

頻度論とベイズをつなぐ統計的信頼度

下平英寿

大阪大学 基礎工学研究科

理化学研究所 革新知能統合研究センター

2016年11月16日

第19回情報論的学習理論ワークショップ (IBIS2016)

最近の様子@下平研

深層クロスドメインマッチング相関分析の提案とその応用

羽田 哲也

heda@sigmath.es.osaka-u.ac.jp

福井 一輝

fukui@sigmath.es.osaka-u.ac.jp

下平 英寿

shimo@sigmath.es.osaka-u.ac.jp

大阪大学大学院 基礎工学研究科 システム創成専攻 数理科学領域 下平研究室

クロスドメインマッチング相関分析 (CDMCA) [1]

問題設定

分種のドメインにわたるデータを共通の K 次元空間へ射影する。このとき、関連するデータ間の対応関係(類似度)を最大化する(マッピング関数)によって、対応した次元を持つような類似度関数を提案する。



データ行列 (類似): $X^1 \in \mathbb{R}^{n_1 \times K}$
 マッピング関数 (類似): $W^1 \in \mathbb{R}^{n_1 \times K}$
 類似度行列 (類似): $A^1 \in \mathbb{R}^{n_1 \times n_1}$

$$\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n W_{ij}^1 (A_{ij}^1 - (W^1)^T X_{ij}^1)^2 \text{ subject to } \sum_{i=1}^n (A^1)^T G^1 A^1 = I_K$$

マッピング関数

$$G^1 = (X^1)^T X^1 + \epsilon I_{n_1}, \quad M^1 = \text{diag}(G^1, \dots, G^1)$$

解決

$D=3$ の場合について説明する。次のような設定を導入する。

$$X = \begin{pmatrix} X^1 \\ X^2 \\ X^3 \end{pmatrix}, \quad W = \begin{pmatrix} W^1 \\ W^2 \\ W^3 \end{pmatrix}, \quad M = \begin{pmatrix} M^1 & 0 & 0 \\ 0 & M^2 & 0 \\ 0 & 0 & M^3 \end{pmatrix}$$

$$A = \begin{pmatrix} A^1 \\ A^2 \\ A^3 \end{pmatrix}, \quad H = X^T W X, \quad G = X^T M X$$

このとき、CDMCA は次のような最適化問題に帰着する。

$$\text{maximize } \text{tr}(A^T H A) \text{ subject to } A^T G A = I_K$$

$$\text{最適解 } \sum_{i=1}^n \lambda_i, \quad \lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \dots \geq \lambda_K \geq 0 \text{ の } G^{-1/2} H G^{-1/2} \text{ の最大固有値}$$

$$\text{最適解 } A = G^{-1/2} (U_1, \dots, U_K), \quad U_1, \dots, U_K \in \mathbb{R}^{n_1 \times 1} \text{ の } G^{-1/2} H G^{-1/2} \text{ の最大固有値}$$

提案法: 深層クロスドメインマッチング相関分析 (DCDMCA)

CDMCA をよりよくしたい。ニューラルネットワークを用いた非線形射影と CDMCA を組み合わせることで、マッピング関数を柔軟に学習できるような共通空間への射影が得られる。DCDMCA はニューラルネットワークを用いたデータの射影を非線形射影として CDMA を行うことを提案する。



各ドメインにわたるニューラルネットワーク $f^D: \mathbb{R}^{n_D} \rightarrow \mathbb{R}^K$

非線形射影されたデータを CDMCA で射影変換

$$Z^D = f^D(X^D) A^D$$

線形変換 A^D と非線形変換 f^D の最適化

- $f^D(X^1), \dots, f^D(X^K)$ を入力としたときの CDMCA の最適化を目的関数として f^D を最適化する。
- CDMCA の最適化は、計算コストの低い固有値問題を求解 Frobenius ノルムに依存する。
- Frobenius ノルム $\|S\|_F = \sqrt{\sum_{i=1}^n \sum_{j=1}^n S_{ij}^2}$ は平方根を取った絶対値の総和 $\sum_{i=1}^n \sum_{j=1}^n |S_{ij}|$ に近い。
- $f^D(X^1), \dots, f^D(X^K)$ のドメインが決定したら、通常の CDMCA と同様に $\text{maximize } \text{tr}(S^T S)$ を計算する。

$$\text{maximize } \text{tr}(S^T S) \text{ subject to } S^T S = I_K$$

人工データを用いた実験



MCCA [2] CDMCA DCMCA DCDMCA

- 人工的に生成した3つのドメインにおけるデータ共通空間への射影結果。
- CDMCA と非線形射影を組み合わせると、マッピング関数がより小さくなった。

画像-マルチメディア間の相互検索への応用

実験の目的

- 画像と、それに関連付けられているタグデータのドメインとして、画像ドメインとタグドメインを共通空間に射影し、ドメインをまたいだ相互検索を行う。

NUS-WIDE: マルチメディア検索データセット

- Flickr を使ったデータセット: 画像数 221000枚、タグ数 13194枚
- 画像種別ごとの分類数: 134種
- タグの出現頻度に基づいて Top30 (3000回) を用いた。

ニューラルネットワークの設定

- 画像ドメイン: 単語検索 GooNet (4) + ノード数 1024, 1024, 300 の逆伝播学習 (即座2層)
- タグドメイン: ノード数 300, 1024, 300 の逆伝播学習
- 学習は両方ともすべて ReLU を用いた。

実験結果

method	mAP	random
random	1.0	
CCA	48.4	
CDMCA	60.0	
DCCA (5)	100.6	
DCDMCA	102.5	

共有空間の次元 = 300 として、画像クエリからタグへの相互検索を行った。

DCDMCA が DCCA を上回る性能を達成することが確認された。

クエリ画像: bear, panda, forest, pandas, xoo

クエリ画像: wonderful, flowers, flower, petals, blooms

クエリタグ: "car"

クエリタグ: "cat"



図解

CDMCA は NVIDIA から提供された GPU Tesla M40, GeForce GTX T100 を用いて行われた。

参考文献

[1] Hoshino, T., Shimizu, T., Chen, K., Corrado, G. S., & Shiohata, S. (2015). Distributed learning representations with phrases and their composability. In Advances in neural information processing systems (pp. 3113-3121).

D2-17

クロスドメインマッチング相関分析を用いた単語と画像の同時埋め込み

大阪大学大学院 基礎工学研究科 システム創成専攻 数理科学領域

福井 一輝

押切 孝将

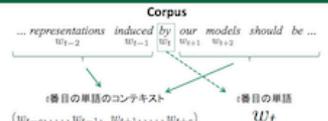
下平 英寿

fukui@sigmath.es.osaka-u.ac.jp oshikiri@sigmath.es.osaka-u.ac.jp shimo@sigmath.es.osaka-u.ac.jp

1. はじめに

単語の分類的意味表現では、コーパス内の単語の共起情報から、コーパス内の各単語を、その意味を表現する実ベクトルに変換する。分類的意味表現の既存手法には skip-gram model [1] などがあり、これらによって埋め込まれた単語ベクトルは、画像ベクトル zero-shot learning やテキスト分類等に使用されている。

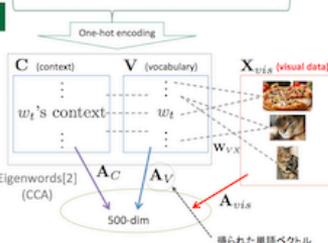
本研究では、正準相関分析(CCA)にもとづいた単語の埋め込み手法を提案することにより、豊富な情報を反映した単語ベクトルを得る手法を提案する。提案手法によって得られた単語ベクトルを用いることで、単語からの画像検索や画像・単語間の演算結果にもとづいた検索などを行うことが可能となる。



2. 画像情報を組み合わせた単語埋め込み

コーパスから得られる単語ベクトルの場合、テキストから得られる情報のみが反映される。ここに、視覚的な情報を加える手法は [Lazaridou et al. IJCAI 2015] や [Kotler et al. CVPR 2016] において提案されている。これらの手法では skip-gram model の拡張や、neural network にもとづいたモデルが提案されている。

本研究では Dhillon [6] によって提案された Eigenwords [2] を拡張する。Eigenwords では one-hot encoding で表現された単語と単語のコンテキストを、CCA を用いて低次元空間へ埋め込むことで単語ベクトルが得られる。本研究の提案手法では、CCA の一般化である任意サイズの単語ベクトル間の対応関係の情報は、CDMCA は任意の種類のデータ間で多対多の対応関係を扱うことができるため、Eigenwords のモデルに画像と画像・単語間の対応関係の情報を付加することができる。単語のデータ行列を V 、コンテキストを C 、画像特徴量を X とし、画像と単語の対応関係の強さを表す行列を $W_{V \times X}$ とする。また、以下の行列を定義する。



$$A^T = \begin{pmatrix} A^C & A^V & A^X \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad G = \begin{pmatrix} V^T \text{diag}(I + W_{V \times X}) V & 0 & 0 \\ 0 & C^T C & 0 \\ 0 & 0 & X^T \text{diag}(W_{V \times X}) X \end{pmatrix}$$

$$H = \begin{pmatrix} C^T V & V^T C & V^T W_{V \times X} X \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \square \text{だけだと Eigenwords [2] と等価}$$

このとき、CDMCA における最適化問題は下記のようになる。この式は一般化固有値問題として解くことができ、得られた A が単語ベクトルとなる。同時に、画像の特徴空間から単語ベクトルの空間への線形変換 A_{vis} が得られる。

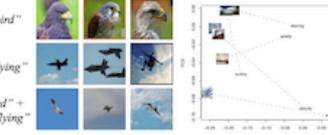
$$\max_A \text{Tr}(A^T H A) \text{ s.t. } A^T G A = I$$

2.1. 実験1: Word Similarity Task

提案手法によって得られた単語ベクトルを評価するために、word similarity task の実験を行った。実験には text コーパスと Flickr 画像データセット (NUS-WIDE [4]) からランダムにサンプリングした 10 枚の画像データを用いた。画像特徴量には AlexNet [5] の出力レイヤから抽出した画像特徴量 (4096 次元) を使用した。また、画像と単語との対応関係の情報は、CDMCA に付随する数万種類の Flickr タグの情報を利用した。このタスクでは、MEN [6] と (Silberer & Lapata 2014) [7] に提供されている。人間によって付けられた単語ペアの類似度のスコアと単語ベクトルを用いて得られた単語ペアの類似度とで、スピアマンの順位相関係数 [8] に近いほど良いと評価する。

Method	MEN [6]	Semantic [7]	Visual [7]
Eigenwords	0.583	0.455	0.345
提案手法	0.289	0.157	0.168

全てのベンチマークにおいて、提案手法の順位相関係数が Eigenwords よりも高かった。これは、単語のドメインに画像を組み込むことで、単語間の類似性を反映した構造が得られたという点と、多義性のある単語と関連付けられたカテゴリの異なる画像による影響が原因として考えられる。



3. 参考文献

[1] Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In Advances in neural information processing systems (pp. 3111-3119).

[2] Dhillon, P. S., Foster, D. J., & Shalizi, C. (2011). Eigenwords: Spectral word embeddings. The Journal of Machine Learning Research, 12(1), 805-828.

[3] Dhillon, P. S. (2016). Cross-validation of matching correlation analysis by resampling matching weights. Neural Networks, 75, 126-140.

[4] Chek, T. S., Hong, S. H., Lee, J., Lim, S., & Sheng, Y. (2006, July). NUS-WIDE: A real-world web image database from National University of Singapore. In Proceedings of the ACM international conference on image and video retrieval (pp. 482-491).

[5] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. In Advances in neural information processing systems (pp. 1097-1105).

[6] Brown, P., Pei, K. L., & Barzilay, M. (2014). Multiscale Distributed Semantics. AACL/IJCAI, 482-491.

[7] Silberer, C., & Lapata, M. (2014, June). Learning Distributed Meaning Representations with Autoencoders. In ACL (pp. 712-721).

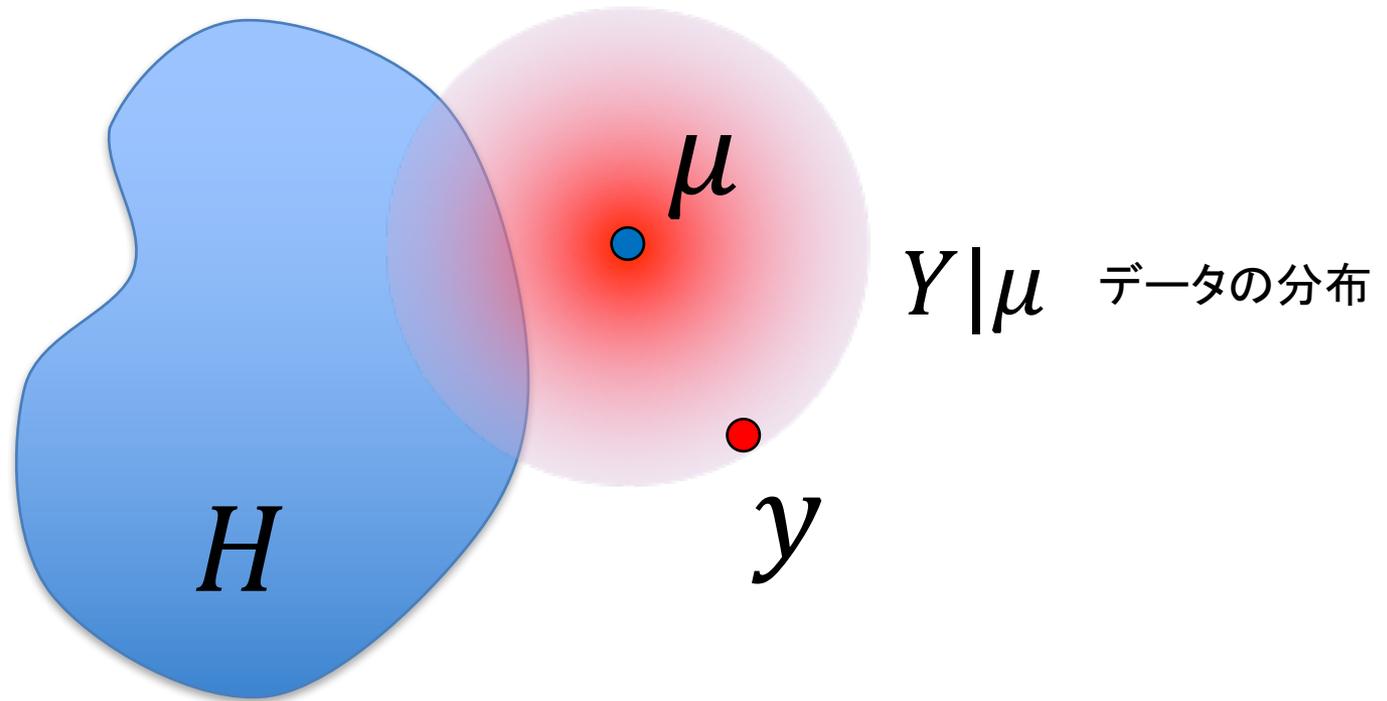
[8] Spearman, K., Galambos, R. A., & Munn, S. (1951). Computing rank correlations from multimodal neural language models. IJCL.

本日のトーク： 地味にやってる話

- Multiple comparisons of log-likelihoods with applications to phylogenetic inference. H Shimodaira, M Hasegawa. *Molecular biology and evolution* 16, 1114-1116, 1999
- An approximately unbiased test of phylogenetic tree selection. H Shimodaira. *Systematic biology* 51 (3), 492-508, 2002
- Approximately unbiased tests of regions using multistep-multiscale bootstrap resampling. H Shimodaira. *The Annals of Statistics* 32, 2004
- Testing regions with nonsmooth boundaries via multiscale bootstrap. H Shimodaira. *Journal of Statistical Planning and Inference* 138, 1227-1241, 2008
- Higher-order accuracy of multiscale-double bootstrap for testing regions. H Shimodaira. *Journal of Multivariate Analysis* 130, 208-223, 2014

