

ブートストラップ法の幾何学と スケール変換

東工大 情報理工
数理・計算科学専攻
下平英寿

日本数学会 2004年秋季総合分科会

概要

- 現実の複雑な応用(バイオインフォマティクスなど) ⇒ 精度の高い確率値の必要性
- 近似的に不偏な検定
 - k 次の精度 バイアス = $O(n^{-k/2})$
- ブートストラップ確率
 - Felsenstein (1985) 進化系統樹推定に応用
 - 1次の精度
- スケール変換(サンプルサイズを変更)
 - ブートストラップ確率の変化率
 - 幾何学的な量を推定 (BCa法に相当)
 - 3次の精度
- これまでに提案されたアルゴリズム
 - 精度, 計算量, 実装
- ブートストラップ法 (Efron 1979)
 - 精度1次, 計算量 $O(B)$, 射影不要
- p^* -formula (Barndorff-Nielsen 1986)
 - 精度3次, モデル既知, 射影必要
- ダブルブートストラップ法 (Hall 1992)
 - 精度3次, 計算量 $O(B^2)$, 射影必要
- 2レベル・ブートストラップ法 (Efron et al 1996)
 - 精度2次, 計算量 $O(B)$, 射影必要
- マルチスケール・ブートストラップ法 (Shimodaira 2002, 2004)
 - 精度3次, 計算量 $O(B)$, 射影不要

分子系統樹の統計推測

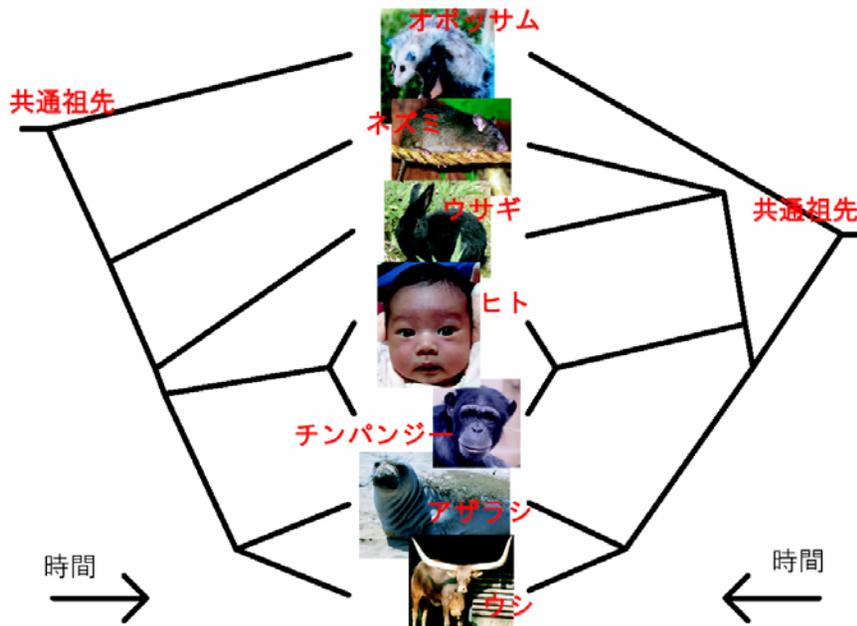
哺乳類のミトコンドリアDNA

(オポッサム, マウス, ウサギ, ヒト, チンプ, アザラシ, ウシ)



オポッサム	ATTAACCTTATTAATATATATATCCCTATCCTCCTAGCTGTAGCATTTTAACTCTAGTAGA
マウス	ATTAATATCCAACTCCTCGTCCCCATTCTAATCGCCATAGCCTTCCTAACATTAGTAGA
ウサギ	ATTAATACACTCCTTTTAACTTACCTGTACTTTAGCCATAGCATTCCTCACCTTAGTCGA
ヒト	GCCAACTCCTACTCCTCATTGTACCCATTCTAATCGCAATGGCATTCCCTAAATGCTTACCGA
チンプ	ACCAACTCCTACTCCTCATTGTACCCATCCTAATCGCAATAGCATTCCTAAATGCTAACCGA
アザラシ	ATTAATATCACTCACTAATTAATCCCAATTCTCCTCGCCGTAGCTTTCCTAACATTAGTAGA
ウシ	ATTAACATCTTAACTACTAATTAATCCCATCCTATTGGCCGTAGCATTCCTTACGTTAGTGGA

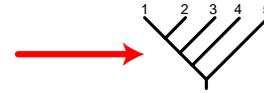
1.....10.....20.....30.....40.....50.....60...



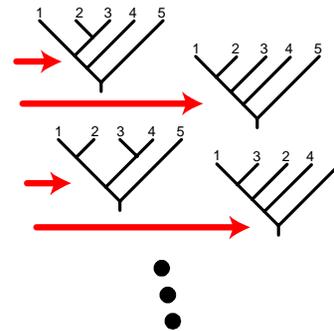
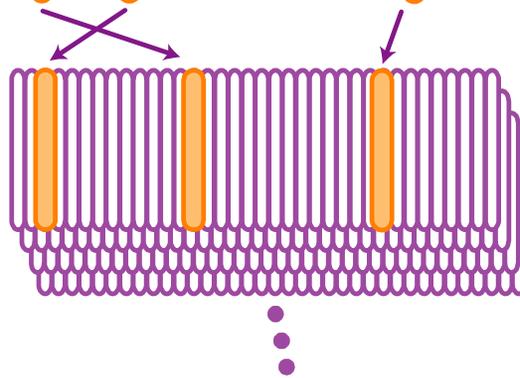
ブートストラップ法の計算

データ

```
GCACCTCTACTCCTCATTGTACCCATCTAATCSCAATGG
ACACCTCTACTCCTCATTGTACCCATCTAATCSCAATAG
ATTAATATCATCTCACTAATTATCCCAATCTCTCCGCCGTAG
ATTACATCTAATACTAATTATCCCATCCTATTGCCGTAG
ATTATACACTCTTTTAATCCTACCTGTACTTTAGCCATAG
ATTAATATCTTACACTCCTCGTCCCAATCTTAATCSCATAG
ATTACTTATTAAATATATTATCCCTATCCTCTAGCTGTAG
```

