

Approximate Bayesian Computation: 基礎と方法

福水健次

統計数理研究所／総合研究大学院大学

「ABCとその周辺」勉強会

Feb. 16, 2015.

主催： 科研費・新学術領域研究
「スペースモデリングの深化と高次元データ駆動科学の創成」

ベイズ計算

■ Bayesの定理

$$q(\theta|y) = \frac{p(y|\theta)\pi(\theta)}{\int p(y|\theta)\pi(\theta)d\theta}$$

$\pi(\theta)$: 事前確率 (prior)

$p(y|\theta)$: 尤度 (likelihood)



Thomas Bayes (1701-1761)

■ ベイズ計算の代表的手法

大問題「分布の積分をどう計算するか？」

- サンプリング：

Markov Chain Monte Carlo (MCMC), Sequential Monte Carlo (SMC), etc

- 近似計算：

Laplace近似, 確率伝搬法 (Belief propagation, BP) , 期待値伝搬法 (Expectation propagation, EP), 変分ベイズ, etc

■ Beyond these

上記の方法は, 尤度値の計算が可能であることを前提とする.

「尤度値の計算が陽にできない」場合どうするか?

→ ABC, aka. “without likelihood” , “likelihood-free” ,

尤度無しのBayes推論

■ 集団遺伝学の例 (Marjoram et al PNAS2003)

Nutka族（バンクーバー周辺の先住民族）63名の
ミトコンドリアDNAのデータ

現集団の遺伝的多様性から、
遺伝的なパラメータを
推論する。

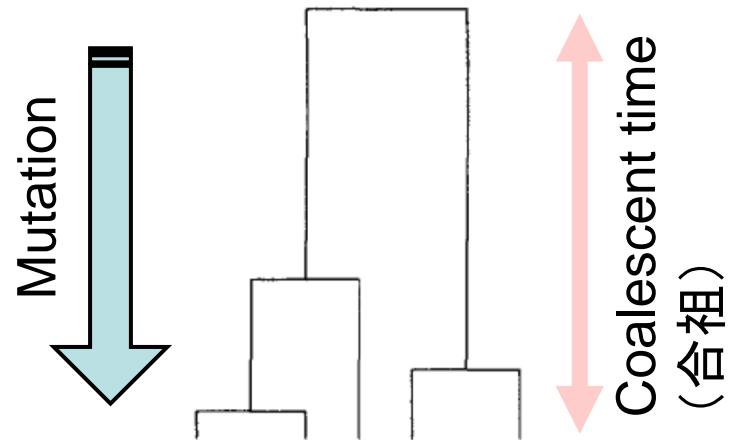


- 観測 $S_n :=$ 多型サイトの数
- 推定したいパラメータ
 - ・突然変異率 : θ
 - ・MRCAまでの時間 : L_n
- 尤度関数

$$P(S_n = k | L_n, \theta)$$

は、モデルが複雑になると
陽にかけない

Most Recent Common Ancestor
(MRCA)



現在の集団の遺伝子の多様性

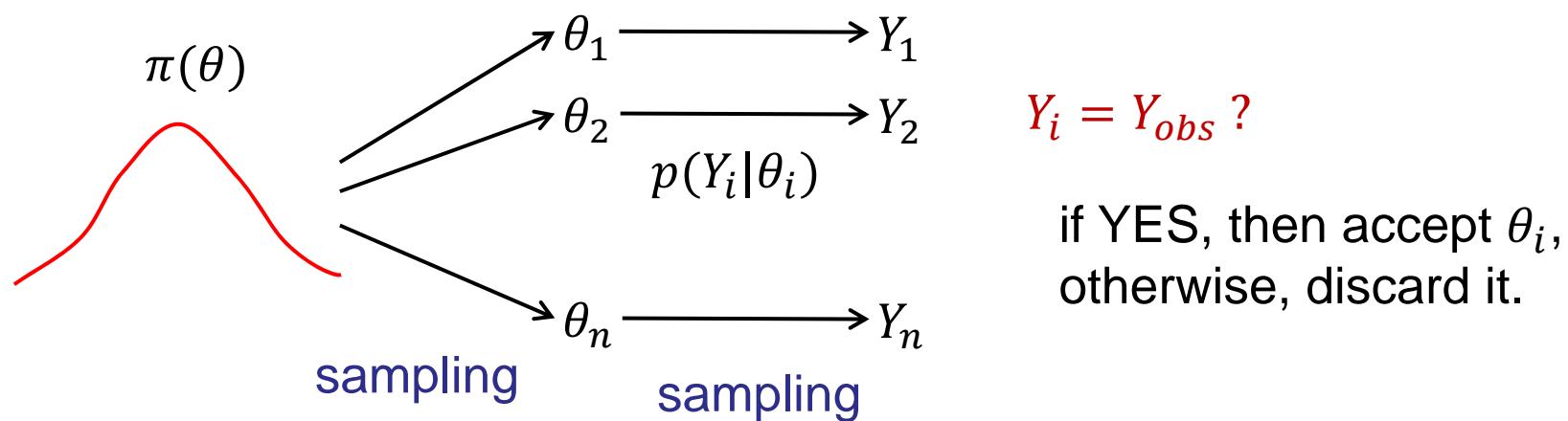
しかし、サンプリングは容易（分岐過程）。

■ ABC：サンプリングによるBayes計算

- 仮定： $p(y|\theta)$ からのサンプリングは可能
尤度値の計算は不可能、またはコストが高い
- サンプリングに基づく事後確率の近似計算の方法
- Tavaré, Balding, Griffiths et al. Inferring Coalescence Times From DNA Sequence Data, (1997, Genetics) が、現在の形を最初に提案
 - ・近い考えは以前にも存在 (Rubin 1984, Diggle & Gratton 1984)
- 集団遺伝学の問題などを中心に発展

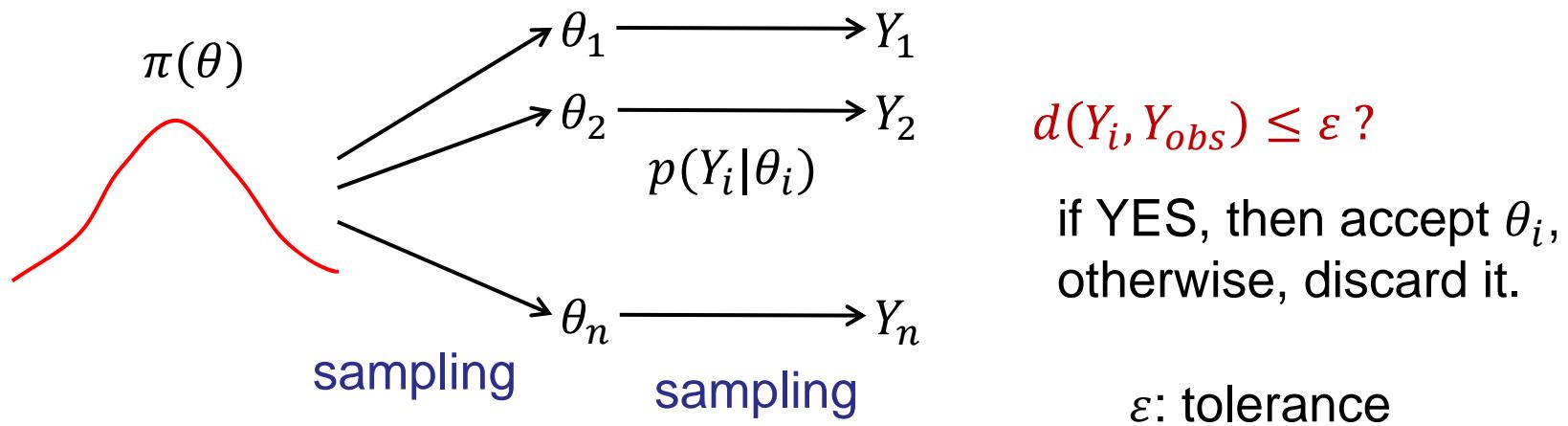
ABC Rejection法：基本的方法

■ Tavaré et al. 1997



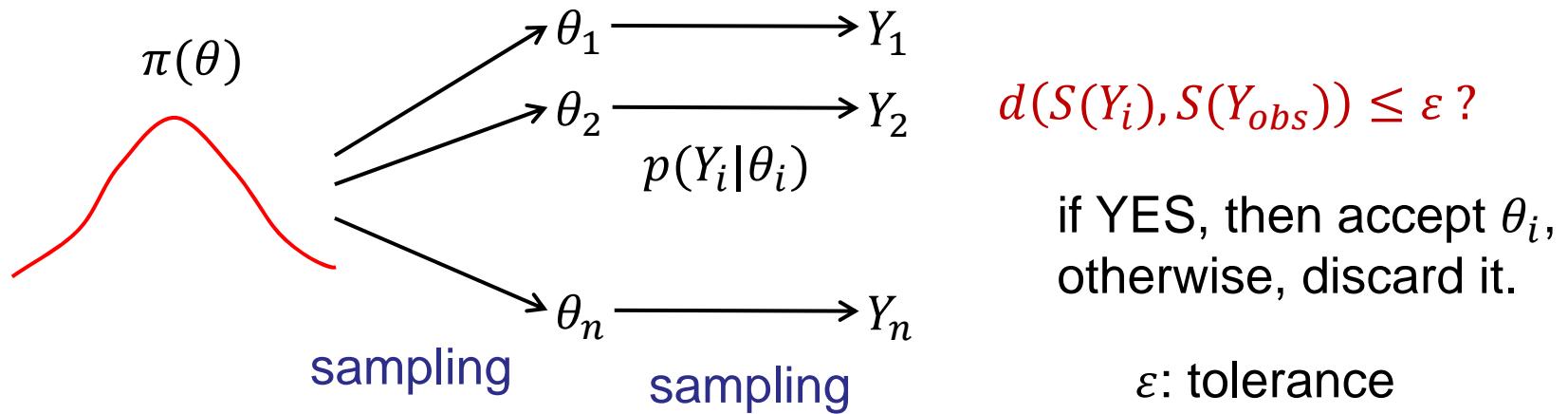
- Acceptされた (θ_i) の分布は事後確率 $q(\theta|Y_{obs})$

- $Y_i = Y_{obs}$ となることは非常に稀 (Y が連続なら確率〇)
そこで,



- Acceptされた (θ_i) の分布は $p(\theta | d(Y, Y_{obs}) \leq \varepsilon) \approx q(\theta | Y_{obs})$
- $\varepsilon \rightarrow 0$ のとき事後確率に収束。

- $d(Y_i, Y_{obs}) \leq \varepsilon$ を満たすことは、 Y が高次元では稀.
- 要約統計量 (summary statistics) $S(Y)$ の導入. $S(Y)$ は Y よりも低次元.



- Acceptされた (θ_i) の分布は $p(\theta | d(S(Y), S(Y_{obs})) \leq \varepsilon) \approx q(\theta | Y_{obs})$
- $S(Y)$ が十分統計量であれば、 $\varepsilon \rightarrow 0$ のとき事後確率に収束.

