative sample size).

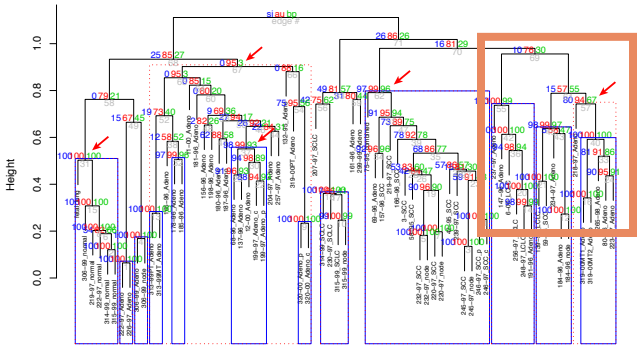
Multiscale bootstrap has been used for p-values of trees and clusters

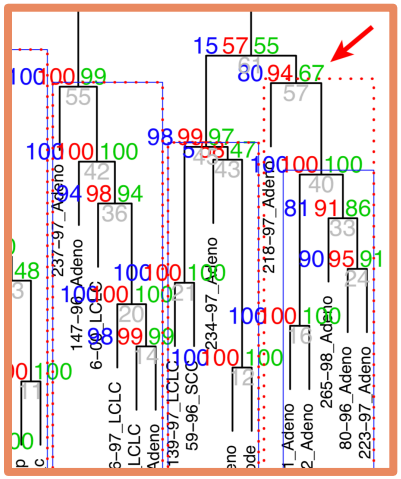
Hidetoshi Shimodaira

FOLLOW

Kyoto University
Verified email at i.kyoto-u.ac.jp - Homepage
Statistics Machine Learning

TITLE	CITED BY	YEAR
Multiple comparisons of log-likelihoods with applications to phylogenetic inference H Shimodaira, M Hasegawa Molecular biology and evolution 16 (8), 1114-1114	3816	1999
An approximately unbiased test of phylogenetic tree selection H Shimodaira Systematic biology 51 (3), 492-508	1997	2002
CONSEL: for assessing the confidence of phylogenetic tree selection H Shimodaira, M Hasegawa Bioinformatics 17 (12), 1246-1247	1919	2001
Pvclust: an R package for assessing the uncertainty in hierarchical clustering R Suzuki, H Shimodaira Bioinformatics 22 (12), 1540-1542	1522	2006
Improving predictive inference under covariate shift by weighting the log-likelihood function H Shimodaira Journal of statistical planning and inference 90 (2), 227-244	958	2000
Mitochondrial genome variation in eastern Asia and the peopling of Japan M Tanaka, VM Cabrera, AM González, JM Larruga, T Takeyasu, N Fuku, ... Genome research 14 (10a), 1832-1850	505	2004
Approximately unbiased tests of regions using multistep-multiscale bootstrap resampling H Shimodaira The Annals of Statistics	300	2004





Today's talk

- Multiscale bootstrap has been used for computing (non-selective) approximately unbiased p-values, and applied to phylogenetic inference and hierarchical clustering
- It is based on a scaling-law of bootstrap probability with geometric view of hypothesis testing
- We are going to extend the theory and method of multiscale bootstrap for selective inference (post-selection inference)

Model Selection

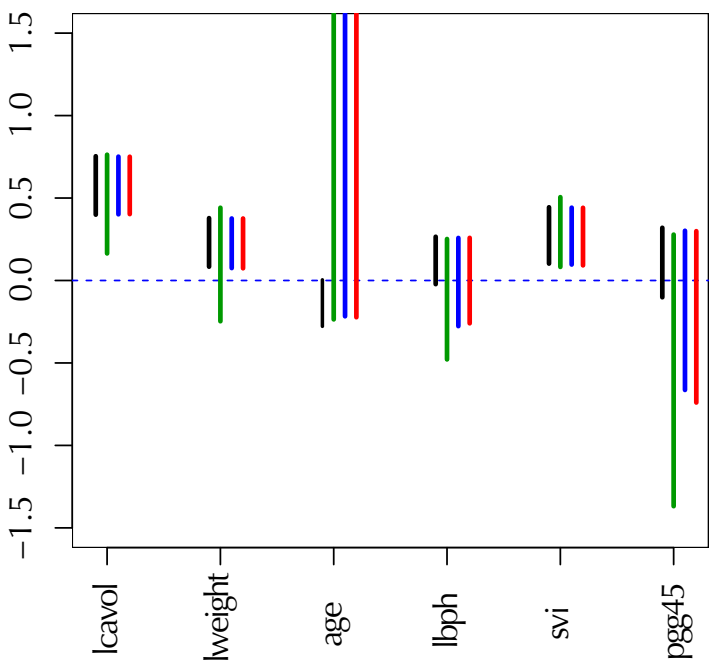
Post-selection inference of variable selection by Lasso

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p + \epsilon$$

6 selected variables (coefficients) out of 8 variables:

$$\{\beta_{\mathrm{lcavol}}, \beta_{\mathrm{lweight}}, \beta_{\mathrm{age}}, \beta_{\mathrm{lbph}}, \beta_{\mathrm{svi}}, \beta_{\mathrm{pgg45}}\}$$

95% Confidence Intervals



Black: non-selective CI

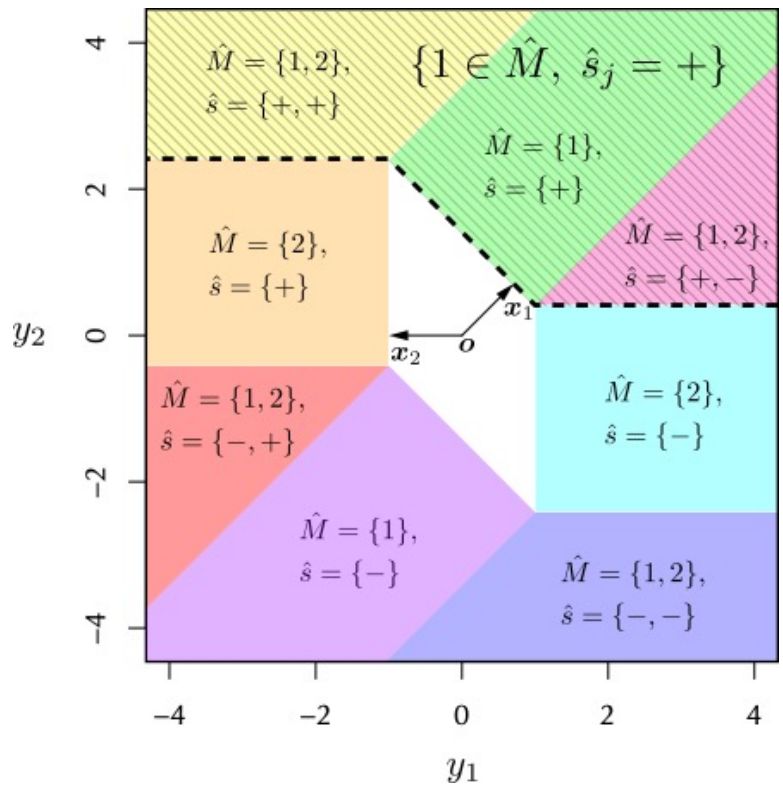
Selective Inference CI

Green: Lee et al. (2016) for model

Blue: Liu et al. (2018) for variable

Red: Terada and Shimodaira (2019) for variable (via multiscale bootstrap)

Selective sets are polyhedra and unions of polyhedra

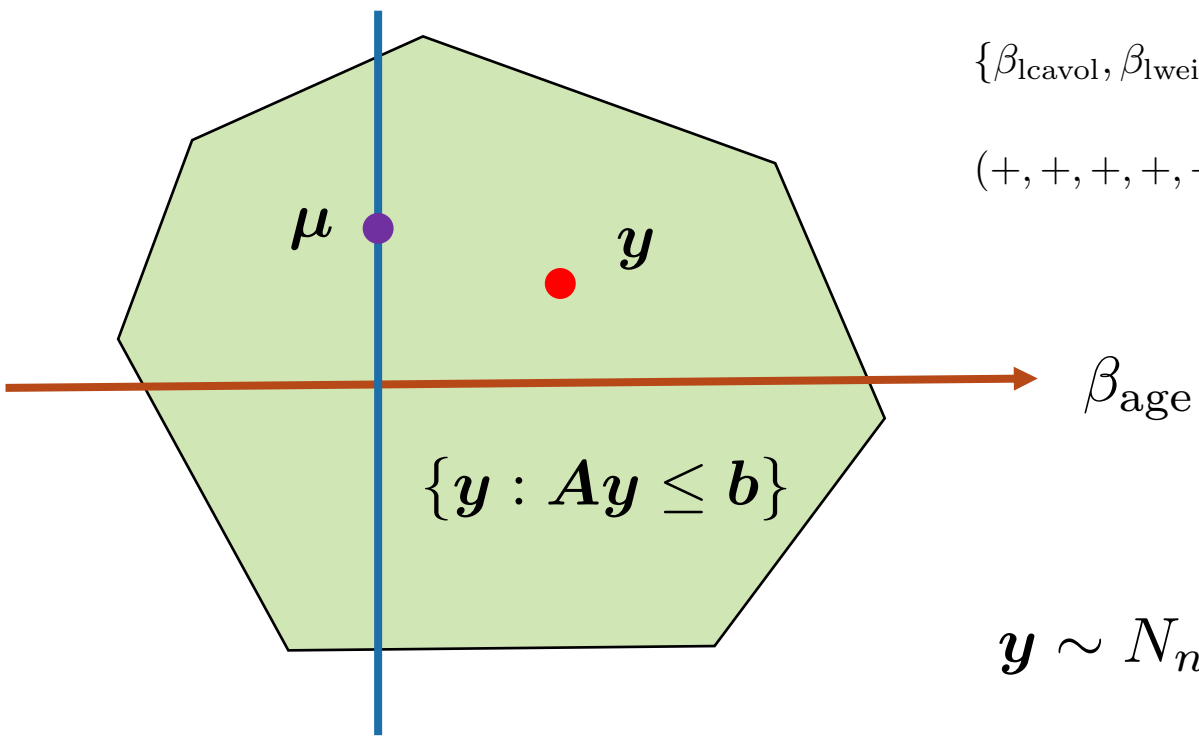


Hypothesis set and selective region

$$\arg\min_{\boldsymbol{\beta}} \|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}\|^2 + \lambda\|\boldsymbol{\beta}\|$$

$$\{\beta_{\text{lcavol}}, \beta_{\text{lweight}}, \beta_{\text{age}}, \beta_{\text{lbph}}, \beta_{\text{svi}}, \beta_{\text{pgg45}}\}$$

$$(+, +, +, +, +, +)$$



$$\{\boldsymbol{\mu} : \beta_{\text{age}} = c\}$$

$$\boldsymbol{y} \sim N_n(\boldsymbol{\mu}, \sigma^2 \boldsymbol{I}_n)$$

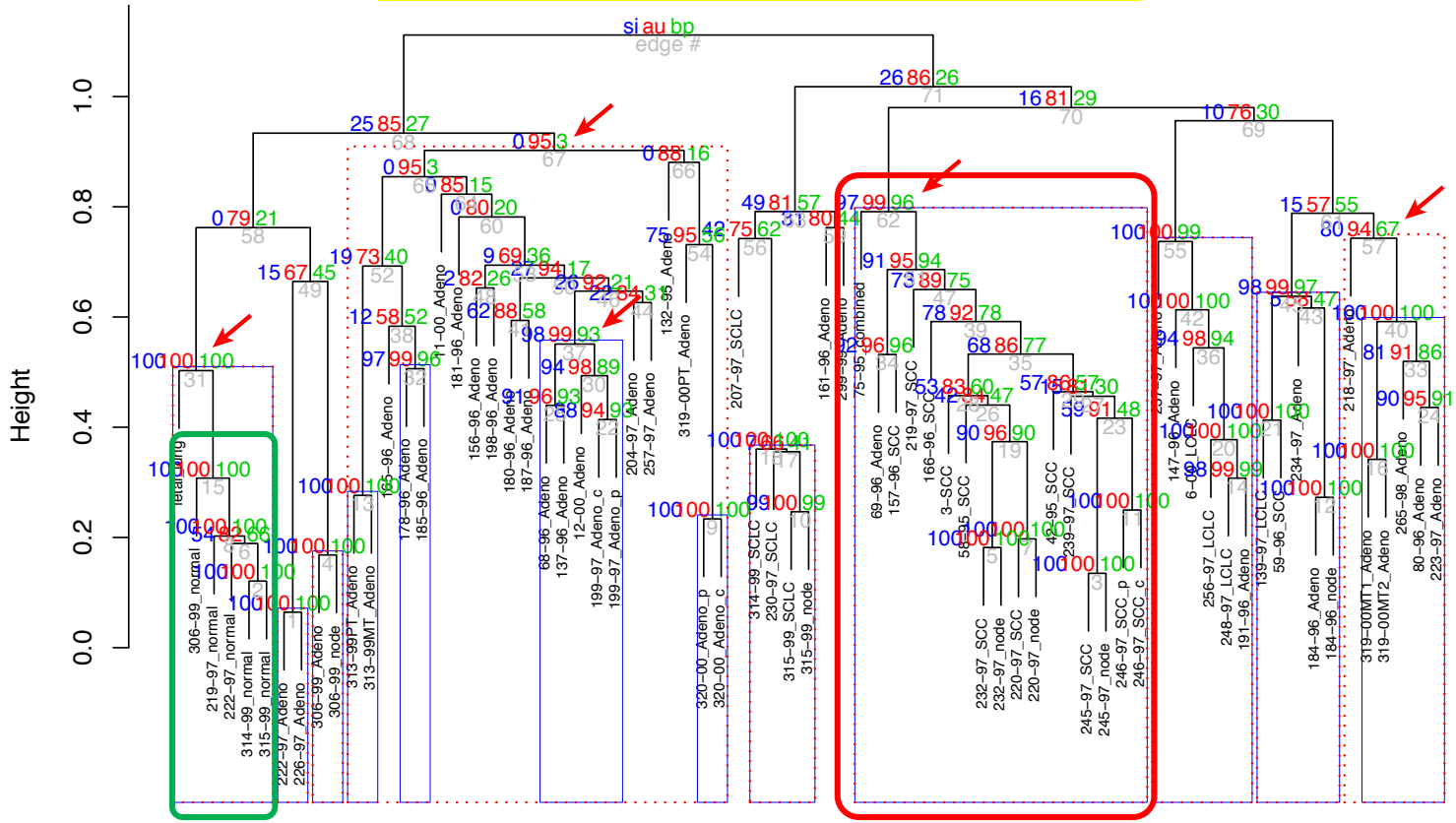
$$\boldsymbol{\beta} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{\mu}$$

Hierarchical Clustering (pvclust)

BP: Bootstrap Probability. (Felsenstein 1985)

AU: Approximately Unbiased. Shimodaira (2002, 2004, 2008) (via multiscale bootstrap)

SI: Selective Inference. Terada and Shimodaira (arXiv 2017) (via multiscale bootstrap)



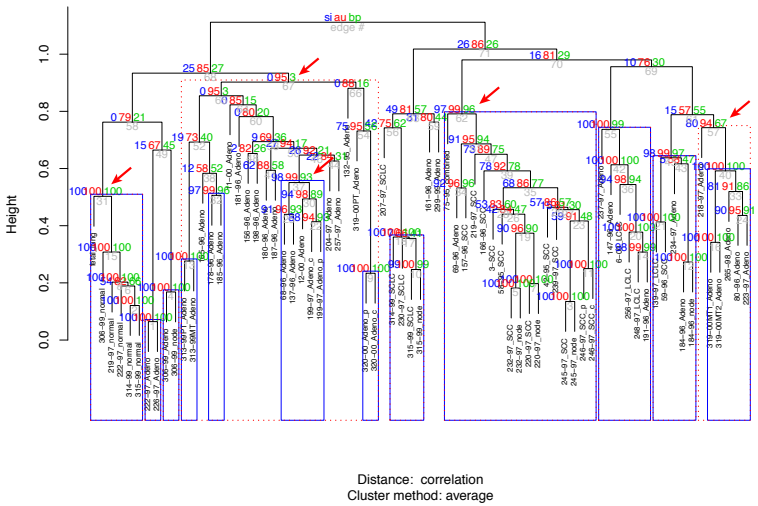
Selected in advance

Selected by clustering

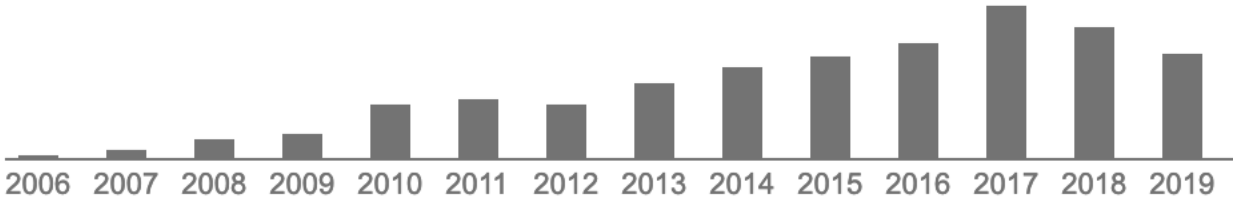
Distance: correlation
Cluster method: average

pvclust is used in biology

- install.packages("pvclust")
- library(pvclust); data(lung)
- result <- pvclust(lung, nboot=1000)
- plot(result)



Cited by 1522



Pvclust: an R package for assessing the uncertainty in hierarchical clustering
R Suzuki, H Shimodaira - Bioinformatics, 2006
Cited by 1522 Related articles All 14 versions

The “Problem of Regions”

Correction Proc. Natl. Acad. Sci. USA 93 (1996) 13429

Evolution. The following article, which appeared in number 14, July 1996, of *Proc. Natl. Acad. Sci. USA* (**93,** 7085–7090) is reprinted in its entirety with the author’s corrections incorporated.

Bootstrap confidence levels for phylogenetic trees

BRADLEY EFRON, ELIZABETH HALLORAN‡, AND SUSAN HOLMES†§

†Department of Statistics, Stanford University, Stanford, CA 94305; and ‡Department of Biostatistics, Rollins School of Public Health, Emory University, Atlanta, GA 30322

Contributed by Bradley Efron, January 26, 1996

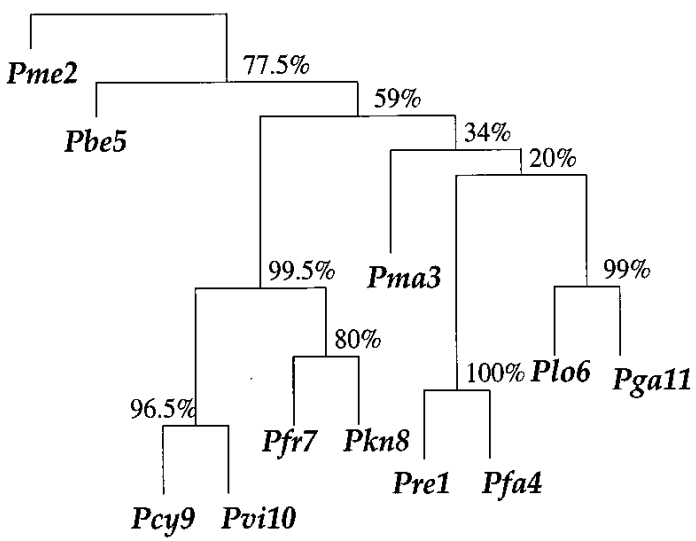


FIG. 2. Phylogenetic tree based on the malaria data matrix; species are numbered as in Fig. 1. The numbers at the branches are confidence values based on Felsenstein’s bootstrap method. $B = 200$ bootstrap replications.

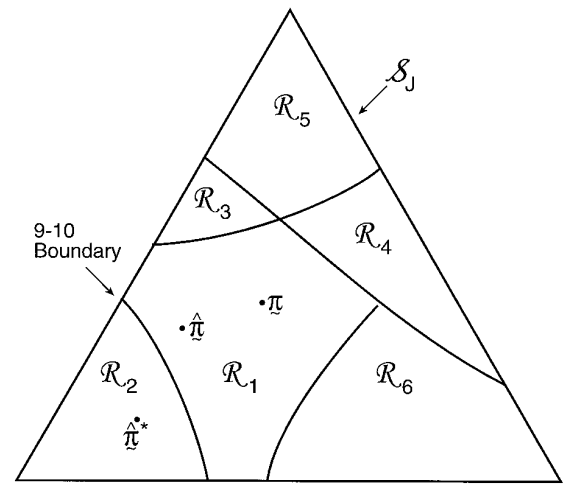


FIG. 3. Schematic diagram of tree estimation; triangle represents the space of all possible π vectors in the multinomial probability model; regions $\mathcal{R}_1, \mathcal{R}_2 \ldots$ correspond to the different possible trees. In the case shown π and $\hat{\pi}$ lie in the same region so TREE = $\widehat{\text{TREE}}$, but $\hat{\pi}^*$ lies in a region where $\widehat{\text{TREE}}^*$ does not have the 9-10 clade.