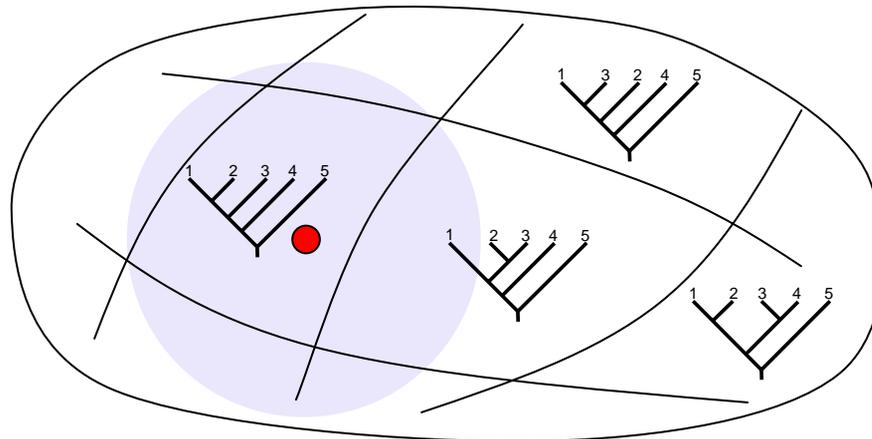


多数の複製データ
(10000セット)



出現頻度を
数える

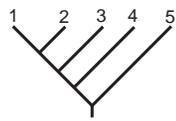
データ空間の分割



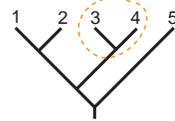
各領域の確率

ブートストラップ確率

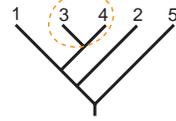
1=ヒト, 2=(アザラシ, ウシ), 3=ウサギ, 4=マウス, 5=オポッサム



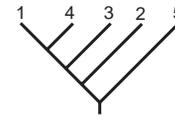
0.58



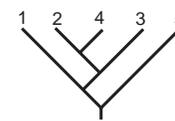
0.01



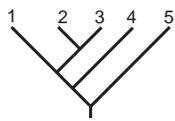
0.02



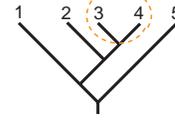
0.00



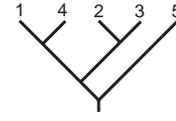
0.00



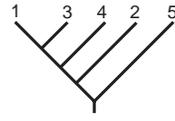
0.31



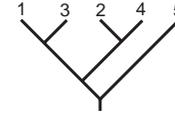
0.04



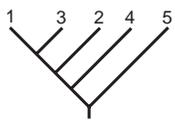
0.00



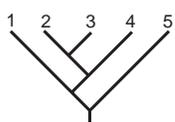
0.00



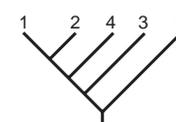
0.00



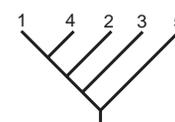
0.04



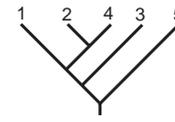
0.01



0.00



0.00



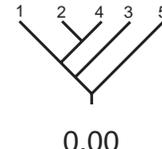
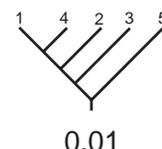
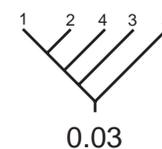
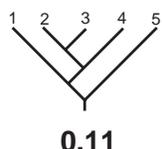
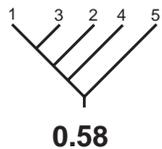
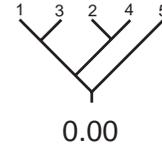
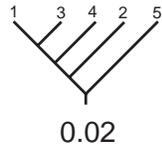
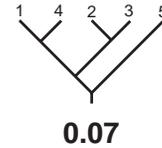
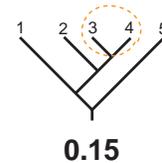
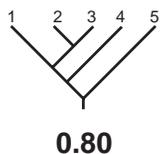
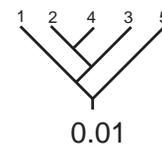
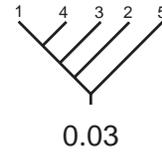
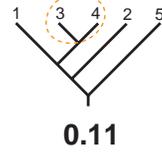
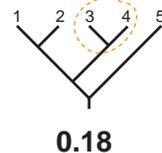
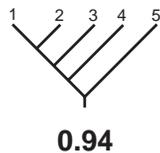
0.00

正のバイアス ⇒ false positiveが多く危険
 仮説の領域が近似的に凸(曲率が正)であることが関係

1次の精度の近似的に不偏な検定

Shimodaira-Hasegawa test

1=ヒト, 2=(アザラシ, ウシ), 3=ウサギ, 4=マウス, 5=オポッサム



負のバイアス ⇒ false negativeが多い. でも安全.

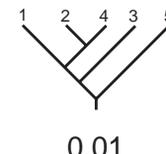
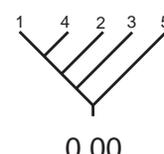
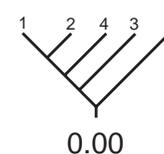
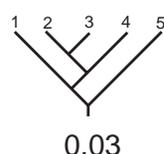
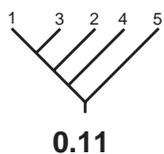
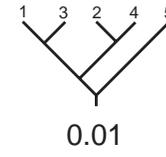
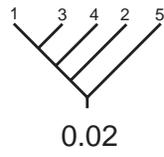
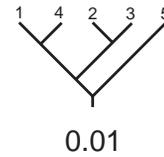
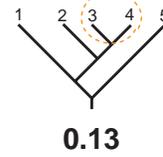
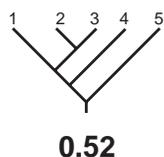
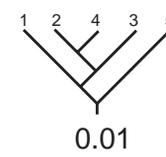
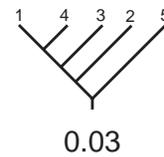
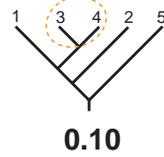
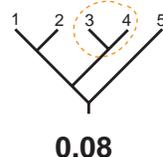
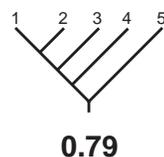
多重比較法の保守性(最悪ケースの評価)

Shimodaira and Hasegawa (1999) *Mol. Biol. Evol.*

SCI論文被引用回数2004年7月までに400回 (年間150回のペース)