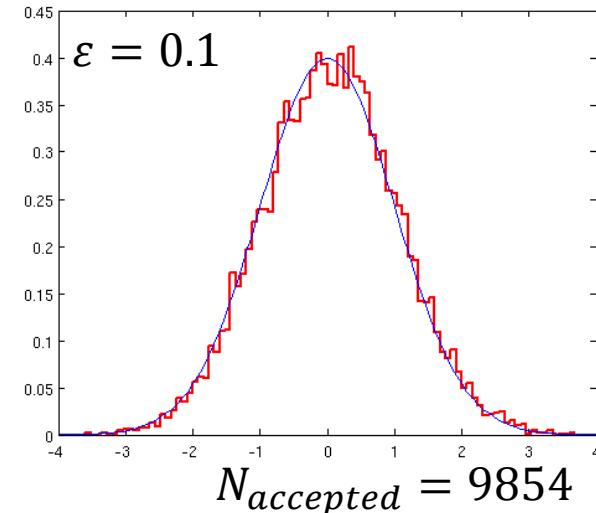
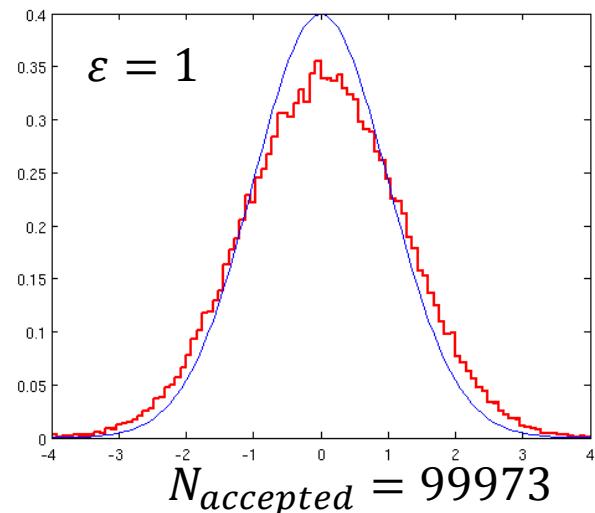
■ ABC Rejection Algorithm

- INPUT: 要約統計量 $S(y)$, 距離 $d(s, s')$, tolerance ε
- 収束条件 (アクセプトされた総数 or 総サンプル数) を満たすまで, 以下を繰り返す.
 1. Generate $\theta^* \sim \pi(\theta)$
 2. Generate $y^* \sim p(y|\theta^*)$
 3. Accept θ^* if $d(S(y^*), S(y_{obs})) \leq \varepsilon$
 4. If rejected, go to 1.

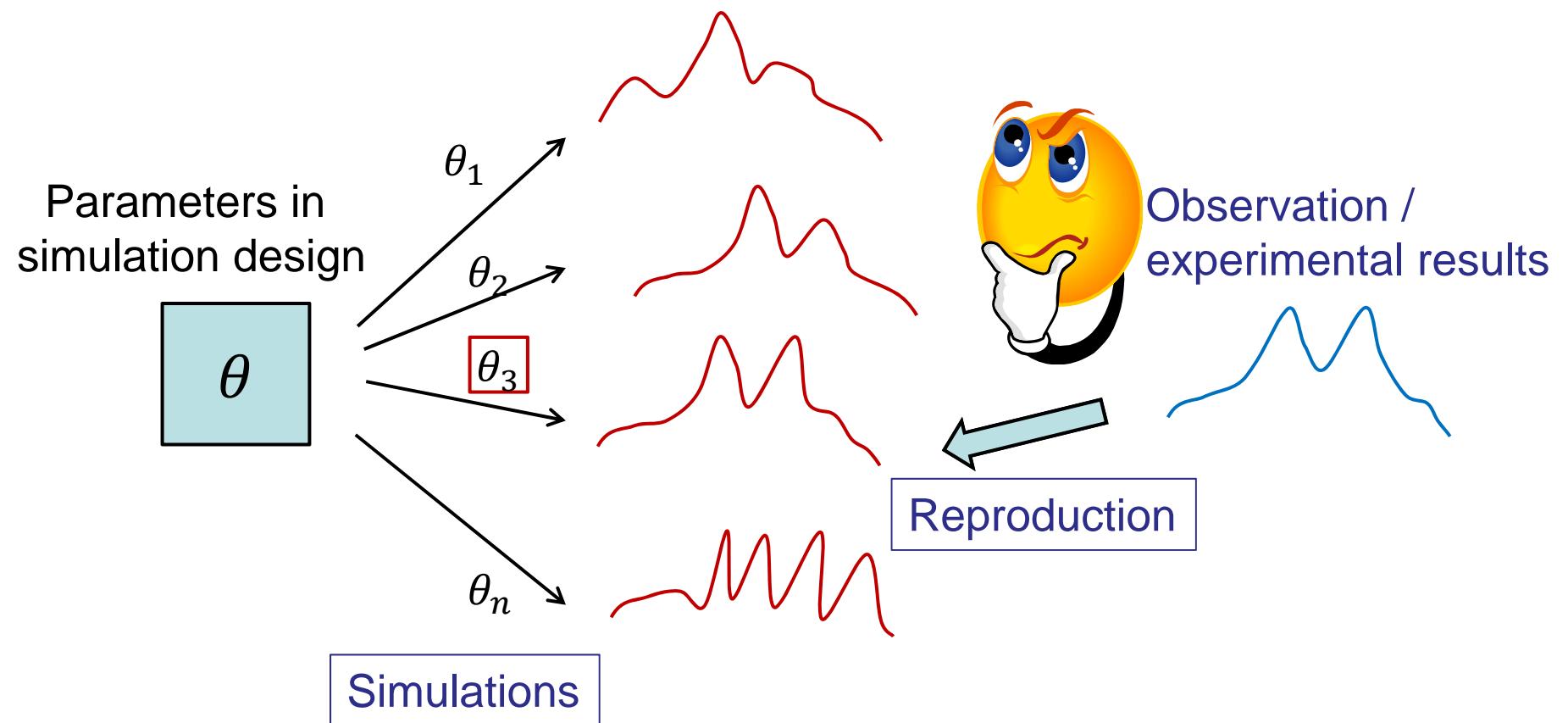
実験1(Rejection method)

$$p(y|\theta) = \frac{1}{2} normal(y; \theta, 1), \quad \pi(\theta) = Unif[-10, 10]$$

サンプリング数 10^6



多くの自然科学・工学の問題に応用できそう！！

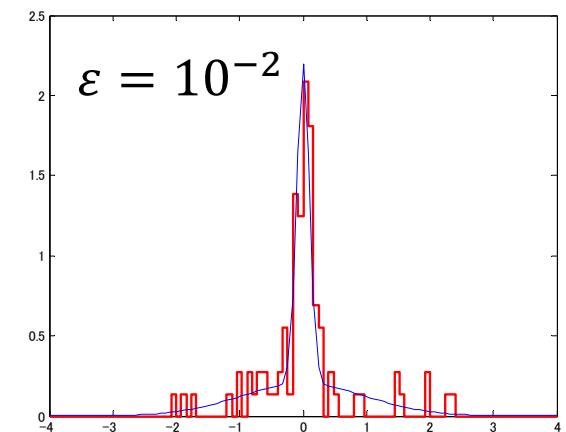
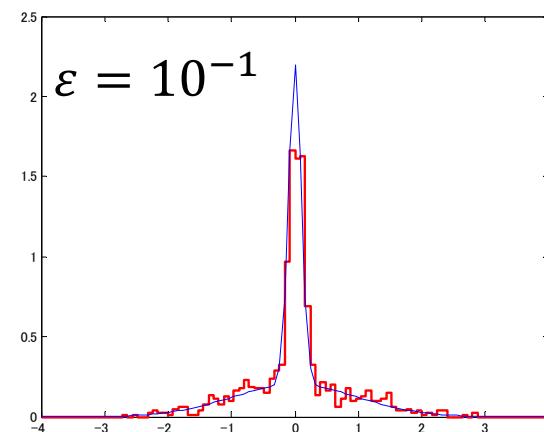
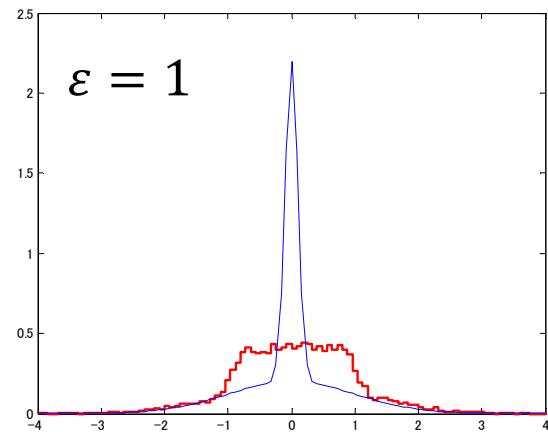


これで解決！ . . . か？

実験2(Rejection method)

$$p(y|\theta) = \frac{1}{2} normal(y; \theta, 1) + \frac{1}{2} normal(y; \theta, 0.01), \quad \pi(\theta) = Unif[-10, 10]$$

サンプリング数 10^5



■ 問題点

- サンプリング数と推定精度のトレードオフ
 - 精度を高めるためには、 ε をなるべく小さくしたい。
 - 一方、 ε を小さくすると acceptされにくく、膨大なサンプリングが必要。
- 要約統計量 $S(y)$ をどう設定するか?
 - 精度を高めるためには、なるべく情報（次元）を残したい
 - 次元が高いと accept されにくい。

ABCの改良

■ Summary statisticsの構成

- 次元削減
- 変数選択

■ サンプリングの改良

- MCMC
- SMC
- Parallel Tempering MCMC

■ その他

- Regression adjustment
- カーネル法

Regression Adjustment

(Beaumont et al., Genetics 2002)

回帰を用いて、近傍の情報を補正する。

1. Sample $(\theta_i, Y_i) \sim p(y|\theta)\pi(\theta)$
2. Accept if $\|s(Y_i) - s(y_{obs})\| \leq \delta$ for a larger tolerance $\delta > 0$.
3. Acceptされた (θ_i, S_i) に対し局所線形回帰を施す
($S_i = s(Y_i)$, 要約統計量)