Efron, Halloran and Holmes (1996)

The “Problem of Regions”

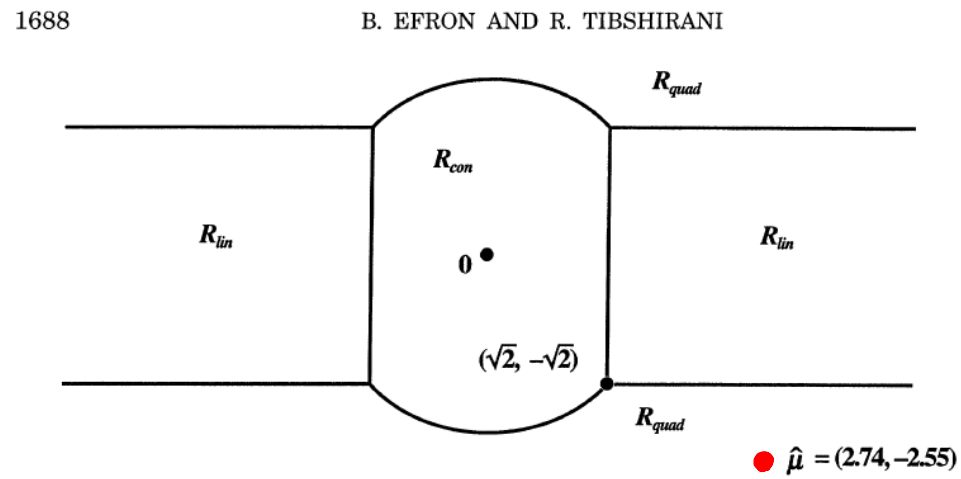


FIG. 1. *An example of the problem of regions: a normally distributed vector $y = \hat{u}$, with covariance I, is observed to lie in the region $\mathscr{R}_{\text{quad}}$. With what confidence can we say that the true expectation vector μ lies in $\mathscr{R}_{\text{quad}}$? This example, which concerns the choice of a polynomial regression model using the C_p criterion, is discussed in Section 5.*

$$y = \beta_0 \text{ (constant)}$$
$$y = \beta_0 + \beta_1 x \text{ (linear)}$$
$$y = \beta_0 + \beta_1 x + \beta_2 x^2 \text{ (quadratic)}$$

Efron and Tibshirani (1998)

The “Problem of Regions”

The regions in multiple comparisons are polyhedral convex cones (considers the least favorable distribution at the vertex)

Table 1
Log-Likelihood Differences and p-Values for the 15 Bifurcating Topologies of the Mammal Data Set

		p-Values				
α	$L_{\hat{\alpha}} - L_{\hat{\alpha}}$	BP	KH	MC	MS	Topology
1.....	0.0	0.583	0.640	0.941	0.944	(((H, (P, B)), O), M, D)
2.....	2.7	0.317	0.360	0.811	0.805	((H, ((P, B), O)), M, D)
3.....	7.4	0.038	0.121	0.577	0.422	(((H, O), (P, B)), M, D)
4.....	17.6	0.012	0.040	0.169	0.203	((H, (P, B)), (O, M), D)
5.....	18.9	0.030	0.066	0.139	0.296	(H, ((P, B), (O, M)), D)
6.....	20.1	0.006	0.050	0.109	0.100	(H, (((P, B), O), M), D)
7.....	20.6	0.011	0.048	0.107	0.248	((H, (O, M)), (P, B), D)
8.....	22.2	0.001	0.032	0.070	0.048	((H, M), ((P, B), O), D)
9.....	25.4	0.000	0.001	0.029	0.013	(((H, (P, B)), M), O, D)
10.....	26.3	0.002	0.018	0.032	0.124	(((H, M), O), (P, B), D)
11.....	28.9	0.000	0.008	0.017	0.069	(((H, O), M), (P, B), D)
12.....	31.6	0.000	0.003	0.006	0.032	(((H, M), (P, B)), O, D)
13.....	31.7	0.000	0.003	0.006	0.035	(H, (((P, B), M), O), D)
14.....	34.7	0.000	0.001	0.002	0.012	((H, O), ((P, B), M), D)
15.....	36.2	0.000	0.000	0.001	0.007	((H, ((P, B), M)), O, D)

Note.—BP is the bootstrap selection probability of Felsenstein (1985) estimated by the RELL method (Kishino, Miyata, and Hasegawa 1990), KH is the p-value of the KH test, MC is the p-value of the multiple-comparisons method with $w_{\alpha \cdot \beta} = 1$, and MS is that with $w_{\alpha \cdot \beta} = \hat{\sigma}_{\alpha \cdot \beta}^{-1}$. See Remark 4 in text. The number of replicates is $N = 10^4$. The labels for the taxa are as follows: H = *Homo sapiens* (human), P = *Phoca vitulina* (harbor seal), B = *Bos taurus* (cow), O = *Oryctolagus cuniculus* (rabbit), M = *Mus musculus* (mouse), and D = *Didelphis virginiana* (opossum).

“Shimodaira-Hasegawa test” in phylogenetics

Shimodaira and Hasegawa (1999)

Multiscale bootstrap is applied to multiple comparisons to improve the power

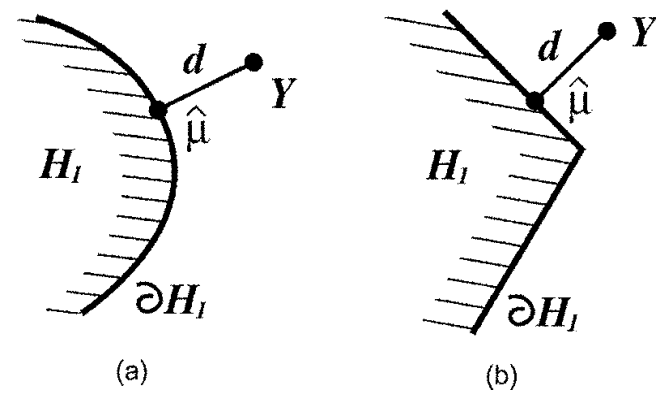
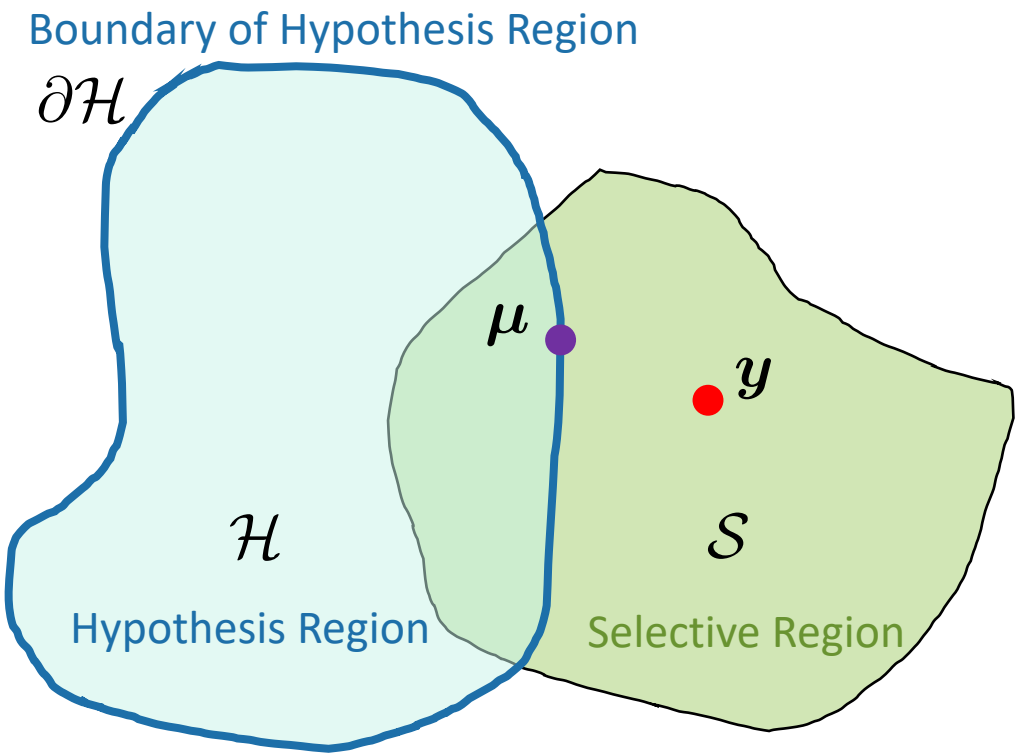


FIGURE 1. Region H_1 with boundary ∂H_1. Y is the data point, $\hat{\mu}$ is the projection, and d is the signed distance. The asymptotic theory behind the AU test assumes (a) a smooth boundary, which approximates (b) a nonsmooth boundary. In fact, the boundary is not smooth for the selection problem, where region H_1 forms a polyhedral convex cone in M-dimensional space. The curvature is zero everywhere but becomes infinite at the vertex and the edges where $\mu_j = \mu_1$ for more than one $j \neq 1$.

Shimodaira (2002)

Hypothesis Testing

A general setting



$$\boldsymbol{y} \sim N_{m+1}(\boldsymbol{\mu}, \boldsymbol{I}_{m+1})$$

If $\boldsymbol{y} \in \mathcal{S}$ then

test null hypothesis $\boldsymbol{\mu} \in \mathcal{H}$

against alternative hypothesis

$\boldsymbol{\mu} \notin \mathcal{H}$

Unbiased p-value:

$$p(\mathcal{H}|\mathcal{S}, \boldsymbol{y}) \mid \boldsymbol{y} \in \mathcal{S}, \boldsymbol{\mu} \in \partial\mathcal{H} \sim \mathrm{Uniform}(0, 1)$$

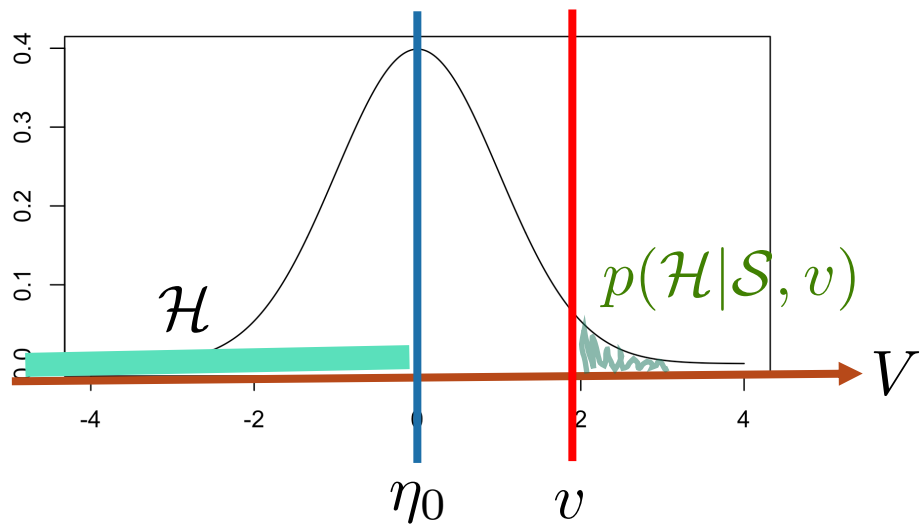
This p-value can be computed by multiscale bootstrap

A trivial case: Normal distribution

$$V \sim N(\eta, 1)$$

$$\mathcal{H} = \{\eta : \eta \leq \eta_0\}$$

$$\mathcal{S} = \{v : -\infty < v < +\infty\}$$



$$p(\mathcal{H}|\mathcal{S}, v) = \mathbb{P}(V > v) = 1 - \Phi(v - \eta_0)$$ one-tailed test

$$p(\mathcal{H}|\mathcal{S}, V) \mid \eta = \eta_0 \sim \text{Uniform}(0, 1)$$

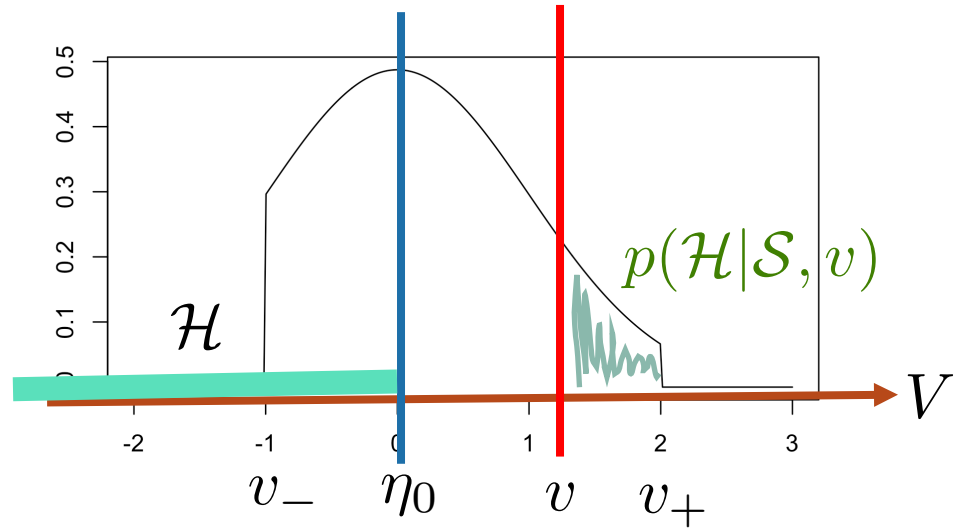
$$\mathbb{P}(p(\mathcal{H}|\mathcal{S}, V) < \alpha) = \mathbb{P}(V - \eta_0 > \Phi^{-1}(1 - \alpha)) = 1 - (1 - \alpha) = \alpha$$

Truncated normal distribution

$$V \sim N(\eta, 1)$$

$$\mathcal{H} = \{\eta : \eta \leq \eta_0\}$$

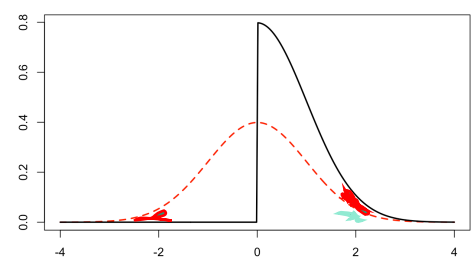
$$\mathcal{S} = \{v : v_- \leq v \leq v_+\}$$



$$\mathbb{P}(V \leq v) = F_\eta^{v_-, v_+}(v) = \frac{\Phi(v - \eta) - \Phi(v_- - \eta)}{\Phi(v_+ - \eta) - \Phi(v_- - \eta)}$$

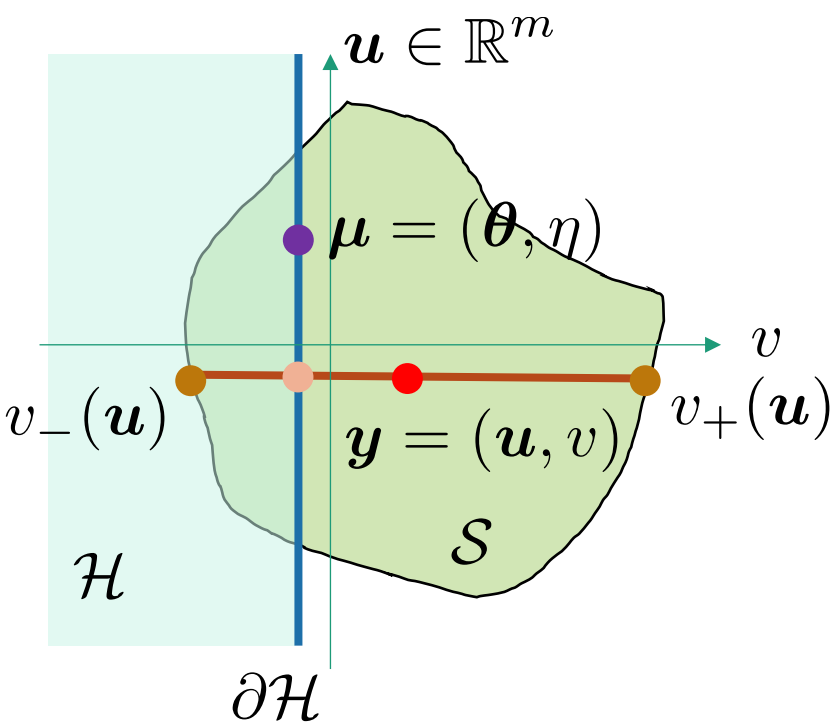
$$p(\mathcal{H}|\mathcal{S}, v) = 1 - F_{\eta_0}^{v_-, v_+}(v)$$

$$p(\mathcal{H}|\mathcal{S}, V) \mid V \in \mathcal{S}, \eta = \eta_0 \sim \text{Uniform}(0, 1)$$



Example: $v_- = 0, v_+ = +\infty$ $\quad p(\mathcal{H}|\mathcal{S}, v) = 2 \times (1 - \Phi(v - \eta_0))$ two-tailed test

The "polyhedral lemma"



$$\boldsymbol{y} \sim N_{m+1}(\boldsymbol{\mu}, \boldsymbol{I}_{m+1})$$

$$\mathcal{H} = \{\boldsymbol{\mu} : \eta \leq \eta_0\}$$

$$\mathcal{S} = \{\boldsymbol{y} : v_-(\boldsymbol{u}) \leq v \leq v_+(\boldsymbol{u})\}$$

Lee et al. (2016) showed it for a polyhedron and a union of polyhedra;
This is used for the lasso confidence intervals.

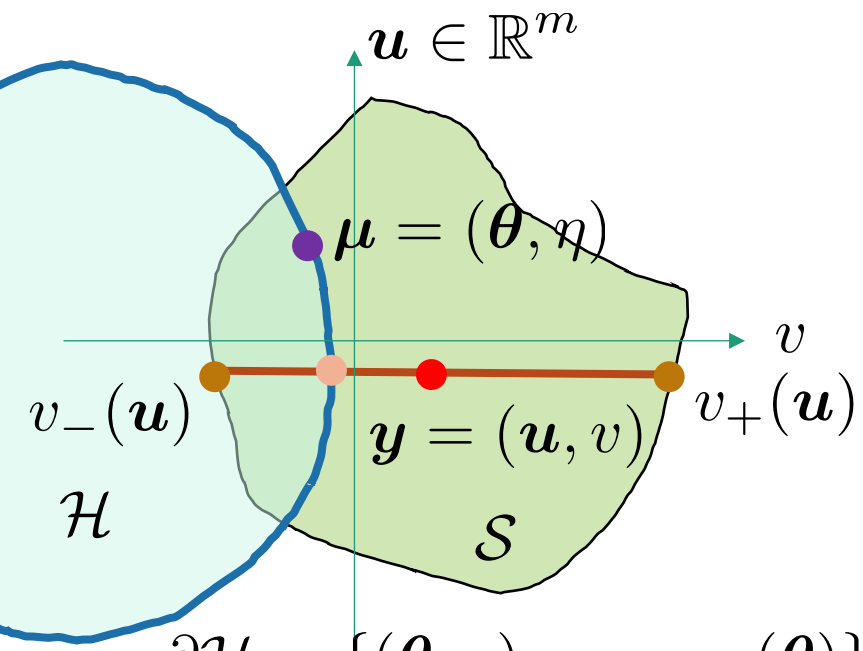
It works for more general regions of S (say, a union of convex regions), but its computation cost can be very large.

$$p(\mathcal{H}|\mathcal{S}, \boldsymbol{u}, v) = 1 - F_{\eta_0}^{v_-(\boldsymbol{u}), v_+(\boldsymbol{u})}(v)$$

$$p(\mathcal{H}|\mathcal{S}, \boldsymbol{U}, V) \mid (\boldsymbol{U}, V) \in \mathcal{S}, \eta = \eta_0, \underline{\boldsymbol{U} = \boldsymbol{u}} \sim \text{Uniform}(0, 1)$$

$$\underline{p(\mathcal{H}|\mathcal{S}, \boldsymbol{U}, V) \mid (\boldsymbol{U}, V) \in \mathcal{S}, \eta = \eta_0 \sim \text{Uniform}(0, 1)}$$

Curved boundary surface of H



$$\boldsymbol{y} \sim N_{m+1}(\boldsymbol{\mu}, \boldsymbol{I}_{m+1})$$

$$\mathcal{H} = \{(\boldsymbol{\theta}, \eta) : \eta \leq \eta_0(\boldsymbol{\theta})\}$$

The polyhedral lemma is not valid for general regions H with curved boundary surface, because we do not know eta0.