マルチスケール法の確率値

1=ヒト, 2=(アザラシ, ウシ), 3=ウサギ, 4=マウス, 5=オポッサム



近似的にバイアスはゼロ(3次の精度) ⇒ 理想的

Shimodaira and Hasegawa (2001) *Bioinformatics*

ISI New Hot Papers in Computer Science (2003-07)

Shimodaira (2002) *Syst.Biol.* これらの被引用回数も100回程度(2004年7月)

確率モデルと帰無仮説

指数型分布族 (p 次元)

$$Y \sim f(y; \eta)$$

観測値

$$y = (y_1, \dots, y_p)$$

期待値パラメタ

$$\eta = E(Y)$$

例: 正規モデル

$$Y \sim N_p(\eta, I_p)$$

興味のある仮説

$$\eta \in \mathcal{R}$$

領域

\mathcal{R} (体積>0, 任意の形状)

境界

$\partial\mathcal{R}$ (なめらかな曲面)

例: 球体

$$\mathcal{R} = \{\eta : \|\eta\| \leq \sqrt{n}\}$$

球面

$$\partial\mathcal{R} = \{\eta : \|\eta\| = \sqrt{n}\}$$

サンプルサイズを変更

$$Y = \sqrt{n} \frac{X_1 + \dots + X_n}{n}$$

$$Y \sim f(y; \eta)$$

$$n \Rightarrow n'$$

スケール変換

$$\tau = \sqrt{\frac{n}{n'}}$$

$$Y = \sqrt{n} \frac{X_1 + \dots + X_{n'}}{n'}$$

$$Y \sim f(y; \eta, \tau)$$

ポテンシャル関数 $\phi(\eta, \tau) = \phi(\eta) / \tau^2$

例: 正規モデル $Y \sim N_p(\eta, \tau^2 I_p)$

$$\phi(\eta, \tau) = \|\eta\|^2 / (2\tau^2)$$

確率値と不偏な検定

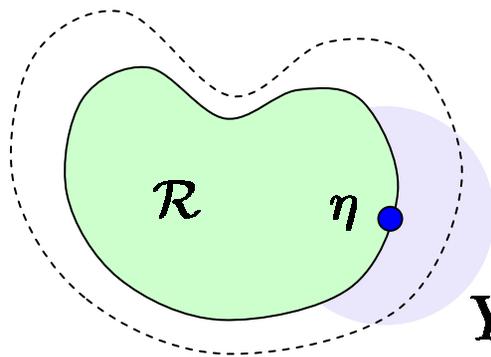
確率値 $\hat{\alpha}(y) \Rightarrow \mathcal{R}$ から離れるほど小さな値をとる

有意水準 $0 < \alpha < 1$

k 次の精度の近似的に不偏な検定

$\eta \in \partial\mathcal{R}$ の任意の点に対して

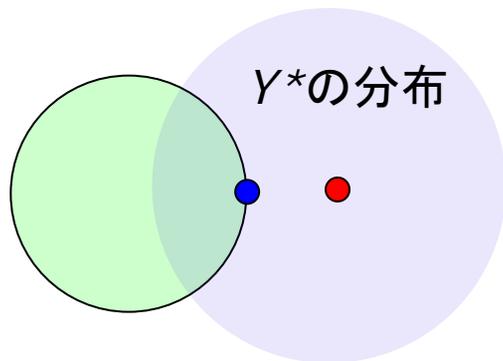
$$\Pr\{\hat{\alpha}(Y) < \alpha; \eta, \mathbf{1}\} = \alpha + O(n^{-k/2})$$



$$Y \sim f(y; \eta, \mathbf{1})$$

点線は $\hat{\alpha}(y) = 0.05$ の等高線
これより外側に出る確率が0.05
になるように等高線をデザインする

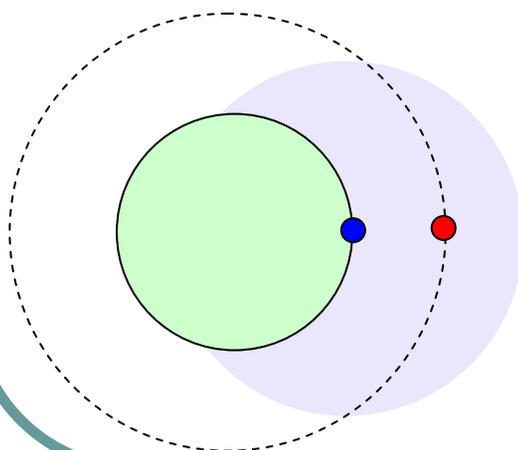
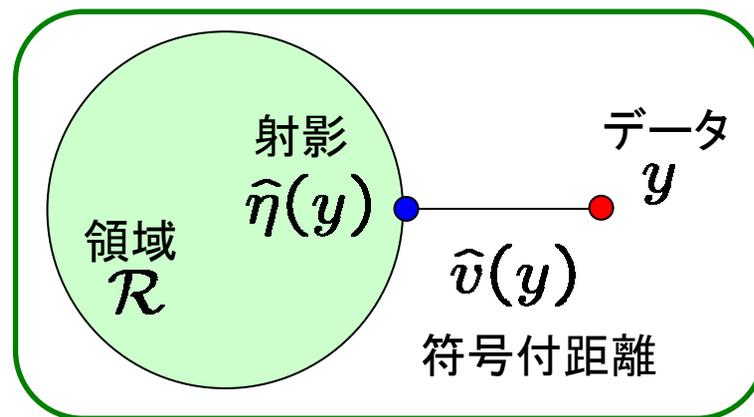
ブートストラップ確率の幾何学



$$Y^* \sim f(y^*; y, 1)$$

$$\hat{\alpha}_0(y) = \Pr\{Y^* \in \mathcal{R}; y, 1\}$$

ブートストラップ確率は1次の精度

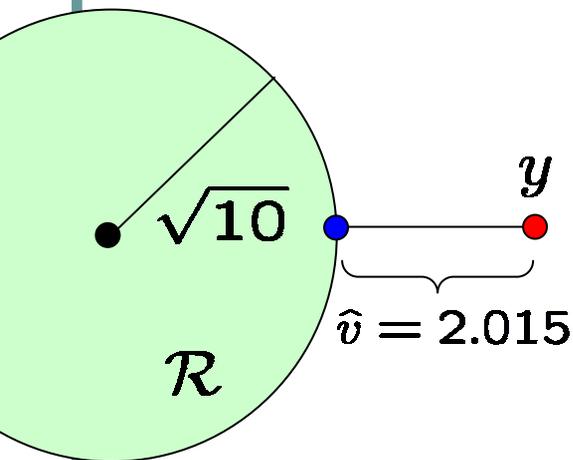


$$Y^* \sim f(y^*; \hat{\eta}(y), 1)$$

$$\hat{\alpha}_\infty(y) = \Pr\{\hat{v}(Y^*) \geq \hat{v}(y); \hat{\eta}(y), 1\}$$