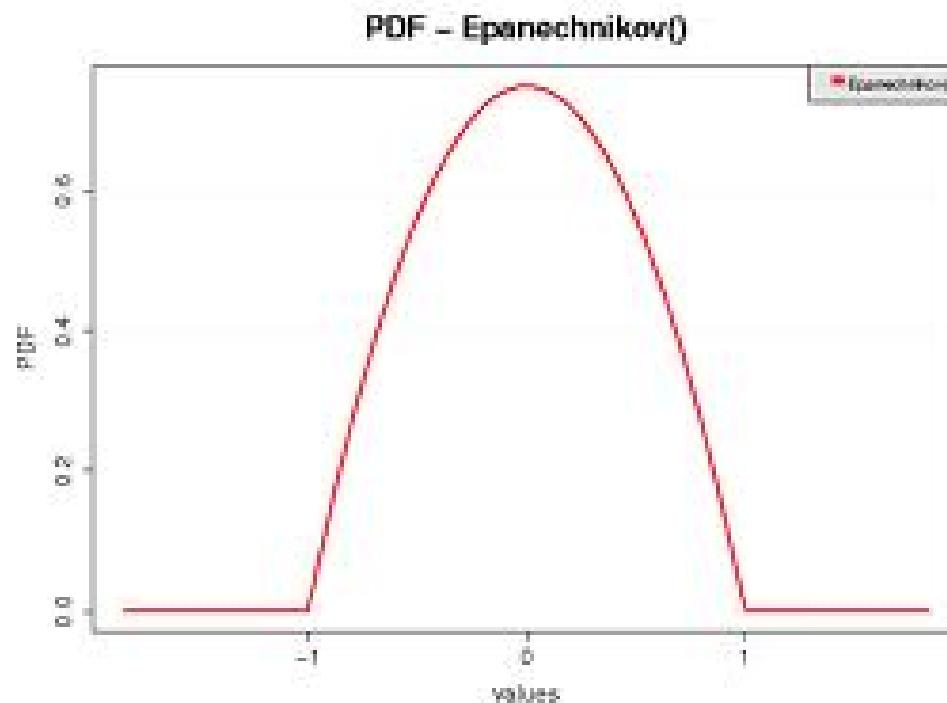
$$\min_{\alpha, \beta} |\theta_i - \{\alpha + (S_i - s_{obs})^T \beta\}|^2 K_\delta(\|S_i - s_{obs}\|)$$

(K_δ : Epanechnikov kernel with bandwidth δ)

4. サンプルの補正 : $\theta_i^* = \theta_i - \hat{\alpha} - (S_i - s_{obs})^T \hat{\beta}$.
5. 補正したサンプルに対して, ABCを施す

- Epanechnikovカーネル

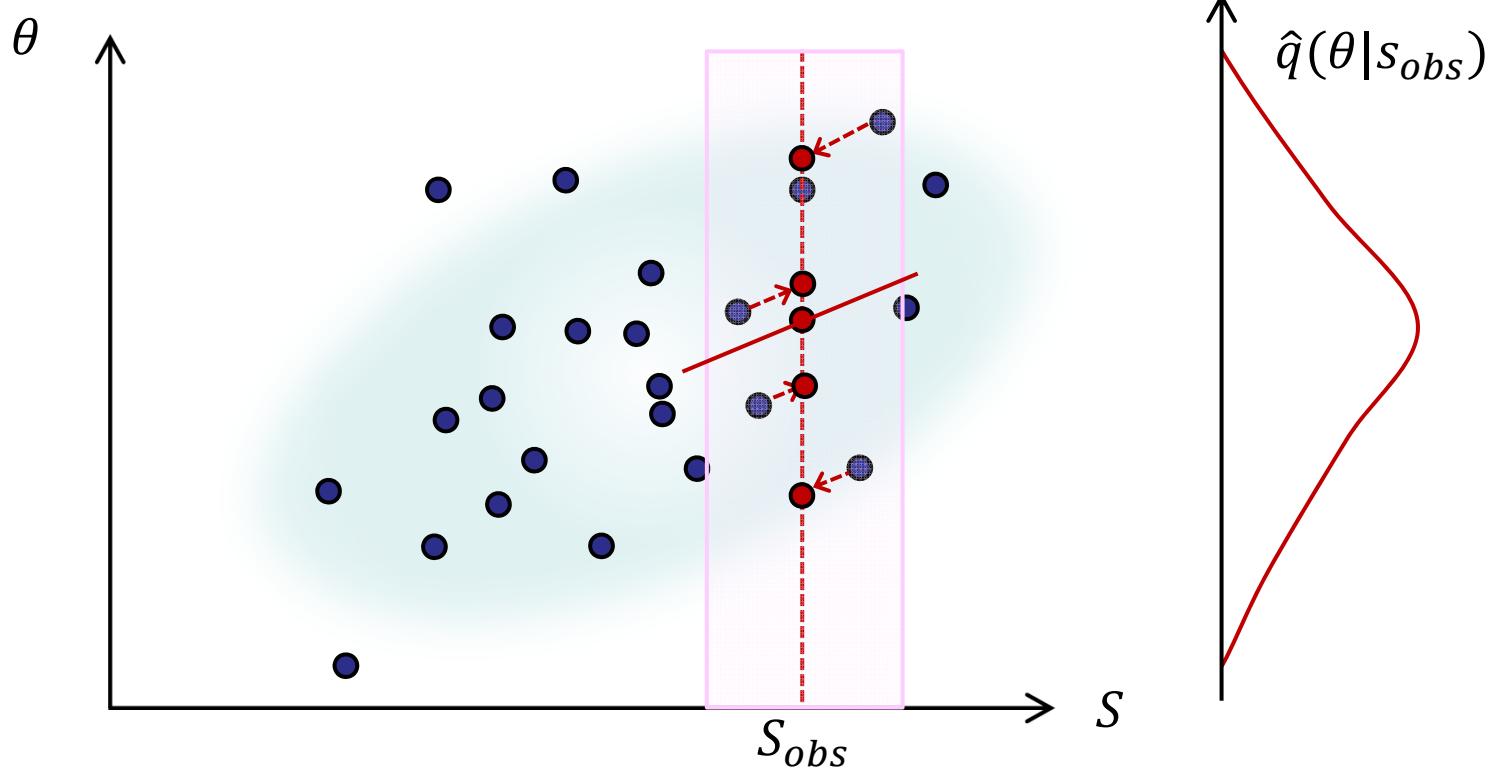
$$K(u) = \frac{3}{4} (1 - u^2) I_{[-1,1]}(u)$$



$$(\theta_i, y_i) \sim q(\theta, y)d\theta dy = p(y|\theta)\pi(\theta)d\theta dy$$

$$s_i = S(y_i)$$

$$\theta_i^* = \theta_i - \hat{\alpha} - (S_i - s_{obs})^T \hat{\beta}.$$



■ 拡張 : heteroscedastic noise (Blum & François 2010)

$$\theta_i = m(y_i) + \sigma(s_i)e_i$$

$$\sigma(s_i)^2 = \text{Var}[\theta | S(Y) = s_i]$$

Estimation:

1. $\hat{m}(s_i) = \hat{\alpha} + (s_i - s_{obs})^T \hat{\beta}$ を局所回帰で求める.
2. 次の局所回帰により $\hat{\sigma}^2(s_i)$ を求める

$$\log(\theta_i - \hat{m}(s_i))^2 \approx \log \sigma^2(s_i)$$

Blum & François 2010では、ニューラルネットを用いている

Summary Statistics の構成

(Blum, Nunes, Prangle, Sisson, Statistical
Science 2013 より抜粋)

概観

■ 変数選択 ($y = (y_1, \dots, y_p)$ から少數変数を選択)

$$S(y) = (y_{i_1}, \dots, y_{i_d}) : \mathbf{R}^p \rightarrow \mathbf{R}^d, \quad d < p$$

- ε -sufficiency (Joyce & Marjoram 2008)
変数を付け加えても情報が増えないかを判定する
- Entropy criterion (Nunes & Balding 2010)
 - ABCによる近似事後確率 $q_{ABC}(\theta|S(y_{obs}))$ のエントロピー最小化.
 - エントロピーの推定は k 近傍法で行う.
 - 変数の組は（原則）総当たりで探す.
- 情報量規準：AIC, BIC (Blum et al 2013)
 - 尤度が書けない → 局所線形回帰における誤差を用いる.

■ 次元削減（線型/非線形射影による低次元化）

$$S(y) = (h^1(y), \dots, h^d(y)) : \mathbb{R}^p \rightarrow \mathbb{R}^d, \quad d < p$$

- Partial least square (PLS) 回帰 (Wegmann et al 2009)
 y の線形射影, 分散と θ との相関がともに高くなるように射影
- 局所回帰 (Fearnhead & Prangle 2012)
Semi-automatic ABCの一部. 後で詳述.
- ニューラルネット (Blum & Fran ois 2010)
3層パーセプトロンの中間層表現を用いる. 非線形射影.
- その他の次元削減法もABCに使えると考えられる
 - dMAVE (Xia 2007) , gKDR (Fukumizu & Leng 2013) など

■ 正則化

- L2正則化
 - Y が共線性（成分間の線形関係）を持つときに有効
(Beaumont et al 2002, Blum & François 2010)
- LASSO回帰 (Picchini 2013)

■ その他

- 次元削減／変数選択しない。高次元に強い方法を使う
 - (θ, y) の空間でガウス混合モデルを当てはめ, $p(\theta|y_{obs})$ を近似
(Bonassi et al 2011)
 - $p(\theta|s_{obs})$ を Mixture of Expert によって推定 (Fan et al 2012)
 - EP-ABC : 期待値伝搬法を用いるABC (Barthelme & Chopin 2011)
 - Kernel ABC : カーネル法によるABC (Nakagome, Mano, Fukumizu 2013)

Semi-Automatic ABC

(Fernhead & Prangle 2012)

- Noisy ABC: “Calibration” する (今日は扱わない)
- 線形回帰による要約統計量の構成:
 1. 初期 $S(y)$ を決めてパイロットABC
→ 事後確率のラフな近似を得る.
 2. パイロット posterior の高密度領域に制約して θ_i をサンプル. $y_i \sim p(y|\theta_i)$ をサンプル. $\{(\theta_i, y_i)\}_{i=1}^M$ を得る.
 3. $\{(\theta_i, y_i)\}_{i=1}^M$ を用いて, 線形回帰により $E[\theta|\phi(y)]$ を推定.
 $\phi(y)$ は基底関数. 例えば $\phi(y) = (y_a, y_a^2, y_a^3, y_a^4)_{a=1}^P$.
$$\theta_i = c + \beta^T \phi(y_i) + n_i$$
 4. $S(y) = \hat{E}[\theta|\phi(y)] = \hat{c} + \hat{\beta}^T \phi(y)$ を要約統計量としてABCを行う.

Note : 2乗誤差の意味で最良 (しかし未知) の要約統計量は $E[\theta|y]$.