$$\partial\mathcal{H} = \{(\boldsymbol{\theta}, \eta) : \eta = \eta_0(\boldsymbol{\theta})\}$$

$$p(\mathcal{H}|\mathcal{S}, \boldsymbol{u}, v) = 1 - F_{\underline{\eta_0(\boldsymbol{u})}}^{v_-(\boldsymbol{u}), v_+(\boldsymbol{u})}(v)$$

$$p(\mathcal{H}|\mathcal{S}, \boldsymbol{U}, V) \mid (\boldsymbol{U}, V) \in \mathcal{S}, \eta = \eta_0(\boldsymbol{\theta}), \boldsymbol{U} = \boldsymbol{u} \not\sim \text{Uniform}(0, 1)$$

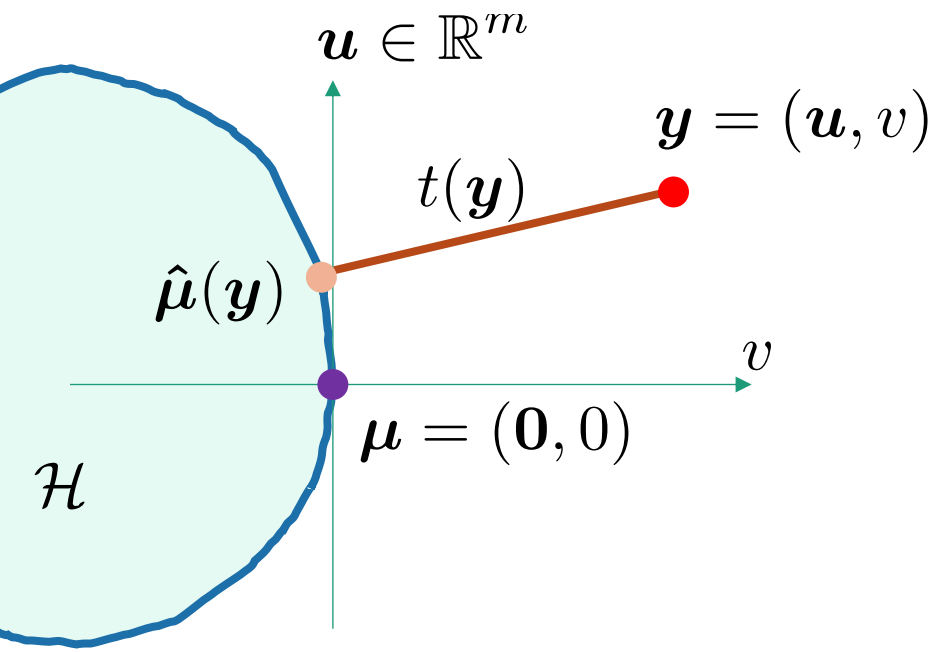
It is valid only when $\eta_0(\boldsymbol{u}) \equiv \eta_0(\boldsymbol{\theta})$

Limitation of “polyhedral lemma”

Hypothesis region H	Selective region S	“polyhedral lemma”
Flat boundary	Polyhedron	Good (used for Lasso)
Flat boundary	Union of polyhedra, or General regions	Slow
Curved boundary	either of above	Invalid

We develop a new method based on multiscale bootstrap

Signed distance and pivot statistic



$$\partial\mathcal{H} = \{(\boldsymbol{u}, v) : v = -\boldsymbol{u}^\top \boldsymbol{D}\boldsymbol{u} + \cdots\}$$

$$\gamma = \mathrm{tr}(\boldsymbol{D}) = O(n^{-1/2}) \text{ at } \boldsymbol{\mu}$$

$$\hat{\gamma} = \mathrm{tr}(\hat{\boldsymbol{D}}) = O_p(n^{-1/2}) \text{ at } \hat{\boldsymbol{\mu}}(\boldsymbol{y})$$

Mean curvature of the boundary surface

Efron (1986), Efron & Tibshirani (1998)

$$\boldsymbol{y} \sim N_{m+1}(\boldsymbol{\mu}, \boldsymbol{I}_{m+1})$$

Assume $\boldsymbol{\mu} \in \partial\mathcal{H}$ then

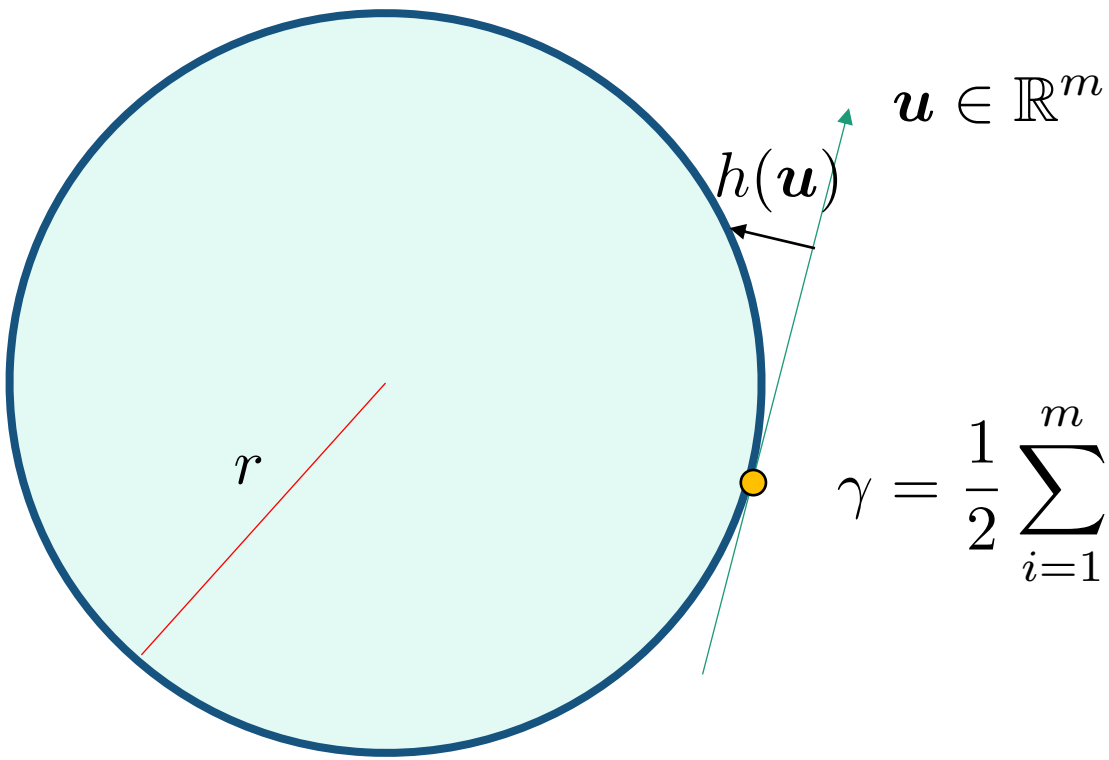
$$t(\boldsymbol{y}) \sim N(\gamma, 1)$$

$$t(\boldsymbol{y}) - \hat{\gamma} \sim N(0, 1)$$

Second order accurate: asymptotic error is $O_p(n^{-1})$

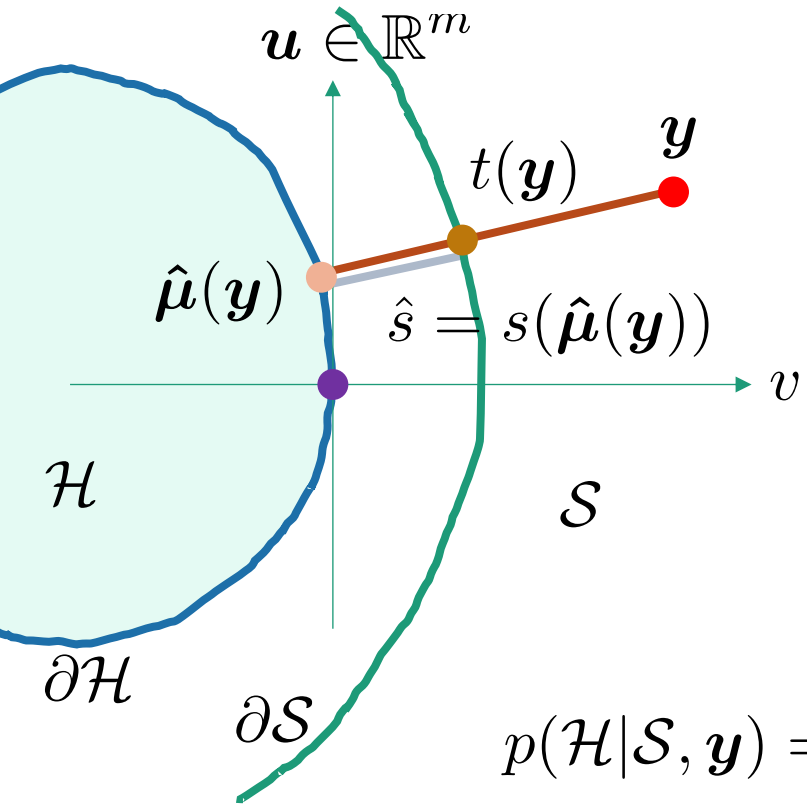
mean curvature

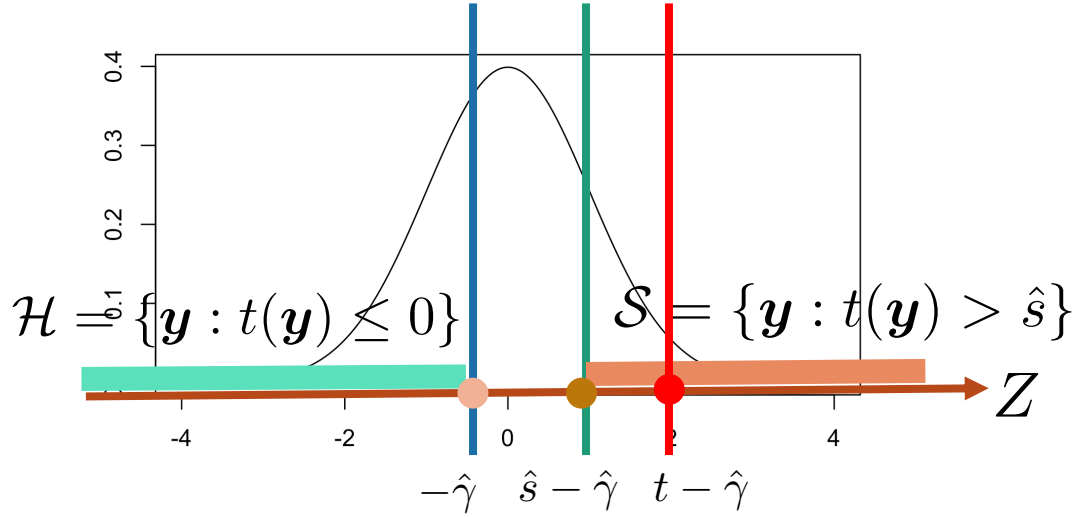
$\gamma = \frac{m}{2r}$ if the surface is sphere of radius r in $\mathbb{R}^{m+1}$



$$\gamma = \frac{1}{2}\sum_{i=1}^{m} \frac{\partial^2 h(\boldsymbol{u})}{\partial u_i^2} = \frac{m}{2r}$$

Selective inference for curved ∂H





$$Z = t(\boldsymbol{y}) - \hat{\gamma} \sim N(0, 1)$$

$$p(\mathcal{H}|\mathcal{S}, \boldsymbol{y}) = \frac{\mathbb{P}(Z > t - \hat{\gamma})}{\mathbb{P}(Z > \hat{s} - \hat{\gamma})} = \frac{\bar{\Phi}(t - \hat{\gamma})}{\bar{\Phi}(\hat{s} - \hat{\gamma})}$$

$$p(\mathcal{H}|\mathcal{S}, \boldsymbol{y}) \mid \boldsymbol{y} \in \mathcal{S}, \boldsymbol{\mu} \in \partial\mathcal{H} \sim \mathrm{Uniform}(0, 1)$$

Terada and Shimodaira (arXiv 2017)

Approximately unbiased p-values

Selective inference: Terada and Shimodaira (arXiv 2017)

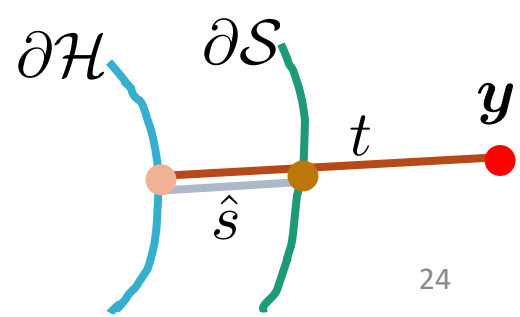
$$p(\mathcal{H}|\mathcal{S},\boldsymbol{y}) = \frac{\bar{\Phi}(t-\hat{\gamma})}{\bar{\Phi}(\hat{s}-\hat{\gamma})} \qquad \boldsymbol{y} \in \mathcal{S}$$

Non-selective inference: Shimodaira (2002, 2004, 2008)

$$p(\mathcal{H}|\boldsymbol{y}) = \bar{\Phi}(t-\hat{\gamma}) \qquad \mathcal{S} = \mathbb{R}^{m+1}$$

Geometric quantities:

t is signed distance from y to H
$\hat{s}$ is signed distance from ∂S to H
$\hat{\gamma}$ is mean curvature of ∂H



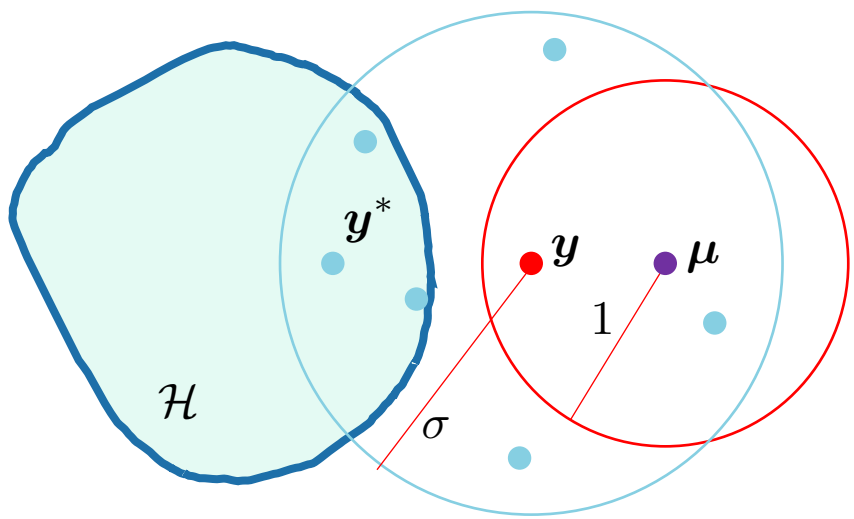
Multiscale bootstrap

Multiscale Bootstrap Probability

$$\boldsymbol{y}|\boldsymbol{\mu} \sim N_{m+1}(\boldsymbol{\mu}, \boldsymbol{I}_{m+1})$$

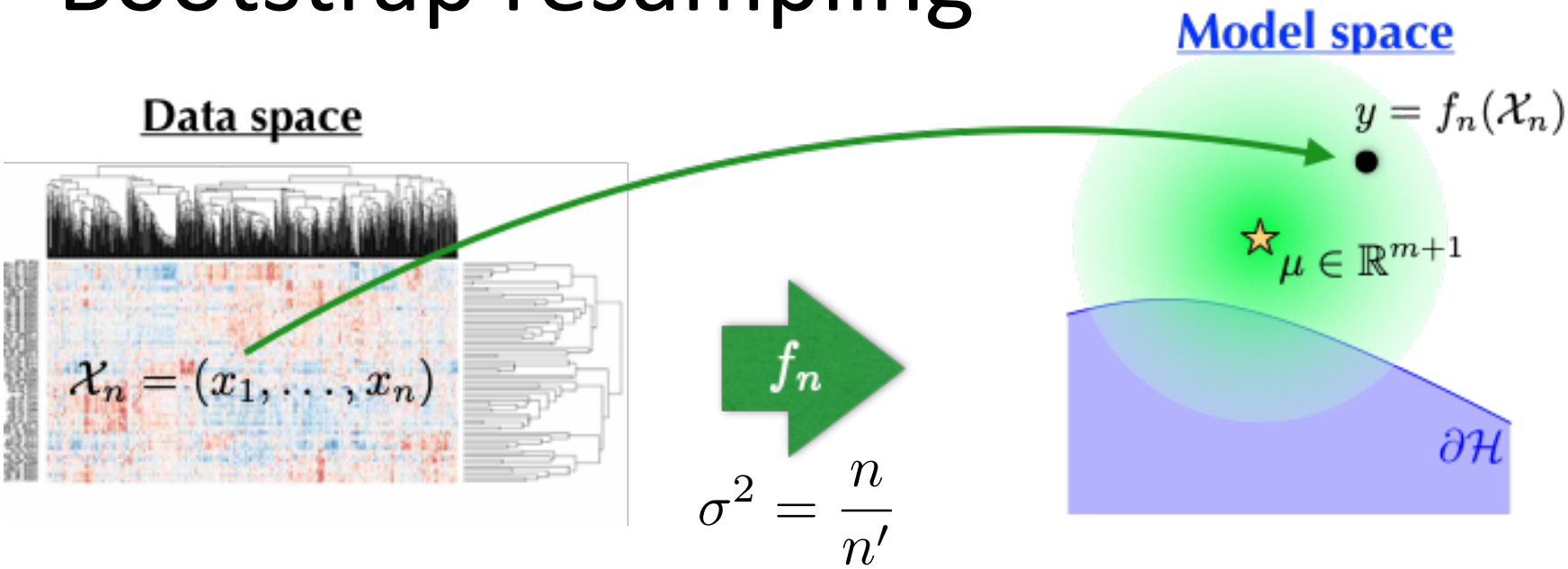
$$\boldsymbol{y}^*|\boldsymbol{y} \sim N_{m+1}(\boldsymbol{y}, \sigma^2 \boldsymbol{I}_{m+1}), \quad \sigma^2 > 0$$

$$\alpha_{\sigma^2}(\mathcal{H}|y) = \mathbb{P}_{\sigma^2}(\boldsymbol{y}^* \in \mathcal{H}|\boldsymbol{y}) = \frac{\#\{\boldsymbol{y}^{*1}, \ldots, \boldsymbol{y}^{*B} \in \mathcal{H}\}}{B}$$



$$\alpha_{\sigma_1^2}(\mathcal{H}|y), \alpha_{\sigma_2^2}(\mathcal{H}|y), \alpha_{\sigma_3^2}(\mathcal{H}|y), \ldots$$

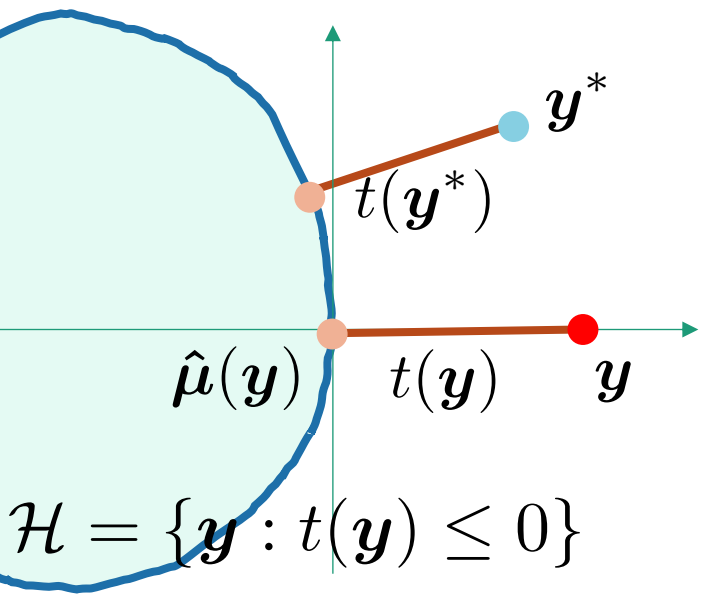
Bootstrap resampling



(example)

$\mathcal{X}_\infty = (x_1, \ldots)$	$f_n(\mathcal{X}_\infty) = \sqrt{n}\boldsymbol{A}\mathbb{E}(x)$	$\boldsymbol{\mu}$
$\mathcal{X}_n = (x_1, \ldots, x_n)$	$f_n(\mathcal{X}_n) = \sqrt{n}\boldsymbol{A}\bar{x}$	$\boldsymbol{y}$
$\mathcal{X}^*_{n'} = (x^*_1, \ldots, x^*_{n'})$	$f_n(\mathcal{X}^*_{n'}) = \sqrt{n}\boldsymbol{A}\bar{x}^*$	$\boldsymbol{y}^*$

Bootstrap probability ($\sigma^2 = 1$)