この確率値は3次の精度

数値例1 (正規モデル)



正規モデル $Y \sim N_p(\eta, \tau^2 I_p)$

球体 $\mathcal{R} = \{\eta : \|\eta\| \leq \sqrt{n}\}$

$p = 4, \quad n = 10$

観測値 $\hat{v} = \|y\| - \sqrt{n} = 2.015$

非心カイ二乗分布から厳密に不偏な確率値が計算できる

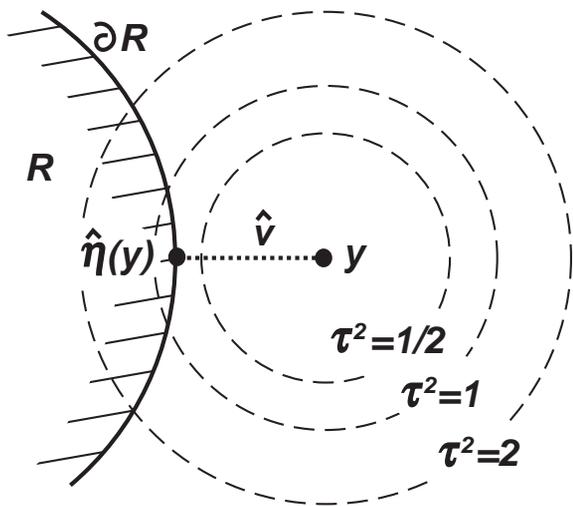
$$\hat{\alpha}_\infty(y) = \Pr\{\|Y\| \geq \|y\|; \eta \in \partial\mathcal{R}\} = 0.05$$

ブートストラップ確率 (スケール 1)

$$\hat{\alpha}_0(y) = 0.0085$$



ブートストラップ確率 (スケール τ)



$$Y^* \sim f(y^*; y, \tau)$$

$$\tilde{\alpha}_1(y, \tau) = \Pr\{Y^* \in \mathcal{R}; y, \tau\}$$

z 値で表現すると都合が良い $z = -\Phi^{-1}(\alpha)$
 \approx は $O(n^{-3/2})$ の誤差を表す(3次の精度)

正規モデルの場合

ブートストラップ確率 $\tilde{z}_1(y, \tau) \approx \hat{v}/\tau + \hat{c}\tau$

不偏な確率値 $\tilde{z}_\infty(y) \approx \hat{v} - \hat{c}$

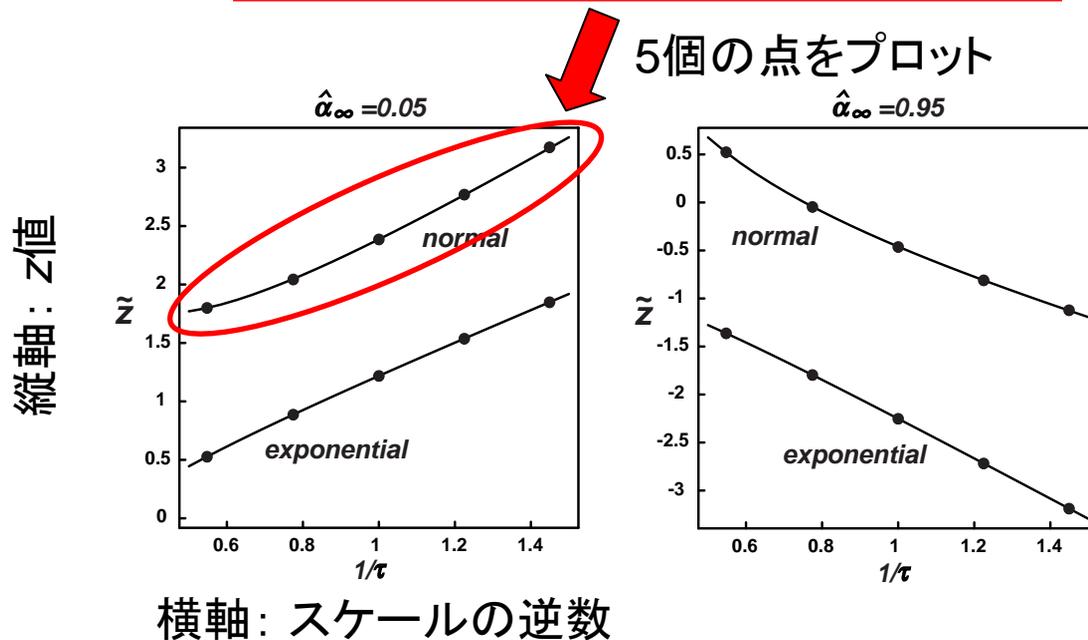
\hat{d}^{ab} = 曲率行列

$$\hat{c} = \hat{d}^{aa} - (\hat{d}^{ab})^2 \hat{v}$$

数値例1 (スケールの変化)

$$n = 10 \Rightarrow n_1 = 3, 6, 10, 15, 21 \quad \tau_1 = \sqrt{\frac{n}{n_1}}$$

$$\tilde{\alpha}_1(y, \tau_1) = \underline{0.0359, 0.0205, 0.0085, 0.0028, 0.0008}$$



理論式の当てはめ $\tilde{z}_1(y, \tau) \approx \hat{v}/\tau + \hat{c}\tau$ 未知係数 \hat{v}, \hat{c}

数値例1 (確率値の計算)

ブートストラップ確率

$$\tilde{z}_1(y, \tau) \approx \hat{v}/\tau + \hat{c}\tau$$

\hat{v} と \hat{c} を回帰係数として推定

未知の幾何的量(符号付距離, 曲率)

$$\hat{v} = 2.015, \quad \hat{c} = 0.323$$

マルチスケール・ブートストラップ法
による推定値

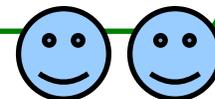
$$\underline{\hat{v} = 2.002, \quad \hat{c} = 0.385}$$