性能評価の方法

多くの場合、真の事後確率は未知なので、誤差を直接はかれない。

■ RSSE (Nunes & Balding 2010)

- $(\theta_i, w_i)_{i=1}^n$: ABCによる重み付サンプル
ABC-rejection の場合、 $w_i = 1 \Leftrightarrow \text{accepted.}$

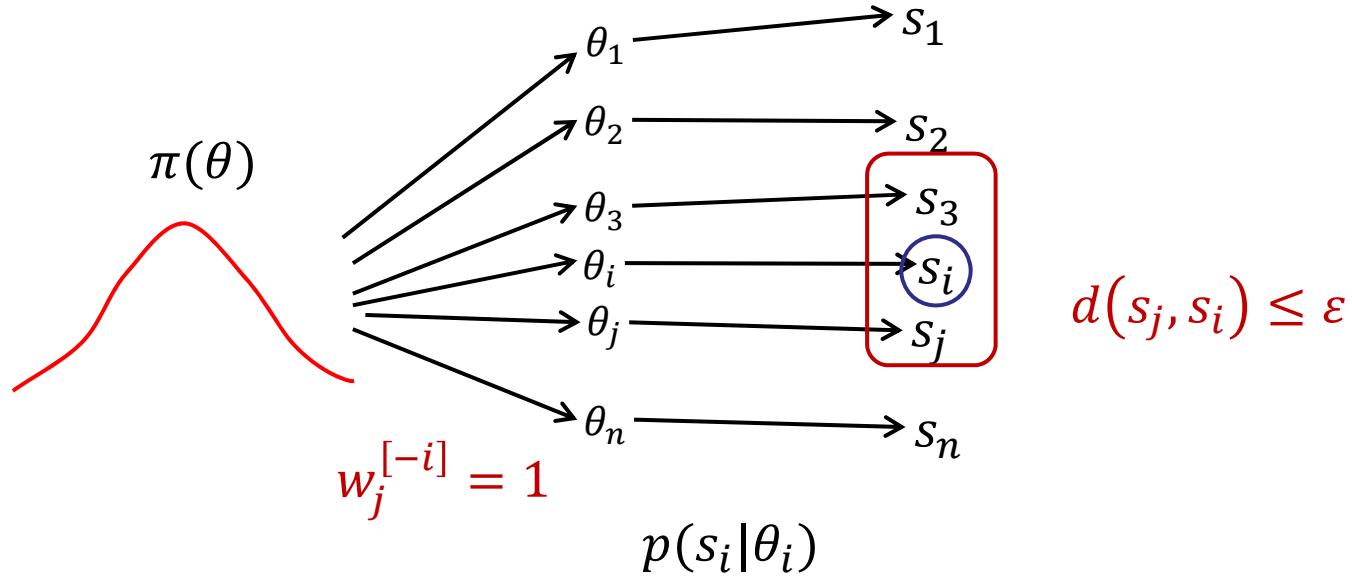
- θ_{true} がわかっていたとすると、

$$RSSE := \left(\sum_{i=1}^n w_i \|\theta_i - \theta_{true}\|^2 \right)^{1/2}.$$

- 通常は θ_{true} は未知なので、Cross-validationする。

$$RSSE_i := \left(\sum_{j \neq i}^n w_j^{[-i]} \|\theta_j - \theta_i\|^2 \right)^{1/2}$$

$w_j^{[-i]}$: $s_{obs} = s_i, \theta_{true} = \theta_i$ としたときの、ABCによる重み。



- s_{obs} が与えられたときの性能を評価する際には、 s_{obs} の近傍のABCサンプルだけを使うとよい。

さまざまな方法の比較

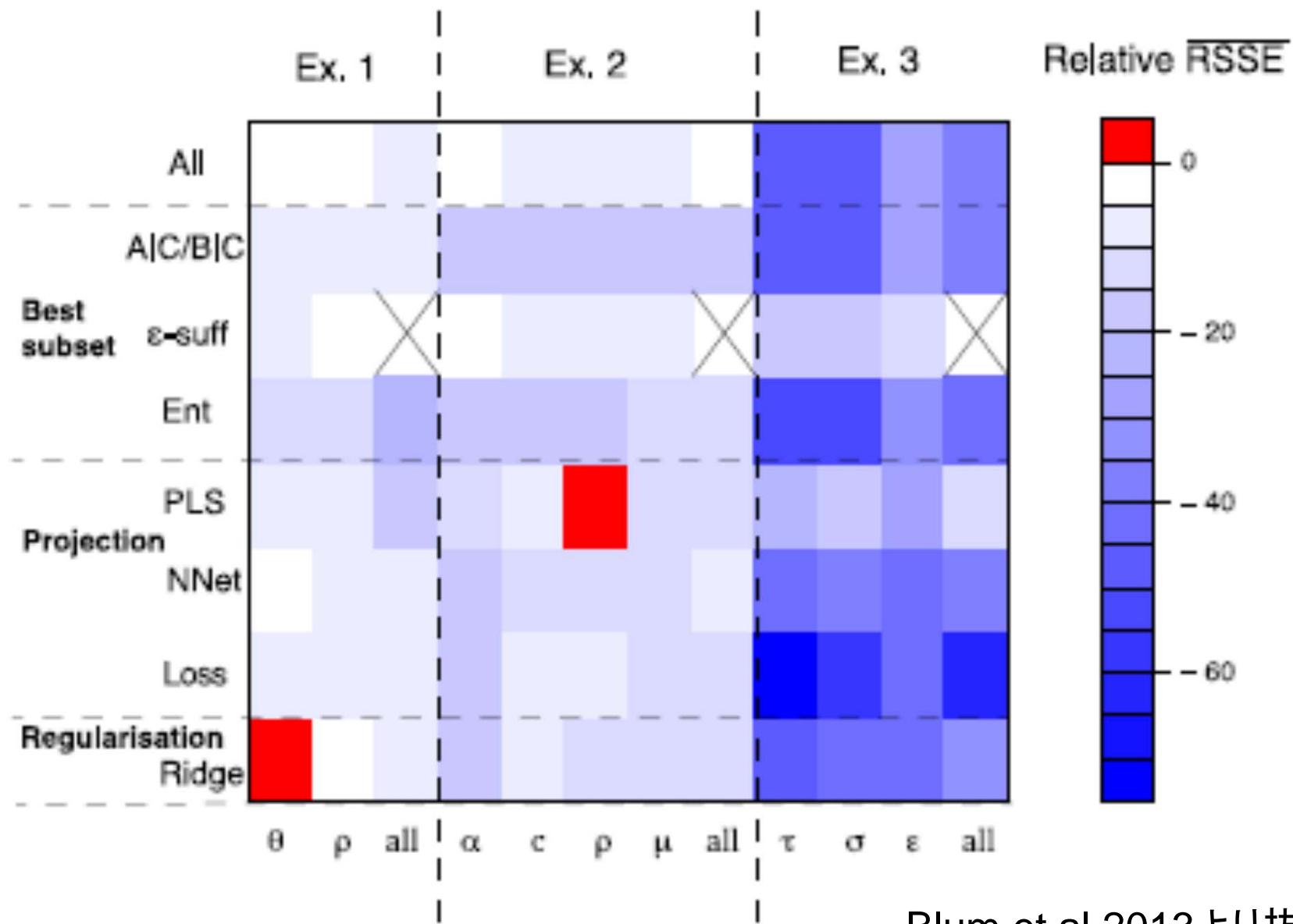
TABLE 2

Relative RSSE for examples 1–3 for different parameter combinations using each method of dimension reduction, and under heteroscedastic regression adjustment. Columns denote no dimension reduction (All), BIC, AIC, AICc, the ε -sufficiency criterion (ε -suff), the two-stage entropy procedure (Ent), partial least squares (PLS), neural networks (NNet), minimum expected posterior loss (Loss) and ridge regression (Ridge). All RSSE are relative to the RSSE obtained when using no regression adjustment with all summary statistics. The score of the best method in each analysis (row) is emphasized in boldface

Semi-automatic

	Best subset selection						Projection techniques			Regularization	
	All	BIC	AIC	AICc	ε -suff	Ent	PLS	NNet ¹	Loss	Ridge ¹	
\emptyset	-3	-5	-5	-5	-6	-11	-6	-4	-7	1	
ρ	-4	-6	-6	-6	0	-12	-7	-8	-7	-3	
$(\tilde{\theta}, \rho)$	-7	-7	-7	-7	-	-24	-16	-7	-6	-6	
α	-3	-15	-15	-15	0	-17	-13	-15	-17	-15	
c	-5	-15	-15	-15	-8	-15	-8	-12	-9	-9	
ρ	-8	-16	-16	-16	-8	-16	1	-12	-9	-10	
μ	-6	-18	-18	-18	-8	-13	-10	-13	-12	-12	
(α, c, ρ, μ)	-4	-19	-19	-19	-	-13	-10	-9	-12	-11	
τ	-49	-47	-47	-48	-19	-52	-22	-20/-42	-75	-48/-48	
σ	-45	-46	-47	-46	-15	-50	-15	-21/-37	-56	-43/-43	
ξ	-27	-29	-29	-28	-13	-32	-28	-7/-41	-41	-26/-44	
(τ, σ, ξ)	-39	-39	-40	-39	-	-42	-11	-4/-38	-60	-39/-32	

¹For the third example, the first value is found by integrating out the regularization parameter, whereas the second one is found by choosing an optimal regularization parameter with cross-validation. In examples 1 and 2, integration over the regularization parameter is performed.



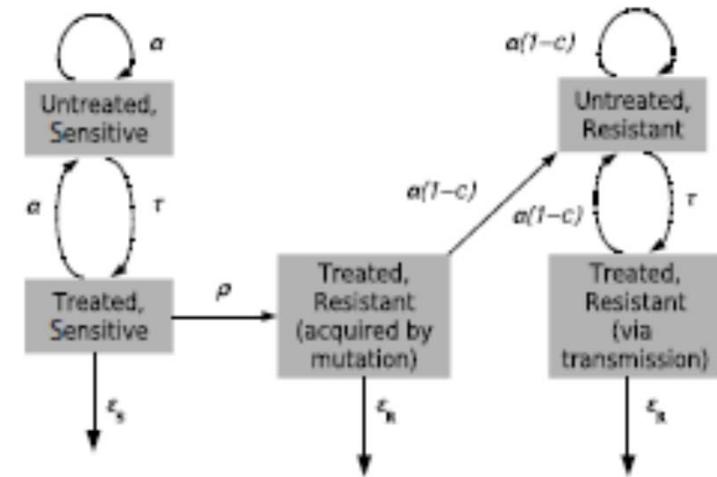
Blum et al 2013より抜粋

■ Ex.1 : 集団遺伝

- パラメータ $\theta = (\bar{\theta}, \rho)$
 - $\bar{\theta}$: mutation rate
 - ρ : recombination rate
- 観測 Y : 5001 塩基のDNA系列, 50 個体。
Coalescence model によるソフトウェア(ms) シミュレーション
- 要約統計量 S : 6 次元
 - s_1 : Number of segregating sites
 - s_2 : Pairwise mean of nucleotidic differences
 - s_3 : Mean R^2 across pairs separated by < 10%
 - s_4 : Number of distinct haplotypes
 - s_5 : Frequency of most common haplotype
 - s_6 : Number of singleton haplotypes