マルチスケール・ブートストラップ法による信頼度計算の研究

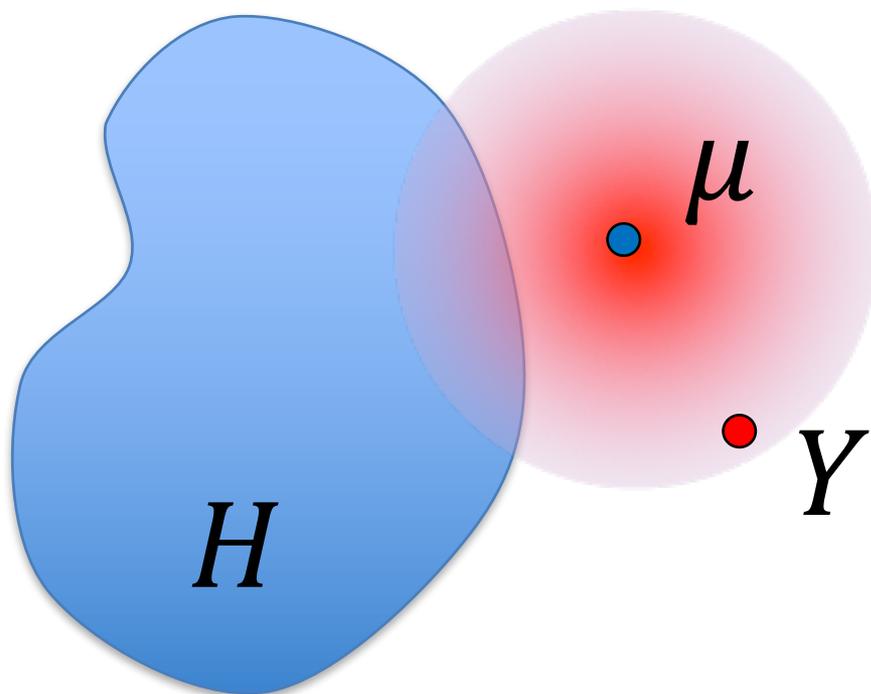
領域の検定と信頼度



$$\text{confidence}(\mu \in H | y)$$

信頼度が小さい時, 仮説を棄却する

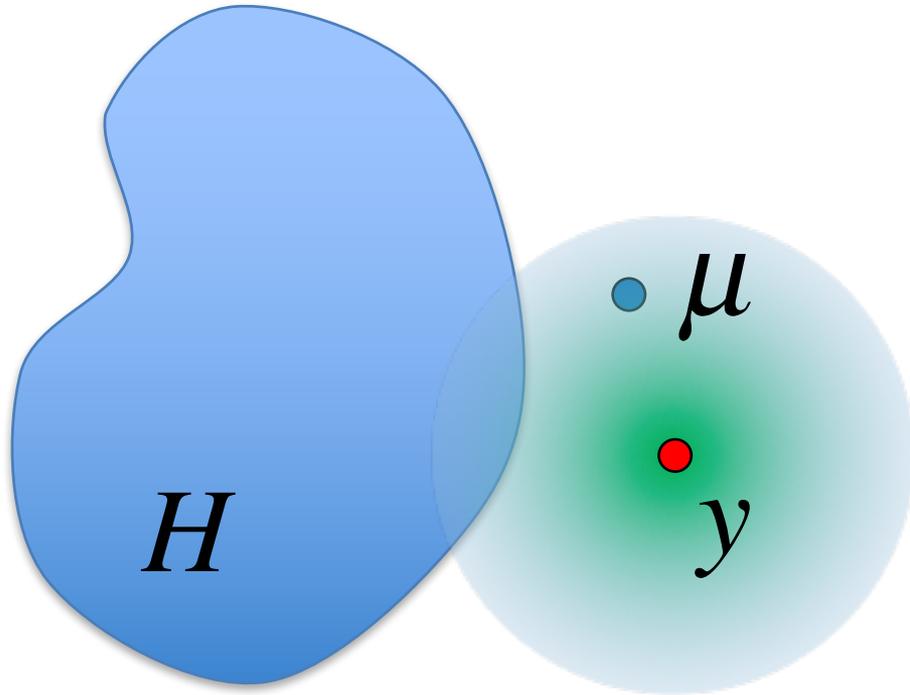
多変量正規モデルで考える



$$Y | \mu \sim N(\mu, I)$$

$$\text{density} \propto \exp(-\|y - \mu\|^2 / 2)$$

ベイズ事後確率は容易に計算



$$Y|\mu \sim N(\mu, I)$$

$$\mu \sim \text{uniform}$$

$$\text{density} \propto \exp(-\|y - \mu\|^2 / 2)$$

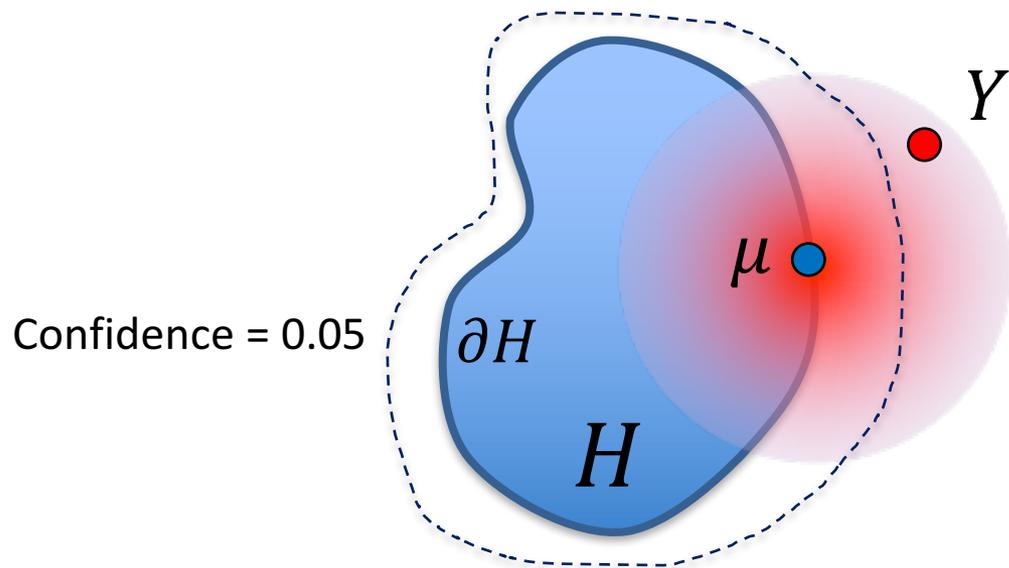
$$\mu | y \sim N(y, I)$$

muの事後分布 = ブートストラップ分布

$$\text{confidence}_{\text{Bayes}}(\mu \in H | y) = P(\mu \in H | y)$$

Hの事後確率 = ブートストラップ確率

頻度論のp-値が満たすべき性質



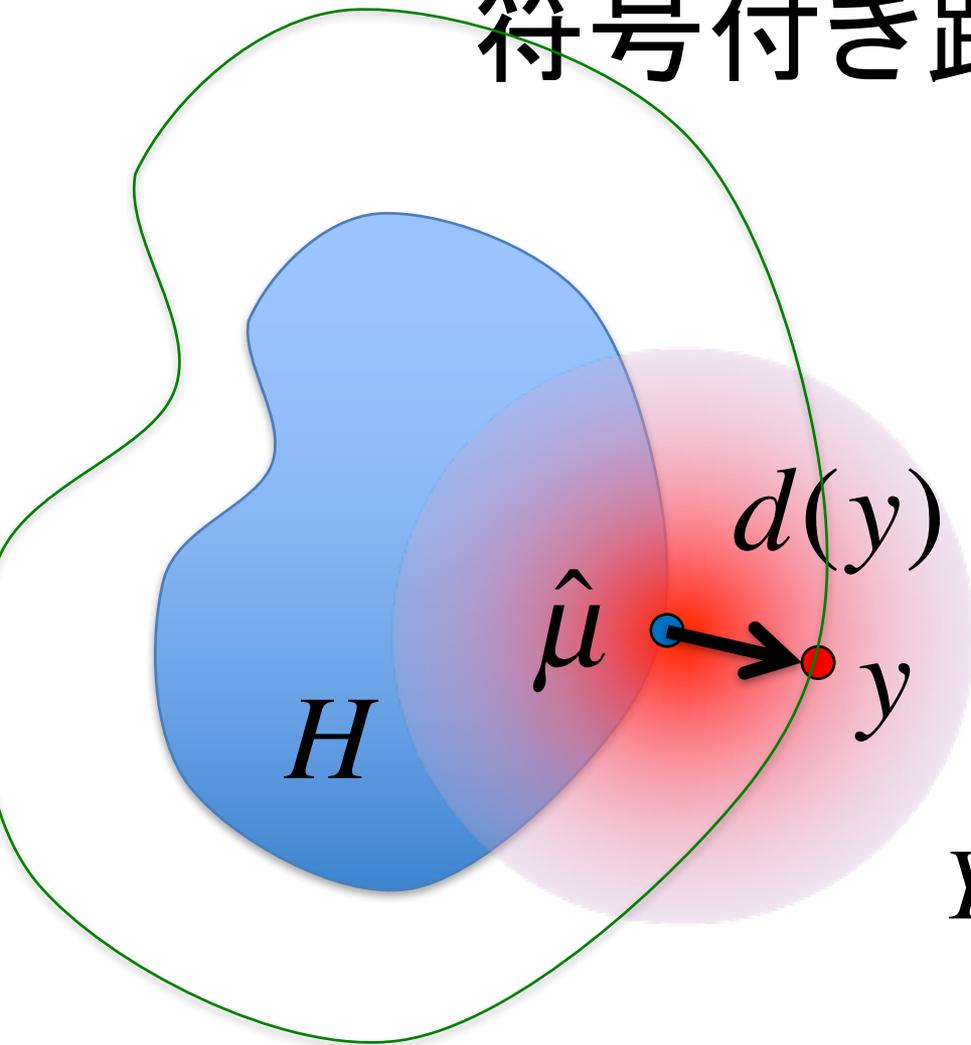
Type-I errorを有意水準(=0.05)で抑えたい。しかし検出力を上げるには等号に近い方が良い。

$$P[\text{confidence}_{freq}(\mu \in H|Y) < 0.05] \leq 0.05 \text{ under } \underline{\mu \in H}$$

不偏検定では、仮説境界上でType-I errorを有意水準と同じにしたい

$$P[\text{confidence}_{freq}(\mu \in H|Y) < 0.05] = 0.05 \text{ under } \underline{\mu \in \partial H}$$

符号付き距離を使う

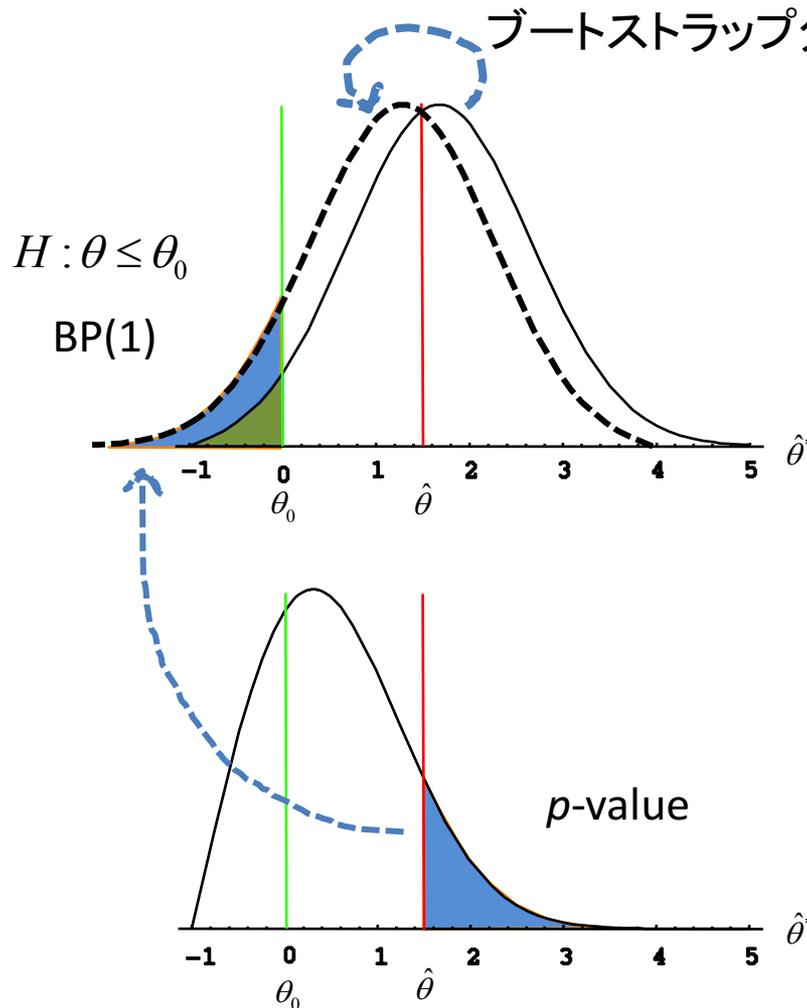


符号付き距離
(signed distance)

$$Y | y \sim N(\hat{\mu}, I)$$

$$\text{confidence}_{freq}(\mu \in H | y) = P(d(Y) > d(y) | \hat{\mu}) + O_p(n^{-\frac{3}{2}})$$

もしパラメータにアクセスすれば簡単



データからブートストラップ分布

帰無仮説からシミュレーション

ベイズ事後確率はp-値を近似する

多変量正規モデル, 平坦な事前分布, 仮説領域の境界が滑らかであることを仮定すると

$$\text{confidence}_{freq}(\mu \in H|Y) = \text{confidence}_{Bayes}(\mu \in H|Y) + O_p(n^{-\frac{1}{2}})$$

たしかに近似してるけど, 誤差が大きい. 誤差の主な原因は, 仮説境界の「平均曲率」.

頻度論の意味でp-値を追求してみる.

ブートストラップ法で計算したい. カウント値だけをつかう.

多重比較との関連.

たとえば、モデル選択

- 多項式回帰モデルの次数選択
- 適切なモデルをAIC最小化法で選ぶ

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon$$

各領域がモデルに相当

多項式回帰の次数選択: 次数=0, 1, 2から選ぶ

1688

B. EFRON AND R. TIBSHIRANI

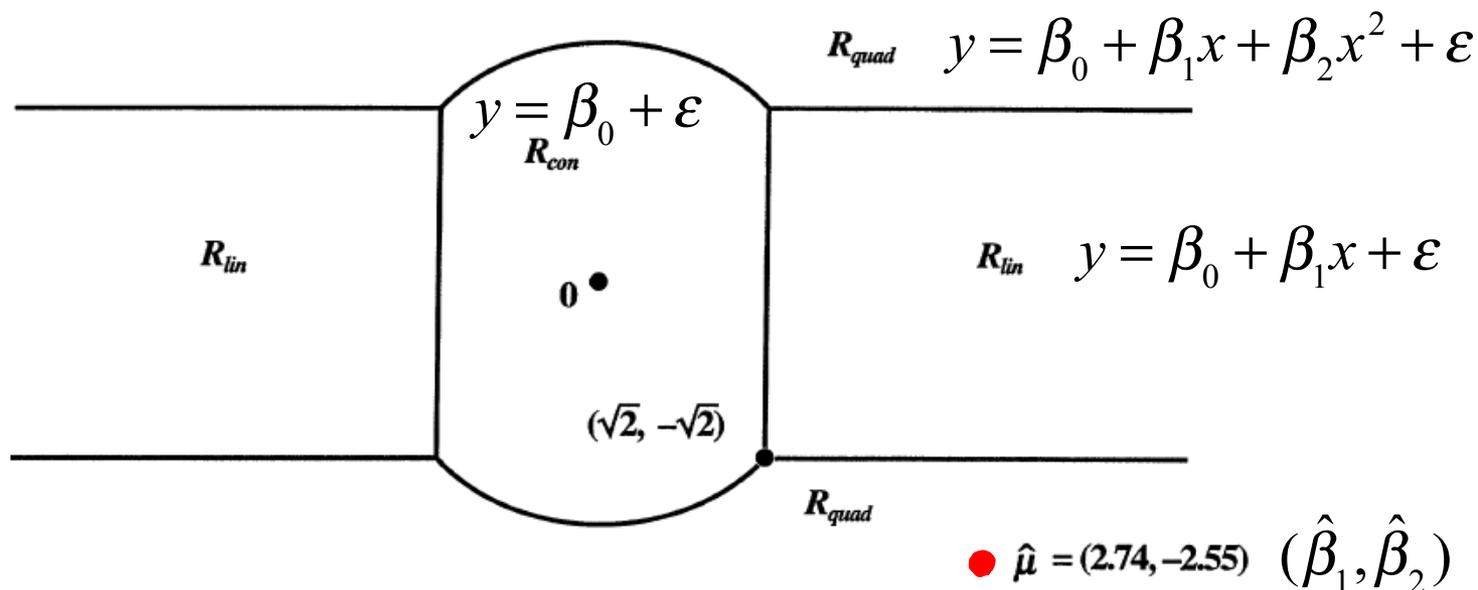


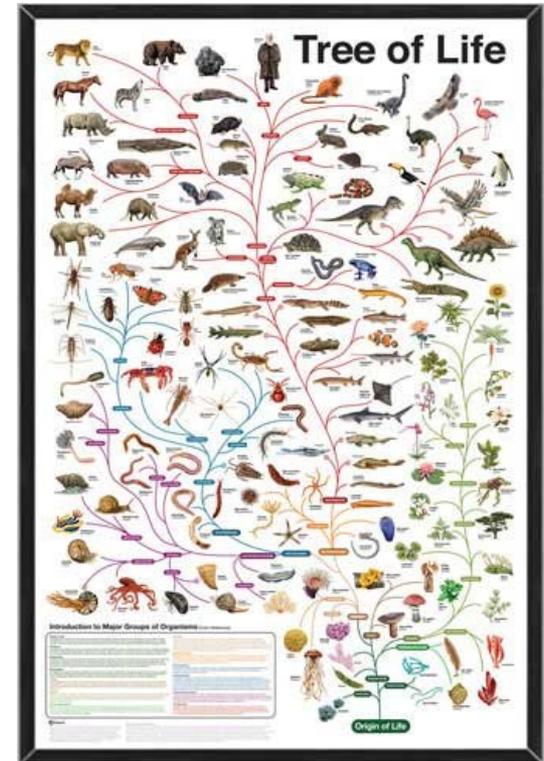
FIG. 1. An example of the problem of regions: a normally distributed vector $y = \hat{u}$, with covariance I , is observed to lie in the region $\mathcal{R}_{\text{quad}}$. With what confidence can we say that the true expectation vector μ lies in $\mathcal{R}_{\text{quad}}$? This example, which concerns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

Efron and Tibshirani (1998)

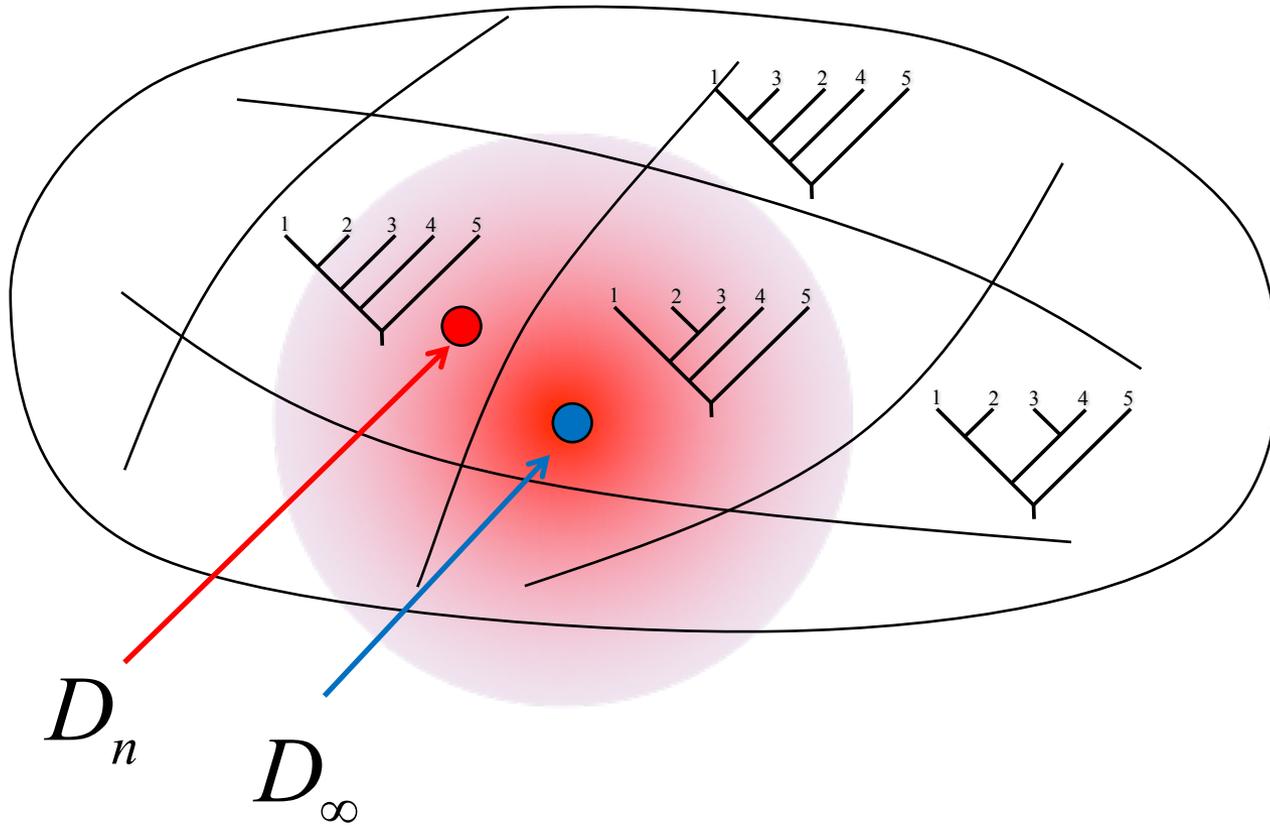
“The problem of regions”

たとえば、進化系統樹の推定

- DNA配列データ
- 系統樹上のマルコフ過程
- 最尤法でモデル選択する



各領域が系統樹に相当



Pvclust: Suzuki and Shimodaira (2006)

Pvclust

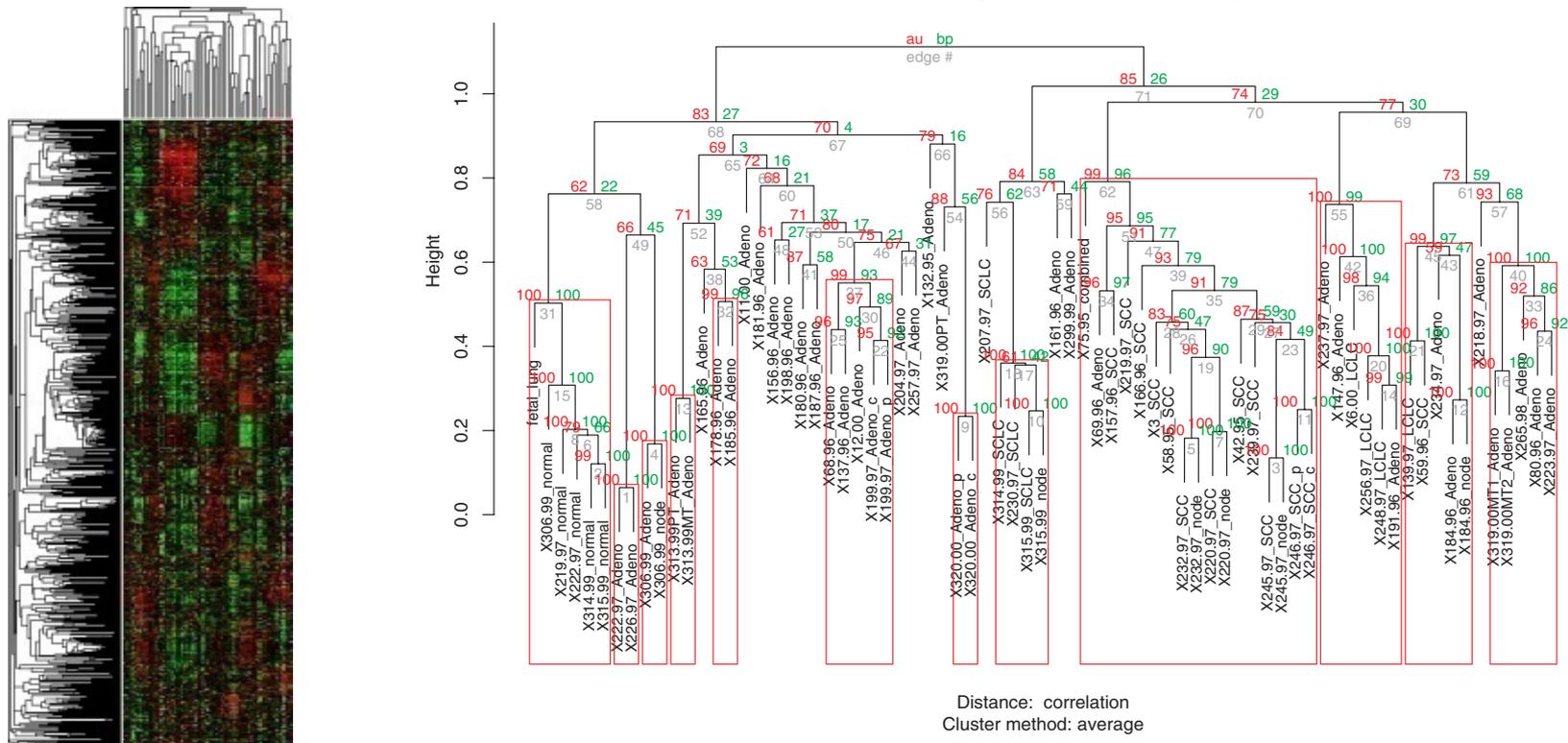


Fig. 1. Hierarchical clustering of 73 lung tumors. The data are expression pattern of 916 genes of Garber *et al.* (2001). Values at branches are AU *p*-values (left), BP values (right), and cluster labels (bottom). Clusters with AU \geq 95 are indicated by the rectangles. The fourth rectangle from the right is a cluster labeled 62 with AU = 0.99 and BP = 0.96.