不偏な確率値

$$\hat{z}_1(y) = \underline{\hat{v}} - \hat{c}$$

$$\hat{\alpha}_1(y) = \Phi(-\hat{z}_1(y))$$

$$\hat{\alpha}_1(y) = \Phi(-2.002 + 0.385) = 0.0529$$



ブートストラップ
確率の変化率

$$\left. \frac{\partial(\hat{v}/\tau + \hat{c}\tau)}{\partial(1/\tau)} \right|_{\tau=1} = \hat{v} - \hat{c}$$

数値例2 (指数モデル)

指数分布 $X_1, \dots, X_n \sim \exp(-x/\mu - \log \mu)$

$Y \sim$ ガンマ分布 ($p=1$, 指数 n , 期待値 η)

仮説 $\eta \leq \sqrt{n}$

観測値 $y - \sqrt{n} = 1.806$

厳密に不偏な確率値 $\hat{\alpha}_\infty(y) = \Pr\{Y \geq y; \eta = \sqrt{n}\} = 0.05$

ブートストラップ確率 $n=10 \Rightarrow n_1 = 3, 6, 10, 15, 21$

$\tilde{\alpha}_1(y, \tau_1) = 0.2990, 0.1875, 0.1115, 0.0622, 0.0322$

$\hat{\alpha}_0(y) = 0.1115$  

幾何的量の推定値 $\hat{v} = 1.328, \hat{c} = -0.110$

確率値の計算

$\hat{\alpha}_1(y) = \Phi(-1.328 - 0.110) = 0.0753$ 

加速定数 $\hat{a} = -\frac{1}{6} \frac{\partial^3 \phi(\eta)}{\partial \eta_p^3} \Big|_{\hat{\eta}}$

一般の指数型分布族の場合, 加速定数は「空間のゆがみ」を表す(正規モデルは0)

$$\text{ブートストラップ確率 } \tilde{z}_1(y, \tau_1) \doteq \frac{\hat{v} - 2\hat{a}\hat{v}^2}{\tau_1} + (\hat{c} - \hat{a})\tau_1$$

$$\text{不偏な確率値 } \hat{z}_\infty(y) \doteq \hat{v} - \hat{c} + \hat{a}(1 - \hat{v}^2)$$

(\doteq は $O(n^{-1})$ の誤差を表す)

マルチスケール・ブートストラップ法で計算した確率値

$$\hat{z}_1(y) \doteq \hat{v} - \hat{c} + \hat{a}(1 - 2\hat{v}^2) \text{ は1次の精度になってしまう}$$

2ステップ法 $Y_1^* \sim f(y_1^*; y, \tau_1), \quad Y_2^* \sim f(y_2^*; y_1^*, \tau_2)$

$$\tilde{\alpha}_2(y, \tau_1, \tau_2) = \Pr\{Y_2^* \in \mathcal{R}; y, \tau_1, \tau_2\}$$

$$\tilde{z}_2(y, \tau_1, \tau_2) - \tilde{z}_1(y, \sqrt{\tau_1^2 + \tau_2^2}) \doteq \frac{\hat{a}\tau_1^2\tau_2^2(\hat{v}^2 - (\tau_1^2 + \tau_2^2))}{(\tau_1^2 + \tau_2^2)^{5/2}}$$

数値例2 (2ステップ法)

$\hat{v}, \hat{c}, \hat{a}$ の代わりに以下の量を係数として推定する

$$\hat{\gamma}_1 \doteq \hat{v} - 2\hat{a}\hat{v}^2, \quad \hat{\gamma}_2 \doteq \hat{v}(\hat{a} - \hat{c}), \quad \hat{\gamma}_3 \doteq \hat{v}\hat{a}$$

$$\hat{z}_2(y, \tau_1, \tau_2) \doteq s_1 \hat{\gamma}_1 (1 + s_2 \hat{\gamma}_3) - \frac{\hat{\gamma}_2 + s_2 \hat{\gamma}_3}{s_1 \hat{\gamma}_1}$$

$$\text{ただし } s_1 = (\tau_1^2 + \tau_2^2)^{-1/2}, \quad s_2 = \tau_1^2 \tau_2^2 s_1^4$$

 $\hat{\gamma}_1 = 1.328, \quad \hat{\gamma}_2 = 0.144, \quad \hat{\gamma}_3 = 0.137$

2次の精度の確率値を計算 $\hat{z}_2(y) = \hat{\gamma}_1(1 + \hat{\gamma}_3) + \frac{\hat{\gamma}_2}{\hat{\gamma}_1}$

$$\hat{\alpha}_2(y) = 1 - \Phi \left\{ 1.328(1 + 0.137) + \frac{0.144}{1.328} \right\}$$

$$= 0.0528 \text{ 😊 😊}$$

3ステップ＝マルチスケール法

$$\left\{ \begin{array}{l} \tilde{z}_3(y, \tau_1, \tau_2, \tau_3) \\ \approx \hat{\gamma}_1 s_1 (1 + \hat{\gamma}_3 s_2 + 4\hat{\gamma}_3^2 s_2^2 + \hat{\gamma}_5 s_3 + \hat{\gamma}_6 s_4) \\ - (\hat{\gamma}_1 s_1)^{-1} (\hat{\gamma}_2 + \hat{\gamma}_3 s_2 + 7\hat{\gamma}_3^2 s_2^2 + \hat{\gamma}_4 s_2 + 3\hat{\gamma}_5 s_3 + 3\hat{\gamma}_6 s_4) \end{array} \right.$$

↓ $\hat{\gamma}_1 = 1.328, \quad \hat{\gamma}_2 = 0.145, \quad \hat{\gamma}_3 = 0.127,$
 $\hat{\gamma}_4 = -0.018, \quad \hat{\gamma}_5 = -0.0004, \quad \hat{\gamma}_6 = -0.036$

$$\left\{ \begin{array}{l} \tilde{z}_3(y) = \hat{\gamma}_1 (1 + \hat{\gamma}_3 + 4\hat{\gamma}_3^2 + \hat{\gamma}_6) + \hat{\gamma}_1^{-1} (\hat{\gamma}_2 + \hat{\gamma}_3^2/2 + \hat{\gamma}_4 + \hat{\gamma}_5) \\ \hat{\alpha}_3(y) = 1 - \Phi \left\{ \begin{array}{l} 1.328(1 + 0.127 + 0.065 - 0.036) \\ + \frac{0.145 + 0.008 - 0.018 - 0.0004}{1.328} \end{array} \right\} \\ = 0.0509 \end{array} \right. \quad \text{3次の精度の確率値}$$