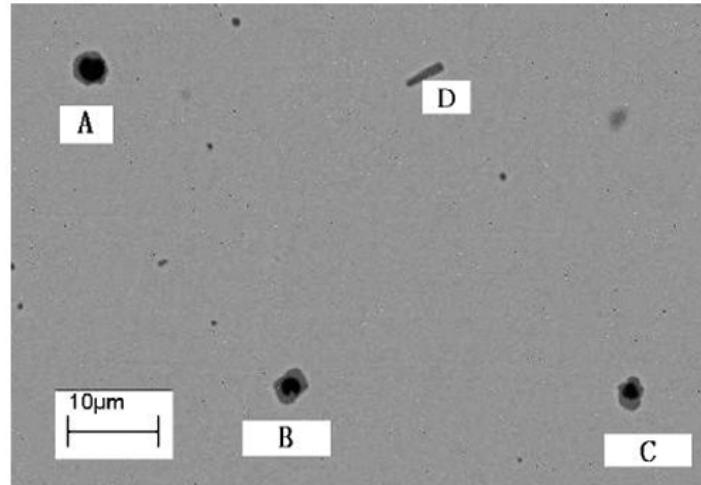
■ Ex. 2 疫学（薬剤耐性結核菌）

- 結核菌の進化と伝播のMarkovモデル
- パラメータ（4次元）
 - α : Transmission rate
 - c : Transmission cost of drug resistant strains
 - ρ : Rate of evolution of resistance
 - μ : Mutation rate of IS6110 marker
- 要約統計量：11次元



■ Ex. 3 品質管理（鉄鋼生産）

- 鋼中の介在物 (inclusion) の個数とサイズを、2次元スライスデータから推定する
- モデル（3パラメータ）
 - 介在物の場所 : Poisson点過程 (rate τ)
 - 介在物の半径 : Pareto分布 (scale $\sigma > 0$, shape ξ)
- 要約統計量 (113次元)
 - 112断面の介在物の半径 ($\geq 5 \mu\text{m}$)
 - 介在物の総数



サンプリングの改良

MCMC without likelihood

(Marjoram et al PNAS 2003)

■ Metropolis-Hastings

Repeat the following:

1. Current state θ . Propose a move $\theta' \sim q(\theta'|\theta)$
2. Accept θ' with probability

$$\min\left(1, \frac{p(y_{obs}|\theta')\pi(\theta')q(\theta|\theta')}{p(y_{obs}|\theta)\pi(\theta)q(\theta'|\theta)}\right),$$

else stay θ .

(緩い仮定のもと) 事後確率 $q(\theta|y_{obs})$ が定常分布となる.

尤度値 $p(y_{obs}|\theta)$ が計算できないと適用できない.

■ MCMC without likelihood

Repeat the following:

1. Current state θ . Propose a move $\theta' \sim q(\theta'|\theta)$
2. Generate $y' \sim p(y|\theta')$
3. If $d(S(y'), S(y_{obs})) < \varepsilon$, go to 4, otherwise stay θ and go to 1.
4. Accept θ' with probability

$$\min\left(1, \frac{\pi(\theta')q(\theta|\theta')}{\pi(\theta)q(\theta'|\theta)}\right),$$

else stay θ .

- θ_i の生成が独立ではなく Markov chain
- 定常分布は $p(\theta | d(S(y), S(y_{obs})) < \varepsilon)$.

■ 問題点

- いったん θ が事後確率の低い領域に入ると, $d(S(y'), S(y_{obs})) < \varepsilon$ が満たされることが稀になる → reject され続ける.
- State-space augmentation を用いたMCMC-ABC (Bortot et al 2007)
 - ε もパラメータとみなして, (θ, ε) の空間でMCMCを行う
→ 大きい ε を経由して早い mixing が期待される.
 - 事後確率 $q(\theta | y_{obs})$ の近似には, 小さい ε に対する θ_i のみを用いる.

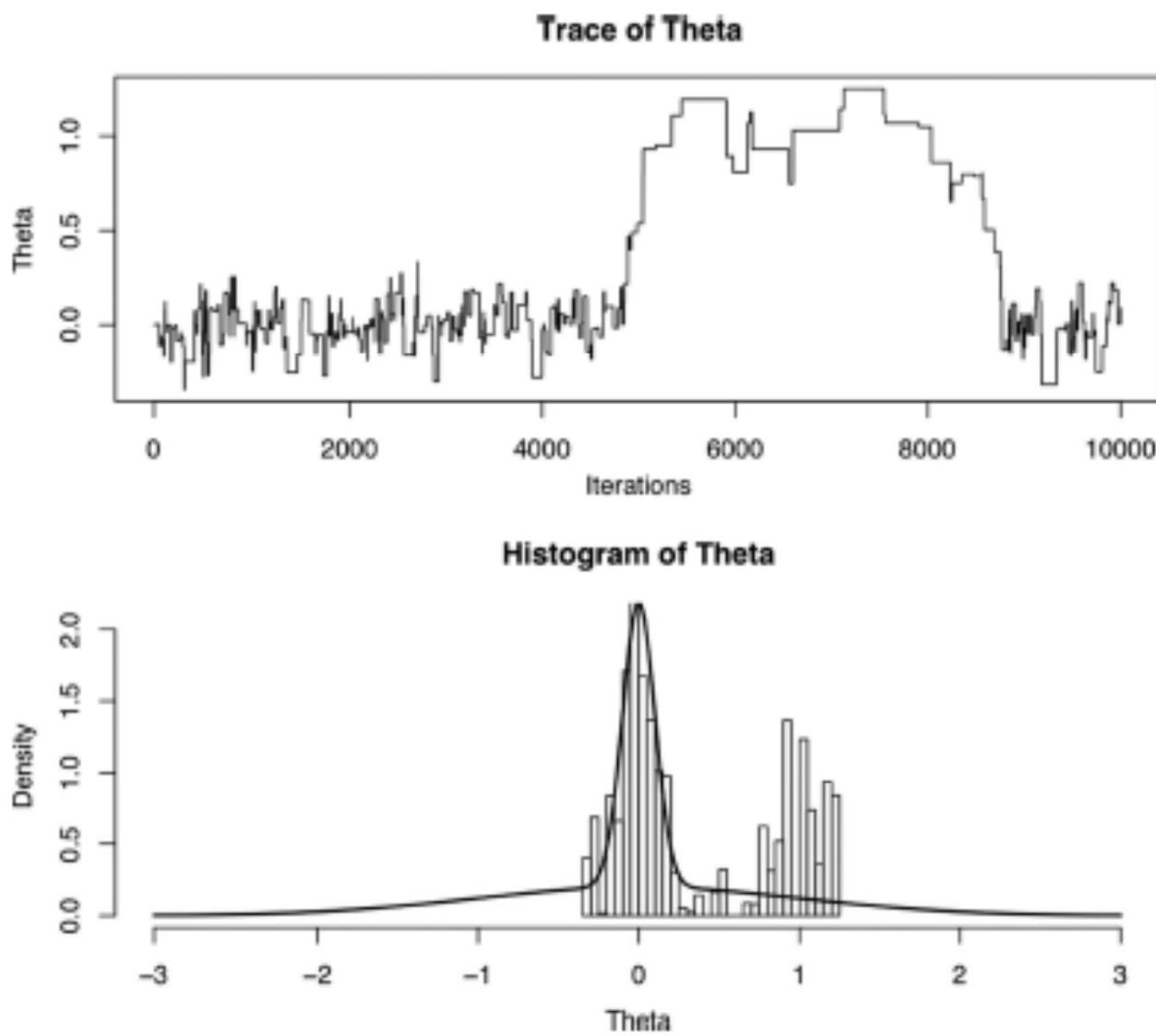


Fig. 1. Trace and histogram of θ . (Upper) Trace of 10,000 ABC-MCMC sampler iterations. (Lower) Target mixture distribution (solid line) and histogram of ABC-MCMC sampler output.

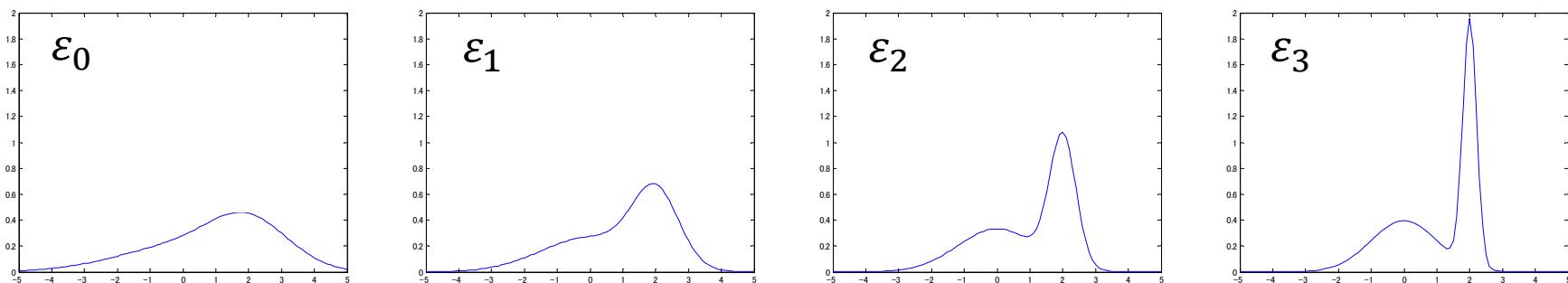
Sisson et al
2007より

Sequential Monte Carlo によるABC

$$\varepsilon_0 > \varepsilon_1 > \varepsilon_2 > \dots > \varepsilon_T = \varepsilon$$

$$q_t(\theta) := q(\theta | d(S(y), S(y_{obs})) \leq \varepsilon_t)$$

より滑らかな近似から、徐々に求めるABC近似事後確率に近づく
系列。



Sequential Monte Carlo Sampler (Del Moral et al 2006)
による近似が可能

Sisson et al 2007, Beaumont et al 2009, Toni et al 2009,
Del Moral et al 2011

■ 復習：SMCサンプラー (Del Moral et al 2006)

$\{p_t\}_{t=0}^T$: 確率密度関数の系列

- 目的: p_t に従う粒子 (重み付サンプル) $\{(Z_t^{(i)}, w_t^{(i)})\}_{i=1}^N$ をシーケンシャルに構成する ($t = 1, 2, \dots, T$)
- Importance Samplingが基礎

$$Z^{(1)}, \dots, Z^{(N)} \sim \eta(x)dx$$

$$E_p[\varphi] = E_\eta \left[\varphi(x) \frac{p(x)}{\eta(X)} \right] \approx \sum_{i=1}^N \varphi(Z^{(i)}) \frac{p(Z^{(i)})}{\eta(Z^{(i)})}$$

重み付サンプル $\{(Z^{(i)}, w^{(i)})\}_{i=1}^N$, $w^{(i)} := \frac{p(Z^{(i)})}{\eta(Z^{(i)})}$ が p を表現