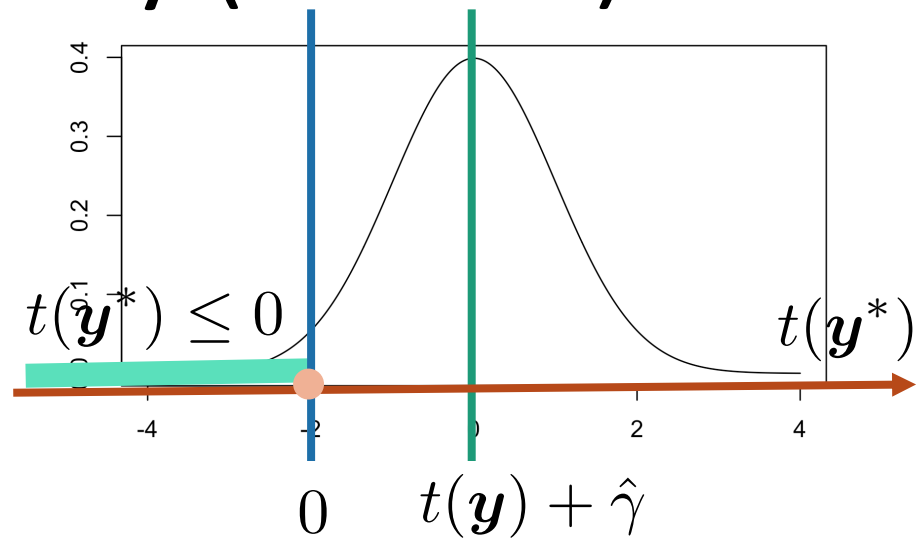


$$\mathcal{H} = \{\boldsymbol{y} : t(\boldsymbol{y}) \leq 0\}$$

$$\boldsymbol{y}^* | \boldsymbol{y} \sim N_{m+1}(\boldsymbol{y}, \boldsymbol{I}_{m+1})$$

$$\bar{\Phi}(1.64) = 1 - \Phi(1.64) = 0.05$$

$$\bar{\Phi}^{-1}(0.05) = 1.64$$



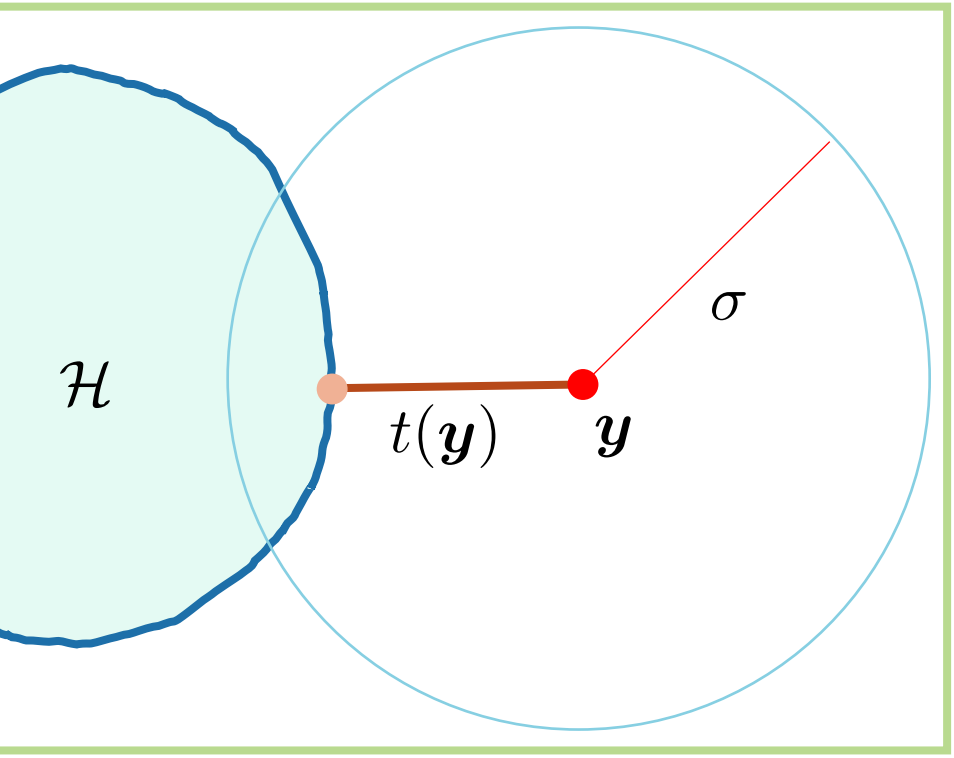
Efron (1986), Efron & Tibshirani (1998)

$$t(\boldsymbol{y}^*)|\boldsymbol{y} \sim N(t(\boldsymbol{y}) + \hat{\gamma}, 1)$$

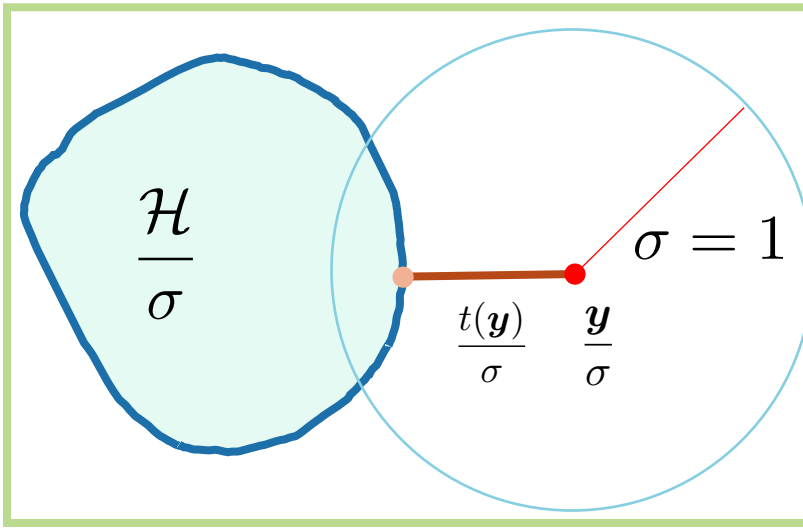
$$\alpha_1(\mathcal{H}|\boldsymbol{y}) = \mathbb{P}_1(t(\boldsymbol{y}^*) \leq 0|\boldsymbol{y})$$
$$= \bar{\Phi}(t(\boldsymbol{y}) + \hat{\gamma})$$

Second order accurate: asymptotic error is $O_p(n^{-1})$

Bootstrap probability ($\sigma^2 > 0$)



$\equiv$



$$\boldsymbol{y}^* | \boldsymbol{y} \sim N_{m+1}(\boldsymbol{y}, \sigma^2 \boldsymbol{I}_{m+1})$$

$$\boldsymbol{y}^* | \boldsymbol{y} \sim N_{m+1}(\frac{\boldsymbol{y}}{\sigma}, \boldsymbol{I}_{m+1})$$

$$\alpha_{\sigma^2}(\mathcal{H}|\boldsymbol{y}) = \alpha_1(\frac{\mathcal{H}}{\sigma}|\frac{\boldsymbol{y}}{\sigma}) = \bar{\Phi}(\frac{t(\boldsymbol{y})}{\sigma} + \hat{\gamma}\sigma)$$

Shimodaira (2002, 2004, 2008)

Computing approximately unbiased p-values via multiscale bootstrap

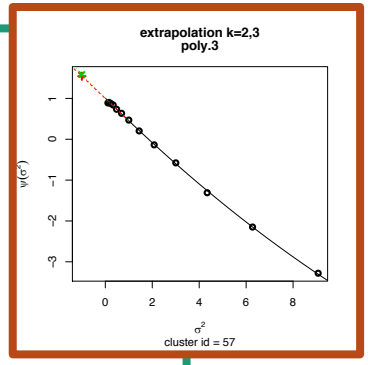
p_{AU}

Shimodaira (2002, 2004, 2008): non-selective inference

$$\alpha_{\sigma^2}(\mathcal{H}|\boldsymbol{y}) = \bar{\Phi}\left(\frac{t(\boldsymbol{y})}{\sigma} + \hat{\gamma}\sigma\right)$$

$$\psi_{\sigma^2}(\mathcal{H}|\boldsymbol{y}) = \sigma\bar{\Phi}^{-1}(\alpha_{\sigma^2}(\mathcal{H}|\boldsymbol{y})) = t(\boldsymbol{y}) + \hat{\gamma}\sigma^2$$

$$p(\mathcal{H}|\boldsymbol{y}) = \bar{\Phi}(t - \hat{\gamma}) = \bar{\Phi}(\psi_{-1}(\mathcal{H}|\boldsymbol{y}))$$



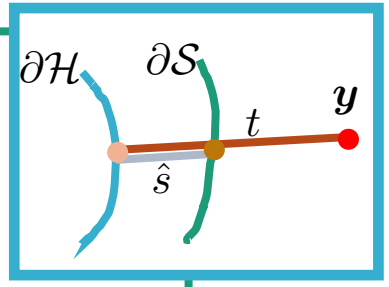
Normalized bootstrap z-value

p_{SI}

Terada and Shimodaira (arXiv 2017): selective inference

$$\psi_0(\mathcal{S}|\boldsymbol{y}) = \hat{s} - t$$

$$p(\mathcal{H}|\mathcal{S}, \boldsymbol{y}) = \frac{\bar{\Phi}(t - \hat{\gamma})}{\bar{\Phi}(\hat{s} - \hat{\gamma})} = \frac{\bar{\Phi}(\psi_{-1}(\mathcal{H}|\boldsymbol{y}))}{\bar{\Phi}(\psi_{-1}(\mathcal{H}|\boldsymbol{y}) + \psi_0(\mathcal{S}|\boldsymbol{y}))}$$



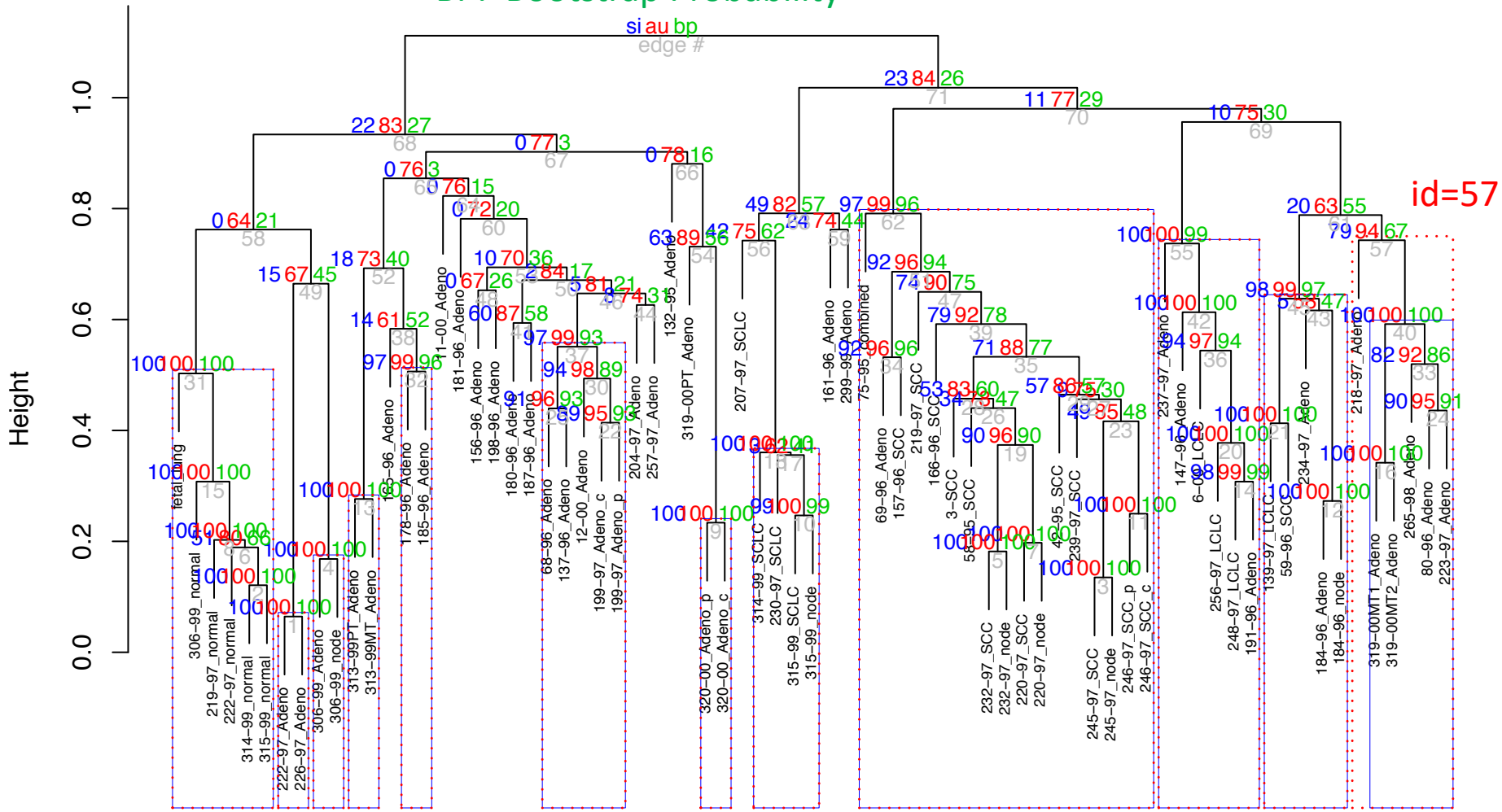
Pvclust example

Terada and Shimodaira (arXiv 2017)

SI: Selective Inference: New in this talk

AU: non-selective inference (Approximately Unbiased)

BP: Bootstrap Probability



pvclust analysis of lung dataset (k=2: default value of pvclust)

Distance: correlation
Cluster method: average

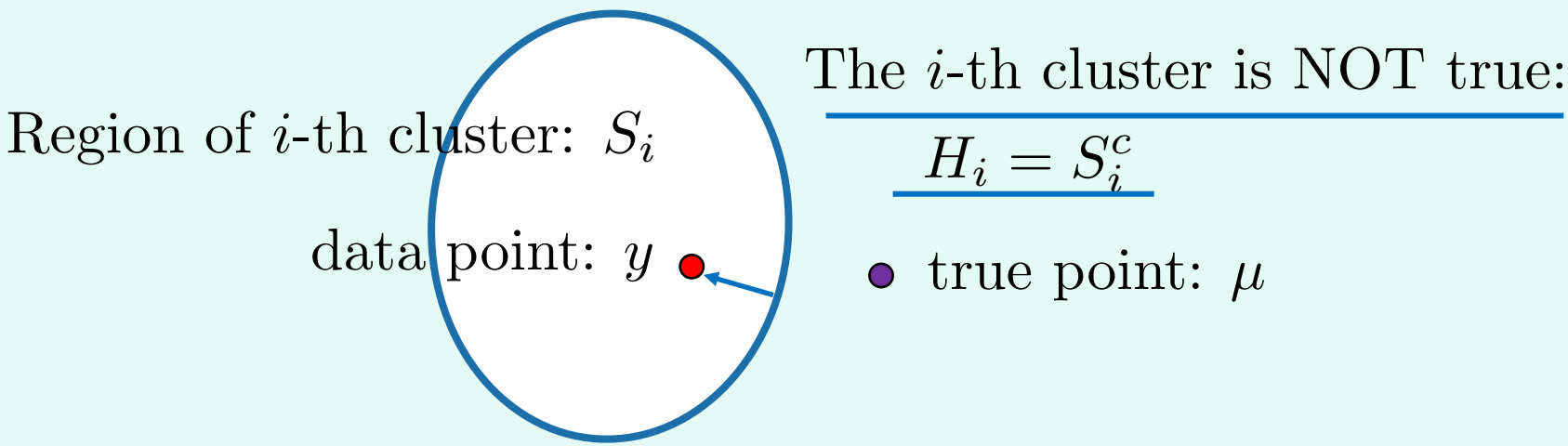
Applying the method to lung data

$n = 916$

$n' = 8244, 5716, 3963, 2748, 1905, 1321, 916, 635, 440, 305, 211, 146, 101$

$\sigma^2 = n/n'$ ranges from $1/9$ to 9 in log-scale

$B = 10000$



Cluster id = 57

Plotting the bootstrap probabilities

$$C_{H^c} = 9962, 9878, 9657, 9271, 8551, 7773, 6807, 5676, 4622, 3695, 2650, 1955, 1381$$

$$\alpha_{\sigma^2}(H|y) = \frac{C_H}{B} = 1 - \frac{C_{H^c}}{B}, \quad \sigma^2 = \frac{n}{n'}$$

$$\psi_{\sigma^2}(H|y) = \sigma\bar{\Phi}^{-1}(\alpha_{\sigma^2}(H|y))$$

Fitting models to psi

$$\varphi_H(\sigma^2|\beta) = \beta_0 + \beta_1\sigma^2$$

$$\hat{\beta}_0 = 0.998, \ \hat{\beta}_1 = -0.545$$

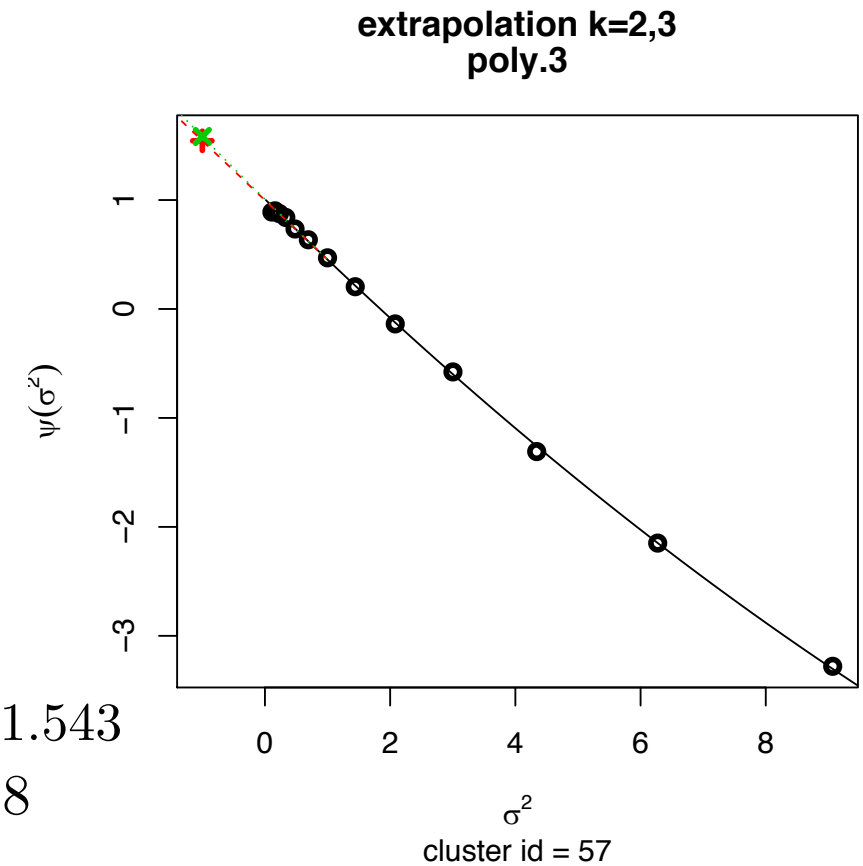
(signed distance and mean curvature)

Extrapolation to $\sigma^2 \leq 0$

$$z_H = \varphi_H(-1|\hat{\beta}) = 0.998 - (-0.554) = 1.543$$

$$z_S = \varphi_{H^c}(0|\hat{\beta}) = -\varphi_H(0|\hat{\beta}) = -0.998$$

$$z_H + z_S = 0.554$$



Cluster id = 57 (cont.)