Bidirectional developmental potential in reprogrammed cells with acquired pluripotency

Haruko Obokata^{1,2,3}, Yoshiki Sasaki⁴, Hitoshi Niwa⁵, Mitsutaka Kadota⁶, Munazah Andrabi⁶, Nozomu Takata⁴, Mikiko Tokoro², Yukari Terashita^{1,2}, Shigenobu Yonemura⁷, Charles A. Vacanti³ & Teruhiko Wakayama^{2,8}

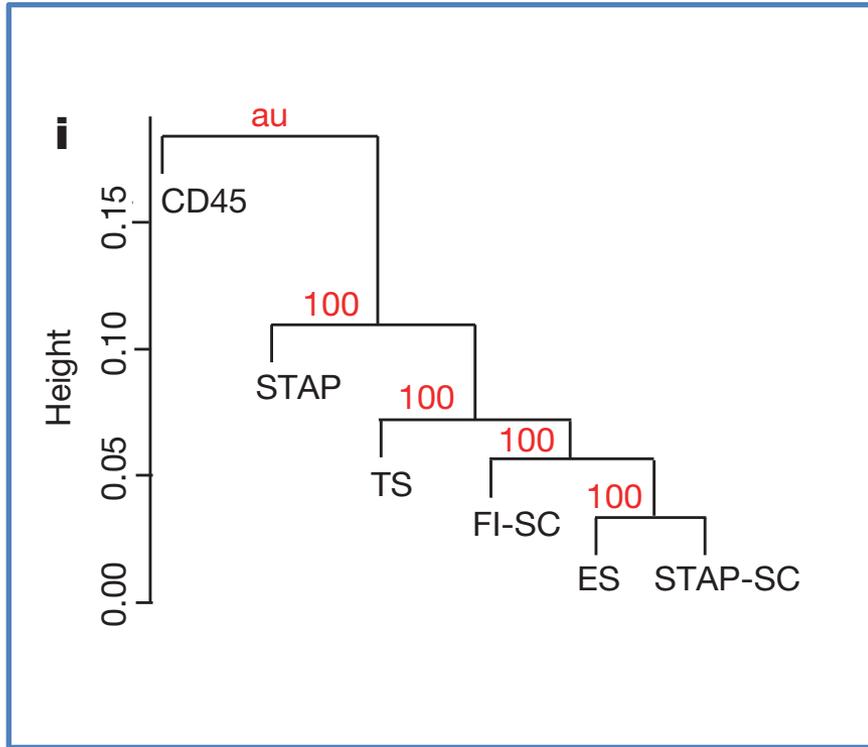
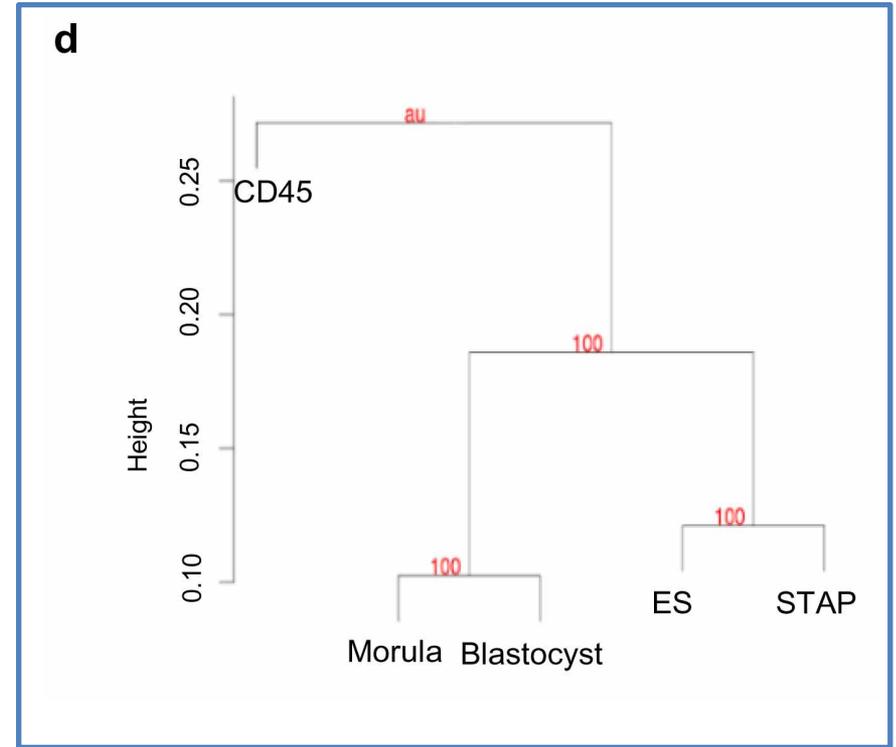


Figure 2 | Fgf4 treatment induces some trophoblast-lineage character in STAP cells.

placental tissues at a detectable level. **i**, Cluster tree diagram from hierarchical clustering of global expression profiles. Red, approximately unbiased P values. **j**, qPCR analysis of Fgf4-induced cells



Extended Data Figure 6 | Characterization of ES-like cells converted from Fgf4-induced stem cells and comparison of STAP cells with early embryos.

Scale bar, 100 μm . **d**, Cluster tree diagram from hierarchical clustering of global expression profiles. Red, AU P values. As this analysis included morula and blastocyst embryos from which only small amounts of RNA could be obtained, we used pre-amplification with the SMARTer Ultra Low RNA kit for Illumina Sequencing (Clontech Laboratories). **e**, **f**, Volcano plot of the expression

各領域がクラスタに相当

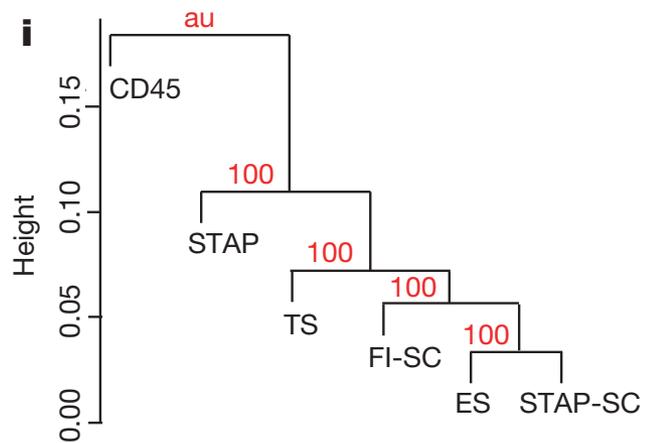
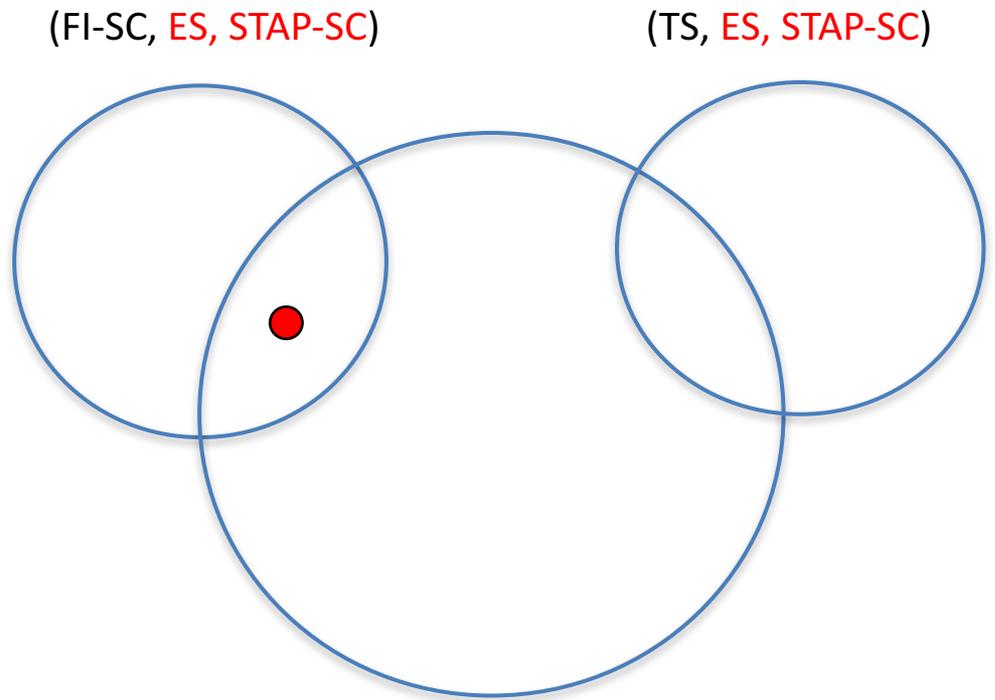


Figure 2 | Fgf4 treatment induces some trophoblast-lineage character in STAP cells.



(ES, STAP-SC)がクラスタになる全てのデンドログラムに相当する領域

系統樹の信頼度は？

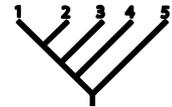
D_n

human 1 GCCAACCTCCTACTCCTCAATGTAACCAATTCCTAATCGCAATGG
cow 2 ACCAACCTCCTACTCCTCAATGTAACCAATTCCTAATCGCAATAG
rabbit 3 ATTAATATCATCTCCTCAATGTAACCAATTCCTCCTCGCCGTAG
mouse 4 ATTAACATCTAATACTAATGTAATCCCAATTCCTAATCGCCGTAG
opossum 5 ATTAATACACTCCATTAATCCTACCTGTAATTTAGCCATAG

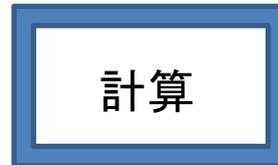
サンプルサイズ n



計算結果

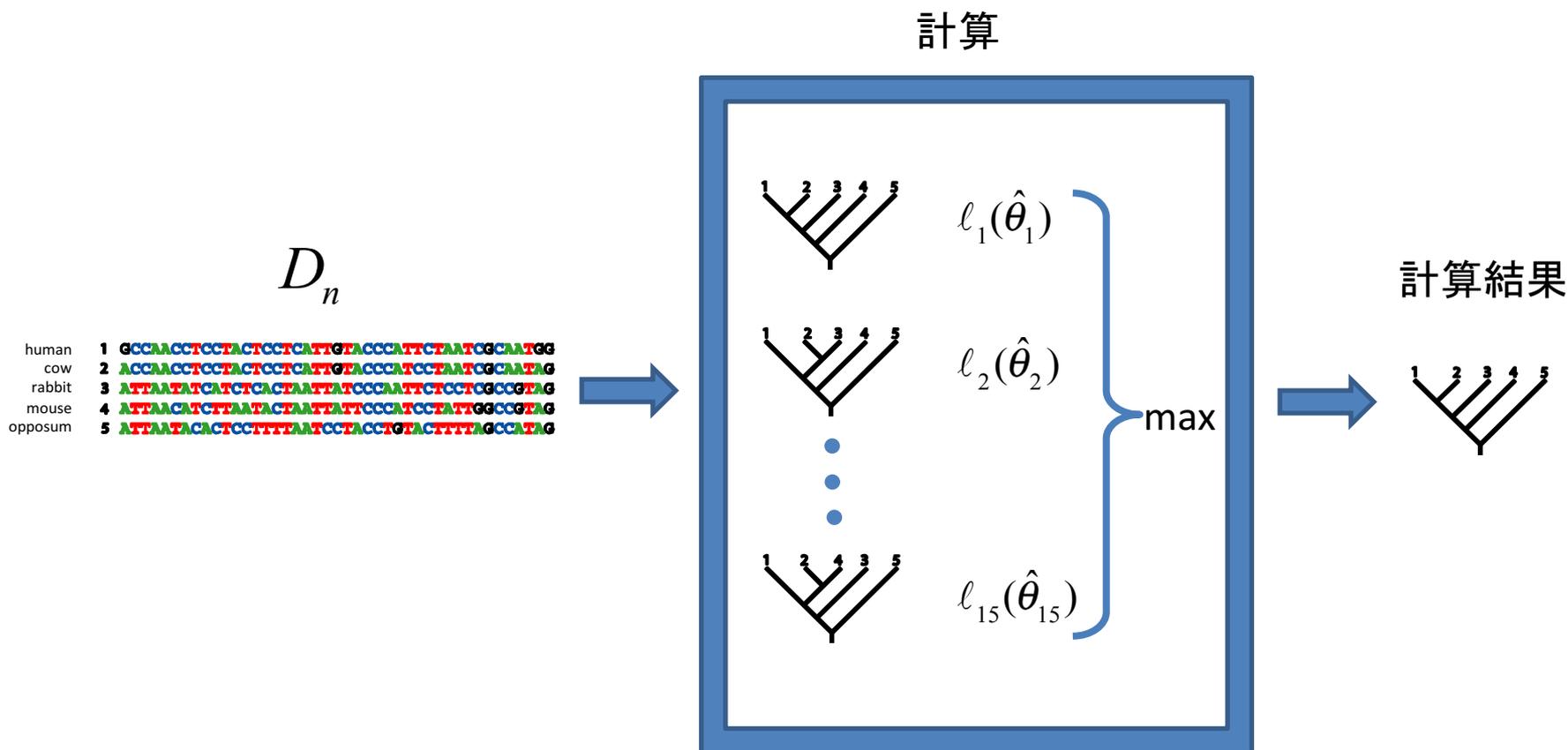


D_∞

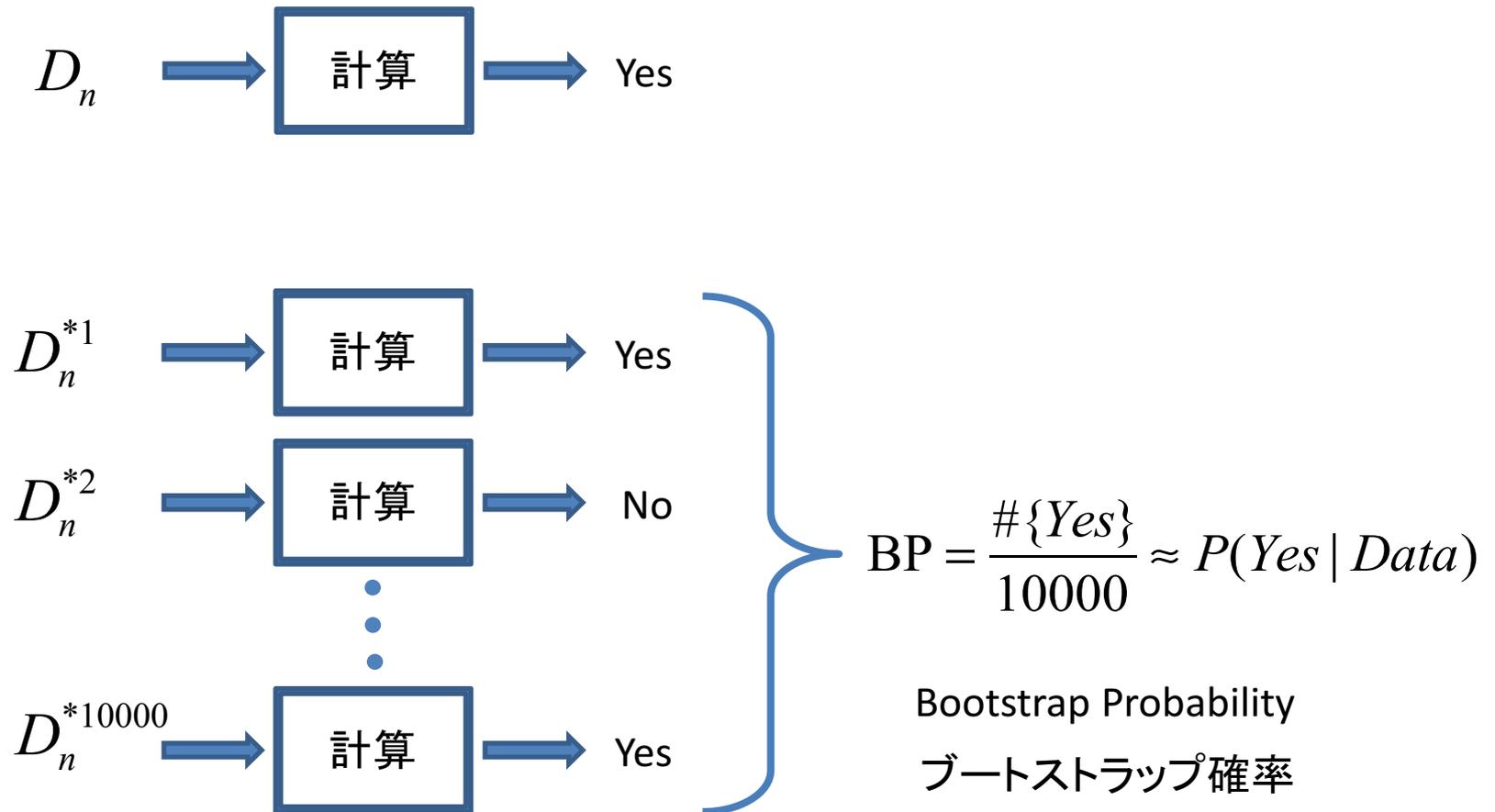


?

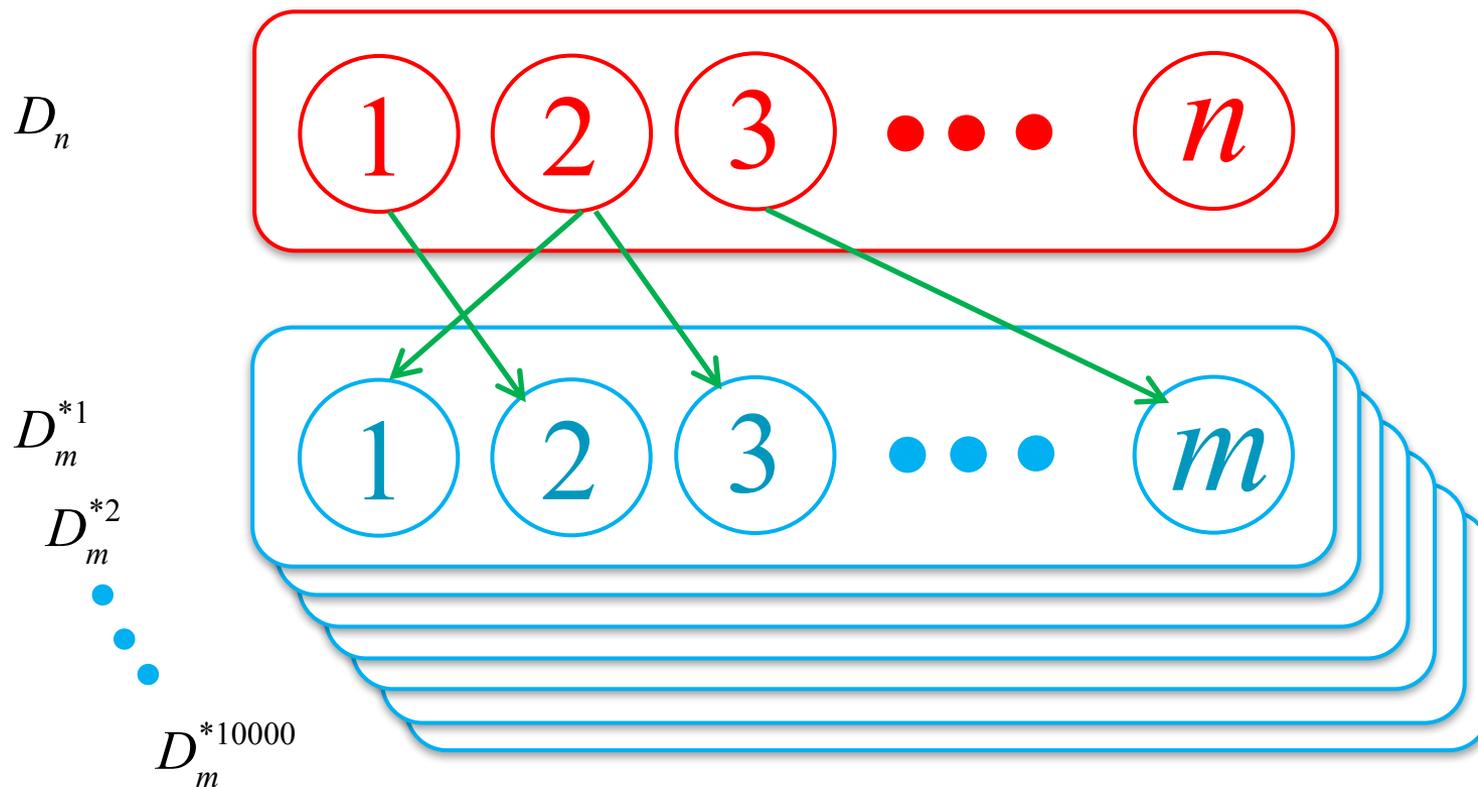
「計算」はブラックボックスとみなす



反復計算してカウントだけ使う



ランダムに要素をコピーする



ブートストラップ・リサンプリング

Efron (1979)
Felsenstein (1985)

Bootstrap: $n=10,242$ columns

human	1	GCCAACCTCCTACTCCTCATTGTACCCATTCTAATCGCAATGG
cow	2	ACCAACCTCCTACTCCTCATTGTACCCATCCTAATCGCAATAG
rabbit	3	ATTAATATCATCTCACTAATTATCCCAATTCTCCTCGCCGTAG
mouse	4	ATTAACATCTTAATACTAATTATTCCCATCCTATTGGCCGTAG
opposum	5	ATTAATACACTCCTTTTAATCCTACCTGTACTTTTAGCCATAG



Assume i.i.d. for columns

$$D_n \xrightarrow{\text{blue arrow}} D_m^{*1}, \dots, D_m^{*100}$$

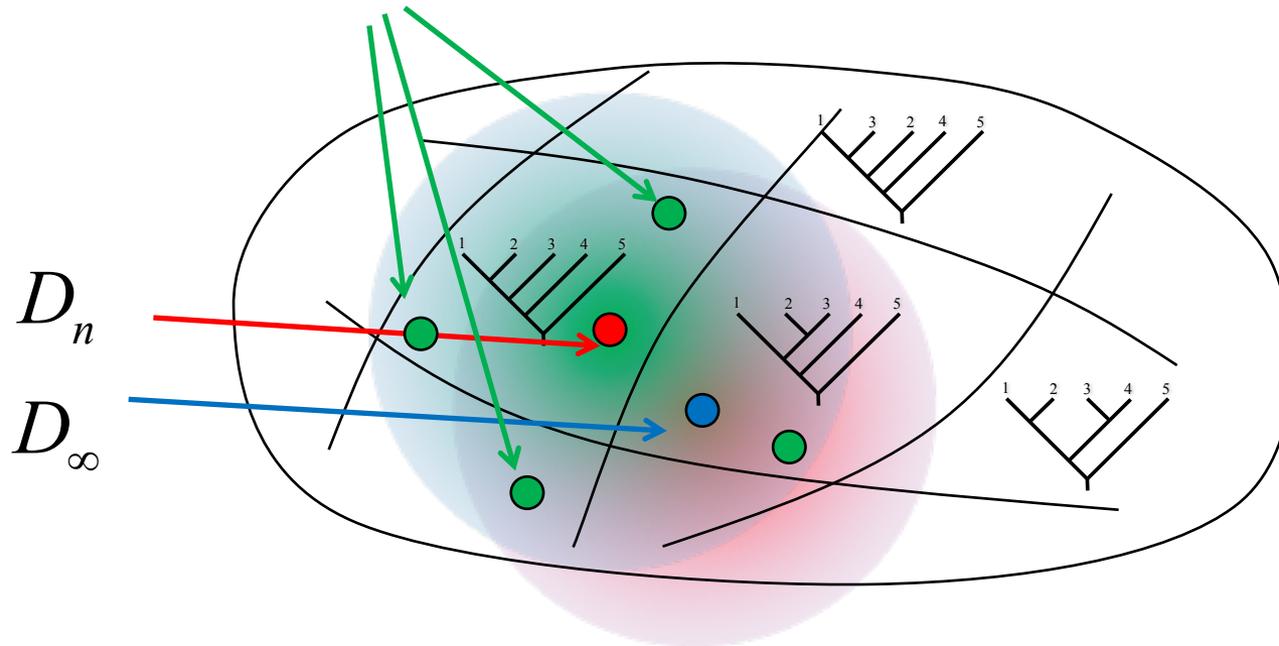
We assume $m = O(n)$, typically $m=10242, 5000, 20000$, say

In the “ m -out-of- n bootstrap” $0 < m \ll n$, typically $m=30$, say

BPはベイズ事後確率

Efron and Tibshirani (1998)

$$D_n^{*1}, D_n^{*2}, D_n^{*3}, \dots \quad BP = P(Tree | data) \approx \frac{\#\{Tree\}}{100}$$



Shimodaira (2002)のマルチスケール・ブートストラップ: D_m^* のバラツキ(分散)は m に反比例する
 m を変化させたBPから, 各仮説領域までの符号付き距離や, 仮説境界の平均曲率がわかる!