- 系列の場合: local move Markov kernel $K_t(x, x')$ を用いて,

$$Z_t^{(i)} \sim K_t(Z_{t-1}^{(i)}, x) dx$$

$$\eta_t^N(x') = \int K_t(x, x') \eta_{t-1}^N(x) dx$$
 しかし、一般には、 $\eta_t^N(x')$ の計算は実行できない

- パス $x_{1:t} := (x_1, \dots, x_t)$ 上の分布で考える。

Backward Markov kernel $L_t(x_{t+1}, x_t)$.

$$\tilde{p}_t(x_{1:t}) := p_t(x_t) \prod_{k=1}^{t-1} L_k(x_{k+1}, x_k)$$

x_t の周辺分布は $p_t(x_t)$ に一致。

- $\tilde{p}_t(x_{1:t})$ に対して Importance Sampling を行う

$$W_t^{(i)}(Z_{1:t}^{(i)}) \propto W_{t-1}^{(i)}(Z_{1:t-1}^{(i)}) \frac{p_t(Z_t^{(i)}) L_{t-1}(Z_t^{(i)}, Z_{t-1}^{(i)})}{p_{t-1}(Z_{t-1}^{(i)}) K_t(Z_{t-1}^{(i)}, Z_t^{(i)})}$$

- Resampling: Effective Sample Size (ESS):= $\frac{1}{\sum_{i=1}^N (W_t^{(i)})^2}$

$$\text{ESS} \leq \text{Thresh} \text{ なら } W_t^{(i)}(Z_{1:t}^{(i)}) = \frac{1}{N}.$$

■ SMC without likelihood (Sisson 2007, Beaumont 2009)

ABC-PMC (Population Monte Carlo)

1. $t = 0$: $\theta_0^{(i)} \sim \eta_0$ i.i.d. (初期分布). $W_0^{(i)} \propto \frac{\pi_0(\theta_0^{(i)})}{\eta_0(\theta_0^{(i)})}$ ($i = 1, \dots, N$)

2. $t = 1, \dots, T$:

$i = 1, \dots, N$ に対して以下を行う.

① Generate $\theta_* \sim \sum_{i=1}^N W_{t-1}^{(i)} \delta_{\theta_{t-1}^{(i)}}$, $\theta_{**} \sim K(\theta_*, \cdot)$. $y_{**} \sim p(y|\theta_{**})$ を $d(S(y_{**}), S(y_{obs})) \leq \varepsilon_t$ が成り立つまで繰り返す.

② $\theta_t^{(i)} := \theta_{**}$,

$$W_t^{(i)} \propto W_{t-1}^{(i)} \frac{\pi_t(\theta_t^{(i)})}{\sum_{i=1}^N W_{t-1}^{(i)} K_t(\theta_{t-1}^{(i)}, \theta_t^{(i)})}, \quad \sum_{i=1}^N W_t^{(i)} = 1.$$

3. [Resampling] If ESS $\leq TH$, then generate N samples from $\sum_{i=1}^N W_t^{(i)} \delta_{Z_t^{(i)}}$ and set $W_t^{(i)} := 1/N$.

- ABC-PMC : 計算量 $O(N^2)$. 大きい N に対して困難.

■ Del Moral et al (2011) のABC-SMC

- Generate $y_1, \dots, y_M \sim p(y|\theta_{**})$

- $\theta_t^{(i)} := \theta_{**}$

$$W_t^{(i)} \propto W_{t-1}^{(i)} \frac{\sum_{j=1}^M \mathbf{1} \left\{ d \left(S(y_j), S(y_{obs}) \right) \leq \varepsilon_t \right\}}{\sum_{j=1}^M \mathbf{1} \left\{ d \left(S(y_j), S(y_{obs}) \right) \leq \varepsilon_{t-1} \right\}}$$

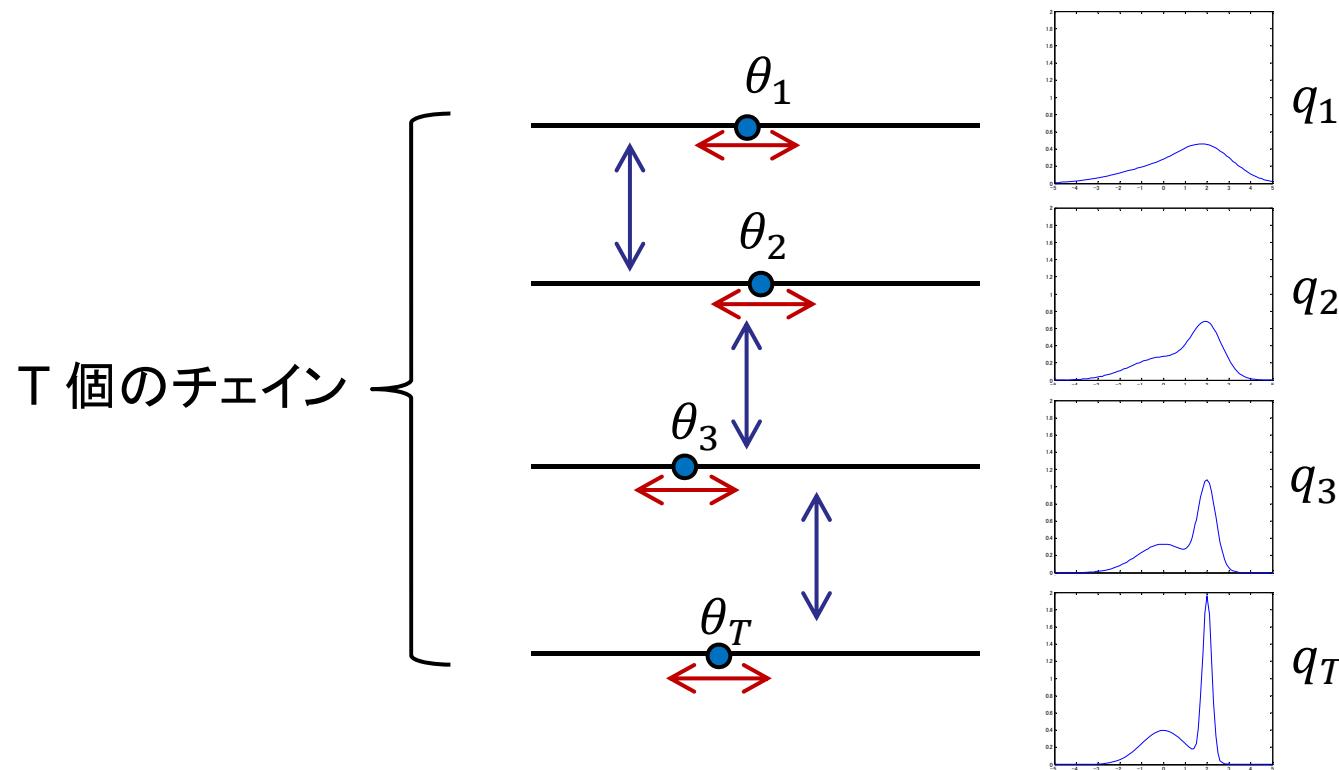
計算量 $O(N)$.

Backward kernel として $L_{t-1}(\theta_t, \theta_{t-1}) := \frac{p_t(\theta_{t-1})K_t(\theta_{t-1}, \theta_t)}{p_t(\theta_t)}$ とおくと導かれる.

- ε_t, K_t の自動設定法も提案されている.

Parallel Tempering-ABC

- Parallel tempering / レプリカ交換 Monte Carloの適用 $\varepsilon_1 > \varepsilon_2 > \dots > \varepsilon_T = \varepsilon$, $q_t(\theta) := q(\theta | d(S(y), S(y_{obs})) \leq \varepsilon_t)$ が自然に tempered posteriorを定義



■ PT-ABC アルゴリズム

1. [初期化] $i = 0$. 各 ε_i において, ABCにより
1サンプルを得る.

$$(\theta_1^{(0)}, y_1^{(0)}), \dots, (\theta_T^{(0)}, y_T^{(0)})$$

2. $n = 2, \dots, N$
 - ① [local move (ABC-MCMC)] 各 t に対し,

Repeat $\theta_t^* \sim q_t(\theta_t^{(i-1)}, \cdot)$, $y_t^* \sim p(y|\theta_t^*)$
until $d(S(y_t^*), S(y_{obs})) \leq \varepsilon_t$.

Accept $(\theta_t^{(i)}, y_t^{(i)}) = (\theta_t^*, y_t^*)$

with prob. $\min \left\{ 1, \frac{\pi(\theta_t^*) q_t(\theta_t^{(i-1)}, \theta_t^*)}{\pi(\theta_t^{(i-1)}) q_t(\theta_t^*, \theta_t^{(i-1)})} \right\}$

- ② [Exchange move]

T 個のペア (t, u) ($t < u$) を一様に選ぶ。

$d(S(y_u), S(y_{obs})) \leq \varepsilon_t$ ならば (θ_t, y_t) と (θ_u, y_u) を交換

比較 (Baragatti et al 2012より)

ABC-MCMC (two runs)

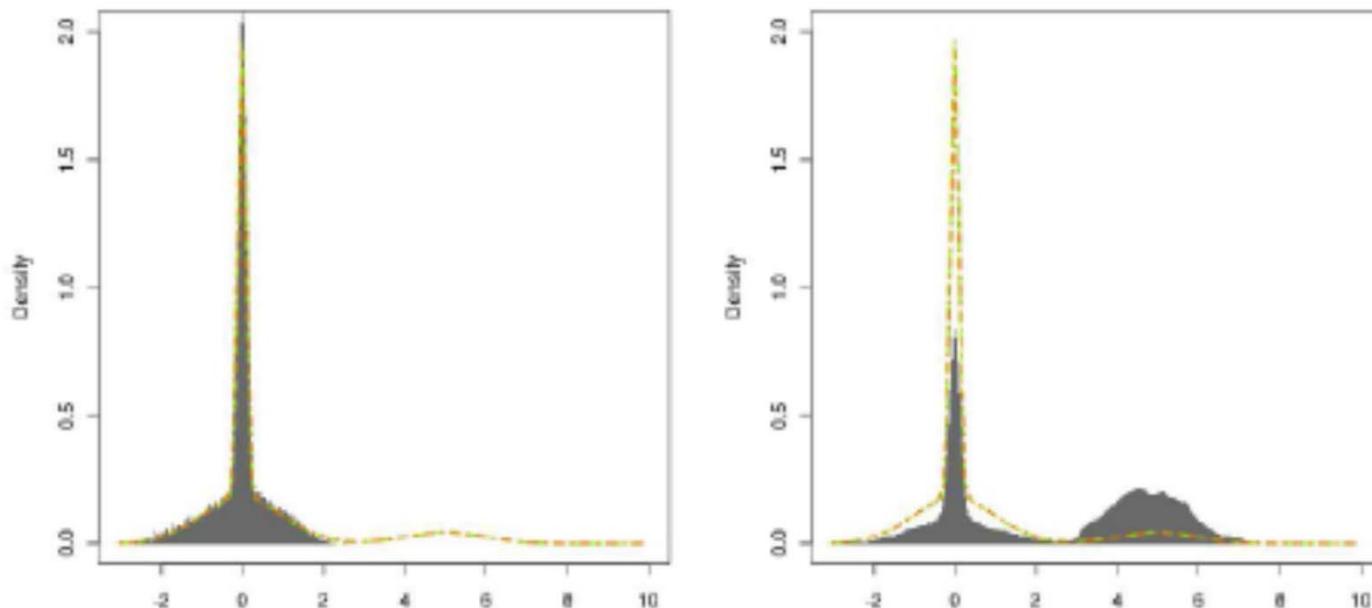


Figure 1: ABC-MCMC for the modified toy example: samples obtained by two runs (20.10^6 iterations with 5.10^6 iterations of burn-in) for tolerance levels of 0.025. The orange dotted curves stand for the exact posterior distribution, the green dotted curves for the target approximation, and densities of obtained samples are represented in gray.