6個の係数 γ と4個の s

幾何的量

$$\left\{ \begin{aligned} \hat{\gamma}_1 &= \hat{v} + \frac{1}{3}\hat{v}^2\hat{\phi}^{ppp} + \hat{v}^3\left\{-\frac{1}{8}(\hat{\phi}^{app})^2 - \frac{1}{18}(\hat{\phi}^{ppp})^2 + \frac{1}{8}\hat{\phi}^{pppp}\right\} \\ \hat{\gamma}_2 &= \hat{v}\left\{-\hat{a}^{aa} - \frac{1}{6}\hat{\phi}^{ppp}\right\} + \hat{v}^2\left\{(\hat{a}^{ab})^2 - \frac{1}{2}\hat{a}^{aa}\hat{\phi}^{ppp}\right. \\ &\quad \left.+ \frac{1}{8}(\hat{\phi}^{app})^2 + \frac{1}{72}(\hat{\phi}^{ppp})^2 - \frac{1}{24}\hat{\phi}^{pppp}\right\} \\ \hat{\gamma}_3 &= -\frac{1}{6}\hat{v}\hat{\phi}^{ppp} + \hat{v}^2\left\{\frac{1}{4}(\hat{\phi}^{app})^2 + \frac{1}{9}(\hat{\phi}^{ppp})^2 - \frac{1}{8}\hat{\phi}^{pppp}\right\} \\ \hat{\gamma}_4 &= \hat{v}^2\left\{-\hat{a}^{ab}\hat{\phi}^{abp} + \frac{1}{3}\hat{a}^{aa}\hat{\phi}^{ppp} + \frac{1}{2}(\hat{\phi}^{abp})^2\right. \\ &\quad \left.+ \frac{1}{2}(\hat{\phi}^{app})^2 + \frac{2}{9}(\hat{\phi}^{ppp})^2 - \frac{1}{4}\hat{\phi}^{aapp} - \frac{1}{6}\hat{\phi}^{pppp}\right\} \\ \hat{\gamma}_5 &= \hat{v}^2\left\{-\frac{1}{8}(\hat{\phi}^{app})^2 - \frac{1}{8}(\hat{\phi}^{ppp})^2 + \frac{1}{12}\hat{\phi}^{pppp}\right\} \\ \hat{\gamma}_6 &= \hat{v}^2\left\{-\frac{1}{8}(\hat{\phi}^{app})^2 - \frac{1}{8}(\hat{\phi}^{ppp})^2 + \frac{1}{24}\hat{\phi}^{pppp}\right\} \end{aligned} \right.$$

スケールの関数

$$\left\{ \begin{aligned} s_1 &= (\tau_1^2 + \tau_2^2 + \tau_3^2)^{-1/2} \\ s_2 &= (\tau_1^2\tau_2^2 + \tau_2^2\tau_3^2 + \tau_3^2\tau_1^2)s_1^4 \\ s_3 &= (\tau_1^2\tau_2^2\tau_3^2 + \tau_2^4\tau_3^2 + \tau_1^4(\tau_2^2 + \tau_3^2))s_1^6 \\ s_4 &= (\tau_1^2\tau_2^2\tau_3^2)s_1^6 \end{aligned} \right.$$

マルチスケール法の導出

- 指数型分布族
- チューブ座標系
- ヤコビアン
- 「修正符号付距離」の分布関数
- 各種ブートストラップ法を統一的に表現
- Edgeworth展開
- *Mathematica*の数式処理
- Shimodaira and Hasegawa (2001) *Bioinformatics* CONSEL: for assessing the confidence of phylogenetic tree selection.
- Shimodaira (2002) *Systematic Biology* An approximately unbiased test of phylogenetic tree selection.
- Shimodaira (2004) *Annals of Statistics* Approximately unbiased tests of regions using multistep-multiscale bootstrap resampling.

詳細は <http://www.is.titech.ac.jp/~shimo/>

指数型分布族

密度関数 $\exp(\theta^i y_i - \psi(\theta) - h(y))$ 観測値 $y = (y_1, \dots, y_p)$

自然パラメタ $\theta = (\theta^1, \dots, \theta^p)$ 期待値パラメタ $\eta = (\eta_1, \dots, \eta_p)$

$$\eta_i = \frac{\partial \psi}{\partial \theta^i}, \quad \theta^i = \frac{\partial \phi}{\partial \eta_i}, \quad i = 1, \dots, p$$

ポテンシャル関数

$$\phi(\eta) = \max_{\theta} \{ \theta^i \eta_i - \psi(\theta) \}$$

計量行列

$$\phi^{ij}(\eta) = \frac{\partial^2 \phi(\eta)}{\partial \eta_i \partial \eta_j}$$

$$\left. \frac{\partial \phi}{\partial \eta_i} \right|_0 = \phi^i = 0$$

$$\left. \frac{\partial^2 \phi}{\partial \eta_i \partial \eta_j} \right|_0 = \phi^{ij} = \delta_{ij}$$

$$\left. \frac{\partial^3 \phi}{\partial \eta_i \partial \eta_j \partial \eta_k} \right|_0 = \phi^{ijk}$$

$$\left. \frac{\partial^4 \phi}{\partial \eta_i \partial \eta_j \partial \eta_k \partial \eta_l} \right|_0 = \phi^{ijkl}$$

チューブ座標系 (u, v)

$\partial\mathcal{R}$ のパラメトリック表現

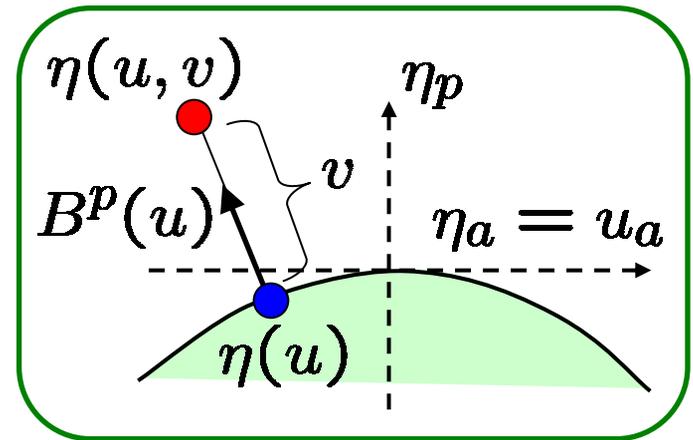
$$\left\{ \begin{array}{l} u = (u_1, \dots, u_{p-1}) \\ \eta_a(u) = u_a, \quad a = 1, \dots, p-1 \\ \eta_p(u) \approx -d^{ab} u_a u_b - e^{abc} u_a u_b u_c \end{array} \right.$$