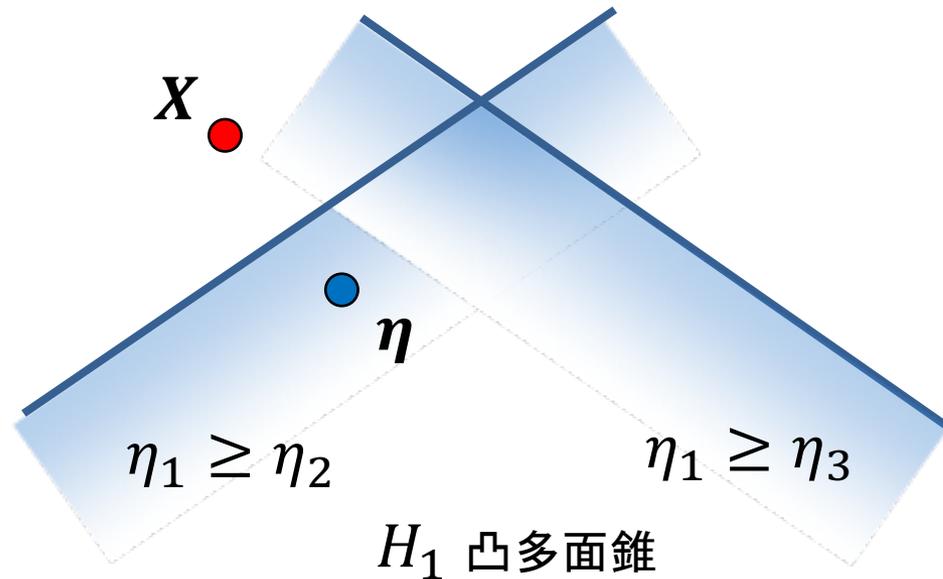
たとえば, 多重比較

$$\mathbf{X} \sim N(\boldsymbol{\eta}, \frac{1}{n}\Sigma)$$

$$H_1: \eta_1 \geq \eta_2, \dots, \eta_1 \geq \eta_p$$

$$\mathbf{X} = \begin{pmatrix} 3.5 \\ 4.0 \\ 0.5 \end{pmatrix}$$

$$\boldsymbol{\eta} = \begin{pmatrix} 3.5 \\ 3.0 \\ 0.5 \end{pmatrix}$$



多重比較の場合

$$\mathbf{X} \sim N(\boldsymbol{\eta}, \frac{1}{n}\Sigma)$$

$$d_1(Y) = \max\{c_2(X_2 - X_1), \dots, c_p(X_p - X_1)\}$$

$$\text{confidence}_{mc}(\mu \in H_1 | y) = P(d_1(Y) > d_1(y) | 0)$$

$$Y = \sqrt{n}\Sigma^{-1/2}\mathbf{X}$$

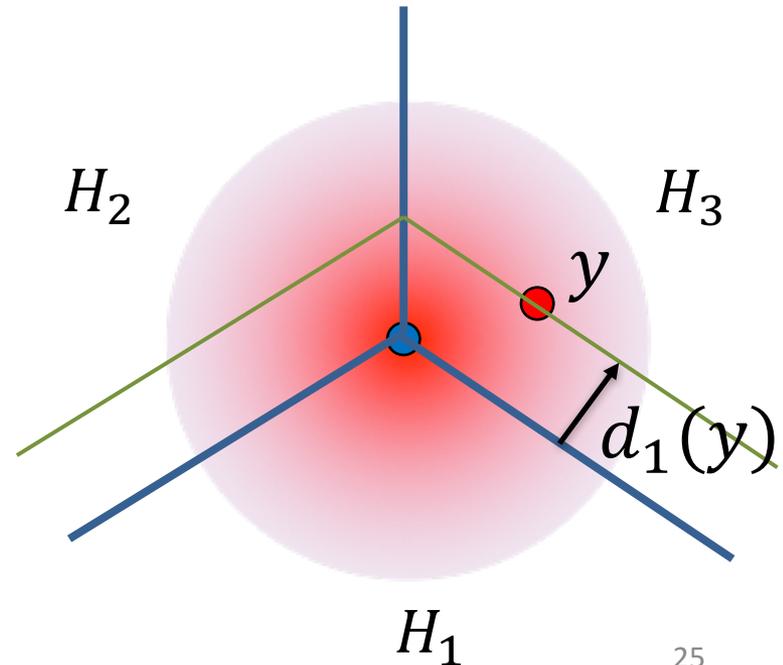
$$\boldsymbol{\mu} = \sqrt{n}\Sigma^{-1/2}\boldsymbol{\eta}$$

$$Y \sim N(\boldsymbol{\mu}, I)$$

Least favorable configuration

$$\eta_1 = \eta_2 = \dots = \eta_p$$

仮説領域(凸錘)の頂点



多重比較は保守的

各領域のType-I error

$$P[\text{confidence}_{mc}(\mu \in H_i | y) < 0.05] \leq 0.05 \text{ under } H_i$$



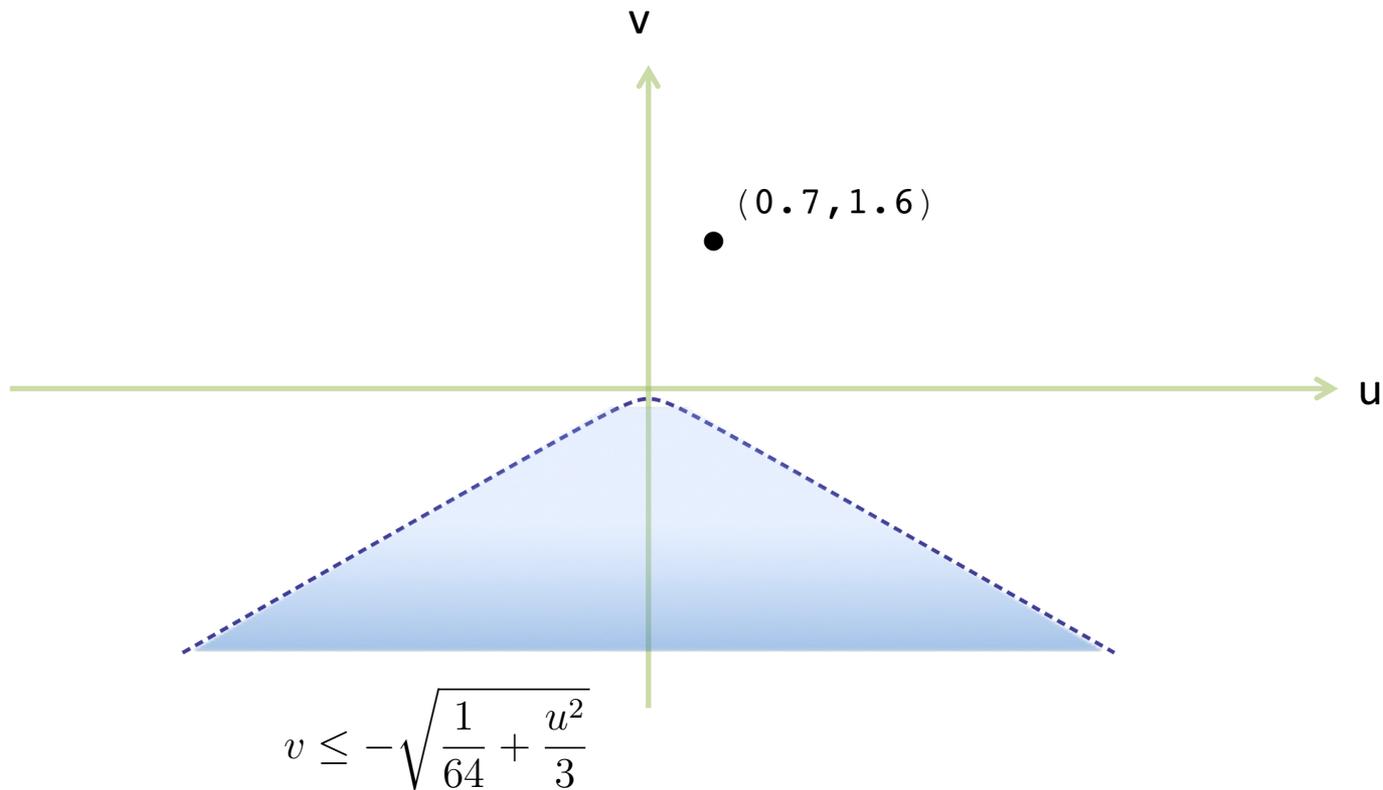
仮説境界上で等号なら嬉しいが、等号になるのはl.f.c.のみ

仮説の信頼集合

$$S(y) = \{H_i: \text{confidence}_{mc}(\mu \in H_i | y) \geq 0.05\}$$

$$P[H_i \notin S(y)] \leq 0.05 \text{ under } H_i$$

Region and a data-point



Multivariate Normal Model

$$y = (u, v) \in \mathbb{R}^{q+1} \quad q = \dim u$$

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

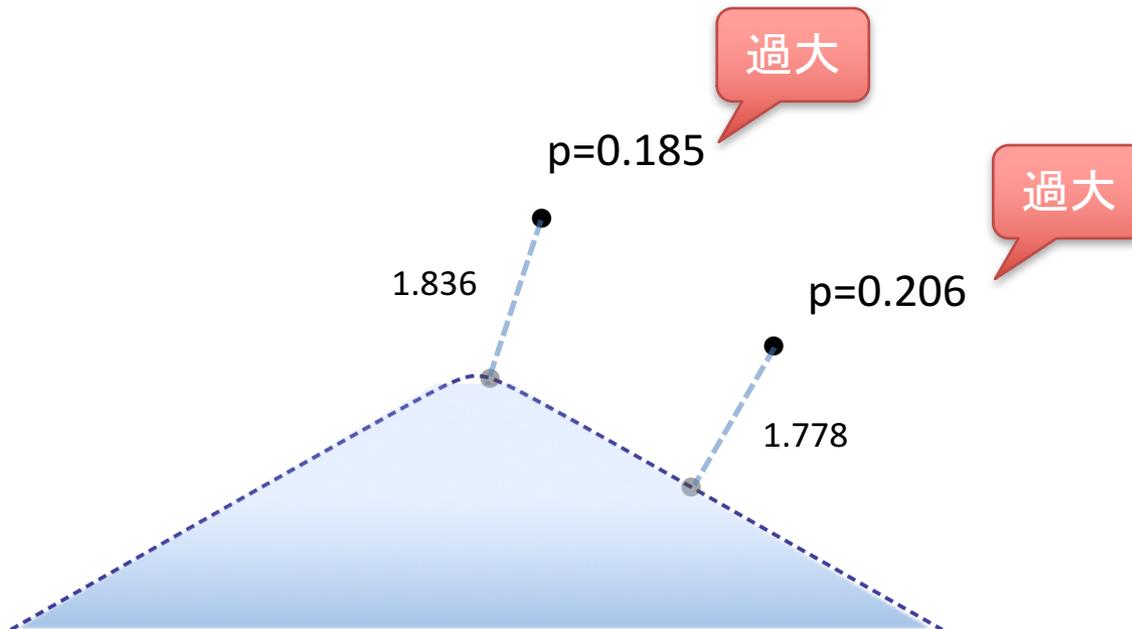
$$H = \left\{ (u, v) \mid v \leq -h(u), u \in \mathbb{R}^q \right\}$$

region: $H = \mathcal{R}(h)$ boundary surface: $\partial H = \mathcal{B}(h)$

Null hypothesis: $\mu \in H$ v.s. Alternative hypothesis: $\mu \notin H$

Chi-square test (very conservative)

distance² ~ chi-square distribution with df=2



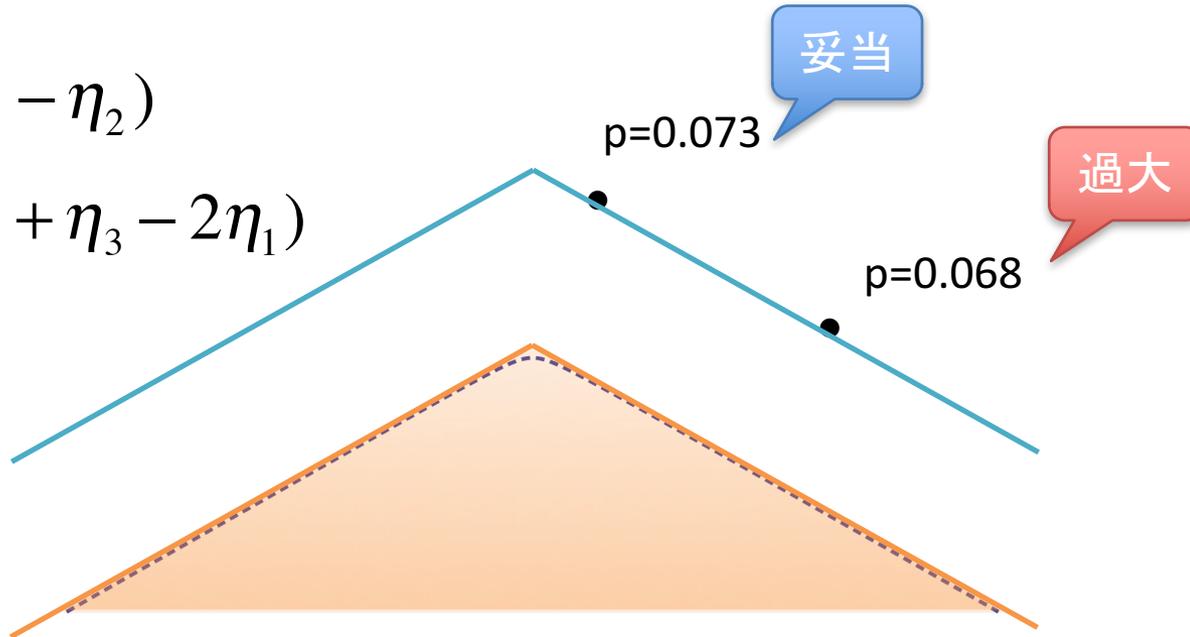
Multiple Comparison (conservative)

$$\bar{X} \sim N_3(\eta, I_3 / n) \Leftrightarrow Y \sim N_2(\mu, I_2)$$

$$\eta_1 \geq \eta_2, \eta_1 \geq \eta_3 \Leftrightarrow v \leq -|u| / \sqrt{3}$$

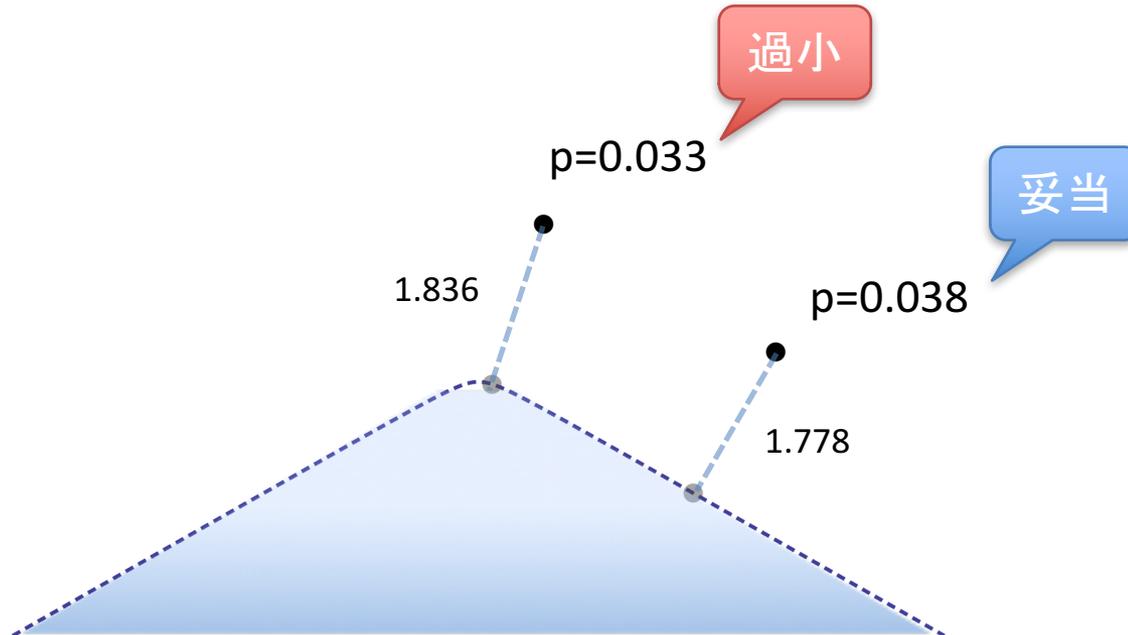
$$u = \sqrt{n/2}(\eta_3 - \eta_2)$$

$$v = \sqrt{n/6}(\eta_2 + \eta_3 - 2\eta_1)$$



Singed LR test (rejecting too much)

distance $\sim N(0,1)$



Bootstrap probability (=Bayesian PP)

$$Y^* \sim N_{q+1}(y, \sigma^2 I_{q+1}) \quad \sigma^2 = 1$$

$$Y^{*1}, \dots, Y^{*B} \quad \rightarrow \quad \widehat{\text{BP}}_{\sigma^2}(H|y) = \frac{\#\{Y^{*b} \in H, b = 1, \dots, B\}}{B}$$

$$\text{BP}_{\sigma^2}(H|y) = P(Y^* \in H|y)$$

BP is interpreted as the Bayesian posterior probability of H if the prior distribution of mu is uniform.