ABC-PT

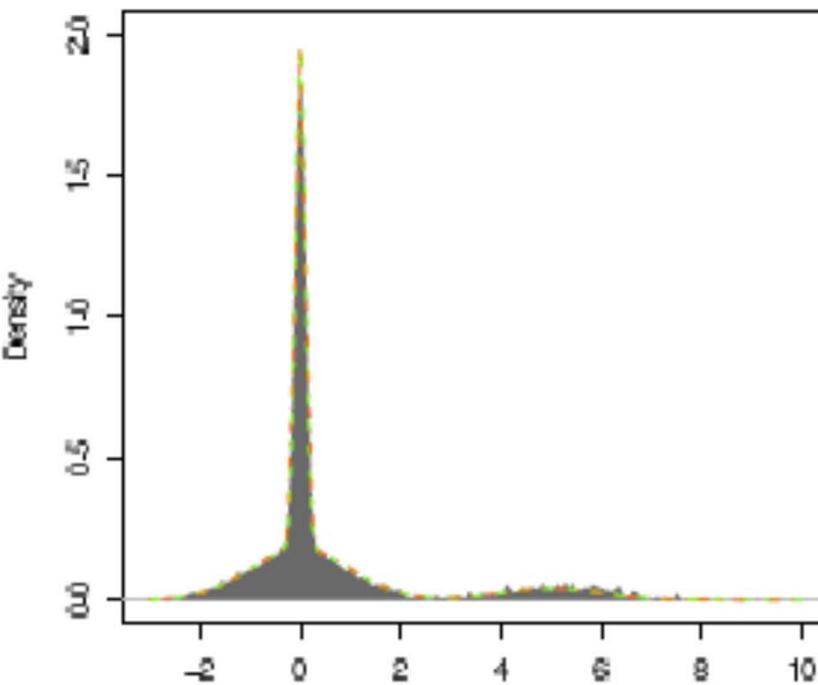


Figure 10: Approximations obtained by the first chain after an ABC-PT algorithm with $N = 15$ chains, with 3 rings of tolerance levels (600 000 iterations with a burn-in of 150 000). The orange dotted curves stand for the exact posterior distribution, the green dotted curves for the target approximation, and densities of obtained samples are represented in gray.

ABC-SMC

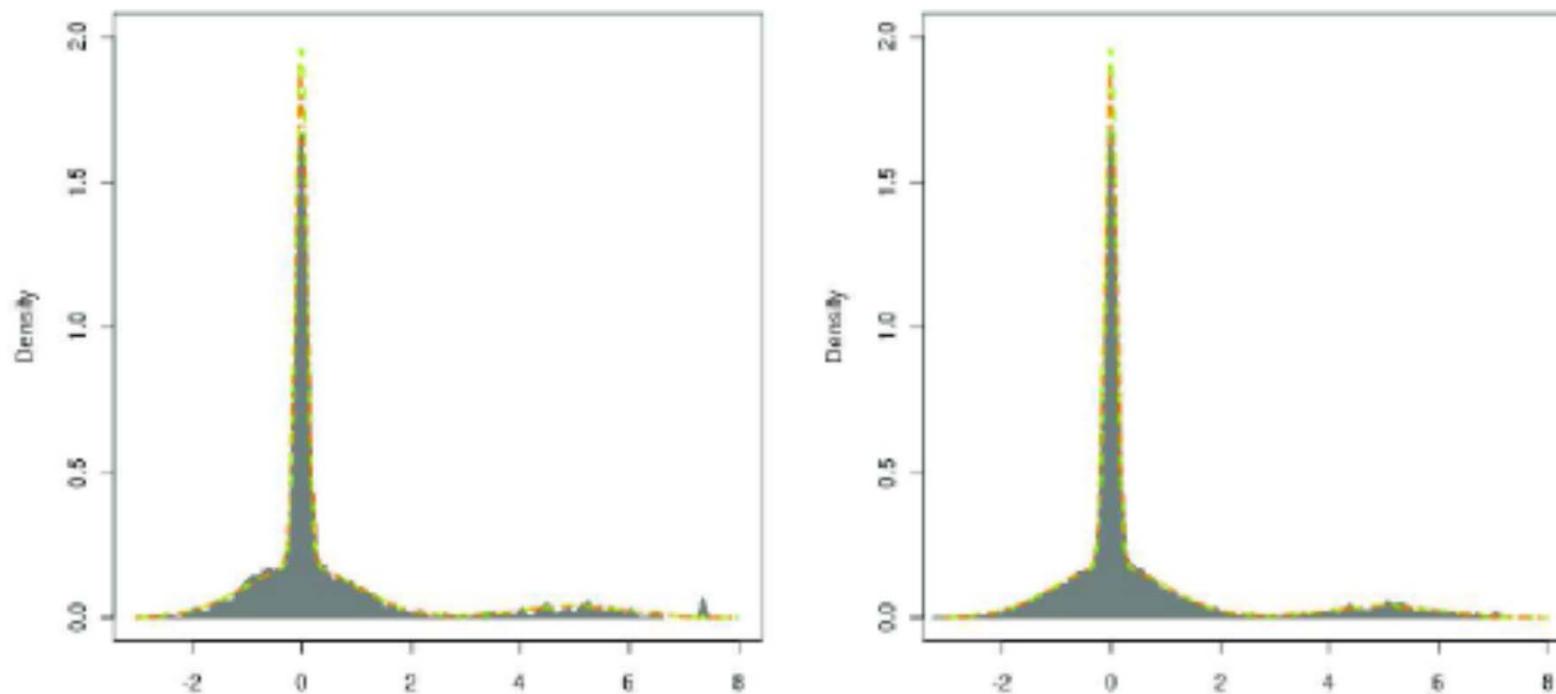


Figure 8: ABC-SMC for the modified toy example, for a tolerance level of 0.025: on the left the results for $M = 1$ and 15 000 particles, on the right the results for $M = 10$ and 20 000 particles. The orange dotted curves stand for the exact posterior distribution, the green dotted curves for the target approximation, and densities of obtained samples are represented in gray.

ABC-PMC

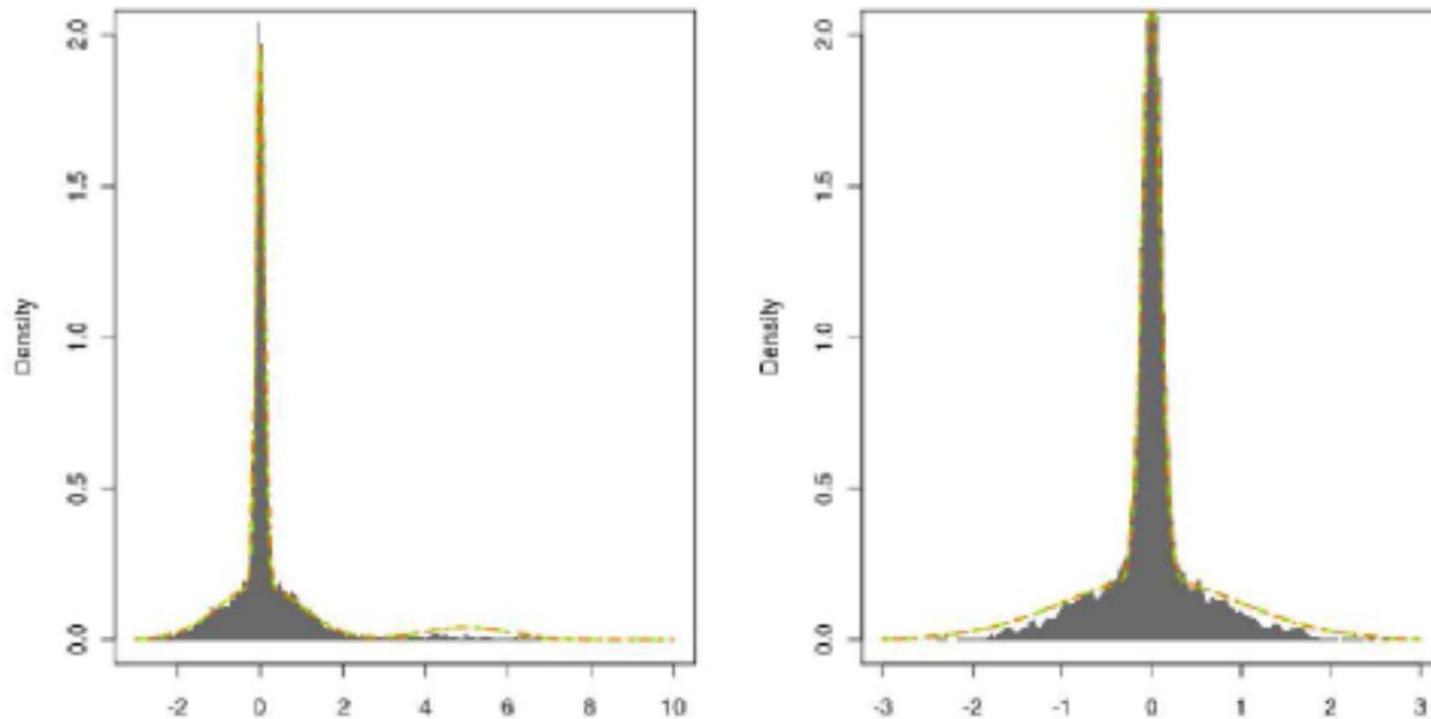


Figure 7: ABC-PMC zoom: populations associated with a tolerance level of 0.025 ([5000 samples](#)), for the modified toy example on the left, and for the standard toy example on the right. The orange dotted curves stand for the exact posterior distributions, the green dotted curves for the target approximation, and the weighted populations are represented in gray.

■ PT-ABC, ABC-SMC, ABC-PMCの比較

- すべて, ABC-MCMCよりよい結果を得られる.
- すべての方法で, カーネルの定義は自明ではない.
- SMCベースのABC-SMC, PMCは, 独立なサンプルを与える.
PT-ABCはMCMCベースなので相関のあるサンプルを与えるので, thinning (間引き) が必要となることもある.
- PT-ABCとABC-SMCはおおよそ同等の結果を与えた (?)