SI: approximately unbiased p-value for selective inference (Terada and Shimodaira 2017)

$$p_{\mathrm{SI}} = \frac{\bar{\Phi}(z_H)}{\bar{\Phi}(z_H + z_S)} = \frac{\bar{\Phi}(1.543)}{\bar{\Phi}(0.554)} = 0.210 \qquad 1 - p_{\mathrm{SI}} = 0.790$$

The numerator is AU: approximately unbiased p-value for non-selective inference (Shimodaira 2002-)

$$p_{\mathrm{AU}} = \bar{\Phi}(z_H) = \bar{\Phi}(1.543) = 0.061 \qquad 1 - p_{\mathrm{AU}} = 0.939$$

The denominator is the selection probability under the null

$$\bar{\Phi}(z_H + z_S) = \bar{\Phi}(0.554) = 0.293$$

BP: bootstrap probability is expressed as

$$p_{\mathrm{BP}} = \bar{\Phi}(\varphi_H(1|\hat{\beta})) = \bar{\Phi}(0.998 + (-0.554)) = \bar{\Phi}(0.453) = 0.325$$

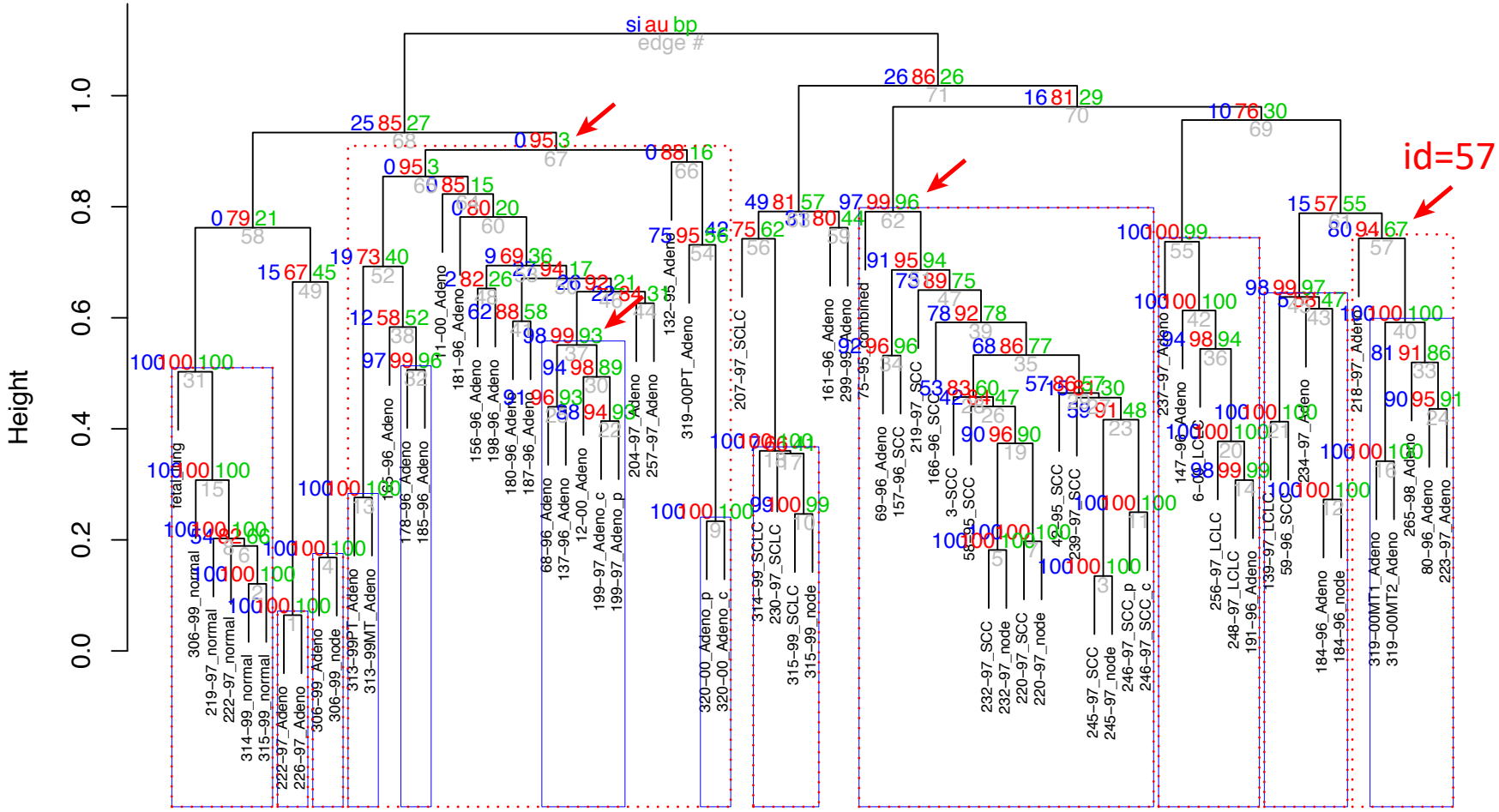
$$1 - p_{\mathrm{BP}} = 0.675$$

1-p is shown for

SI: Selective Inference: New in this talk

AU: non-selective inference (Approximately Unbiased)

BP: Bootstrap Probability

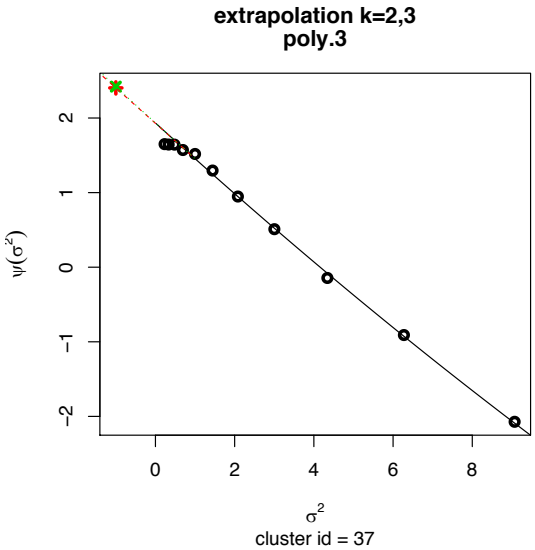


pvclust analysis of lung dataset (k=3)

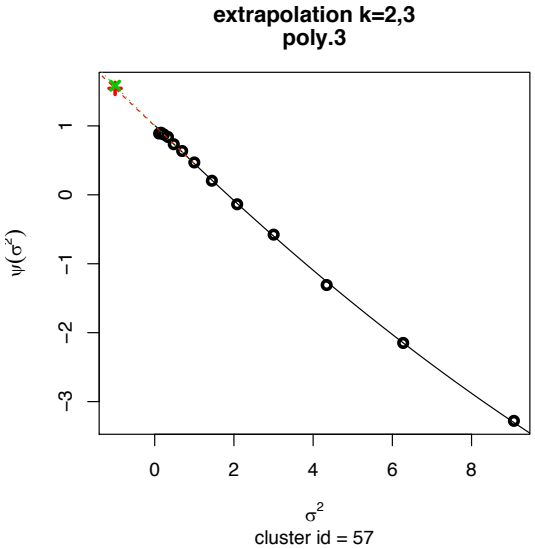
Distance: correlation
Cluster method: average

The linear model (k=2) fits well

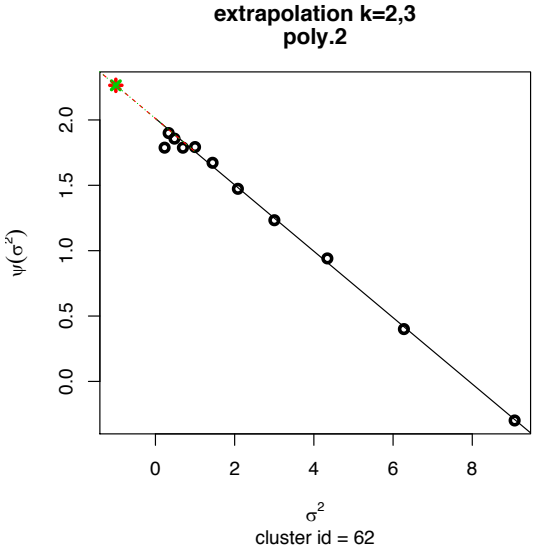
Id=37



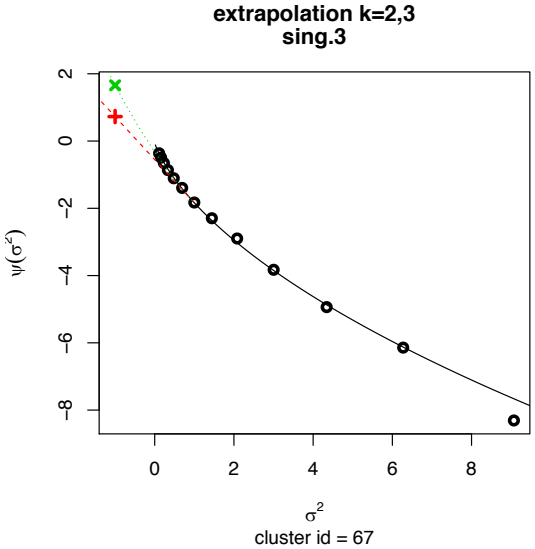
Id=57



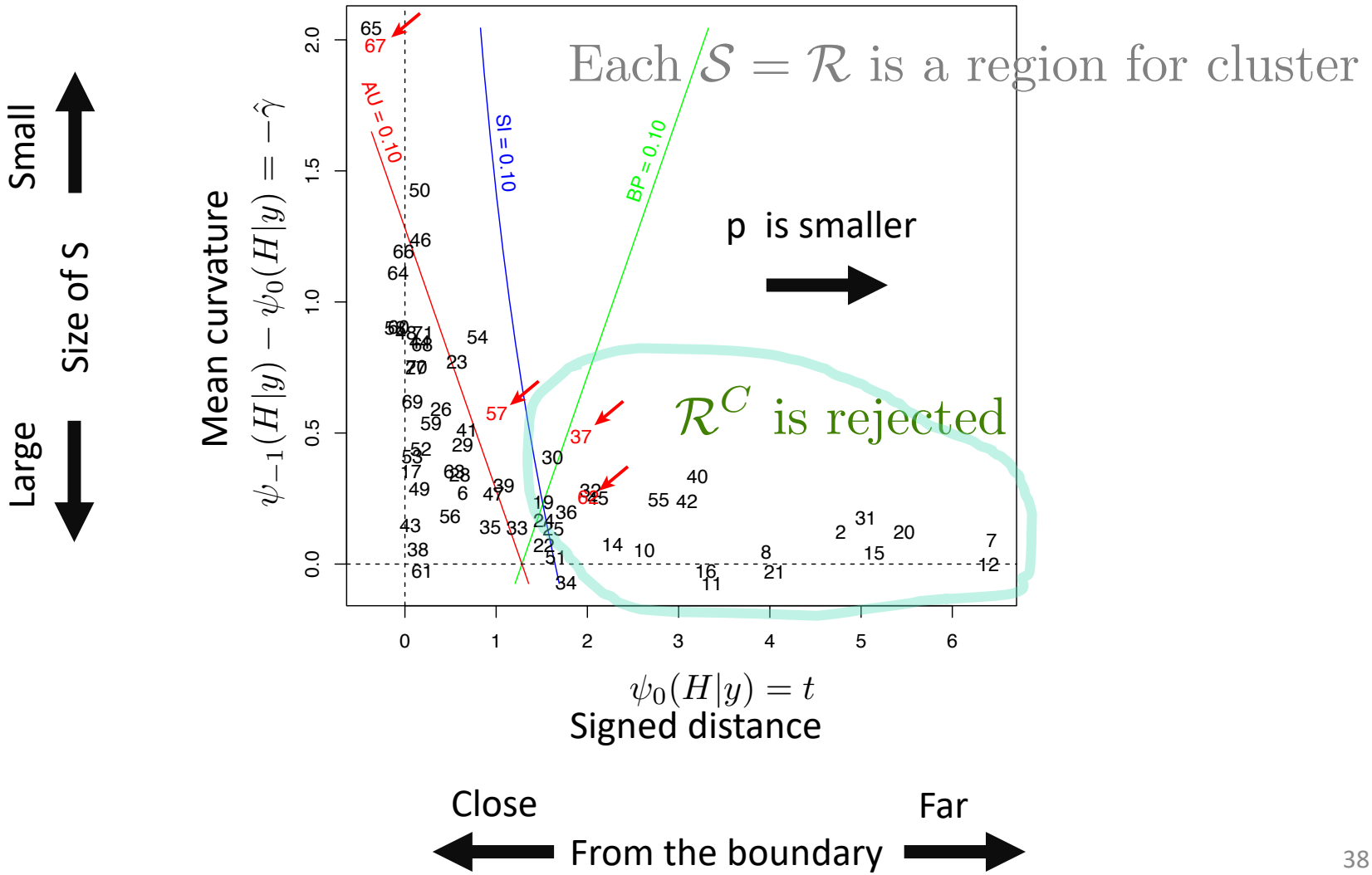
Id=63



Id=67



Signed distance and mean curvature of clusters in lung data (H = S^c)

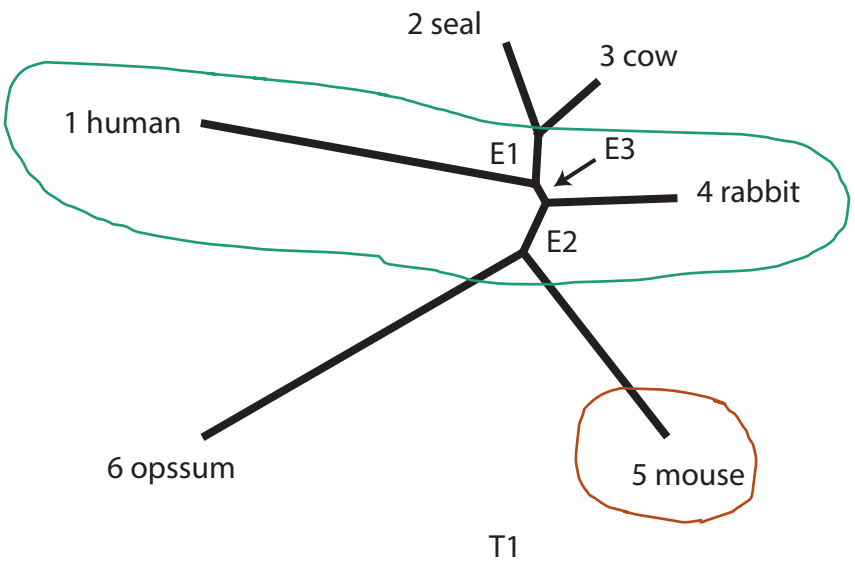


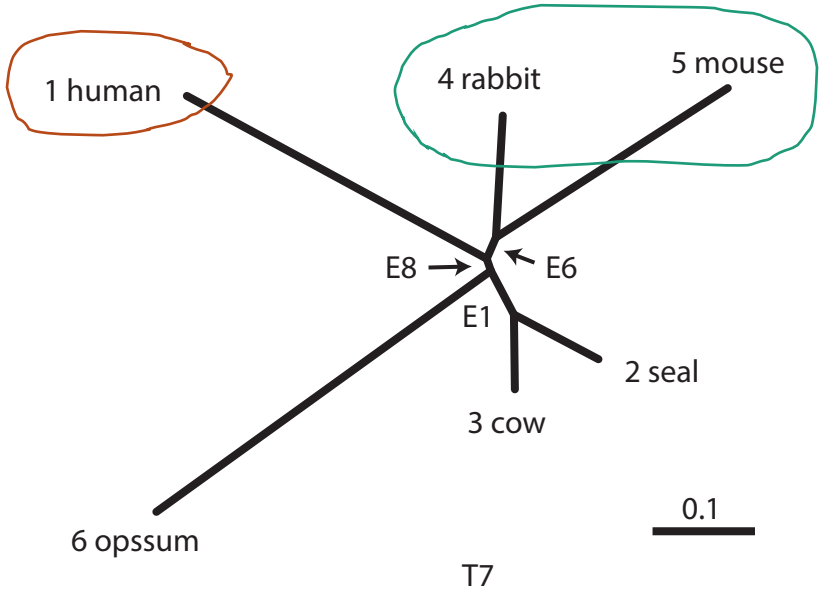
Phylogenetic inference

Shimodaira and Terada (2019)

105 unrooted trees

Mitochondrial protein sequences of six mammalian species (n = 3414 amino acids)
Shimodaira and Hasegawa (1999), Shimodaira (2002), Shimodaira and Terada (2019)





T1: maximum likelihood tree

((human, rabbit), mouse)

T7: “True” tree

((rabbit, mouse), human)

Trees (top 20)

TABLE 1 | Three types of p-values (BP, AU, SI) and geometric quantities (β_0, β_1) for the best 20 trees.

Tree	BP	AU	SI	β_0	β_1	Topology	Edges
T1†	0.559 (0.001)	0.752 (0.001)	0.372 (0.001)	-0.41 (0.00)	0.27 (0.00)	(((1(23))4)56)	E1, E2, E3
T2	0.304 (0.000)	0.467 (0.001)	0.798 (0.001)	0.30 (0.00)	0.22 (0.00)	((1((23)4))56)	E1 ,E2, E4
T3	**0.038** (0.000)	0.126 (0.002)	0.202 (0.003)	1.46 (0.01)	0.32 (0.00)	(((14)(23))56)	E1, E2, E5
T4	**0.014** (0.000)	0.081 (0.002)	0.124 (0.003)	1.79 (0.01)	0.40 (0.01)	((1(23))(45)6)	E1, E3, E6
T5	**0.032** (0.000)	0.127 (0.002)	0.199 (0.003)	1.50 (0.01)	0.36 (0.00)	(1((23)(45))6)	E1, E6, E7
T6	**0.005** (0.000)	**0.032** (0.002)	0.050 (0.002)	2.21 (0.02)	0.35 (0.01)	(1(((23)4)5)6)	E1, E4, E7
T7‡	**0.015** (0.000)	0.100 (0.003)	0.150 (0.003)	1.72 (0.01)	0.44 (0.01)	((1(45))(23)6)	E1, E6, E8
T8	**0.001** (0.000)	**0.011** (0.001)	**0.016** (0.002)	2.74 (0.03)	0.43 (0.02)	((15)((23)4)6)	E1, E4, E9
T9	**0.000** (0.000)	**0.001** (0.000)	**0.001** (0.000)	3.67 (0.09)	0.46 (0.04)	(((1(23))5)46)	E1, E3, E10
T10	**0.002** (0.000)	**0.022** (0.002)	**0.033** (0.002)	2.43 (0.02)	0.42 (0.01)	(((15)4)(23)6)	E1, E8, E9
T11	**0.000** (0.000)	**0.004** (0.001)	**0.006** (0.002)	3.14 (0.07)	0.51 (0.03)	(((14)5)(23)6)	E1, E5, E8
T12	**0.000** (0.000)	**0.000** (0.000)	**0.001** (0.000)	3.78 (0.09)	0.41 (0.04)	(((15)(23))46)	E1, E9, E10
T13	**0.000** (0.000)	**0.000** (0.000)	**0.001** (0.001)	3.96 (0.19)	0.54 (0.09)	(1(((23)5)4)6)	E1, E7, E11
T14	**0.000** (0.000)	**0.000** (0.000)	**0.000** (0.000)	4.66 (0.31)	0.65 (0.12)	((14)((23)5)6)	E1, E5, E11
T15	**0.000** (0.000)	**0.000** (0.000)	**0.000** (0.000)	5.28 (0.34)	0.43 (0.11)	((1((23)5))46)	E1, E10, E11
T16	**0.000** (0.000)	**0.000** (0.000)	**0.001** (0.000)	3.63 (0.04)	0.23 (0.01)	((((13)2)4)56)	E2, E3, E12
T17	**0.000** (0.000)	**0.000** (0.000)	**0.000** (0.000)	3.81 (0.04)	0.22 (0.01)	((((12)3)4)56)	E2, E3, E13
T18	**0.000** (0.000)	**0.000** (0.000)	**0.000** (0.000)	4.33 (0.10)	0.34 (0.03)	(((13)2)(45)6)	E3, E6, E12
T19	**0.000** (0.000)	**0.000** (0.000)	**0.000** (0.000)	4.36 (0.11)	0.32 (0.04)	(((12)3)(45)6)	E3, E6, E13
T20	**0.000** (0.000)	**0.000** (0.000)	**0.000** (0.000)	3.90 (0.12)	0.44 (0.05)	(((1(45))2)36)	E6, E8, E14

Standard errors are shown in parentheses. Boldface indicates significance ($p < 0.05$) for the null hypothesis that the tree is true (outside mode). For the rest of trees (T21, . . . , T105), p-values are very small ($p < 0.001$). † T1 is the ML tree, i.e., the tree selected by the ML method based on the dataset of Shimodaira and Hasegawa (1999). ‡ T7 is presumably the true tree as suggested by later researches; see section 4.3.

Edges (i.e., clusters)

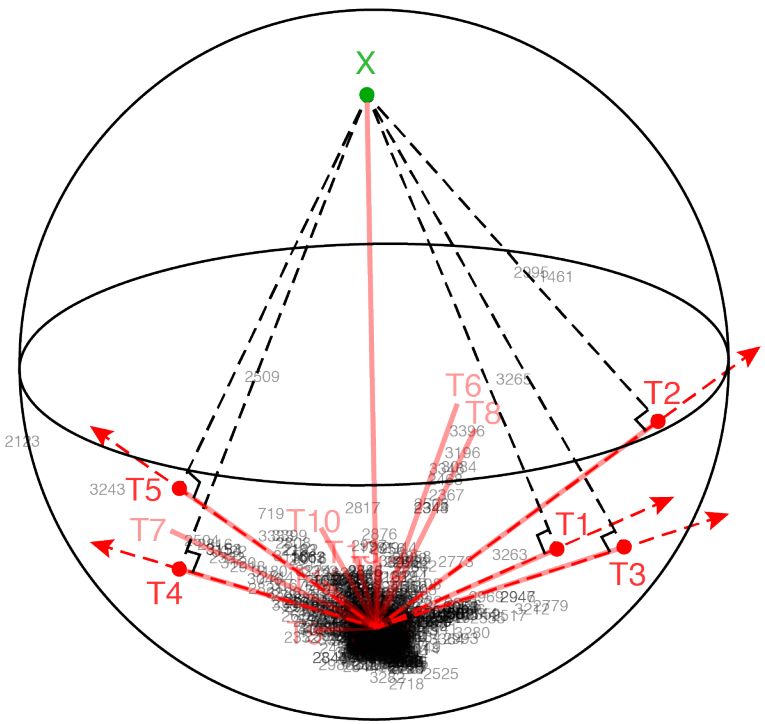
TABLE 2 | Three types of p-values (BP, AU, SI) and geometric quantities (β_0, β_1) for all the 25 edges of six taxa.