Efron and Tibshirani (1998)

BP is even worse

過小

BP=0.019

p=0.033 (signed LR test)

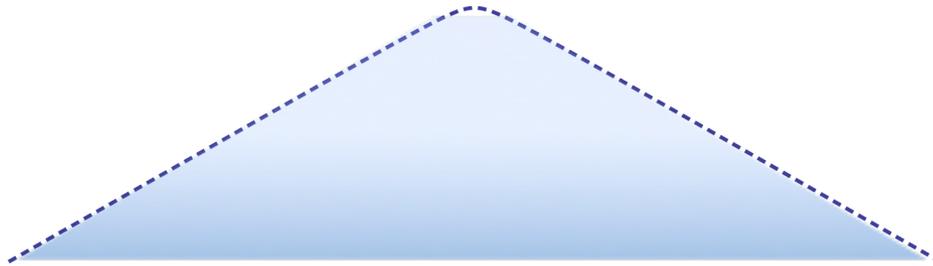
妥当

•

BP=0.037

p=0.038 (signed LR test)

•



Double bootstrap probability

Projection of y onto the boundary surface:

$$\hat{\mu}(H|y) = \arg \min_{\mu \in \partial H} \|y - \mu\|$$

Adjusting BP using resampling from the projection

$$Y^+ \sim N_{q+1}(\hat{\mu}(H|y), \tau^2 I_{q+1}) \quad \tau^2 = 1$$

$$\text{DBP}_{\tau^2, \sigma^2}(H|y) = P_{\tau^2} \left[\text{BP}_{\sigma^2}(H|Y^+) \leq \text{BP}_{\sigma^2}(H|y) \mid \hat{\mu}(H|y) \right]$$

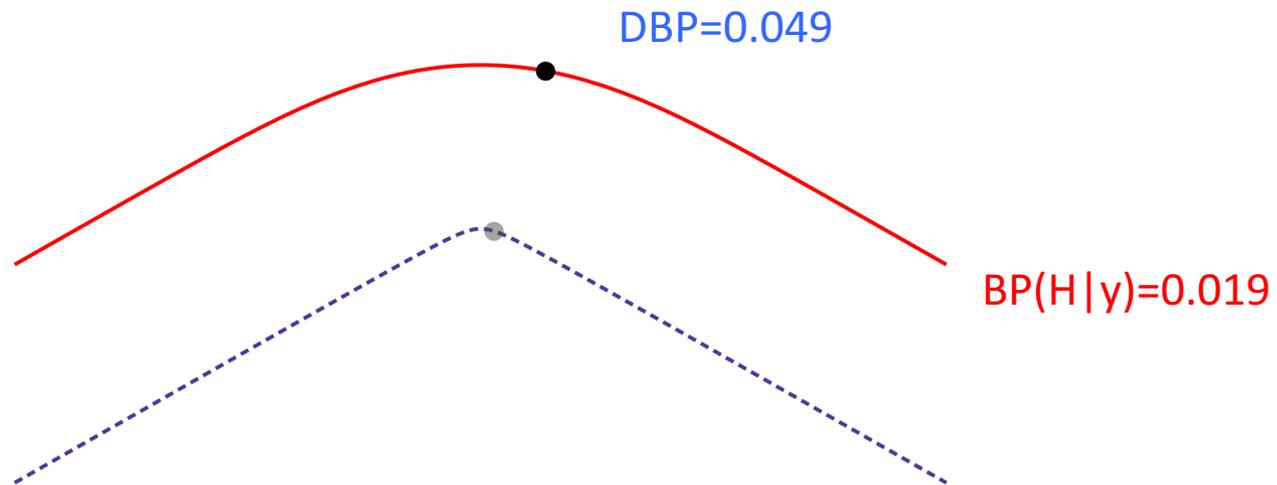
Hall (1992), Efron and Tibshirani (1998)

contour surface of BP=0.019

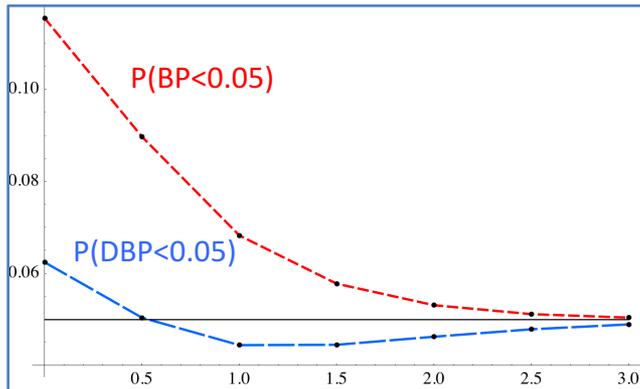
Computing DBP:

$$Y^+ \sim N_2(\hat{\mu}, I_2)$$

$$\text{DBP} = P\left[\text{BP}(H|Y^+) \leq 0.019 \mid \hat{\mu}\right]$$

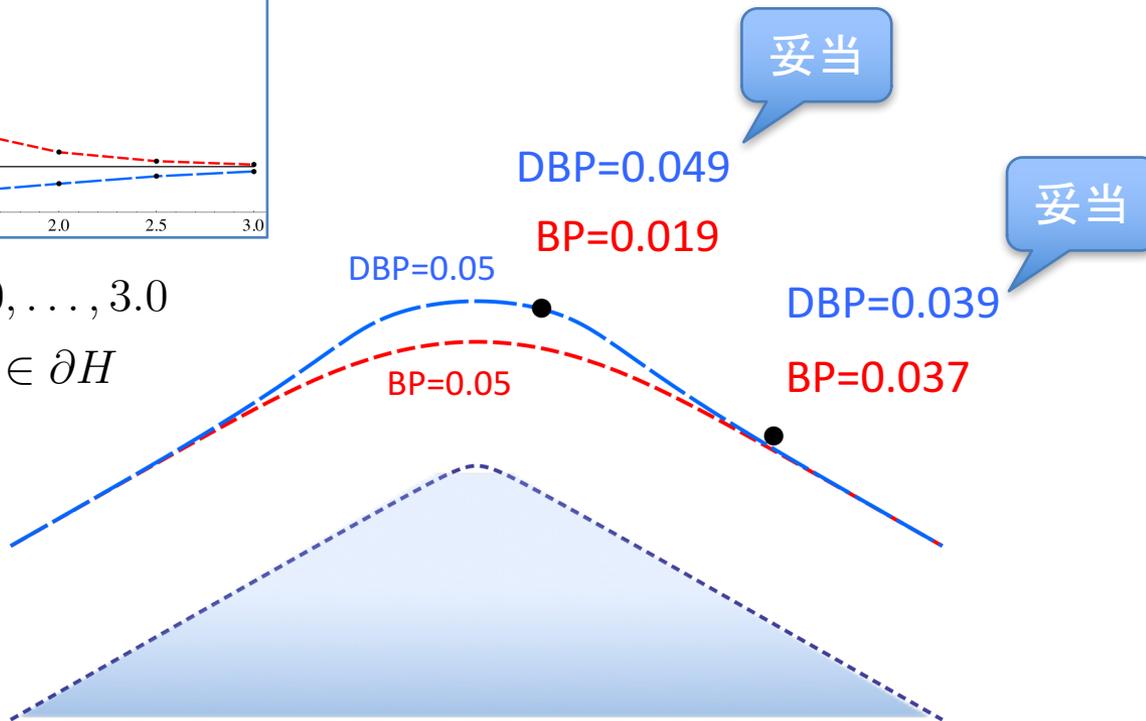


DBP adjusts the bias of BP



$\theta = 0.0, 0.5, 1.0, \dots, 3.0$

$\mu = (\theta, -h(\theta)) \in \partial H$



Approximately unbiased p-values via Multiscale bootstrap

m out of n bootstrap : Politis and Romano (1994), Bickel et al. (1997)

$$\mathcal{X} = \{x_1, \dots, x_n\} \quad \longrightarrow \quad \mathcal{X}^* = \{x_1^*, \dots, x_m^*\}$$

The idea of multiscale bootstrap : Shimodaira (2002, 2004, 2008)

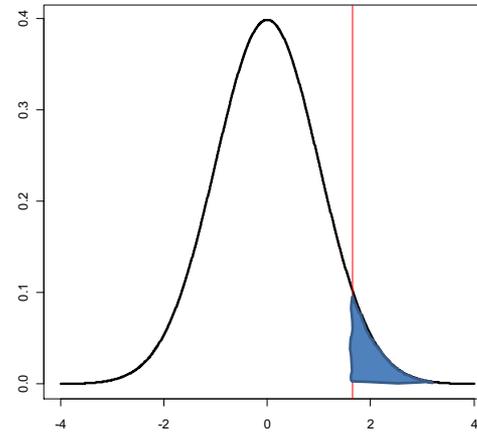
$$\sigma^2 = \frac{n}{m}$$

We compute BP for $\sigma_1^2, \dots, \sigma_S^2$ and extrapolate BP to $\sigma^2 = -1$
(equivalently $m = -n$)

The BP with $m = -n$ is denoted as AU (= Approximately Unbiased)

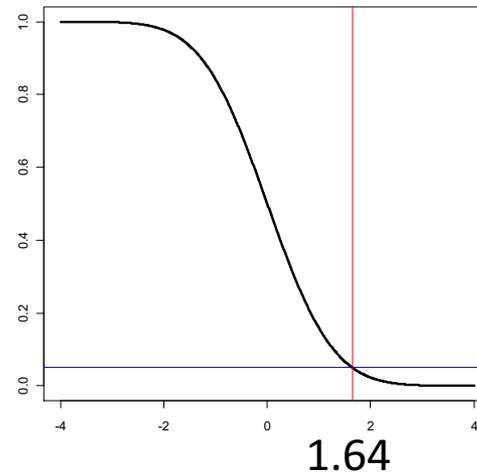
正規分布の上側確率

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$



$$\bar{\Phi}(z) = \int_z^{\infty} \phi(x) dx$$

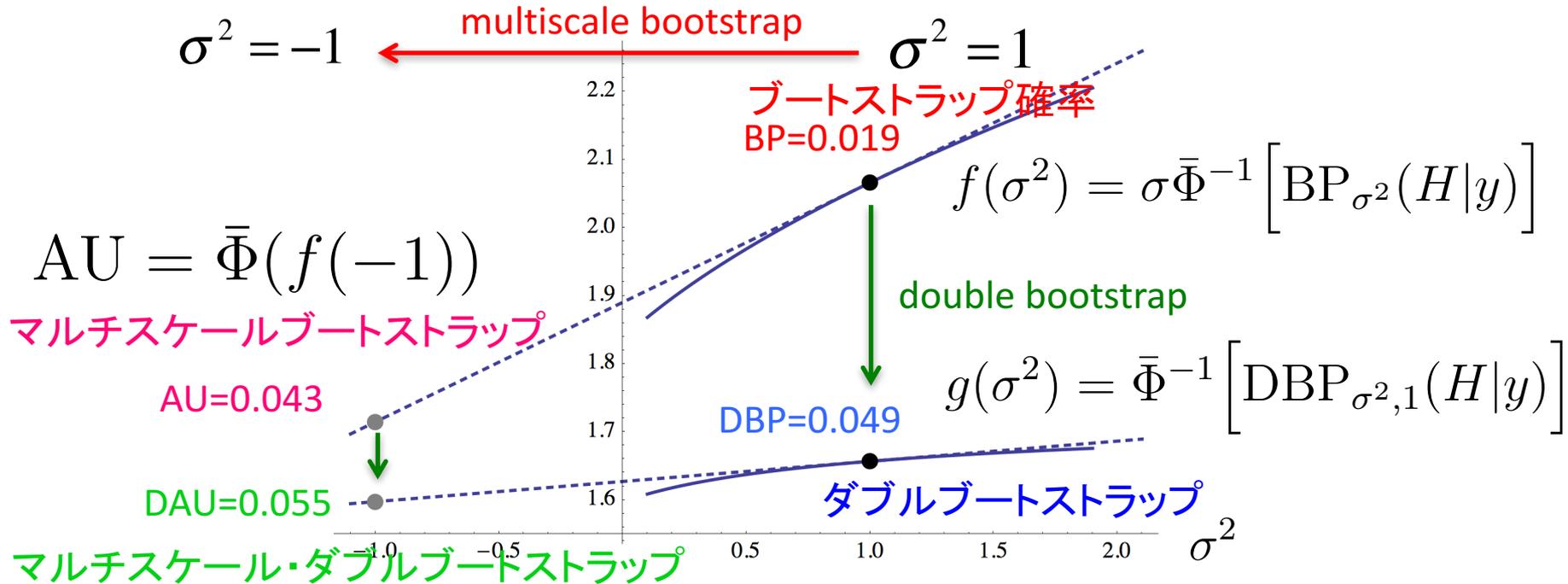
$$\bar{\Phi}(1.64) = 0.05$$



m=-n への外挿

Extrapolation to $\sigma^2 = -1$

$$\bar{\Phi}(z) = 1 - \Phi(z) = \int_z^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$



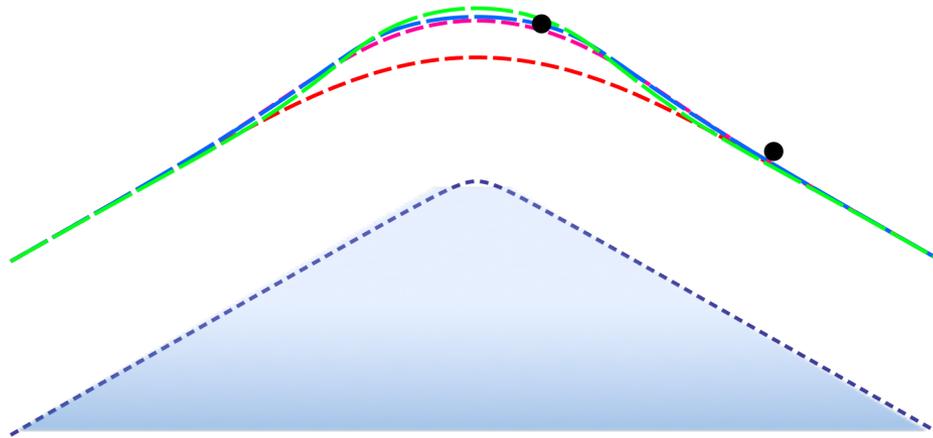
$$\text{DAU} = \bar{\Phi}(g(-1))$$

We apply the **multiscale bootstrap** to DBP for getting DAU

Equivalently, we could say applying **double bootstrap** to AU for getting DAU

contour curves of $p=0.05$

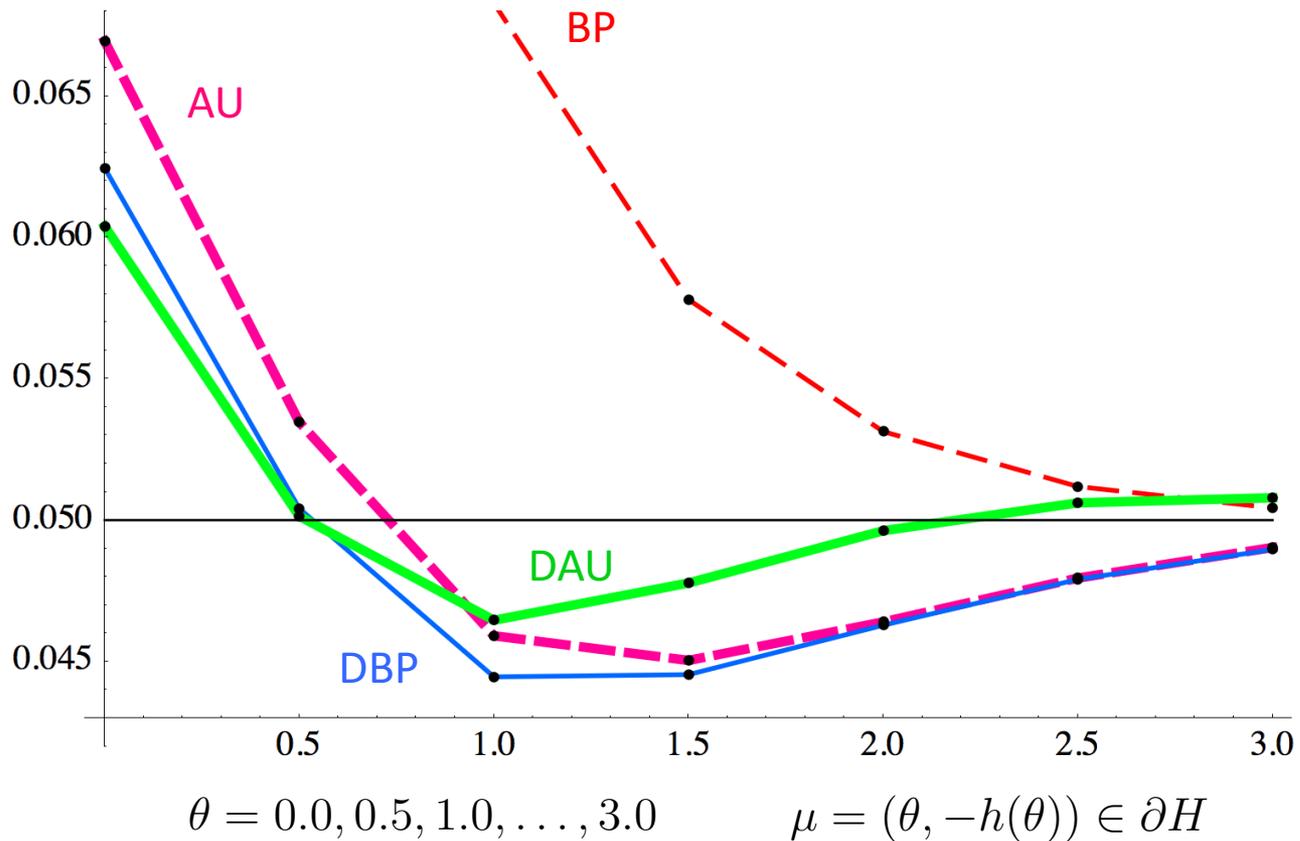
BP, AU, DBP, DAU



DAUの棄却確率は誤差が小さい

Rejection probabilities $P(p < 0.05)$

Error: **DAU** < {**DBP**, **AU**} < **BP**



棄却確率 (0.05に近いほど良い)

TABLE 2

Rejection probabilities (in percent) at significance level $\alpha = 5\%$.

u	0.0	0.5	1.0	1.5	2.0	2.5	3.0
BP	13.39	8.894	6.678	5.676	5.253	5.086	5.027
AU2	7.655	5.171	4.459	4.447	4.628	4.801	4.912
AU3	6.609	4.718	4.493	4.746	4.982	5.080	5.081
DBP	6.619	4.590	4.202	4.364	4.610	4.795	4.905
DAU	6.476	4.660	4.481	4.746	4.981	5.084	5.092
MCB	5.000	3.340	2.880	2.783	2.768	2.766	2.766

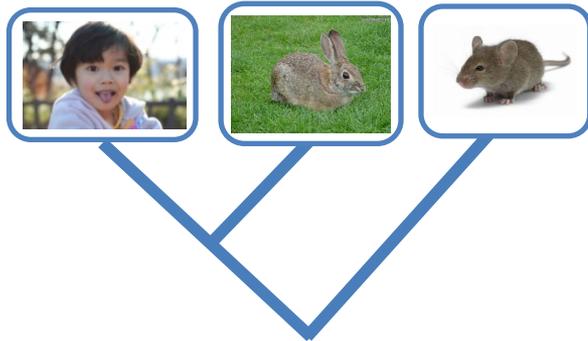
過大

妥当

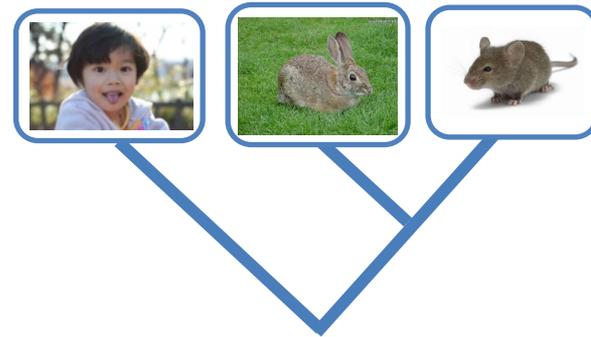
妥当

過小

New Discovery?