単位法線ベクトル

$$B^p(u) = (B_1^p(u), \dots, B_p^p(u))$$

座標変換 $\eta \leftrightarrow (u, v)$

$$\eta_i(u, v) = \eta_i(u) + B_i^p(u)v, \quad i = 1, \dots, p$$



ヤコビアンとスケール変換

$$\frac{\partial \eta}{\partial(u, v)} \approx \exp \left[-\frac{1}{2} \phi^{cpp} u_c + (2d^{aa} - \phi^{aap}) v \right. \\ \left. - \left\{ 2(d^{ab})^2 - 2d^{ab} \phi^{abp} + \frac{1}{2} (\phi^{abp})^2 \right\} v^2 \right. \\ \left. + \left\{ \frac{1}{2} d^{cd} \phi^{ppp} - \frac{1}{4} \phi^{cdpp} \right. \right. \\ \left. \left. + \frac{1}{4} \phi^{cpp} \phi^{dpp} + \frac{1}{2} \phi^{acp} \phi^{adp} + 2d^{ac} (d^{ad} - \phi^{adp}) \right\} u_c u_d \right. \\ \left. + \left\{ 6e^{aac} + d^{aa} \phi^{cpp} + 4d^{ac} \phi^{app} - \phi^{aacp} + \phi^{aad} \phi^{cdp} \right. \right. \\ \left. \left. + \frac{1}{2} \phi^{aap} \phi^{cpp} - 2d^{cd} \phi^{aad} - (2d^{ad} - \phi^{adp}) \phi^{acd} \right\} u_c v \right]$$

$u, v \Rightarrow u/\tau, v/\tau$
 $\phi^{ijk}, d^{ab} \Rightarrow \phi^{ijk}_\tau, d^{ab}_\tau$
 $\phi^{ijkl}, e^{abc} \Rightarrow \phi^{ijkl}_{\tau^2}, e^{abc}_{\tau^2}$

スケール変換

正規モデルなら $\phi^{ijk} = \phi^{ijkl} = 0$ なので

$$\frac{\partial \eta}{\partial(u, v)} \approx \exp \left[2d^{aa} v - 2(d^{ab})^2 v^2 + 2d^{ac} d^{ad} u_c u_d + 6e^{aac} u_c v \right]$$

{	幾何的量	符号付距離	$\hat{v} = O(1)$	$\hat{e}^{abc} = O(n^{-1})$
		仮説の曲面	$\hat{d}^{ab} = O(n^{-1/2})$	$\hat{\phi}^{ijkl} = O(n^{-1})$
		確率モデル	$\hat{\phi}^{ijk} = O(n^{-1/2})$	

修正符号付距離 \hat{w}

座標変換 $(u, v) \leftrightarrow (u, w)$

$$v \approx c_0 + c_1 w + c_2 w^2 + c_3 w^3 - u c b^c(w)$$

$$\text{係数 } c_0, c_2 = O(n^{-1/2}), \quad c_1, c_3, b^c(w) = O(n^{-1})$$

チューブ座標系 (u, v) 又は (u, w) による表示

$$\eta \leftrightarrow (u, v) = \underline{(0, \lambda)}, \quad y \leftrightarrow (\hat{u}, \hat{v}) \leftrightarrow (\hat{u}, \hat{w})$$

(このように座標系をとっても一般性を失わない)

分布関数

$$\Pr\{\hat{W} \leq \hat{w}; \underline{(0, \lambda)}, \tau\} = \Phi(z_c(\hat{w}; \underline{0, \lambda}, \tau))$$

$$z_c(\hat{w}; \underline{0, \lambda}, \tau) \approx \tau^{-1} g_-(\hat{w}, \lambda) + \tau g_+(\hat{w}, \lambda)$$

一般の $z_c(\hat{w}; \underline{u, v}, \tau)$ は幾何的量を $\eta(u, 0)$ で評価

2個の係数(3次の精度)

$$\begin{aligned}g_-(\hat{w}, \lambda) &= (\hat{w} - \lambda) \\ &\quad - c_0 - \frac{1}{3}\phi^{ppp}\lambda^2 + \frac{1}{6}\phi^{ppp}\lambda\hat{w} + \left(\frac{1}{6}\phi^{ppp} - c_2\right)\hat{w}^2 \\ &\quad - \frac{1}{6}c_0\phi^{ppp}\lambda - \left\{c_1 + \frac{1}{3}c_0\phi^{ppp}\right\}\hat{w} \\ &\quad + \left\{\frac{1}{8}(\phi^{app})^2 + \frac{1}{18}(\phi^{ppp})^2 - \frac{1}{8}\phi^{pppp}\right\}\lambda^3 \\ &\quad + \left\{-\frac{1}{8}(\phi^{app})^2 + \frac{1}{24}\phi^{pppp}\right\}\lambda^2\hat{w} \\ &\quad + \left\{-\frac{1}{24}(\phi^{ppp})^2 + \frac{1}{24}\phi^{pppp} - \frac{1}{6}c_2\phi^{ppp}\right\}\lambda\hat{w}^2 \\ &\quad + \left\{-\frac{1}{72}(\phi^{ppp})^2 + \frac{1}{24}\phi^{pppp} - \frac{1}{3}c_2\phi^{ppp} - c_3\right\}\hat{w}^3\end{aligned}$$

$$\begin{aligned}g_+(\hat{w}, \lambda) &= -(d^{aa} + \frac{1}{6}\phi^{ppp}) \\ &\quad + \left\{(d^{ab})^2 - d^{ab}\phi^{abp} + \frac{1}{6}d^{aa}\phi^{ppp} + \frac{1}{2}(\phi^{abp})^2\right. \\ &\quad \quad \left.+ \frac{1}{2}(\phi^{app})^2 + \frac{13}{72}(\phi^{ppp})^2 - \frac{1}{4}\phi^{aapp} - \frac{1}{8}\phi^{pppp}\right\}\hat{w} \\ &\quad + \left\{(d^{ab})^2 - \frac{1}{6}d^{aa}\phi^{ppp} + \frac{1}{8}(\phi^{app})^2\right. \\ &\quad \quad \left.+ \frac{5}{72}(\phi^{ppp})^2 - \frac{1}{24}\phi^{pppp}\right\}\lambda\end{aligned}$$