Edge	BP	AU	SI	β_0	β_1	Clade
E1†‡	**1.000** (0.000)	**1.000** (0.000)	**1.000** (0.000)	-3.87 (0.03)	0.16 (0.01)	−++−−−
E2†	0.930 (0.000)	**0.956** (0.001)	0.903 (0.001)	-1.59 (0.00)	0.12 (0.00)	++++−−
E3†	0.580 (0.001)	0.719 (0.001)	0.338 (0.001)	-0.39 (0.00)	0.19 (0.00)	+++−−−
E4	0.318 (0.000)	0.435 (0.001)	0.775 (0.001)	0.32 (0.00)	0.16 (0.00)	−+++−−
E5	**0.037** (0.000)	0.124 (0.002)	0.198 (0.002)	1.47 (0.01)	0.32 (0.00)	+−−+−−
E6‡	0.060 (0.000)	0.074 (0.001)	0.141 (0.002)	1.50 (0.00)	0.05 (0.00)	−−−++−
E7	**0.038** (0.000)	0.091 (0.002)	0.154 (0.002)	1.56 (0.01)	0.22 (0.00)	−++++−
E8‡	**0.018** (0.000)	0.068 (0.002)	0.110 (0.003)	1.80 (0.01)	0.31 (0.01)	+−−++−
E9	**0.003** (0.000)	**0.014** (0.001)	**0.023** (0.002)	2.48 (0.02)	0.27 (0.02)	+−−−+−
E10	**0.000** (0.000)	**0.000** (0.000)	**0.001** (0.000)	3.72 (0.07)	0.29 (0.03)	+++−+−
E11	**0.000** (0.000)	**0.000** (0.000)	**0.000** (0.000)	4.31 (0.10)	0.35 (0.03)	−++−+−
E12	**0.000** (0.000)	**0.000** (0.000)	**0.000** (0.000)	3.68 (0.05)	0.17 (0.02)	+−+−−−
E13	**0.000** (0.000)	**0.000** (0.000)	**0.000** (0.000)	3.90 (0.04)	0.15 (0.02)	++−−−−
E14	**0.000** (0.000)	**0.000** (0.000)	**0.000** (0.000)	4.03 (0.09)	0.30 (0.04)	++−++−
E15	**0.000** (0.000)	**0.000** (0.000)	**0.000** (0.000)	4.03 (0.13)	0.38 (0.06)	+−+++−
E16	**0.000** (0.000)	**0.000** (0.000)	**0.000** (0.000)	4.44 (0.05)	0.12 (0.01)	−+−+−−
E17	**0.000** (0.000)	**0.000** (0.000)	**0.000** (0.000)	4.70 (0.07)	0.19 (0.02)	++−+−−
E18	**0.000** (0.000)	**0.000** (0.000)	**0.000** (0.000)	3.94 (0.09)	0.26 (0.04)	−+−++−
E19	**0.000** (0.000)	**0.000** (0.000)	**0.000** (0.000)	5.23 (0.43)	0.57 (0.13)	−−++−−
E20	**0.000** (0.000)	**0.000** (0.000)	**0.000** (0.000)	5.66 (0.29)	0.28 (0.09)	+−++−−
E21	**0.000** (0.000)	**0.000** (0.000)	**0.000** (0.000)	6.38 (0.33)	0.24 (0.08)	−−+++−
E22	**0.000** (0.000)	**0.000** (0.000)	**0.000** (0.000)	5.62 (0.21)	0.17 (0.07)	−−+−+−
E23	**0.000** (0.000)	**0.000** (0.000)	**0.000** (0.000)	4.86 (0.43)	0.70 (0.13)	−+−−+−
E24	**0.000** (0.000)	**0.000** (0.000)	**0.000** (0.000)	5.61 (0.17)	0.23 (0.04)	+−+−+−
E25	**0.000** (0.000)	**0.000** (0.000)	**0.000** (0.000)	6.32 (0.71)	0.52 (0.20)	++−−+−

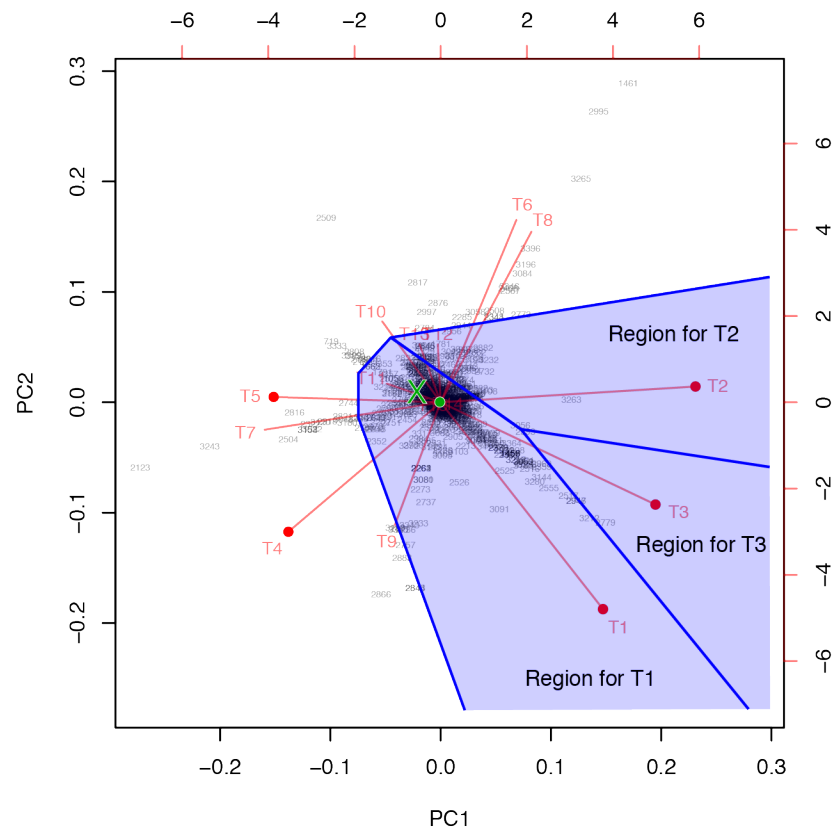
Standard errors are shown in parentheses. Boldface without underline indicates significance ($p < 0.05$) for the null hypothesis that the edge is true (outside mode). Boldface with underline indicates significance ($p > 0.95$) for the null hypothesis that the edge is not *true (inside mode).* † *Edges included in T1.* ‡ *Edges included in T7.*

Regions for trees

A Projection of data point X to tree models



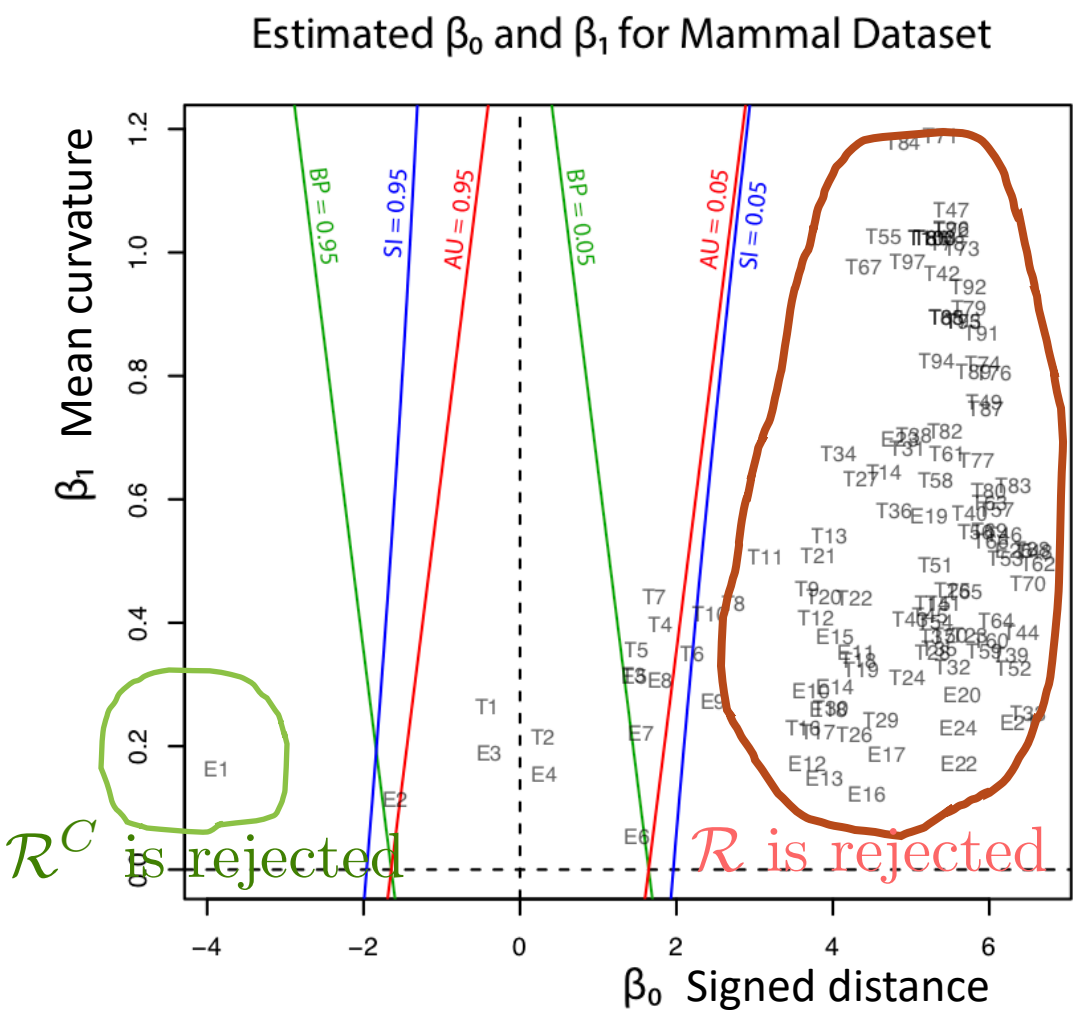
B Regions for trees with data point X



$$\mathcal{R}_{\mathrm{T1}}, \mathcal{R}_{\mathrm{T2}}, \ldots$$

Geometric quantities for regions

Each $\mathcal{R}$ is a region for tree or cluster



$\mathcal{H} = \mathcal{R}^C, \mathcal{S} = \mathcal{R}$ for $\boldsymbol{y} \in \mathcal{R}$

1-p is shown when signed distance < 0

E1 = (seal, cow)
E2 = (human, seal, cow, rabbit)
T1 = ML tree
E3 = (human, seal, cow)

$\mathcal{H} = \mathcal{R}, \mathcal{S} = \mathcal{R}^C$ for $\boldsymbol{y} \notin \mathcal{R}$

p is shown when signed distance > 0

All other regions

Confidence intervals of regression coefficients

Terada and Shimodaira (arXiv 2019)

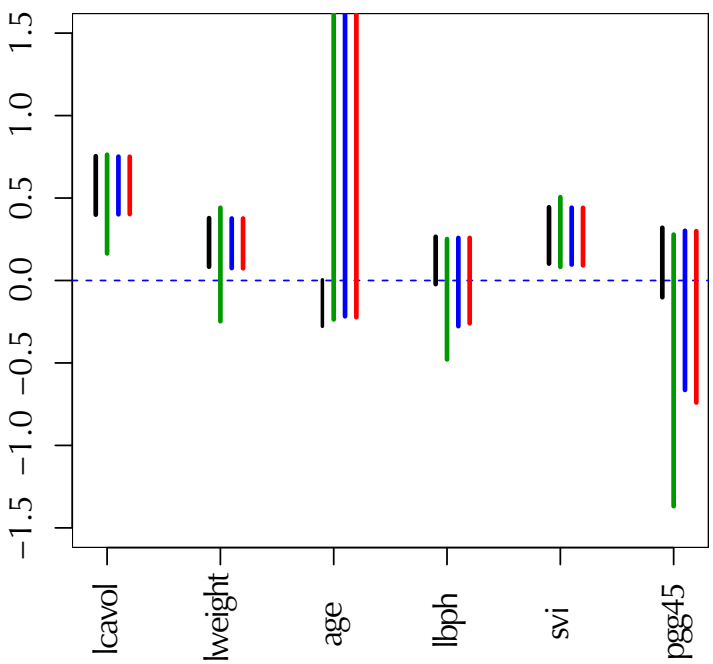
Post-selection inference of variable selection by Lasso

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p + \epsilon$$

6 selected variables (coefficients) out of 8 variables:

$$\{\beta_{\mathrm{lcavol}}, \beta_{\mathrm{lweight}}, \beta_{\mathrm{age}}, \beta_{\mathrm{lbph}}, \beta_{\mathrm{svi}}, \beta_{\mathrm{pgg45}}\}$$

95% Confidence Intervals



Black: non-selective CI

Selective Inference CI

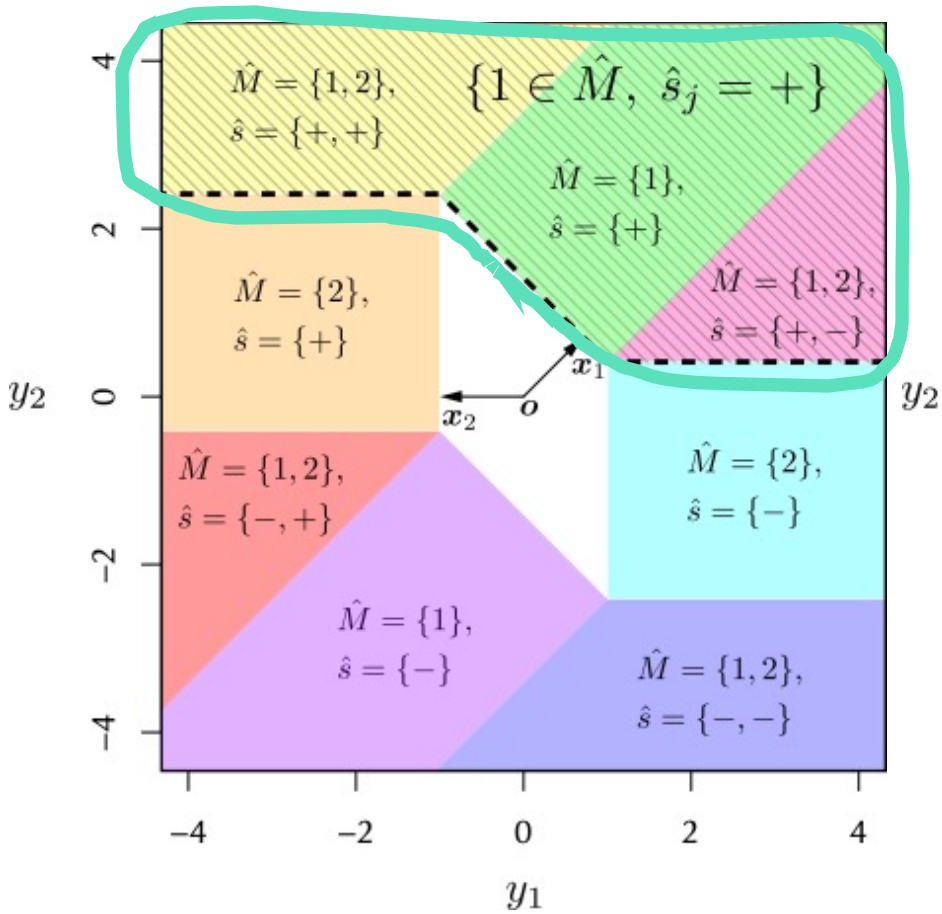
Green: Lee et al. (2016) for model

Blue: Liu et al. (2018) for variable

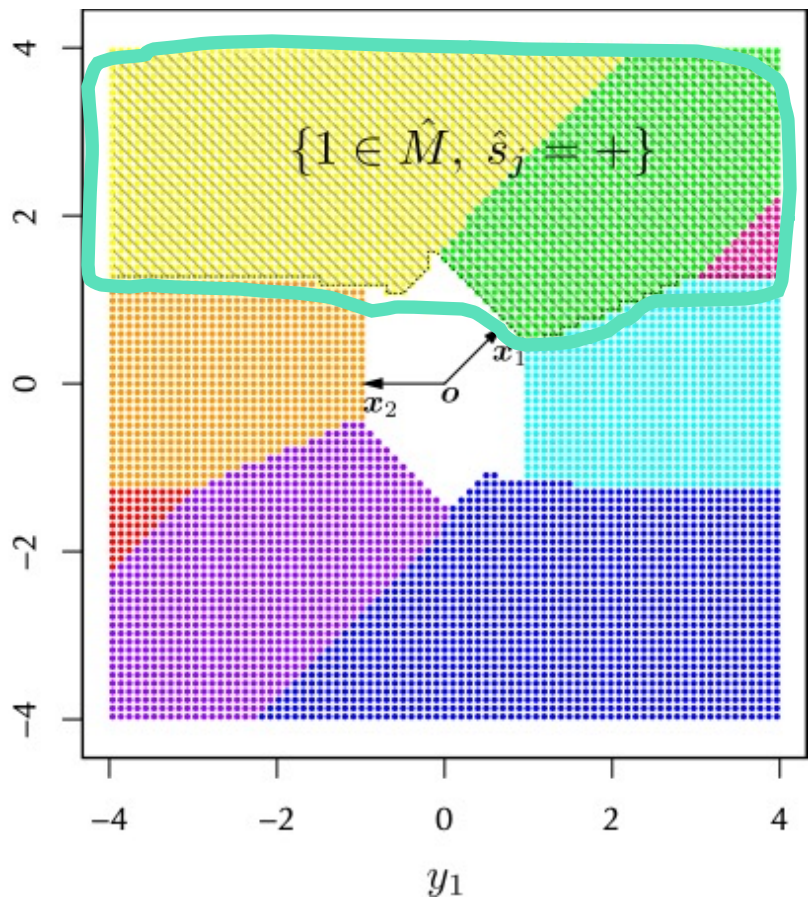
Red: Terada and Shimodaira (2019) (via multiscale bootstrap)

Complicated selective regions

Lasso



MCP



Computing CI

- For hypothesis region H, signed distance t is easily computed by projection. Curvature is $\gamma = 0$.
- For selective region S, signed distance is estimated by multiscale bootstrap of "resampling residuals" with scales.

$$\boldsymbol{y}^* = \boldsymbol{X}\hat{\boldsymbol{\beta}} + \sigma\hat{\boldsymbol{\epsilon}}^*$$

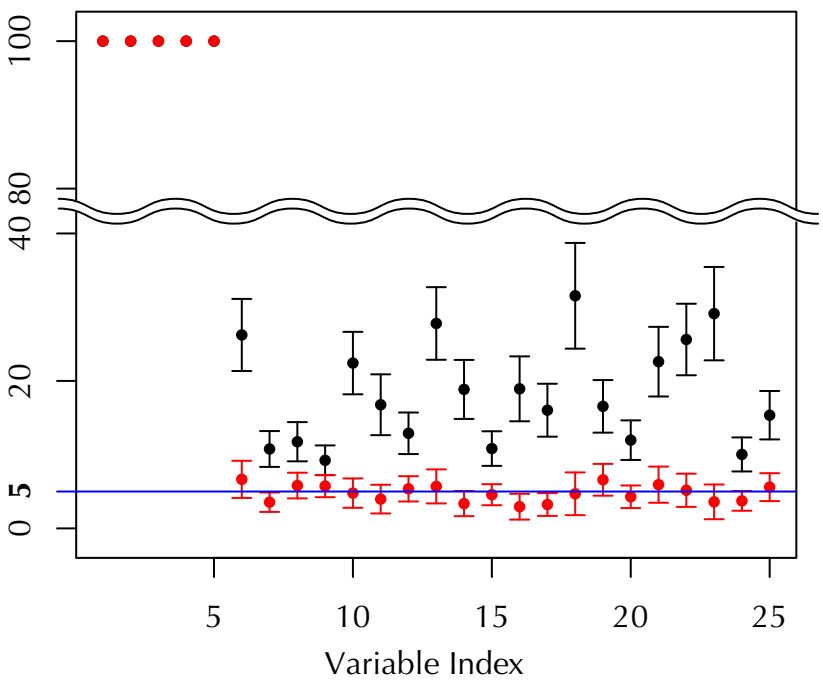
- For each hypothesis $\beta_i = c$, p-values are computed and CI is defined as the set of non-rejected c values.