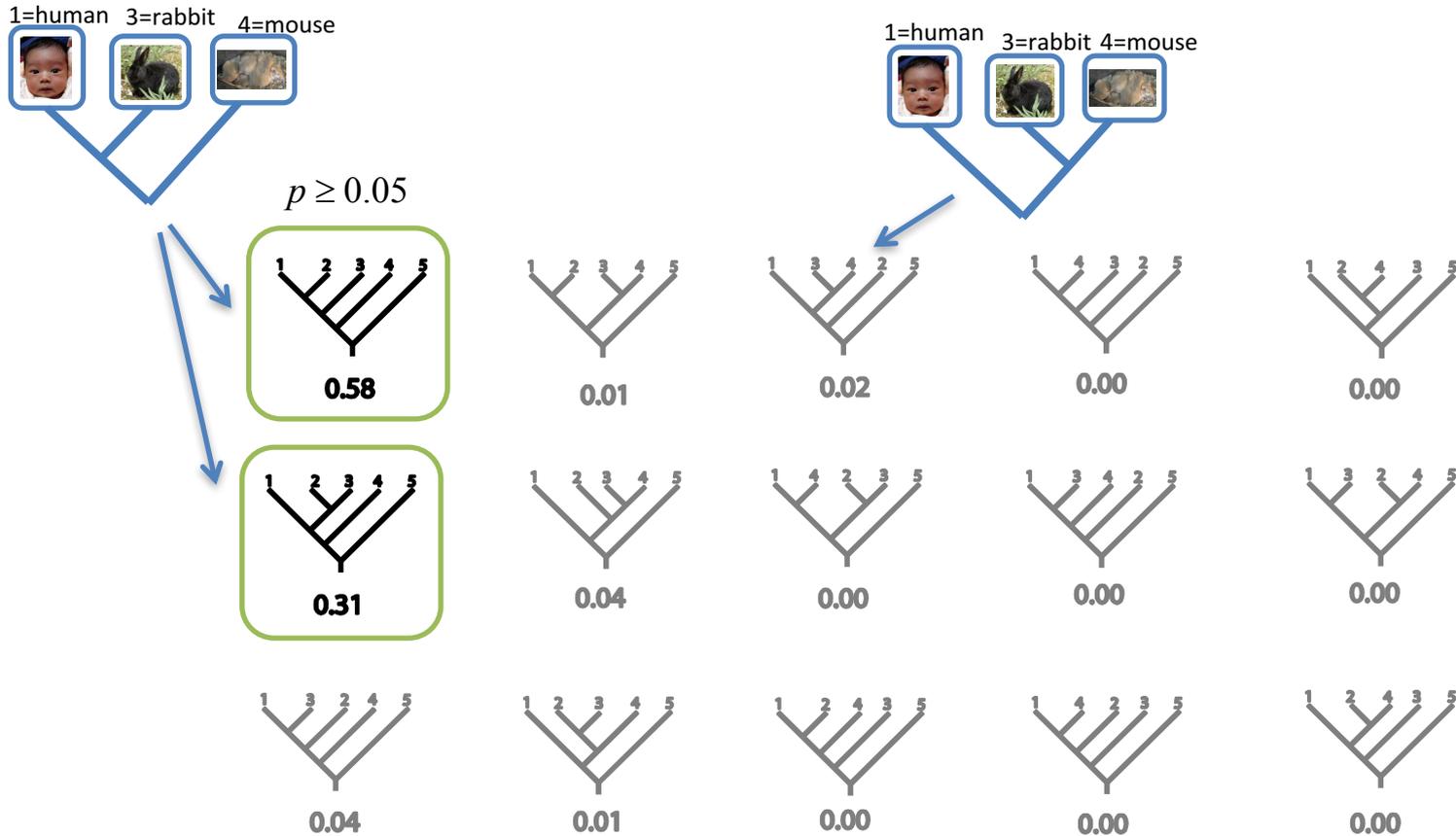


58%



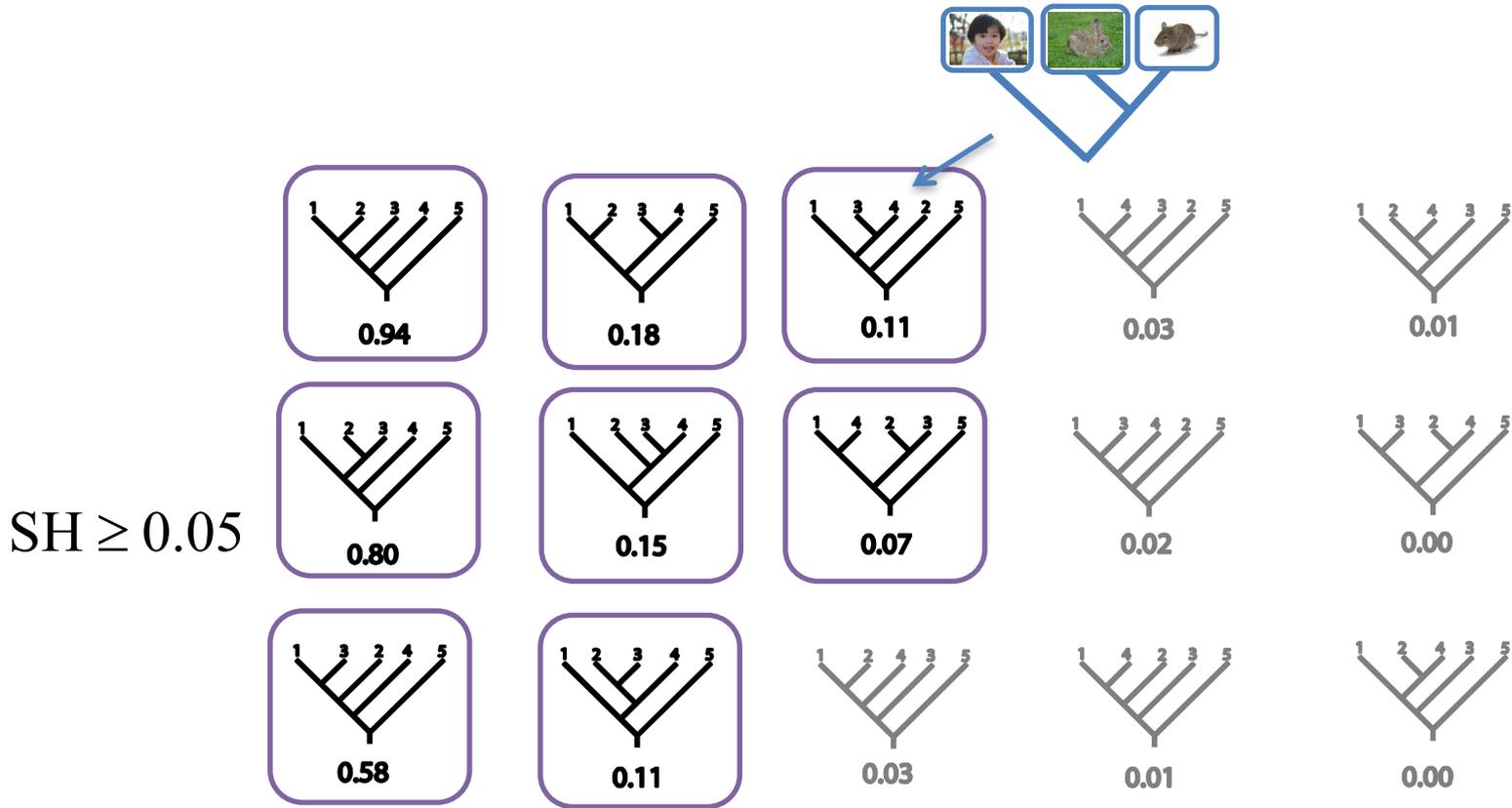
2%

バイズの事後確率



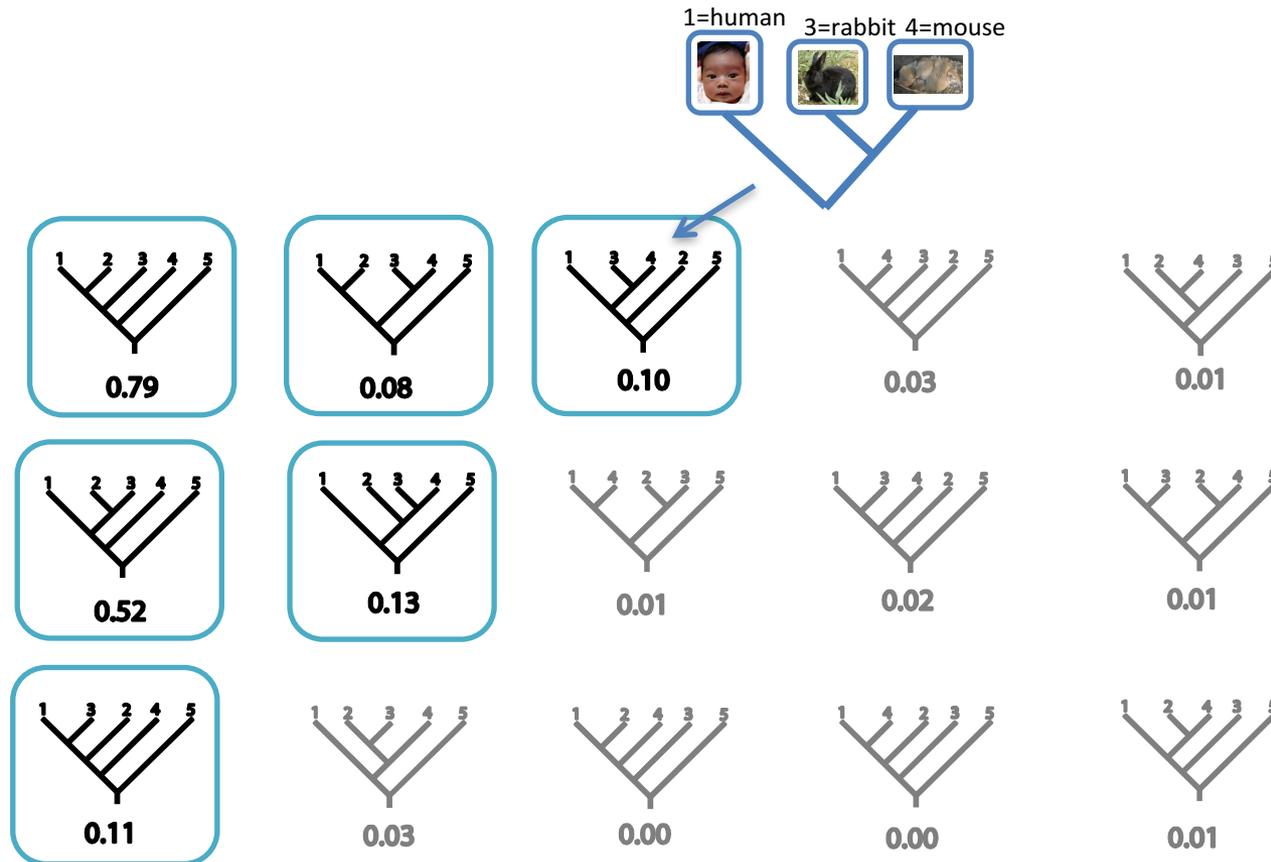
ブートストラップ法 (m=n) で計算

Shimodaira-Hasegawa testは安全すぎる



対数尤度に多重比較法を適用してモデル選択の多重性を調整

頻度論のp-値 (近似的に不偏 AU)



マルチスケール・ブートストラップ法 ($m=n$) で計算

曲率を反転するとベイズは頻度論

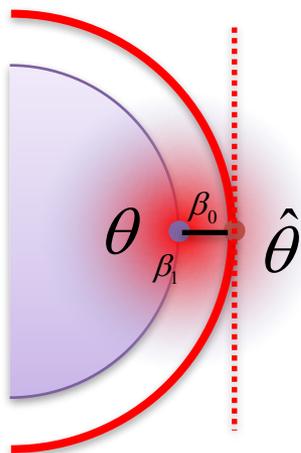
頻度論のp-値

ベイズの事後確率

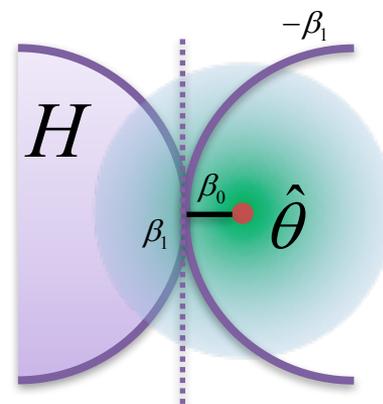


Jerzy Neyman
(1894-1981)

頻度論



$$AU = \bar{\Phi}(\beta_0 - \beta_1) + O(n^{-3/2})$$

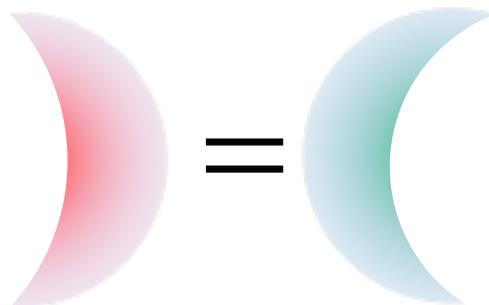


$$BP = \bar{\Phi}(\beta_0 + \beta_1) + O(n^{-3/2})$$



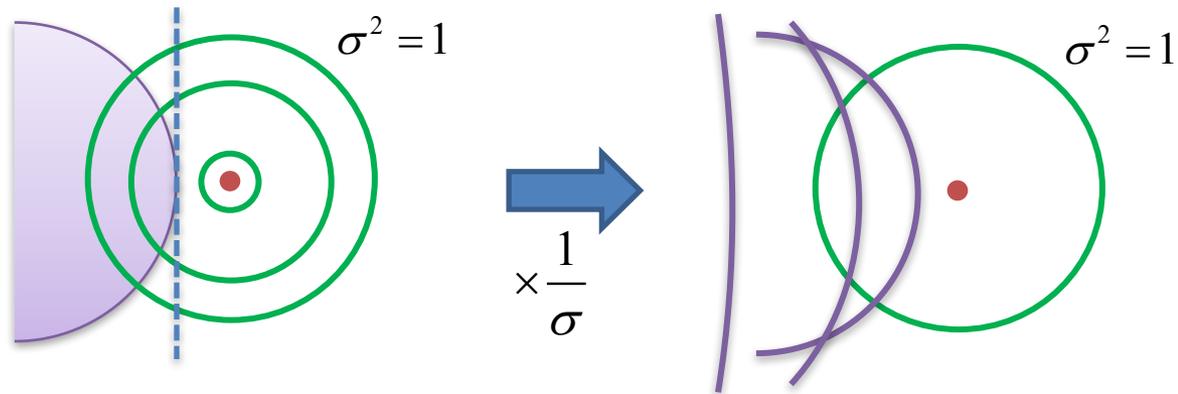
Thomas Bayes
(1702-1761)

ベイズ統計学



Rescaling the whole picture

Shimodaira (2002)



$$BP = \bar{\Phi}(\beta_0 + \beta_1) + O(n^{-3/2})$$

$$BP(\sigma^2) = \bar{\Phi} \left[\frac{\beta_0}{\sigma} + \beta_1 \sigma \right] + O(n^{-3/2})$$

$$NBP(\sigma^2) = \bar{\Phi} \left\{ \sigma \bar{\Phi}^{-1} \left(BP(\sigma^2) \right) \right\} = \bar{\Phi}(\beta_0 + \beta_1 \sigma^2) + O(n^{-3/2})$$

ベイズと頻度論のギャップを埋める

$$NBP(\sigma^2) \triangleq \Phi \left\{ \sigma \Phi^{-1} \left(BP(\sigma^2) \right) \right\}$$

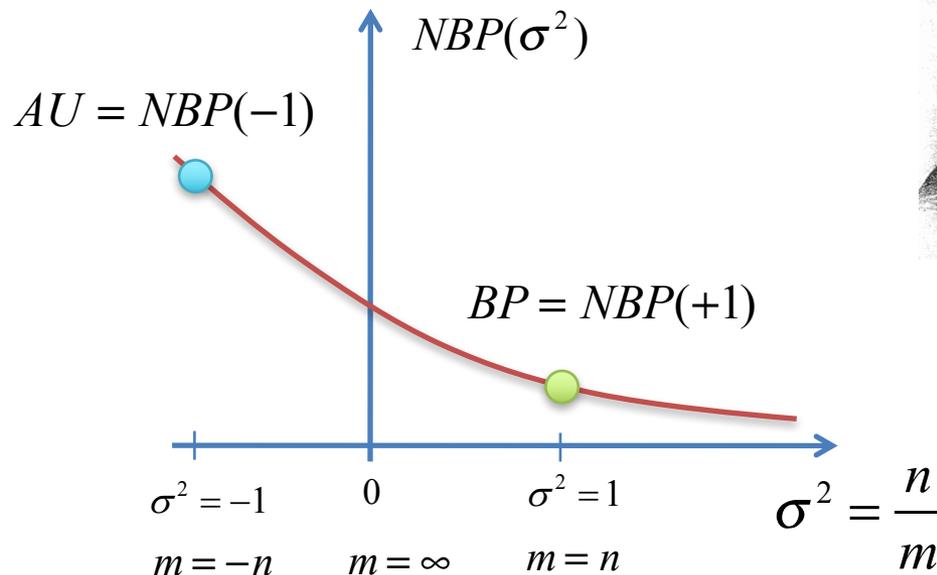
正規化したブートストラップ確率(NBP)

スケーリング則



Jerzy Neyman
(1894-1981)

頻度論

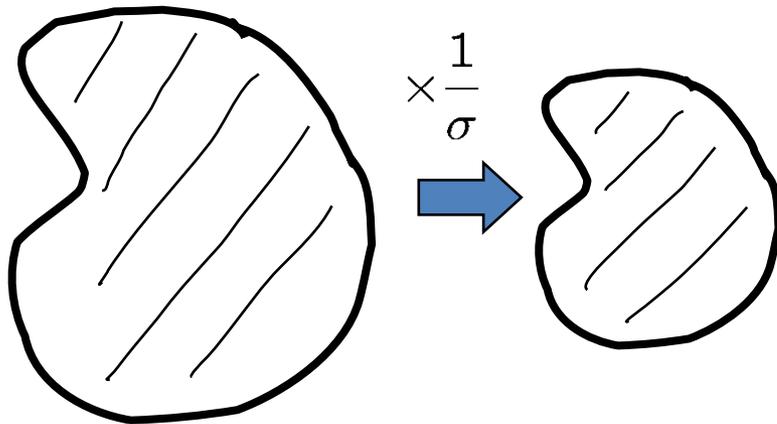


Thomas Bayes
(1702-1761)

ベイズ統計学

AUの計算手順: いくつかの $m > 0$ でBPを計算して, NBPに変換してから $m = -n$ へ外挿

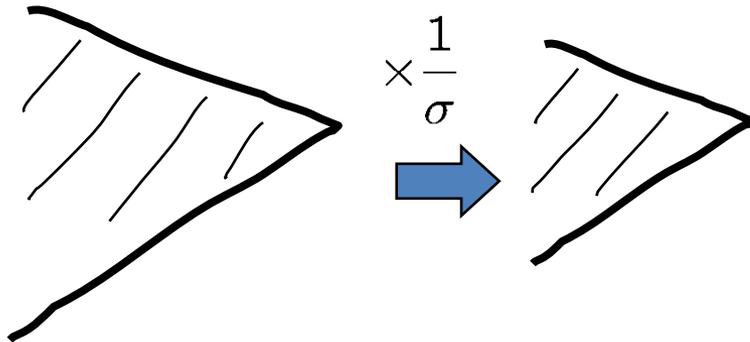
We can't compute $nBP(-1)$ for cones



smooth surface

$$BP(\sigma^2) \approx \bar{\Phi} \left[\frac{\beta_0}{\sigma} + \beta_1 \sigma \right]$$

$$nBP(\sigma^2) \approx \bar{\Phi}(\beta_0 + \beta_1 \sigma^2)$$



cone

$$BP(\sigma^2) \approx \bar{\Phi} \left[\frac{\beta_0}{\sigma} + \beta_1 \right]$$

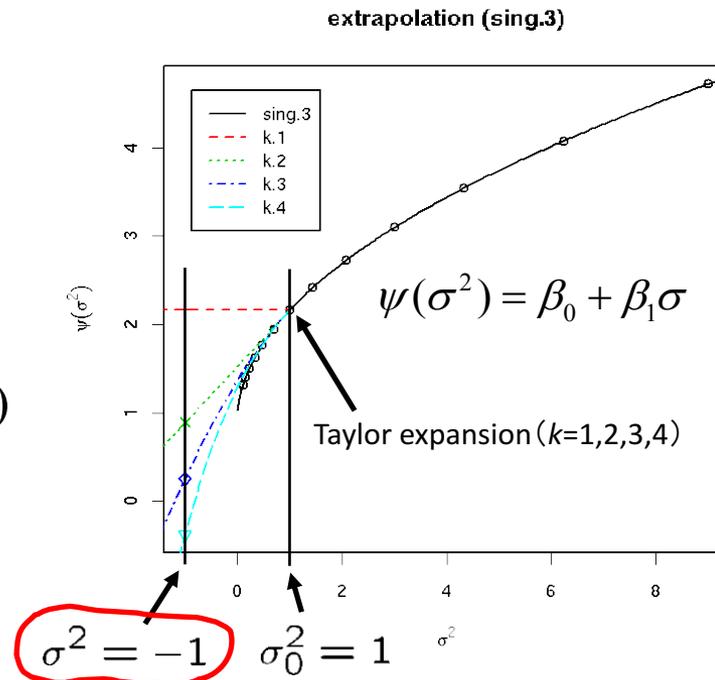
$$nBP(\sigma^2) \approx \bar{\Phi}(\beta_0 + \beta_1 \sigma)$$

Taylor expansion of nBP using k terms

Shimodaira (2008)

$$nBP_k(\sigma^2) = \bar{\Phi} \left[\sum_{j=0}^{k-1} \frac{(\sigma^2 - \sigma_0^2)^j}{j!} \frac{\partial^j \psi(\sigma^2)}{\partial (\sigma^2)^j} \Big|_{\sigma_0^2} \right]$$

$$AU_k = nBP_k(-1)$$



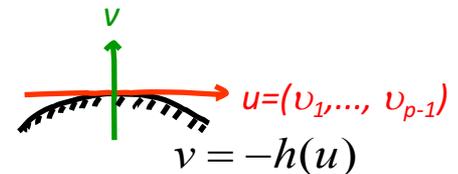
Scaling-law and geometry

Shimodaira (2008)

$$NBP(\sigma^2) = \bar{\Phi}\left(\underset{\substack{\uparrow \\ \text{distance}}}{\beta_0} + \underset{\substack{\uparrow \\ \text{mean curvature}}}{\beta_1}\sigma^2 + \beta_2\sigma^4 + \beta_3\sigma^6 + \dots\right) + O(h^2)$$

$$\text{p-value} = \bar{\Phi}\left(\beta_0 - \beta_1 + \beta_2 - \beta_3 + \dots\right) + O(h^2)$$

$$\beta_j(u) = \frac{1}{2^j j!} \sum_{j_1 + \dots + j_{p-1} = j} \frac{j!}{j_1! \dots j_{p-1}!} \frac{\partial^{2j} h(u)}{\partial u_1^{2j_1} \dots \partial u_{p-1}^{2j_{p-1}}}$$



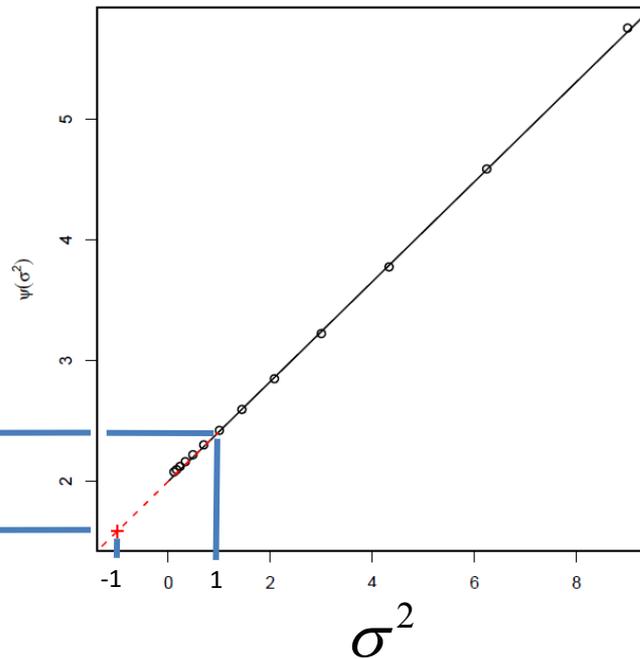
Using a new asymptotic theory with “nearly flat surface h ” instead of large n .