近似的に不偏な確率値の表現

不偏な確率値(3次の精度?) $Y^* \sim f(y^*; \hat{\eta}(y), 1)$

$$\hat{\alpha}_\infty(y) = \Pr\{\hat{v}(Y^*) \geq \hat{v}(y); \hat{\eta}(y), 1\}$$

チューブ座標系で表示

$$y \leftrightarrow (\hat{u}, \hat{v}), \quad y^* \leftrightarrow (\hat{u}^*, \hat{v}^*), \quad \hat{\eta}(y) \leftrightarrow (\hat{u}, 0)$$

$$\hat{\alpha}_\infty(y) = \Pr\{\hat{V}^* \geq \hat{v}; (\hat{u}, 0), 1\}$$

$$= 1 - \Phi(z_c(\hat{v}; \hat{u}, 0, 1)) \quad z_c \text{の係数 } c_0, c_1, c_2, c_3 \text{はゼロ}$$

確率値の表現が得られる

$$\hat{z}_\infty(y) = z_c(\hat{v}; \hat{u}, 0, 1)$$

$$\begin{aligned} &\approx \hat{v} - (d^{aa} + \frac{1}{6}\hat{\phi}^{ppp}) + \frac{1}{6}\hat{\phi}^{ppp}\hat{v}^2 \\ &+ \left\{ (d^{ab})^2 - d^{ab}\phi^{abp} + \frac{1}{6}d^{aa}\phi^{ppp} + \frac{1}{2}(\phi^{abp})^2 \right. \\ &\quad \left. + \frac{1}{2}(\phi^{app})^2 + \frac{13}{72}(\phi^{ppp})^2 - \frac{1}{4}\phi^{aapp} - \frac{1}{8}\phi^{pppp} \right\} \hat{v} \\ &+ \left\{ -\frac{1}{72}(\phi^{ppp})^2 + \frac{1}{24}\phi^{pppp} \right\} \hat{v}^3 \end{aligned}$$

近似的に不偏な確率値の分布関数

少し一般化して

$$\hat{z}_q(y) \approx \hat{z}_\infty(y) + q_0 + q_1 \hat{v} + q_2 \hat{v}^2 + q_3 \hat{v}^3 + \hat{u}_c g^c(\hat{v})$$

$$\begin{aligned} \text{係数 } q_0 &= O(n^{-1/2}), q_1 = O(n^{-1}), q_2 = O(n^{-1/2}), \\ q_3 &= O(n^{-1}) \quad g^c(\hat{v}) = O(n^{-1}) \end{aligned}$$

$\hat{z}_q(y)$ は再び \hat{w} として表現される \Rightarrow もう一度 z_c -formula を適用

$$\begin{aligned} \Pr\{\hat{z}_q(Y) \leq x; 0, 1\} &\approx \Phi \left[x - q_0 - q_2 x^2 \right. \\ &\quad \left. + \left\{ -q_1 - 2q_2 (d^{aa} + \frac{1}{6} \phi^{ppp} - q_0) \right\} x \right. \\ &\quad \left. + \left\{ \frac{1}{3} \phi^{ppp} q_2 + 2q_2^2 - q_3 \right\} x^3 \right] \end{aligned}$$

$q_0 \approx q_1 \approx q_2 \approx q_3 \approx 0$ なら3次の精度

ブートストラップ確率の表現

ブートストラップ確率 $Y^* \sim f(y^*; y, \tau)$

$$\begin{aligned}\tilde{\alpha}_1(y, \tau) &= \Pr\{Y^* \in \mathcal{R}; y, \tau\} \\ &= \Pr\{\hat{V}^* \leq 0; (\hat{u}, \hat{v}), \tau\} \\ &= \Phi(z_c(0; \hat{u}, \hat{v}, \tau)) \quad z_c \text{の係数 } c_0, c_1, c_2, c_3 \text{ はゼロ}\end{aligned}$$

z値の表現が得られる

$$\begin{aligned}\tilde{z}_1(y, \tau) &\approx \tau^{-1} \left[\hat{v} + \frac{1}{3} \hat{\phi}^{pppp} \hat{v}^2 - \left\{ \frac{1}{8} (\hat{\phi}^{app})^2 + \frac{1}{18} (\hat{\phi}^{ppp})^2 - \frac{1}{8} \hat{\phi}^{pppp} \right\} \hat{v}^3 \right] \\ &\quad + \tau \left[(\hat{d}^{aa} + \frac{1}{6} \hat{\phi}^{ppp}) - \left\{ (\hat{d}^{ab})^2 - \frac{1}{6} \hat{d}^{aa} \hat{\phi}^{ppp} \right. \right. \\ &\quad \left. \left. + \frac{1}{8} (\hat{\phi}^{app})^2 + \frac{5}{72} (\hat{\phi}^{ppp})^2 - \frac{1}{24} \hat{\phi}^{pppp} \right\} \hat{v} \right]\end{aligned}$$

2ステップ法はこれを積分する

$$\tilde{z}_2(y, \tau_1, \tau_2) = \Phi^{-1} \left\{ \int \Phi(\tilde{z}_1(y^*, \tau_2)) f(y^*; y, \tau_1) dy^* \right\}$$

まとめと課題

- ブートストラップ確率だけから高精度確率値を計算するマルチスケール・ブートストラップ法の提案
- 「スケール変換」という原理の導入
- 符号付距離, 曲率, 加速定数など幾何的量の推定
- z_C -formulaを導出し, 不偏な確率値, 各種のブートストラップ法を統一的に表現できた
- 実用的には1ステップ法または2ステップ法で十分だろう
- 3ステップ法は数値的安定性の上で検討の余地がある
- 適切なスケールの選択など考慮すべき課題もある
- パラメトリック・ブートストラップの問題
- 仮説の境界が「滑らかな曲面」でない場合の問題