Selective rejection probability for hypotheses $\beta_i = 0$

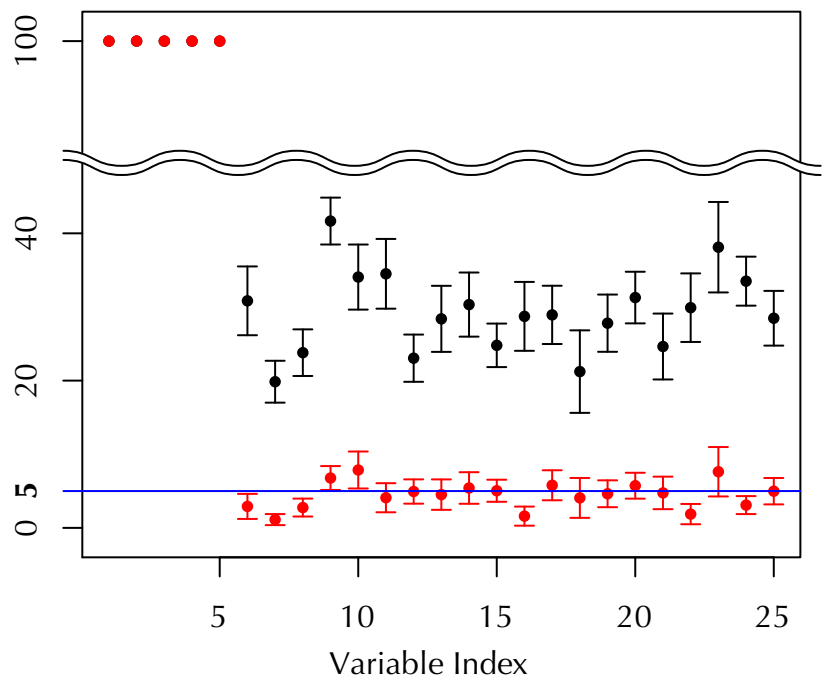
$$\boldsymbol{\beta} = (\underbrace{2, \ldots, 2}_{5}, \underbrace{0, \ldots, 0}_{p-5})^T$$

n=50, p=25

Lasso



MCP



Main results

Terada and Shimodaira (arXiv 2017)

[main result 1] Selective inference via multiscale bootstrap (Terada and Shimodaira 2017)

$$p_{\mathrm{SI}}(H|S,y) = \frac{\bar{\Phi}(\psi_{-1}(H|y))}{\bar{\Phi}(\psi_{-1}(H|y) + \psi_0(S|y))}$$

$$\text{upper tail of } N(0,1)\text{: } \bar{\Phi}(x) = 1 - \Phi(x)$$

$$\text{The normalized bootstrap } z\text{-values for } \sigma^2 > 0 \text{ are}$$

$$\psi_{\sigma^2}(S|y) = \sigma\bar{\Phi}^{-1}(\alpha_{\sigma^2}(S|y))$$

$$\psi_{\sigma^2}(H|y) = \sigma\bar{\Phi}^{-1}(\alpha_{\sigma^2}(H|y))$$

Theorem (large sample theory): If the boundary surfaces of H and S are smooth and "nearly parallel", this p-value is second order accurate with error $O(n^{-1})$.

Theorem (nearly flat surfaces): If the boundary surfaces of H and S are "nearly flat", approaching flat but allowing non-smooth such as cones and polyhedra (Shimodaira 2008), this p-value is justified as unbiased with error $O(\text{lambda}^2)$

"Selective unbiased test" controls the conditional rejection probability

We want to compute a p-value of selective inference for the problem of regions

$$\left.\begin{aligned}
&p\text{-value: } p(H|S,y)\\
&\text{regions: } H,\ S \subset \mathbb{R}^{m+1}\\
&\text{model: } Y \sim N_{m+1}(\mu, I_{m+1})\\
&\frac{P(p(H|S,Y) < \alpha \mid \mu)}{P(Y \in S \mid \mu)} = \alpha, \quad \forall \mu \in \partial H
\end{aligned}\right\}$$

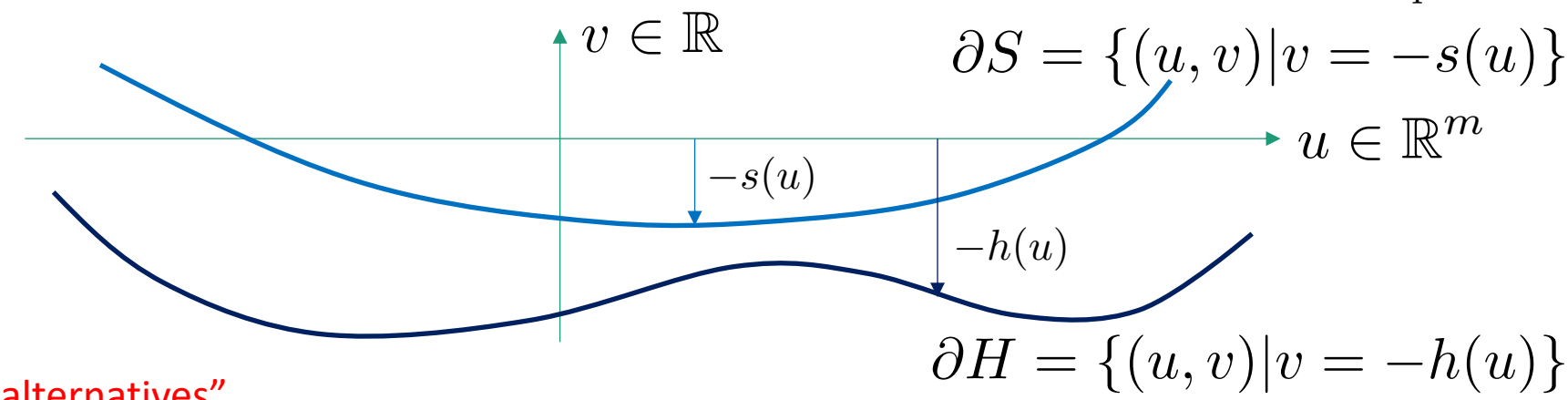
New setting in this work

In the literature, we have been working on non-selective inference for the problem of regions

$$P(p(H|Y) < \alpha \mid \mu) = \alpha, \quad \forall \mu \in \partial H$$

Two asymptotic theories

$\partial S = \partial H$ for pvclust



"local alternatives"

(1) The large sample theory with ``nearly parallel surfaces'' : $n \to \infty$

$$h(u) = h_0 + h_i u_i + h_{ij} u_i u_j + h_{ijk} u_i u_j u_k + \cdots ,$$

$$h_0 = O(1),\ h_i = O(n^{-1/2}),\ h_{ij} = O(n^{-1/2}),\ h_{ijk} = O(n^{-1}), \ldots$$

Quadratic surface

(u,v) are $O(\sqrt{n})$ and $h_0 = O(n^{1/2})$, $h_i = O(1)$, but multiplied by $O(n^{-1/2})$

(2) Another asymptotic theory of ``nearly flat surfaces'' : $\lambda \to 0$

$$\sup_{u \in \mathbb{R}^m} |h(u)| = O(\lambda),\ \int_{\mathbb{R}^m} |h(u)|\, du < \infty,\ \int_{\mathbb{R}^m} |\mathcal{F}h(\omega)|\, d\omega < \infty$$

(Shimodaira 2008)

Higher degree terms, even non-smooth surfaces

Algorithm 1 Computing approximately unbiased p-values

1: Specify several $n' \in \mathbb{N}$ values, and set $\sigma^2 = n/n'$. Set the number of bootstrap replicates B, say, 1000.

2: For each n', perform bootstrap resampling to generate Y^* for B times and compute $\alpha_{\sigma^2}(H|y) = C_H/B$ and $\alpha_{\sigma^2}(S|y) = C_S/B$ by counting the frequencies $C_H = \#\{Y^* \in H\}$ and $C_S = \#\{Y^* \in S\}$. (We actually work on $\mathcal{X}^*_{n'}$ instead of Y^*.) Compute $\psi_{\sigma^2}(H|y) = \sigma\bar{\Phi}^{-1}(\alpha_{\sigma^2}(H|y))$ and $\psi_{\sigma^2}(S|y) = \sigma\bar{\Phi}^{-1}(\alpha_{\sigma^2}(S|y))$.

3: Estimate parameters $\beta_H(y)$ and $\beta_S(y)$ by fitting models

$$\boxed{\psi_{\sigma^2}(H|y) = \varphi_H(\sigma^2|\beta_H)} \text{ and } \boxed{\psi_{\sigma^2}(S|y) = \varphi_S(\sigma^2|\beta_S),}$$

Model fitting to psi

respectively. The parameter estimates are denoted as $\hat{\beta}_H(y)$ and $\hat{\beta}_S(y)$. If we have several candidate models, apply above to each and choose the best model based on AIC value.

4: Approximately unbiased p-values of selective inference (p_{SI}) and non-selective inference (p_{AU}) are computed by one of (A) and (B) below.

$$\sigma^2 = \frac{n}{n'}$$
$$\sigma^2 = 1 \Leftrightarrow n' = n$$
$$\sigma^2 = -1 \Leftrightarrow n' = -n$$
$$\sigma^2 = 0 \Leftrightarrow |n'| = \infty$$

(A) Extrapolate $\psi_{\sigma^2}(H|y)$ and $\psi_{\sigma^2}(S|y)$ to $\sigma^2 = -1$ and 0, respectively, by

$$\boxed{z_H = \varphi_H(-1|\hat{\beta}_H(y))} \text{ and } \boxed{z_S = \varphi_S(0|\hat{\beta}_S(y)),}$$

Extrapolation

and then compute p-values by

Selective p-value $\boxed{p_{\mathrm{SI}}(H|S,y) = \dfrac{\bar{\Phi}(z_H)}{\bar{\Phi}(z_H + z_S)}}$ and $\boxed{p_{\mathrm{AU}}(H|y) = \bar{\Phi}(z_H).}$ Non-selective p-value

(B) Specify $k \in \mathbb{N}$, $\sigma_0^2, \sigma_{-1}^2 > 0$ (e.g., $k = 3$ and $\sigma_{-1}^2 = \sigma_0^2 = 1$). Extrapolate $\psi_{\sigma^2}(H|y)$ and $\psi_{\sigma^2}(S|y)$ to $\sigma^2 = -1$ and 0, respectively, by

$$z_{H,k} = \varphi_{H,k}(-1|\hat{\beta}_H(y), \sigma_{-1}^2) \text{ and } z_{S,k} = \varphi_{S,k}(0|\hat{\beta}_S(y), \sigma_0^2),$$

where the Taylor polynomial approximation of φ_H at $\tau^2 > 0$ with k terms is:

$$\boxed{\varphi_{H,k}(\sigma^2|\hat{\beta}_H(y), \tau^2) = \sum_{j=0}^{k-1} \frac{(\sigma^2 - \tau^2)^j}{j!} \frac{\partial^j \varphi_H(\sigma^2|\hat{\beta}_H(y))}{\partial(\sigma^2)^j}\bigg|_{\sigma^2=\tau^2},}$$

and that of φ_S is defined similarly. Then compute p-values by

$$p_{\mathrm{SI},k}(H|S,y) = \frac{\bar{\Phi}(z_{H,k})}{\bar{\Phi}(z_{H,k} + z_{S,k})} \text{ and } p_{\mathrm{AU},k}(H|y) = \bar{\Phi}(z_{H,k}).$$

For non-smooth surfaces (such as cones and polyhedra)

The theory of nearly flat surfaces for smooth surfaces

There are more terms considered in nearly flat surfaces than the large sample theory (Shimodaira 2008)

$$\psi_{\sigma^2}(H|y) \simeq \beta_0 + \beta_1\sigma^2 + \beta_2\sigma^4 + \beta_3\sigma^6 + \beta_4\sigma^8 + \cdots$$

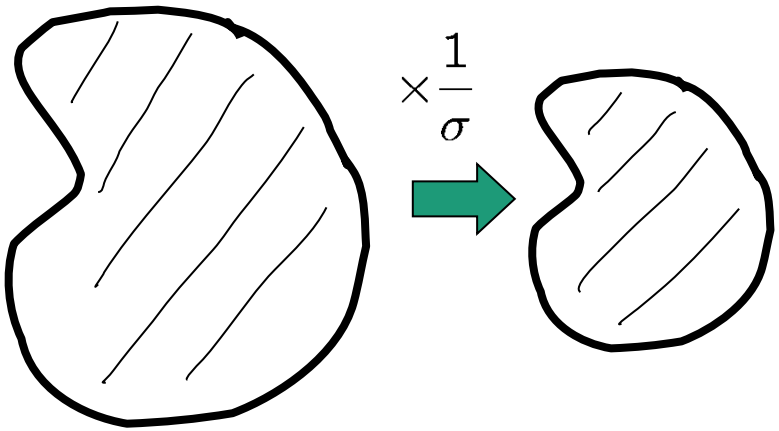
$$p(H|y) \simeq \bar{\Phi}\Big(\beta_0 - \beta_1 + \beta_2 - \beta_3 + \beta_4 + \cdots\Big) \simeq \bar{\Phi}\Big(\psi_{-1}(H|y)\Big)$$

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

'$\simeq$' ignores $O(\lambda^2)$

Note: the expansion series will take a different form for the large sample theory of fourth order-accuracy (Shimodaira 2014)

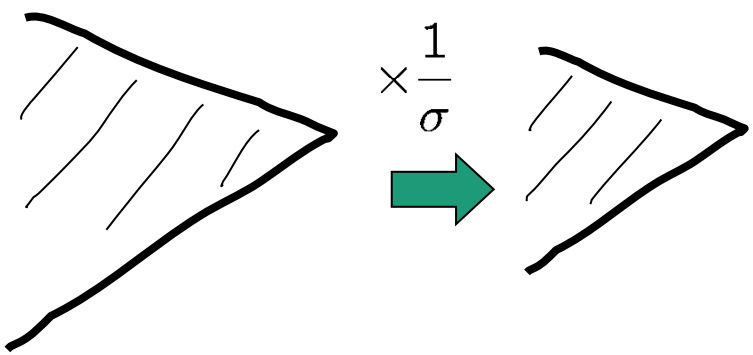
We can't compute psi(-1) for cones



smooth surface

$$\alpha_{\sigma^2}(H|y) = \bar{\Phi}\left(\frac{\beta_0}{\sigma} + \beta_1 \sigma\right)$$

$$\psi_{\sigma^2}(H|y) = \beta_0 + \beta_1 \sigma^2$$



cone

$$\alpha_{\sigma^2}(H|y) = \bar{\Phi}\left(\frac{\beta_0}{\sigma} + \beta_1\right)$$

$$\psi_{\sigma^2}(H|y) = \beta_0 + \beta_1 \sigma$$

$$= \beta_0 + \beta_1 \sqrt{\sigma^2}$$

So, the Taylor series approximation is used

Models of the psi (normalized bootstrap z-value)

$$\varphi_{\text{poly},1}(\sigma^2|\beta) = \beta_0$$

$$\varphi_{\text{poly},2}(\sigma^2|\beta) = \beta_0 + \beta_1\sigma^2$$

$$\varphi_{\text{poly},3}(\sigma^2|\beta) = \beta_0 + \beta_1\sigma^2 + \beta_2(\sigma^2)^2$$

Smooth surface

$$\varphi_{\text{sing},3}(\sigma^2|\beta) = \beta_0 + \frac{\beta_1\sigma^2}{1+\beta_2(\sigma-1)}$$

$$\varphi_{\text{sing},3}(\sigma^2|\beta) = \beta_0 + \beta_1\sqrt{\sigma^2}, \quad \beta_2 = 1$$

Nonsmooth surface

(This form comes from the fact that cones are scale invariant)

Extrapolation using Taylor polynomial approximation

Model fitting

$$\varphi_H(\sigma^2) = \varphi_H(\sigma^2|\hat{\beta}_H(y)) \text{ is fitted to } \psi_{\sigma^2}(H|y)$$

Taylor expansion

$$\varphi_{H,k}(\sigma^2|\tau^2) = \sum_{j=0}^{k-1} \frac{(\sigma^2-\tau^2)^j}{j!} \frac{\partial^j \varphi_H(\sigma^2)}{\partial(\sigma^2)^j}\bigg|_{\sigma^2=\tau^2}$$

$$k = 3, \sigma_0^2 = \sigma_{-1}^2 = 1$$

Extrapolation to sigma^2 = -1 or 0

$$\varphi_{H,k}(-1|\sigma_{-1}^2) \text{ replaces } \varphi_H(-1)$$

$$\varphi_{S,k}(0|\sigma_0^2) \text{ replaces } \varphi_S(0)$$

[main result 2] Taylor series extrapolation and iterated bootstrap

Extrapolation using Taylor polynomial approximation; B bootstrap replicates

$$p_{\mathrm{SI},k}(H|S,y) = \frac{\bar{\Phi}(\varphi_{H,k}(-1|\sigma^2_{-1}))}{\bar{\Phi}(\varphi_{H,k}(-1|\sigma^2_{-1}) + \varphi_{S,k}(0|\sigma^2_0))}$$

Bootstrap iteration; B^k bootstrap replicates

$$p_{\mathrm{BP},1}(H|S,y) = \frac{\bar{\Phi}(\psi_{\sigma^2}(H|y))}{P_1(Y^* \in S \mid \hat{\mu})}, \quad \sigma^2 > 0$$

$$p_{\mathrm{BP},k+1}(H|S,y) = \frac{P_1(p_{\mathrm{BP},k}(H|S,Y^*) < p_{\mathrm{BP},k}(H|S,y) \mid \hat{\mu}(H|y))}{P_1(Y^* \in S \mid \hat{\mu}(H|y))}$$

Theorem (large sample theory): p_{SI} is equivalent to $p_{BP,2}$ with error $O_p(n^{-1})$

Theorem (nearly flat surfaces): $p_{SI,k}$ and $p_{BP,k}$ are asymptotically unbiased as k goes infinity. They are unbiased if h and s are polynomials of degree 2k-1 or less. Here unbiasedness allows error of O(lambda^2).

Related works

- The selective inference is a direct extension of the previous result on multiscale bootstrap developed in Shimodaira (2002, 2004, 2008, 2014) for non-selective inference
- No result of selective inference for the problem of regions in the literature. In selective inference literature, variable selection in regression has been discussed
- Our large sample theory is based on Efron (1985), Efron and Tibshirani (1998) with geometry of signed distance and curvature
- Multiscale bootstrap adjusts the bias caused by curvature of the boundary surface. This is easier than the previous methods. BCa (Efron 1987) needs access to parameter. ET1998 needs the null distribution.

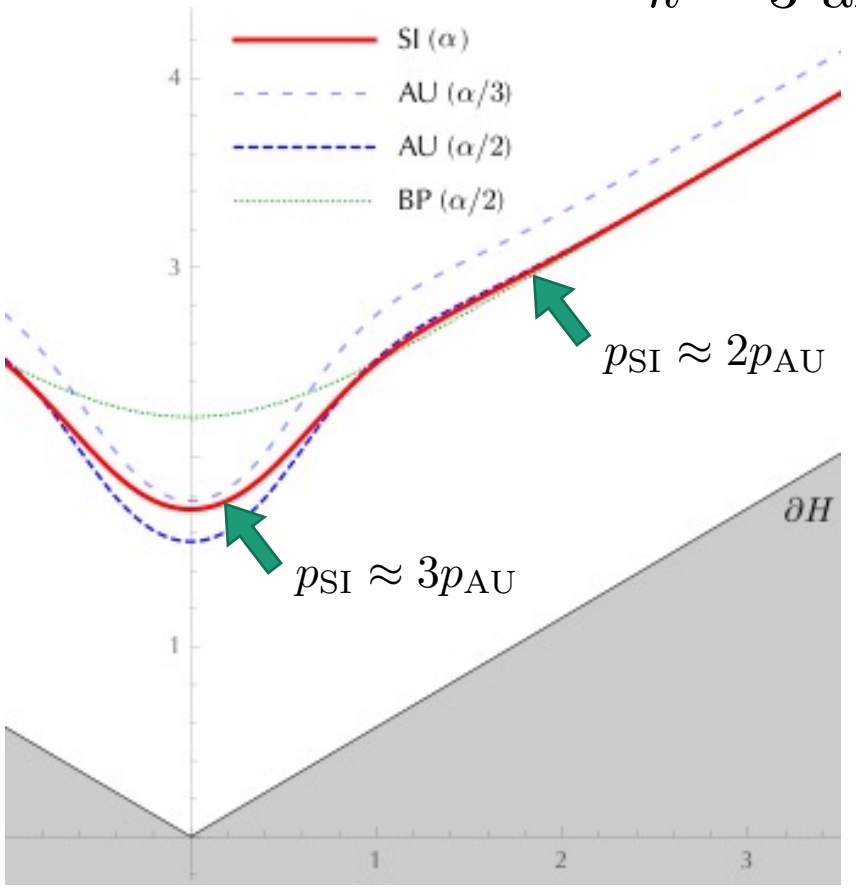
Related works (cont.)