Extrapolating NBP to m=-n

$$NBP(\sigma^2) \triangleq \bar{\Phi} \left\{ \sigma \bar{\Phi}^{-1} \left(BP(\sigma^2) \right) \right\}$$

Plotting : $\sigma \bar{\Phi}^{-1} \left(BP(\sigma^2) \right)$

Fitting a model : $\psi(\sigma^2) = \beta_0 + \beta_1 \sigma^2 + \beta_2 \sigma^4 + \dots + \beta_{k-1} \sigma^{2(k-1)}$



BP=0.0078

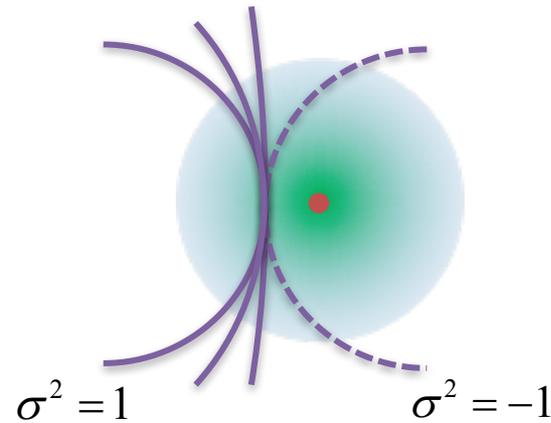
k=2: AU=0.057

k=3: AU=0.052

$p=0.05$

What NBP is calculating ?

Shimodaira (2008)



$$NBP(\sigma^2) = \bar{\Phi}\left(\beta_0 + \beta_1 \sigma^2\right) + O(n^{-3/2})$$

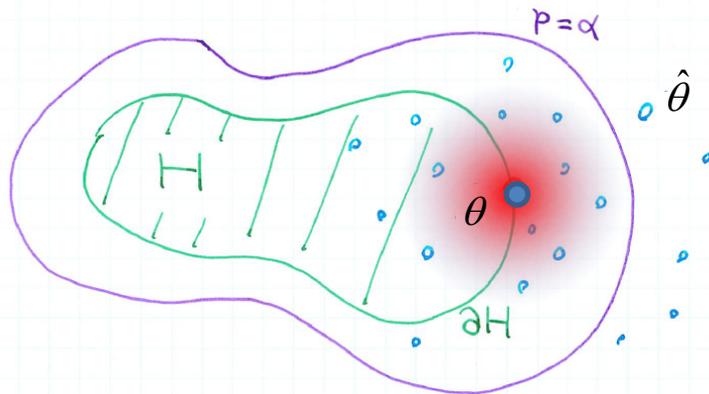
curvature

$$NBP(-1) = p\text{-value} + O(n^{-3/2})$$

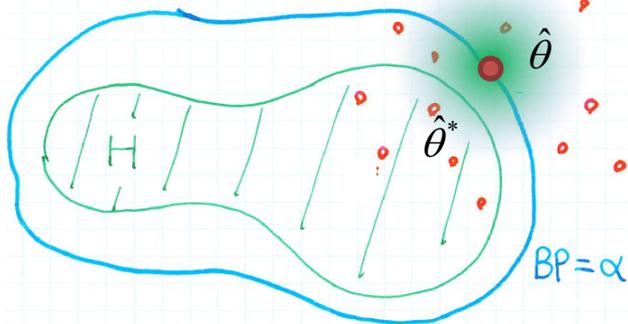
What is an unbiased test?

The p-value of an Unbiased Test (“similar” on the boundary)

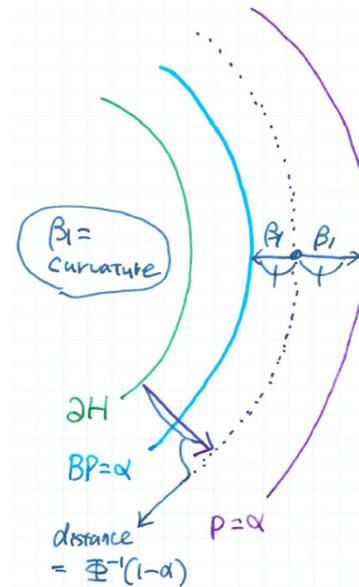
$$P\{p(\hat{\theta}) < \alpha \mid \theta\} = \alpha, \quad \forall \theta \in \partial H, \quad 0 < \alpha < 1$$



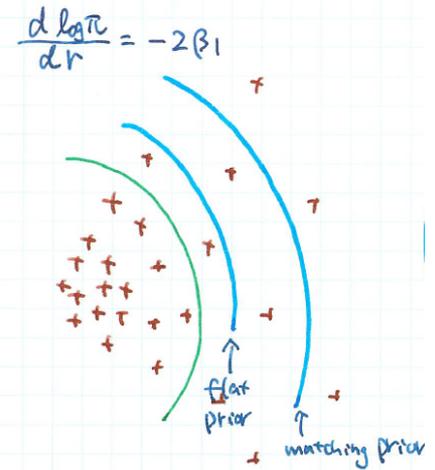
Bootstrap Probability



Geometry



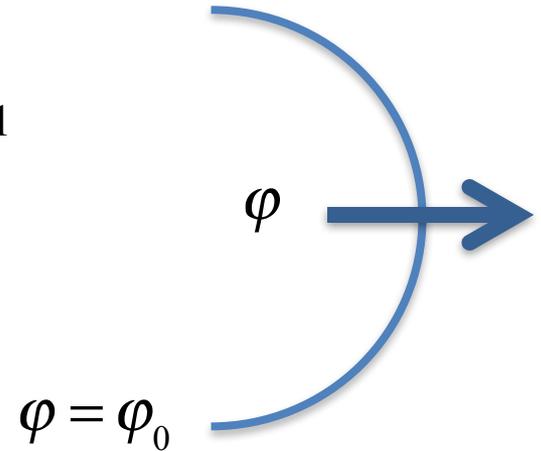
Probability Matching Prior



Matching priors

$$\theta \leftrightarrow (\varphi, \lambda) \quad \varphi \in \mathbb{R}, \lambda \in \mathbb{R}^{p-1}$$

$$H = \{ \theta : \varphi \leq \varphi_0 \}$$



Tibshirani (1989) assumes orthogonality $I_{\varphi\lambda}(\varphi, \lambda) = 0$

$$\pi(\theta) = g(\lambda) I_{\varphi\varphi}^{1/2}(\varphi, \lambda) |J(\varphi, \lambda)|^{-1}$$

$$\frac{d \log \pi(\theta)}{d\varphi} = -2\beta_1 \quad (\text{curvature of the boundary surface of } H)$$

$$\beta_1 \propto 1 / \text{radius}$$

Reweighting the bootstrap by this matching prior makes BP=p-value

Correcting a deviation from normality

Efron (1987), Efron and Tibshirani (1998), Shimodaira (2004)

$$a = \frac{1}{6} \frac{E((\hat{\theta} - \theta)^3)}{E((\hat{\theta} - \theta)^2)^{3/2}} \quad \text{acceleration constant}$$

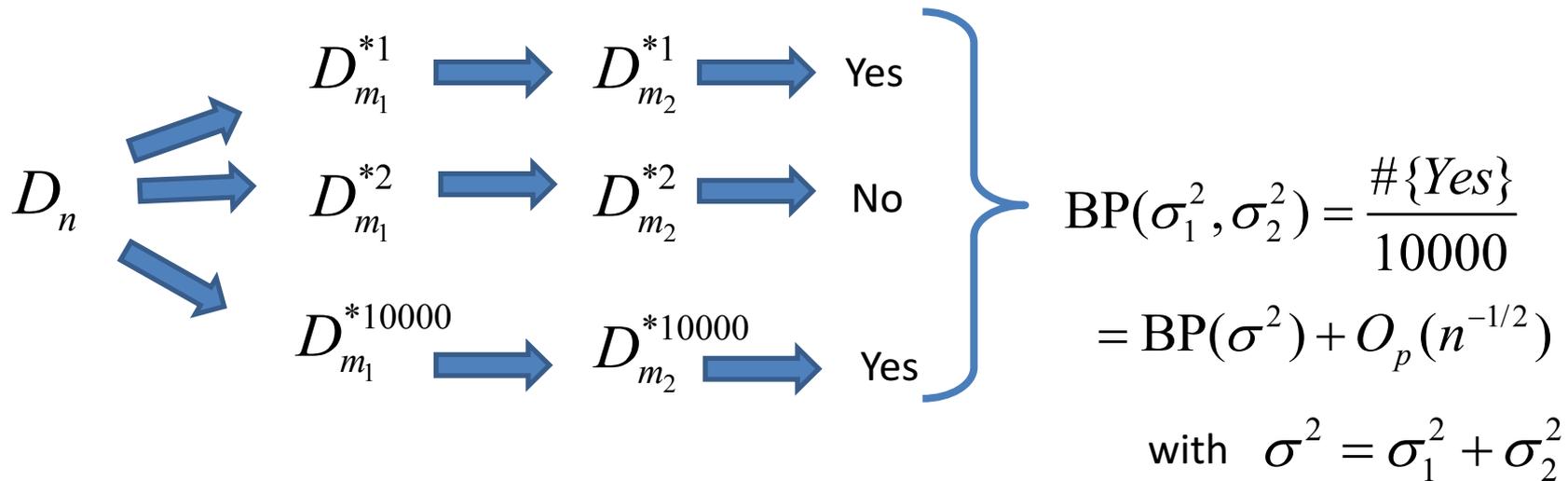
$$nBP(\sigma^2) = \bar{\Phi} \left[(\beta_0 - \beta_1 \sigma^2) - \underline{a(2\beta_0^2 + \sigma^2)} \right] + O(n^{-1})$$

$$p\text{-value} = \bar{\Phi} \left[\beta_0 - \beta_1 + \underline{a(1 - \beta_0^2)} \right] + O(n^{-1})$$

$$\bar{\Phi}^{-1}(p\text{-value}) - \bar{\Phi}^{-1}(\text{AU}(-1)) = a\beta_0^2 + O(n^{-1})$$

Two-step multiscale bootstrap

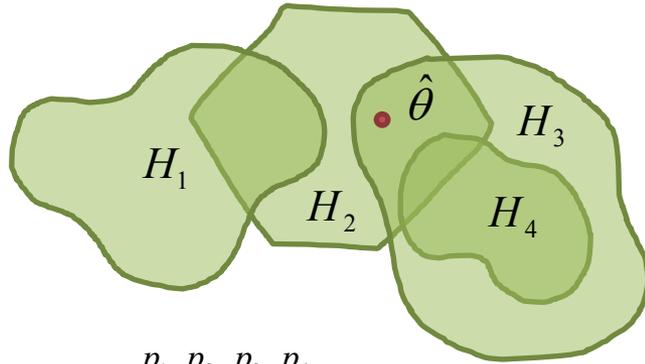
Shimodaira (2004)



$$\sigma \bar{\Phi}^{-1}(\text{BP}(\sigma_1^2, \sigma_2^2)) = \sigma \bar{\Phi}^{-1}(\text{BP}(\sigma^2)) + n^{-1/2} a \sigma_1^2 \sigma_2^2 \sigma^{-4} (\beta_0 - \sigma^2) + O_p(n^{-1})$$

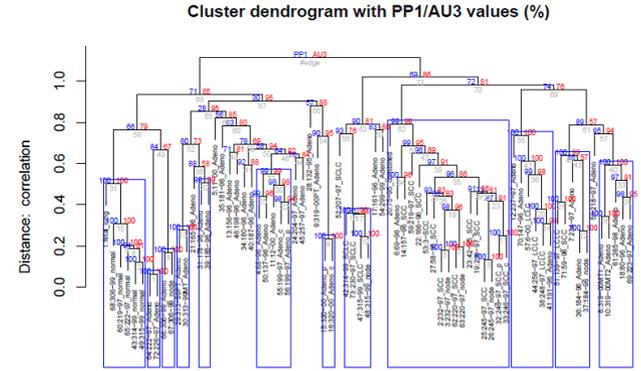
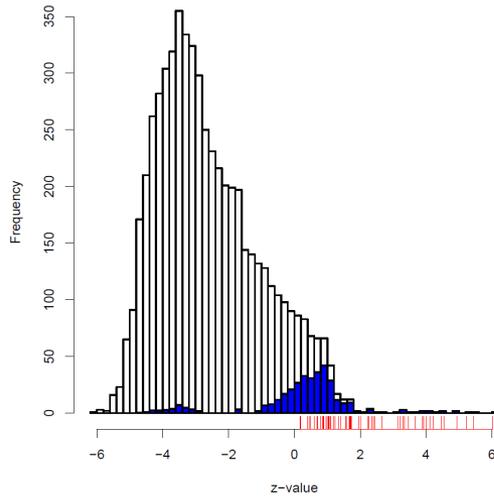
c.f. Three-step multiscale bootstrap estimates six geometrical parameters and gives third-order accuracy

Many hypotheses: FDR, visualization

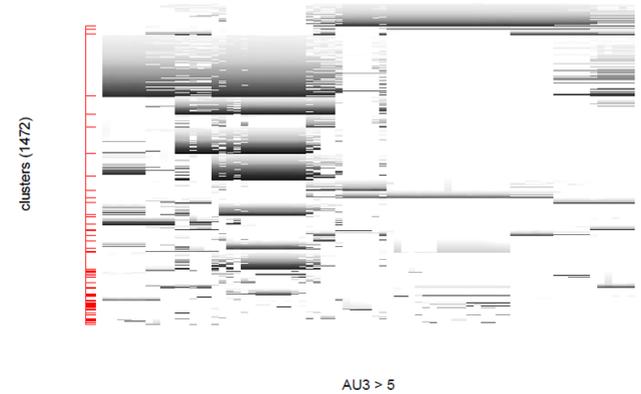


$P_1, P_2, P_3, P_4, \dots$

Histogram of z-values with ticks of critical values (AU3)



PP1 > 95
Cluster method: average



Asymptotic theory of 4th order accuracy

$$h(u) = \sum_{i=1}^q \sum_{j=1}^q h_{ij} u_i u_j + \sum_{i=1}^q \sum_{j=1}^q \sum_{k=1}^q h_{ijk} u_i u_j u_k + \dots$$

$$h_{ij} = \frac{1}{2} \frac{\partial^2 h(u)}{\partial u_i \partial u_j} \Big|_0 = O(n^{-1/2}) \quad h_{ijk} = \frac{1}{6} \frac{\partial^3 h(u)}{\partial u_i \partial u_j \partial u_k} \Big|_0 = O(n^{-1})$$

曲面の微分係数
のスケーリング則

The k -th order derivatives are $O(n^{-(k-1)/2})$ for $k \geq 1$, because the coordinates u_1, \dots, u_q as well as $h(u)$ are scaled by the factor \sqrt{n}

(Class \mathcal{S})

We take care of terms up to $O(n^{-3/2})$ ignoring $O(n^{-2})$

$$h(u) \simeq h_0 + h_i u_i + h_{ij} u_i u_j + h_{ijk} u_i u_j u_k + h_{ijkl} u_i u_j u_k u_l$$

$$h_0 = O(1), h_i = O(n^{-1}), h_{ij} = O(n^{-1/2}), h_{ijk} = O(n^{-1}), h_{ijkl} = O(n^{-3/2})$$

Thm: Asymptotic expansion of BP1

$$\text{BP}_1(H|y) \simeq 1 - \Phi(\beta_0 + \beta_1 + \beta_2)$$

Proved by a simple argument of Taylor expansion and integration.

data point $y = (0, \lambda_0 - h_0)$

signed distance $\beta_0 = \lambda_0 = O(1)$

curvature + ... $\beta_1 = \underline{\gamma_1} - \lambda_0 \gamma_2 + \frac{4}{3} \lambda_0^2 \gamma_3 = O(n^{-1/2})$

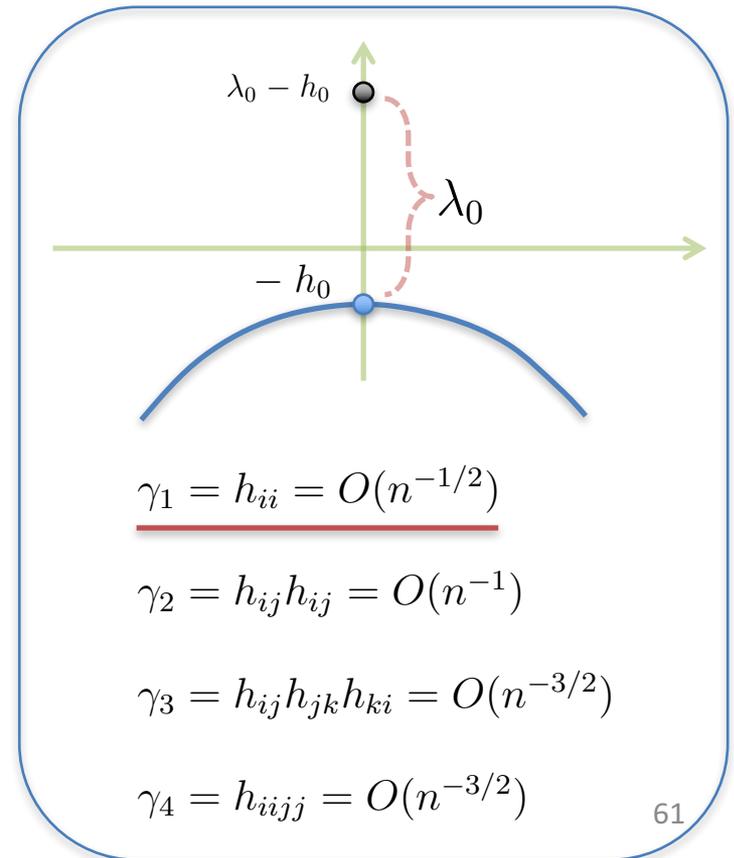
fourth-order terms $\left\{ \begin{array}{l} \beta_2 = 3\gamma_4 - \gamma_1 \gamma_2 - \frac{4}{3} \gamma_3 = O(n^{-3/2}) \\ \beta_3 = 6\gamma_4 - 2\gamma_1 \gamma_2 - 4\gamma_3 = O(n^{-3/2}) \end{array} \right.$

$$\gamma_1 = \frac{1}{2} \frac{\partial^2 h(u)}{\partial u_i \partial u_i} \Big|_0$$

mean curvature of the surface
仮説曲面の平均曲率

$$\beta_3 = \frac{1}{2} \frac{\partial^2 \gamma_1(h, u)}{\partial u_i \partial u_i} \Big|_0$$

mean curvature of the mean curvature
平均曲率の平均曲率



ブートストラップ確率のスケーリング則

Thm: scaling law of BP

$$\text{BP}_{\sigma^2}(H|y) \simeq 1 - \Phi \left[\beta_0 \sigma^{-1} + \beta_1 \sigma + \beta_2 \sigma^3 \right]$$

$$\text{NBP}_{\sigma^2}(H|y) = \Phi \left[\sigma \Phi^{-1}(\text{BP}_{\sigma^2}(H|y)) \right]$$

$$\text{NBP}_{\sigma^2}(H|y) \simeq \bar{\Phi} \left[\beta_0 + \beta_1 \sigma^2 + \beta_2 \sigma^4 \right]$$

Proved by a simple rescaling argument.

$$\text{BP}_{\sigma^2}(H|y) = \text{BP}_1(H/\sigma|y/\sigma)$$

$$\beta_0 \rightarrow \beta_0 \sigma^{-1}, \quad \beta_1 \rightarrow \beta_1 \sigma, \quad \beta_2 \rightarrow \beta_2 \sigma^3$$

$$\lambda_0 \rightarrow \lambda_0 / \sigma$$

$$h_0 \rightarrow h_0 / \sigma, h_i \rightarrow h_i, h_{ij} \rightarrow \sigma h_{ij}, h_{ijk} \rightarrow \sigma^2 h_{ijk}, h_{ijkl} \rightarrow \sigma^3 h_{ijkl}$$

$$\gamma_1 \rightarrow \sigma \gamma_1, \gamma_2 \rightarrow \sigma^2 \gamma_2, \gamma_3 \rightarrow \sigma^3 \gamma_3, \gamma_4 \rightarrow \sigma^3 \gamma_4$$

バイアスの無い（不偏な）p-値

Thm: unbiased p-value

Def: k-th order accurate p-values should satisfy

$$P\left[\text{PV}(H|Y) < \alpha \mid \mu\right] = \alpha + \underbrace{O(n^{-k/2})}_{\text{error}}, \quad \mu \in \partial H.$$

Thm: fourth-order accuracy (k=4) is achieved by

$$\text{PV}(H|y) \simeq 1 - \Phi\left[\beta_0 - \beta_1 - \beta_2 + \beta_3\right]$$

$\begin{array}{cccc} \uparrow & \uparrow & \uparrow & \uparrow \\ O(1) & O(n^{-1/2}) & O(n^{-3/2}) & \end{array}$

Corollary: BP is first-order accurate (k=1), AU is third-order accurate (k=3)

$$\text{NBP}_{\sigma^2}(H|y) \simeq \bar{\Phi}\left[\beta_0 + \beta_1\sigma^2 + \beta_2\sigma^4\right]$$

$$\text{BP} = \text{PV} + O(n^{-1/2}) \quad \text{AU} = \text{PV} + O(n^{-3/2})$$

Rejection probabilities of BP and AU

BP is first-order accurate (k=1)

$$P\left(\text{BP}(H|Y) < \alpha\right) = \Phi(z_\alpha + \underline{2\gamma_1}) + O(n^{-1}) = \alpha + O(n^{-1/2})$$

$$z_\alpha = \Phi^{-1}(\alpha)$$

mean curvature of the surface
仮説曲面の平均曲率

AU is third-order accurate (k=3)

$$P\left(\text{AU}(H|Y) < \alpha\right) \simeq \Phi(z_\alpha + \underline{\frac{4}{3}\gamma_3}) = \alpha + O(n^{-3/2})$$

Using $q \times q$ hessian matrix $D = \left(\frac{\partial^2 h(u)}{\partial u_i \partial u_j} \Big|_0 : i, j = 1, \dots, q \right)$

$$\gamma_1 = h_{ii} = \frac{1}{2} \text{tr}(D) \quad \gamma_3 = h_{ij} h_{jk} h_{ki} = \frac{1}{8} \text{tr}(D^3)$$