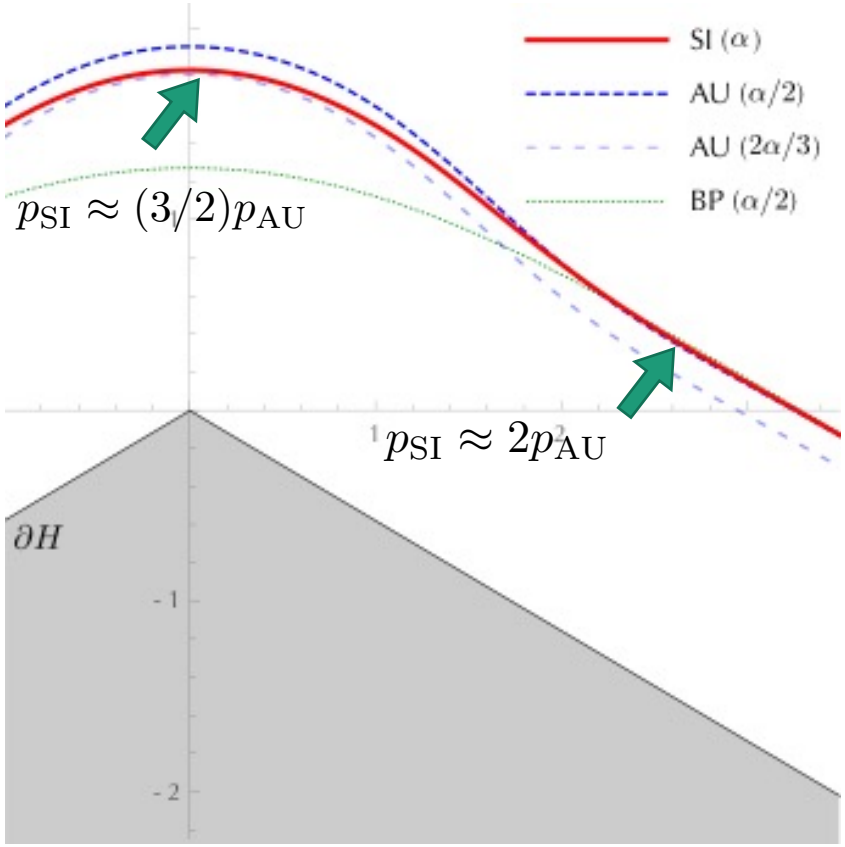
- Exponential family is considered instead of normal in Shimodaira (2004). The “acceleration constant” of BCa is estimated by another mechanism (two-step bootstrap), but the effect is not so large compared to the curvature. So we ignore it in this work (a=0 in normal distribution).
- Multiple comparisons procedures are conservative, and our approximately unbiased approach improves the power (but type-I error is controlled only approximately). No unbiased test exists for cones (Lehmann 1952). Counter-intuitive examples in Perlman and Wu (1999).
- Large sample theory scales all parameters by sqrt(n) which works only for smooth surface. Our nearly flat surface theory (Shimodaira 2008) scales only the normal direction and fixing the tangent directions; it works also for non-smooth surfaces.

Examples in $\mathbb{R}^2$

SI adjusts AU by the selection probability

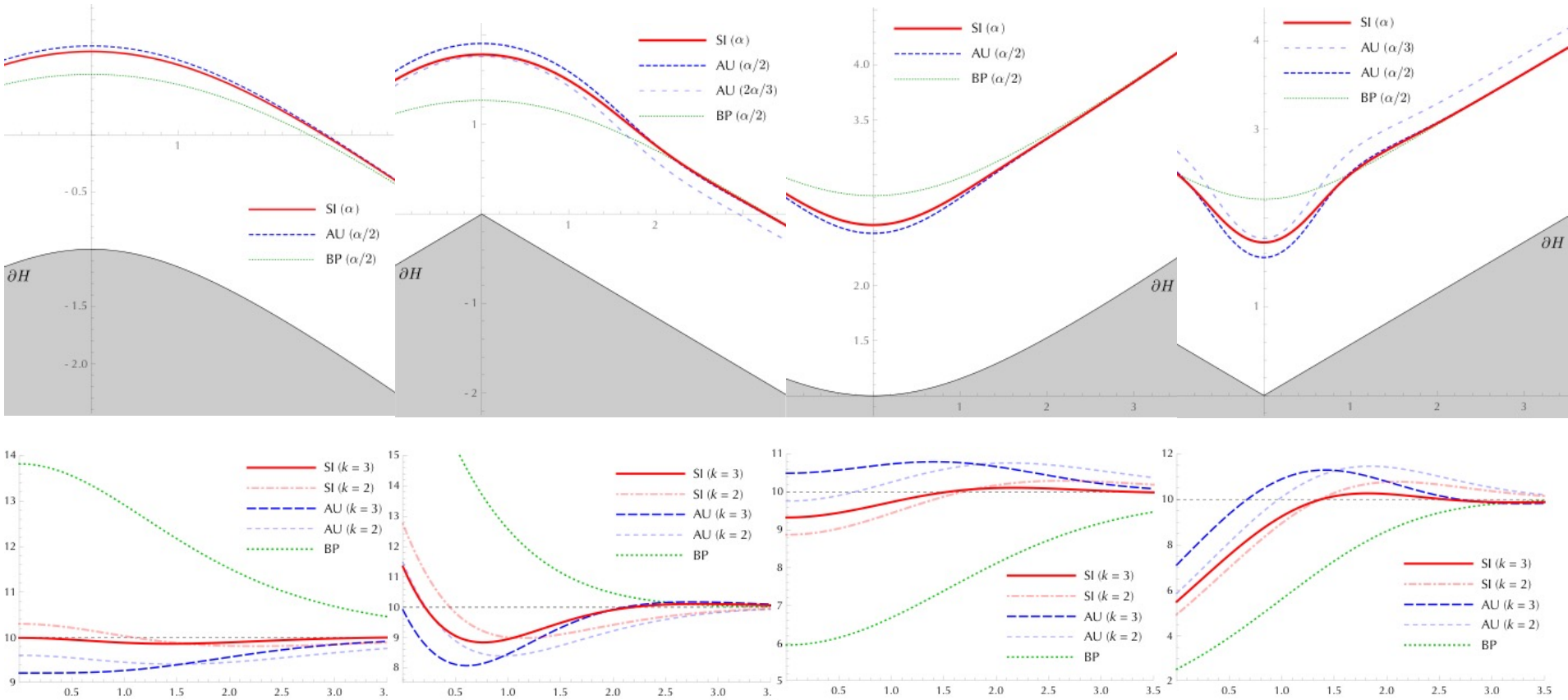
$k = 3$ and $\alpha = 0.1$





Rejection surfaces (p = const)

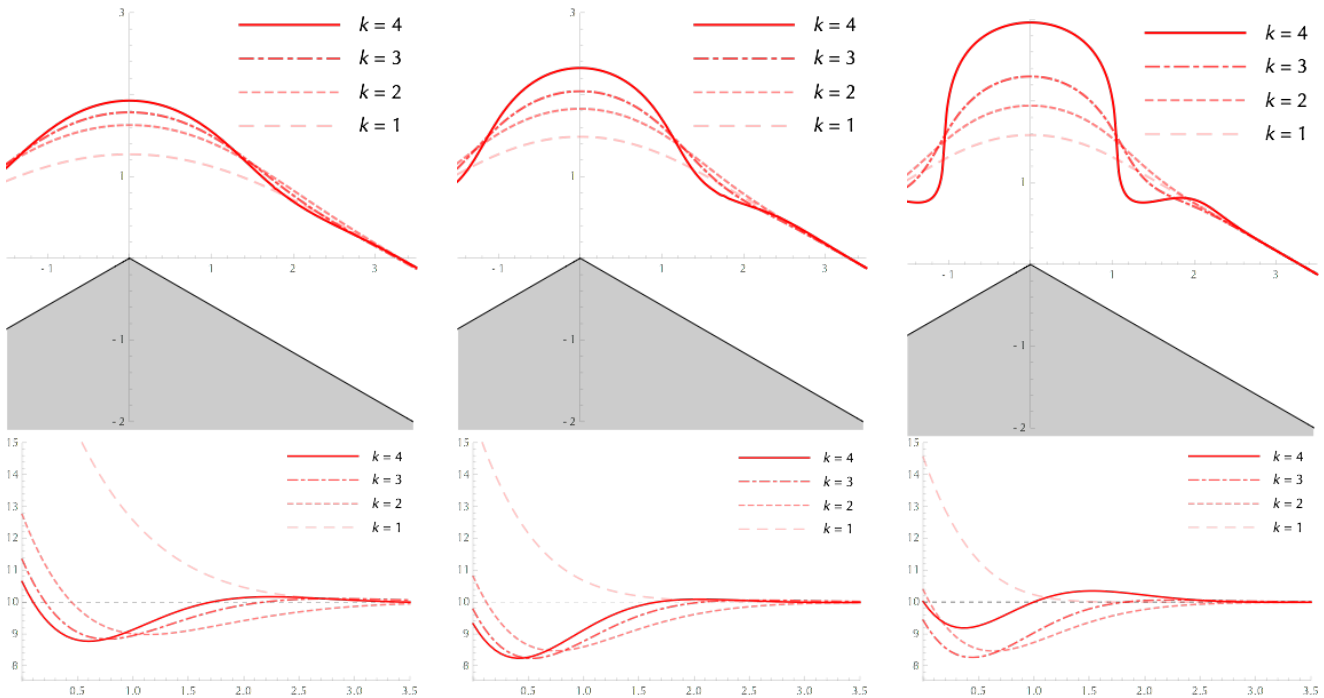
p_{SI} is approximately unbiased, and better than $2p_{AU}$



Note: For p of one-tailed test, 2p is the p-value for two-tailed test

Bias reduces by increasing k, but the shape behaves bad (convex)

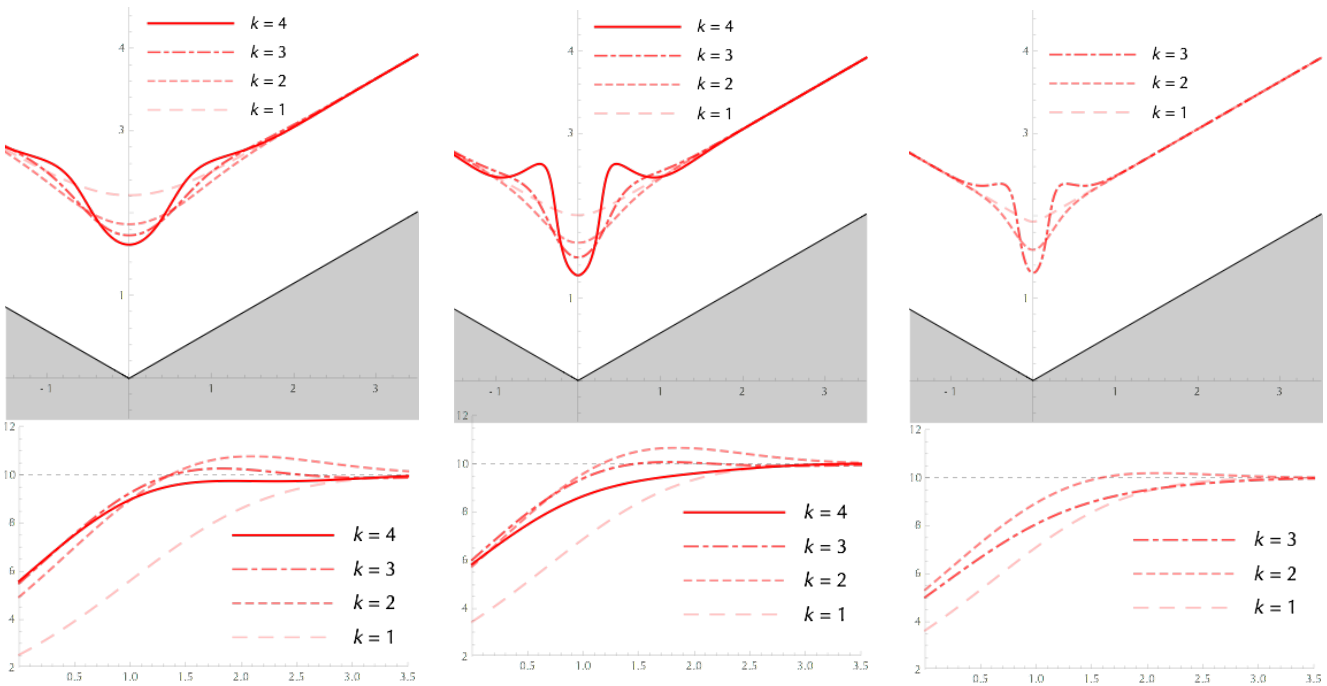
$\sigma_0^2 = \sigma_{-1}^2 = 1$ $\sigma_0^2 = \sigma_{-1}^2 = 1/2$ $\sigma_0^2 = \sigma_{-1}^2 = 1/4$



	$k = 1$	$k = 2$	$k = 3$	$k = 4$
1	2.077	0.618	0.432	**0.380**
1/2	0.825	0.610	0.561	**0.489**
1/4	0.515	0.544	0.505	**0.239**

Bias reduces by increasing k, but the shape behaves bad (concave)

$\sigma_0^2 = \sigma_{-1}^2 = 1$ $\sigma_0^2 = \sigma_{-1}^2 = 1/2$ $\sigma_0^2 = \sigma_{-1}^2 = 1/4$



	$k = 1$	$k = 2$	$k = 3$	$k = 4$
1	2.730	1.239	**0.855**	0.960
1/2	1.971	0.909	**0.690**	1.030
1/4	1.848	**0.942**	1.370	–

The Emperor's New Tests

Perlman and Wu (1999)

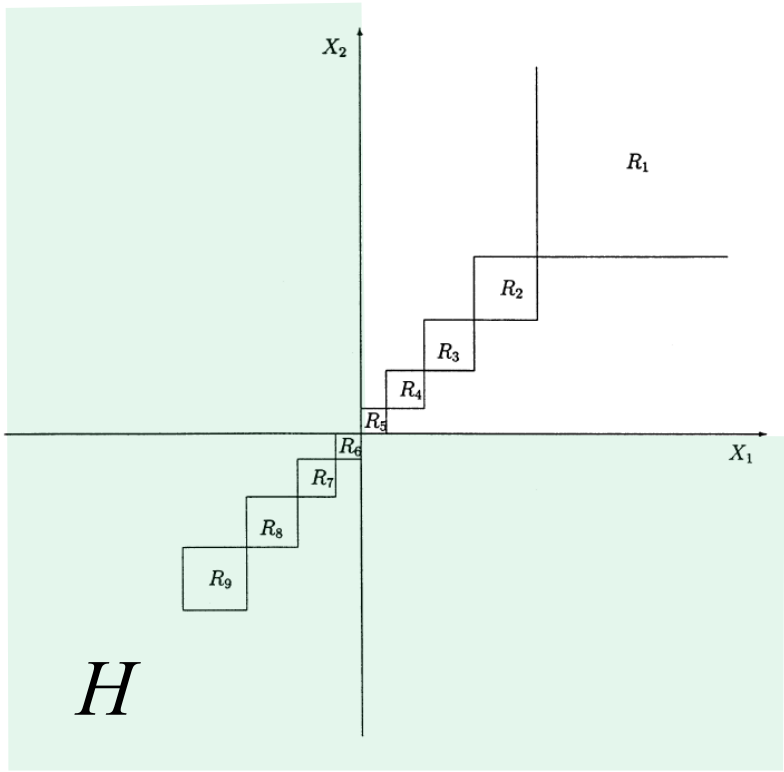


FIG. 1. *The rejection region for the LRT for testing problem* (6) *is* R_1. *The rejection regions for Berger's tests I and II are* $R_1 \cup R_2 \cup \cdots \cup R_5$ *and* $R_1 \cup R_2 \cup \cdots \cup R_9$, *respectively.*

Perlman and Wu (2003)

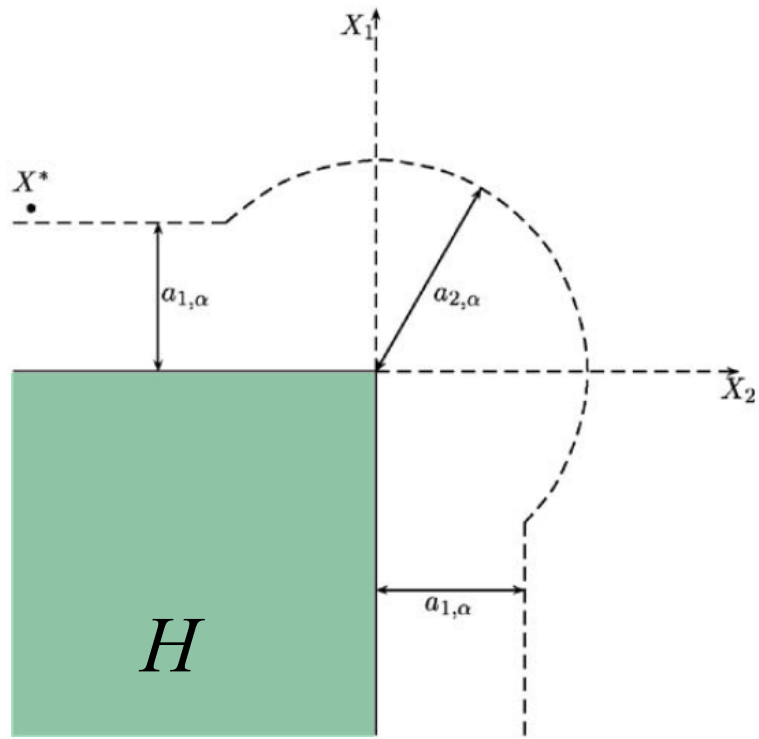


Fig. 11. The size α adaptive test (64) for (54) $\equiv$ (55) ($p = 2$, $\alpha = 0.05$, $a_{1,\alpha} = 1.64$, $a_{2,\alpha} = 2.33$).

Note: No unbiased test exists for cones (Lehmann 1952)

Final Remarks

- We developed selective inference p-values for the problem of regions
- Multiscale bootstrap computes the p-values
- The p-values are justified by two asymptotic theories: the large sample theory, and the nearly flat surfaces
- For non-smooth surfaces, the choice of k and $sigma_0$, $sigma_{-1}$ would be an issue. We recommend k=2 or 3, $sigma_0 = sigma_{-1} = 1$
- The method seems working ok in pvclust and also in the simple simulations
- We have just started the research on selective inference . Any comments are very welcome

Thank you

References

- Y. Terada, H. Shimodaira (2019). Selective inference after variable selection via multiscale bootstrap. arXiv:1905.10573.
- H. Shimodaira, Y. Terada (2019). Selective Inference for Testing Trees and Edges in Phylogenetics. Frontiers in Ecology and Evolution 7, 174.
- K. Liu, J. Markovic, R. Tibshirani (2018). More powerful post-selection inference, with application to the Lasso. arXiv:1801.09037.
- Y. Terada, H. Shimodaira (2017). Selective inference for the problem of regions via multiscale bootstrap. arXiv:1711.00949.
- J. Lee, D. Sun, Y. Sun, J. Taylor (2016). Exact post-selection inference, with application to the lasso. Ann. Statist. 44, 907-927.
- W. Fithian, D. Sun, J. Taylor. (2014). Optimal Inference After Model Selection. arXiv:1410.2597.
- H. Shimodaira (2014). Higher-order accuracy of multiscale-double bootstrap for testing regions. Journal of Multivariate Analysis 130, 208-223.
- H. Shimodaira (2008). Testing regions with nonsmooth boundaries via multiscale bootstrap. Journal of Statistical Planning and Inference 138, 1227-1241.
- R. Suzuki, H. Shimodaira (2006). Pvclust: an R package for assessing the uncertainty in hierarchical clustering. Bioinformatics 22, 1540-1542.
- H. Shimodaira (2004). Approximately unbiased tests of regions using multistep-multiscale bootstrap resampling. The Annals of Statistics 32, 2616–2641.
- H. Shimodaira (2002). An approximately unbiased test of phylogenetic tree selection Systematic biology 51, 492-508.
- H. Shimodaira, M. Hasegawa (2001). CONSEL: for assessing the confidence of phylogenetic tree selection Bioinformatics 17, 1246-1247.
- H. Shimodaira, M. Hasegawa (1999). Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Molecular biology and evolution 16, 1114-1116.
- B. Efron, R. Tibshirani (1998). The problem of regions. Ann. Statist., 26, 1687-1718.
- B. Efron, E. Halloran, S. Holmes (1996) Bootstrap confidence levels for phylogenetic trees. PNAS 93, 13429-13434.