曲面の移動

4.3. Shifting surfaces

We consider shifting $\mathcal{B}(h)$ toward the normal direction. Let $f(u)$ be the normal vector at $(u, -h(u)) \in \mathcal{B}(h)$. For a specified $\lambda \in \mathcal{S}$, we move the point $(u, -h(u))$ by $\lambda(u)$ toward the normal direction. This is expressed as

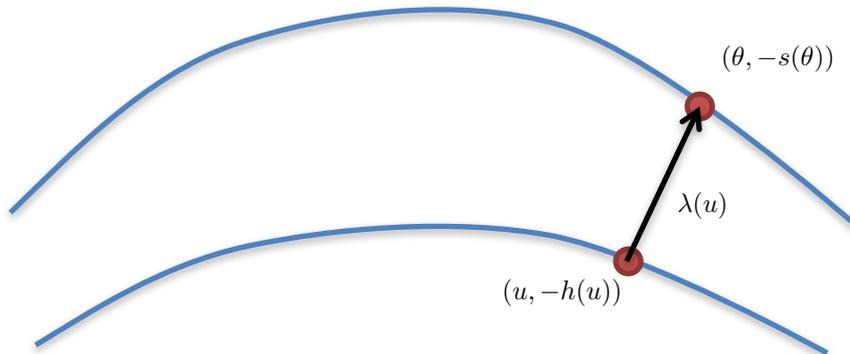
$$(\theta, -s(\theta)) = (u, -h(u)) + \lambda(u) \|f(u)\|^{-1} f(u), \quad (25)$$

where $s(u)$ is some function of $u \in \mathbb{R}^q$, and $\theta \in \mathbb{R}^q$ is used when distinction is needed. We can interpret (25) as

$$\hat{\mu}(H|(\theta, -s(\theta))) = (u, -h(u))$$

with signed distance $\lambda(u)$. For sufficiently large n , such $s(\theta)$ is uniquely defined for each θ , because all the surfaces approach flat as $n \rightarrow \infty$. We denote (25) as

$$s = \mathcal{M}(h, \lambda).$$



Lemma 3. *Let $s = \mathcal{M}(h, \lambda)$ for $h \in \mathcal{S}$, $\lambda \in \mathcal{S}$. If $\lambda(u)$ is expressed as*

$$\lambda(u) \simeq \lambda_0 + \lambda_i u_i + \lambda_{ij} u_i u_j$$

with $\lambda_0 = O(1)$, $\lambda_i = O(n^{-1})$, $\lambda_{ij} = O(n^{-3/2})$, then we have $s \in \mathcal{S}$ with coefficients $s_0 = h_0 - \lambda_0 = O(1)$, $s_i = h_i - \lambda_i - 2\lambda_0 h_{mi}(h_m - \lambda_m) = O(n^{-1})$, $s_{ij} = h_{ij} - \lambda_{ij} - 2\lambda_0 h_{mi} h_{mj} + 4\lambda_0^2 h_{ml} h_{mi} h_{lj} = O(n^{-1/2})$, $s_{ijk} = h_{ijk} - 2\lambda_0 (h_{mi} h_{mjk} + h_{mj} h_{mik} + h_{mk} h_{mij}) = O(n^{-1})$, $s_{ijkl} = h_{ijkl} = O(n^{-3/2})$. The four geometric quantities at $(0, -s(0))$ are $\gamma_1(s, 0) = s_{ii} \simeq \gamma_1 - \lambda_{ii} - 2\lambda_0 \gamma_2 + 4\lambda_0^2 \gamma_3$, $\gamma_2(s, 0) = s_{ij} s_{ij} \simeq \gamma_2 - 4\lambda_0 \gamma_3$, $\gamma_3(s, 0) = s_{ij} s_{jk} s_{ki} \simeq \gamma_3$, $\gamma_4(s, 0) = s_{iijj} \simeq \gamma_4$, where $\gamma_i = \gamma_i(h, 0)$, $i = 1, \dots, 4$.

BPが一定の曲面

5.1. Contour surfaces of bootstrap probability

We consider a surface on which the bootstrap probability remains constant. For $H = \mathcal{R}(h)$ with $h \in \mathcal{S}$, we consider a function $s(u)$ of $u \in \mathbb{R}^q$ satisfying

$$\text{BP}_{\sigma^2}(H|(u, -s(u))) = 1 - \alpha, \quad u \in \mathbb{R}^q,$$

meaning $\text{BP}_{\sigma^2}(H|y) = 1 - \alpha$ is constant for any $y \in \mathcal{B}(s)$. Then, $\mathcal{B}(s)$, as well as s itself, will be called as the *contour surface of the bootstrap probability* of H with variance $\sigma^2 > 0$ at level $1 - \alpha$. In particular, we choose α so that $(0, \lambda_0 - h_0) \in \mathcal{B}(s)$ for a specified $\lambda_0 \in \mathbb{R}$. We denote this contour surface as

$$s = \mathcal{L}_{\sigma^2}(h, \lambda_0).$$

Lemma 4. *Let $s = \mathcal{L}_{\sigma^2}(h, \lambda_0)$ for $h \in \mathcal{S}$, $\lambda_0 \in \mathbb{R}$, and $\sigma^2 > 0$. Then, s is expressed as $s = \mathcal{M}(h, \lambda)$ by specifying $\lambda(u) \simeq \lambda_0 + \lambda_i u_i + \lambda_{ij} u_i u_j$ with $\lambda_0 = O(1)$,*

$$\lambda_i = \sigma^2(-3h_{mmi} + 6\lambda_0 h_{ml} h_{mli}), \quad \lambda_{ij} = \sigma^2(-6h_{mmij} + 2h_{mm} h_{li} h_{lj} + 4h_{ml} h_{mi} h_{lj}). \quad (26)$$

We have $s \in \mathcal{S}$ with coefficients

$$\begin{aligned} s_0 &= h_0 - \lambda_0, \quad s_i = h_i - 2\lambda_0 h_m h_{mi} + \sigma^2(3h_{mmi} - 6\lambda_0 h_{ml} h_{mli} - 6\lambda_0 h_{mi} h_{mli}), \\ s_{ij} &= h_{ij} - 2\lambda_0 h_{mi} h_{mj} + 4\lambda_0^2 h_{ml} h_{mi} h_{lj} + \sigma^2(6h_{ijmm} - 2h_{mm} h_{li} h_{lj} - 4h_{ml} h_{mi} h_{lj}), \\ s_{ijk} &= h_{ijk} - 2\lambda_0(h_{mi} h_{mjk} + h_{mj} h_{mik} + h_{mk} h_{mij}), \quad s_{ijkl} = h_{ijkl}. \end{aligned} \quad (27)$$

The four geometric quantities of s at $(0, -s(0))$ are

$$\begin{aligned} \gamma_1(s, 0) &\simeq \gamma_1 - 2\lambda_0 \gamma_2 + 4\lambda_0^2 \gamma_3 + \sigma^2(6\gamma_4 - 2\gamma_1 \gamma_2 - 4\gamma_3), \\ \gamma_2(s, 0) &\simeq \gamma_2 - 4\lambda_0 \gamma_3, \quad \gamma_3(s, 0) \simeq \gamma_3, \quad \gamma_4(s, 0) \simeq \gamma_4, \end{aligned} \quad (28)$$

where $\gamma_i = \gamma_i(h, 0)$, $i = 1, \dots, 4$.

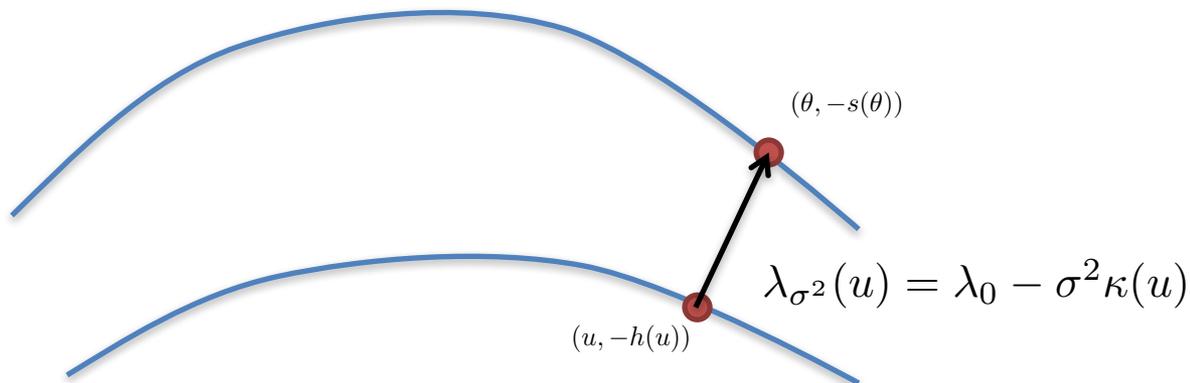
BPの等高線を仮説のシフトと解釈

We denote the $\lambda(u)$ of (26) as $\lambda_{\sigma^2}(u) = \lambda_0 - \sigma^2 \kappa(u)$ with

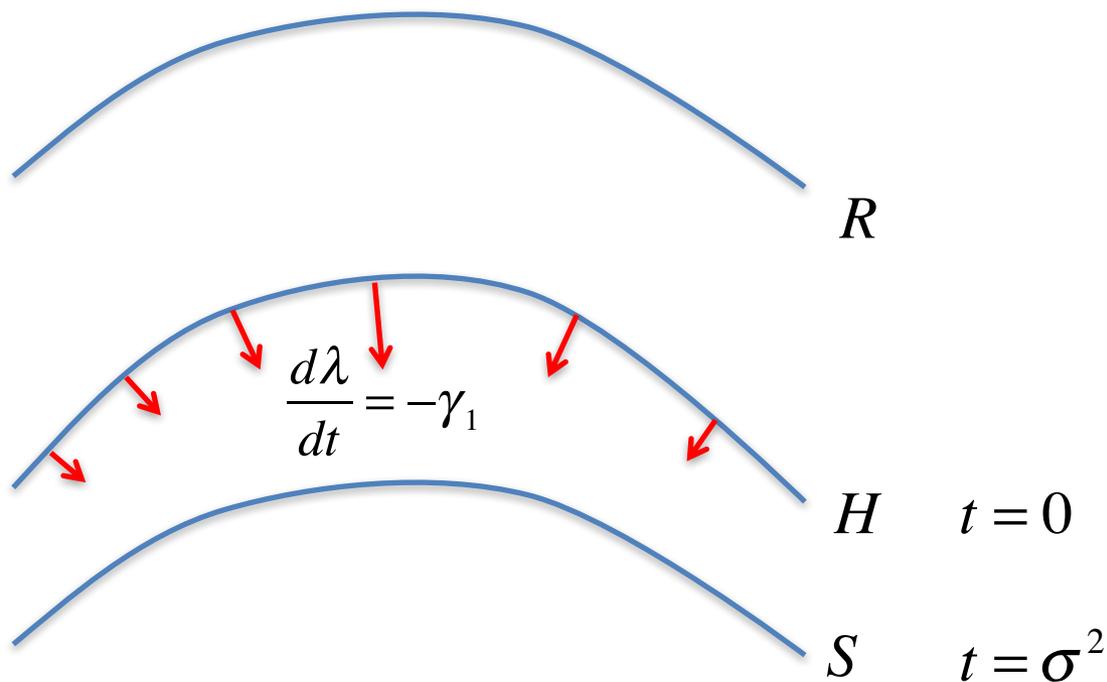
$$\begin{aligned} \kappa(u) &= \gamma_1(h, u) - \gamma_1(h, 0) - \lambda_0(\gamma_2(h, u) - \gamma_2(h, 0)) \\ &\simeq (3h_{mmi} - 6\lambda_0 h_{ml} h_{mli})u_i + (6h_{mmij} - 2h_{mm} h_{li} h_{lj} - 4h_{ml} h_{mi} h_{lj})u_i u_j. \end{aligned} \quad (29)$$

This also relates to (8) as $(1/2)\partial^2 \kappa(u)/\partial u_i \partial u_i|_0 = \beta_3$ or $(1/2)\partial^2 \lambda_{\sigma^2}(u)/\partial u_i \partial u_i|_0 = -\sigma^2 \beta_3$.
The contour surface of $\text{BP}_{\sigma^2}(H|y)$ for $\sigma^2 > 0$ is expressed asymptotically as

$$\mathcal{L}_{\sigma^2}(h, \lambda_0) = \mathcal{M}(h, \lambda_{\sigma^2}),$$

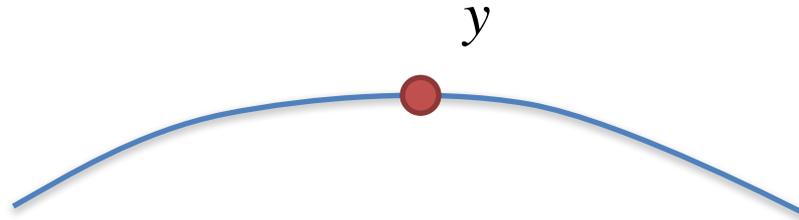


平均曲率流で動かす (3次の議論)



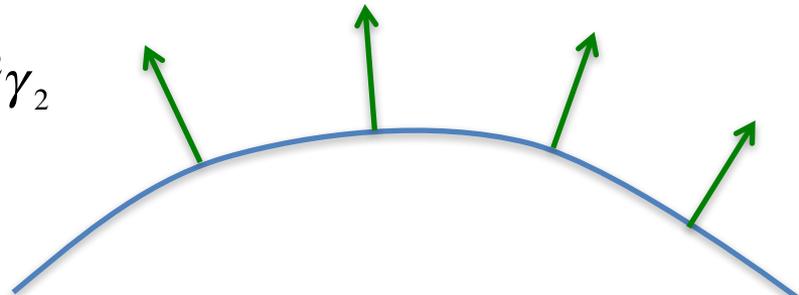
$$NBP_{\sigma^2}(H \mid (u, -s(u))) = 1/2$$

さらに曲面を動かす



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

$$\frac{d\lambda}{dt} = 1 + \sigma^2 \gamma_2$$



$S \quad t = 0$

$$NBP_{\sigma^2}(H | (u, -s(u))) = 1/2$$

BPの等高線(曲面)のオペレータを定義して証明します

sketch of the proof for PV

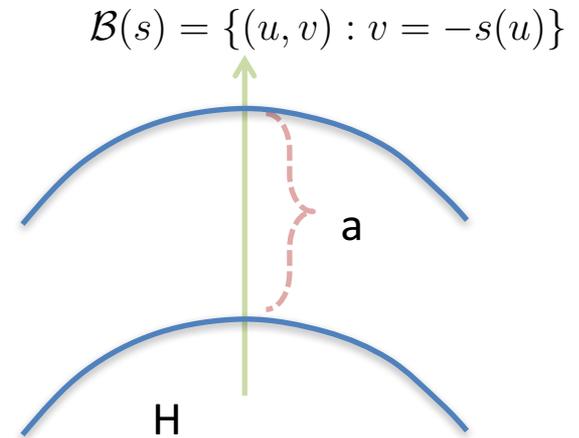
Contour surface of BP $s = \mathcal{L}_{\sigma^2}(h, a)$

is defined by $\text{BP}_{\sigma^2}(H|y) = \text{constant}$ for any $y \in \mathcal{B}(s)$

additivity $\mathcal{L}_{\sigma_2^2}(\mathcal{L}_{\sigma_1^2}(H, a_1), a_2) \doteq \mathcal{L}_{\sigma_1^2 + \sigma_2^2}(H, a_1 + a_2)$

identity $h = \mathcal{L}_0(h, 0)$

inverse $\mathcal{L}_{\sigma^2}^{-1}(\cdot, a) \doteq \mathcal{L}_{-\sigma^2}(\cdot, -a)$



Consider contour surface of PV: $\text{PV}(H|y) = \text{constant}$ for any $y \in \mathcal{B}(s)$

then unbiasedness requires $\text{BP}_1(\mathcal{R}(s)|\mu) = \text{constant}$ for any $\mu \in \mathcal{B}(h)$

$$h = \mathcal{L}_1(s, -\lambda_0) \quad \longrightarrow \quad s = \mathcal{L}_{-1}(h, \lambda_0)$$

Thm: scaling-law of DBP

$$\text{DBP}_{\tau^2, \sigma^2}(H|y) \simeq 1 - \Phi \left[\beta_0 \tau^{-1} - \beta_1 \tau - \beta_2 \tau^3 - \beta_3 \tau \sigma^2 \right]$$

$$\bar{\Phi}^{-1} \left[\text{DBP}_{1, \sigma^2}(H|y) \right] \simeq (\beta_0 - \beta_1 - \beta_2) - \underline{\beta_3 \sigma^2}$$

mean curvature of the mean curvature
平均曲率の平均曲率

$$\beta_3 = \frac{1}{2} \frac{\partial^2 \gamma_1(h, u)}{\partial u_i \partial u_i} \Big|_0$$

Corollary: DBP is third-order accurate (k=3), DAU is fourth-order accurate (k=4)

$$\text{PV}(H|y) \simeq 1 - \Phi \left[\beta_0 - \beta_1 - \beta_2 + \beta_3 \right]$$

$$\text{DBP} = \text{PV} + O(n^{-3/2}) \quad \text{DAU} = \text{PV} + O(n^{-2})$$

sketch of the proof for DBP

$$s = \mathcal{L}_{\sigma^2}(h, \lambda_0) \quad \text{contour surface of BP}$$

$$\widetilde{\text{DBP}}_{\tau^2, \sigma^2}(H|y) = 1 - \text{BP}_{\tau^2}(\mathcal{R}(s)|\tilde{\mu})$$

The proof completes by applying the asymptotic expansion of BP to $\mathcal{R}(s)$

棄却確率

Rejection probabilities of DBP and DAU

$$P\left(\text{DBP}_{1,\sigma^2}(H|Y) < \alpha\right) \simeq \Phi\left[z_\alpha - \underbrace{(1 + \sigma^2)\beta_3}_{\text{mean curvature of the mean curvature}}\right]$$

mean curvature of the mean curvature

平均曲率の平均曲率

DBP is third-order accurate (k=3)

$$P\left(\text{DBP}(H|Y) < \alpha\right) \simeq \Phi\left(z_\alpha - \underbrace{2\beta_3}_{\text{mean curvature of the mean curvature}}\right) = \alpha + O(n^{-3/2})$$

DAU is fourth-order accurate (k=4)

$$P\left(\text{DAU}(H|Y) < \alpha\right) \simeq \Phi(z_\alpha) = \alpha$$

$$\gamma_1 = \frac{1}{2} \frac{\partial^2 h(u)}{\partial u_i \partial u_i} \Big|_0$$

mean curvature of the surface

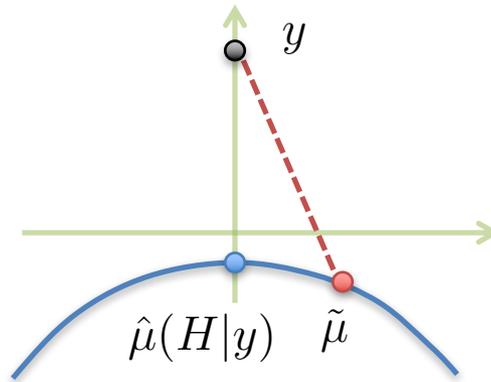
$$\beta_3 = \frac{1}{2} \frac{\partial^2 \gamma_1(h, u)}{\partial u_i \partial u_i} \Big|_0$$

mean curvature of the mean curvature

DAUは射影誤差に影響されにくい

Robustness to projection error

If $\hat{\mu}(H|y) = (0, -h(0)) \in \partial H$ is replaced by $\tilde{\mu} = (\theta, -h(\theta)) \in \partial H$



DBP becomes

$$\widetilde{\text{DBP}}_{\tau^2, \sigma^2}(H|y) \simeq \bar{\Phi} \left[\beta_0 \tau^{-1} - \beta_1 \tau - \beta_2 \tau^3 - \beta_3 \tau \sigma^2 - \underbrace{\tau^{-1}(\tau^2 + \sigma^2) \kappa(\theta)}_{\text{error} = O(n^{-1})} \right]$$

$$\begin{aligned} \kappa(u) &= \gamma_1(h, u) - \gamma_1(h, 0) - \lambda_0(\gamma_2(h, u) - \gamma_2(h, 0)) \\ &\simeq (3h_{mmi} - 6\lambda_0 h_{ml} h_{mli})u_i + (6h_{mmij} - 2h_{mm} h_{li} h_{lj} - 4h_{ml} h_{mi} h_{lj})u_i u_j \end{aligned}$$

Corollary: DBP becomes only second-order accurate (k=2), but
DAU keeps fourth-order accuracy (k=4)

$$\widetilde{\text{DBP}} = \text{DBP} + O(n^{-1})$$

$$\widetilde{\text{DAU}} = \text{DAU} + O(n^{-2})$$

Efron and Tibshirani (1998)の結果

Shimodaira (2014)

まとめ

- ブートストラップ確率はベイズ事後確率
- 頻度論のp-値との差は、仮説境界の「平均曲率」が原因
- ダブルブートストラップ法でも修正したときの誤差は、「平均曲率の平均曲率」
- マルチスケールブートストラップ法で誤差解消
- 検定以外にも使えるかも？（境界までの距離や境界の曲率を推定できる）

Summary and other issues

- DAU = “DBP with $m=-n$ ” is proposed
- The accuracy of BP is first order ($k=1$), AU is third-order ($k=3$), DBP is third-order ($k=3$)
- DAU is fourth-order accurate ($k=4$)
- DAU is robust to the projection error (surprisingly, $k=4$)
- Geometry of surfaces played important roles

- Shimodaira (2008) showed another theory of AU using unusual asymptotic theory of “nearly flat surfaces”

- Shimodaira (2004) discussed deviation from the multivariate normal model, and results for exponential family distributions are given there for multistep-AU

- Future topics may be DAU for nearly flat surfaces, or for exponential family distributions