カーネルABC

Kernel ABC

(Nakagome, Mano, Fukumizu 2013)

■ カーネル法に基づくABC

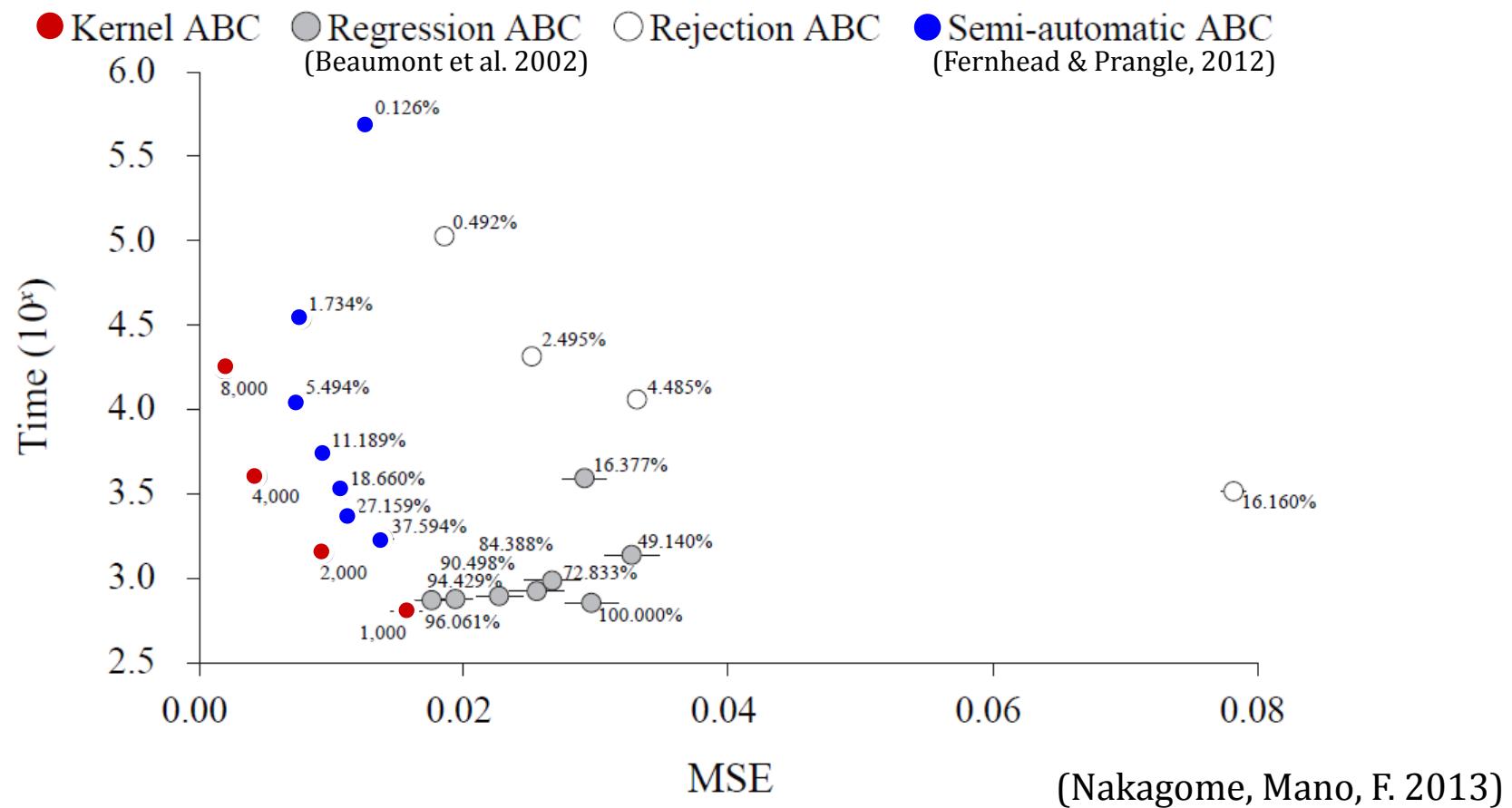
条件付確率のカーネル平均表現（カーネルリッジ回帰）
サンプルをすべて使う。

1. Sample $(\theta_i, Y_i) \sim p(y|\theta)\pi(\theta)$, $i = 1, \dots, N$
2. Apply conditional kernel mean

$$\hat{E}[\Phi_\theta(\theta)|Y = y_{obs}] = \sum_{i=1}^N w_i k(\theta_i, \cdot)$$
$$w = (G_Y + n\varepsilon_n I_n)^{-1} \mathbf{k}_Y(y_{obs})$$

- Good performance for high-dimensional cases is expected (c.f. local linear fitting).

Comparison



- Summary statistics (14 dim):
Site frequency spectrum + Haplotype frequency spectrum
- Gaussian kernel is used.

Kernel ABC has been applied to studies in population genetics.

- Osada, Nakagome, Mano, et al. Finding the factors of reduced genetic diversity on X chromosomes of *Macaca fascicularis*: male-driven evolution, demography, and natural selection. *Genetics* (2013)
- Nakagome S., Mano S., Hasegawa M. (2013) Ancestral polymorphisms and sex-biased migration shaped the demographic history of brown bears and polar bears. *PLOS ONE* 8(11) e78813.

まとめ

■ ABCの改良

- 事後確率を精度よくかつ効率的に近似するためには、低次元の要約統計量の構成が重要。
 - 変数選択
 - 次元削減（線形／非線形特徴）
- サンプリングの高度化
 - MCMC
 - SMC
 - PT
- より直接的なモデリング
 - Mixture モデル
 - カーネルABC

■ ABCの応用の発展

- Population genetics, biologyを超えた幅広い応用には、高次元の観測を扱うための方法の改良が必要と思われる
一般的な応用：数十～数百次元
- 最近の高次元の応用例（生成モデルの活用）
 - 画像処理（Moores et al 2015）
画像の生成モデルとしてPottsモデル（MRF）を用い、パラメトリックモデルで尤度を近似
 - 画像認識（Mansinghka et al 2013）
画像の生成モデル+ABC

References

- Tavaré S, Balding DJ, Griffiths RC et al. (1997) Inferring Coalescence Times From DNA Sequence Data, *Genetics* 145:505-518.
- Beaumont, M. A., W. Zhang, and D. J. Balding. (2002). Approximate Bayesian computation in population genetics. *Genetics* 162:2025--2035.
- Blum, M. G. B. and François, O. (2010). Non-linear regression models for approximate Bayesian computation. *Stat. Comput.* 20 63–73.
- Blum, M. G. B. Nunes, M. A. Prangle D. and Sisson S. A. (2013) A Comparative Review of Dimension Reduction Methods in Approximate Bayesian Computation. *Statistical Science*, 28(2): 189—208.
- Nunes MA, Balding DJ (2010) On optimal selection of summary statistics for approximate Bayesian computation. *Stat Appl Genet Mol Biol* 9: Article 34. doi: 10.2202/1544-6115.1576
- Wegmann, D., Leuenberger, C. and Excoffier, L. (2009). Efficient approximate Bayesian computation coupled with Markov chain Monte Carlo without likelihood. *Genetics* 182 1207–1218.
- Xia, Y. (2007) A constructive approach to the estimation of dimension reduction directions. *Annals of Statistics* 35, 2654-2690

- Kenji Fukumizu and Chenlei Leng (2013) Gradient-based kernel dimension reduction for regression. *Journal of the American Statistical Association*, Volume 109, Issue 505, pages 359-370.
- U. Picchini (2014). Inference for SDE models via Approximate Bayesian Computation. *Journal of Computational and Graphical Statistics*, 23(4), 1080-1100.
- Fan, Y., Nott, D. J. and Sisson, S. A. (2013). Approximate Bayesian computation via regression density estimation. *Stat.* Volume 2, Issue 1, pages 34–48.
- Simon Barthelmé & Nicolas Chopin (2014) Expectation Propagation for Likelihood-Free Inference. *Journal of the American Statistical Association*, Volume 109, Issue 505, pages 315-333
- Paul Fearnhead, Dennis Prangle (2012) Constructing summary statistics for approximate Bayesian computation: semi - automatic approximate Bayesian computation. *Journal of the Royal Statistical Society: Series B*. Volume 74, Issue 3, Pages 419-474.
- Csilléry. K, Blum, M.G., Gaggiotti, O.E., François, O. (2010) Approximate Bayesian Computation (ABC) in practice. *Trends Ecol Evol.* 25(7):410-8.
- Marjoram, P. Molitor, J. Plagnol, V. and Tavaré. S. (2003) Markov chain Monte Carlo without likelihoods. *P. Natl. Acad. Sci. USA*, 100(26):15324–15328.

- P. Bortot, S. G. Coles and S. A. SissonSource: Journal of the American Statistical Association, Vol. 102, No. 477 (Mar., 2007), pp. 84-92
- Sisson, S.A. Fan, Y. and Tanaka. M.M. (2007) Sequential Monte Carlo without likelihood. *P. Natl. Acad. Sci. USA*, 104:1760–1765. (Errata 2009)
- M. Beaumont, J.M. Cornuet, J.M. Marin, and C.P. Robert. (2009) Adaptive approximate Bayesian computation. *Biometrika*, 96(4):983–990.
- T. Toni, D. Welch, N. Strelkowa, A. Ipsen, and M. Stumpf. (2009) Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *Journal of the Royal Society Interface*, 6:187–202.
- Pierre Del Moral, Arnaud Doucet, Ajay Jasra (2011) An adaptive sequential Monte Carlo method for approximate Bayesian computation. *Statistics and Computing*. Volume 22, Issue 5, pp 1009-1020.
- Pierre Del Moral, Arnaud Doucet, Ajay Jasra. (2006). Sequential Monte Carlo samplers. *J. Royal Statist. Society Series B*, 68, 411–436.
- Meïli Baragatti, Agnès Grimaud, Denys Pommeret (2012) Likelihood-free parallel tempering. *Statistics and Computing*, Volume 23, Issue 4, pp 535-549

- Shigeki Nakagome, Shuhei Mano and Kenji Fukumizu (2013) Kernel Approximate Bayesian Computation in Population Genetic Inferences. *Statistical Applications in Genetics and Molecular Biology*, Volume 12, Issue 6, Pages 667—678
- Osada, N., Nakagome, S., Mano, S., Kameoka, Y., Takahashi, I., Terao, K. (2013) Finding the factors of reduced genetic diversity on X chromosomes of Macaca fascicularis: male-driven evolution, demography, and natural selection. *Genetics*, 195, 1007-1025.
- Nakagome S., Mano S., Hasegawa M. (2013) Ancestral polymorphisms and sex-biased migration shaped the demographic history of brown bears and polar bears. *PLOS ONE* 8(11) e78813.
- Matthew T. Moores, Christopher C. Drovandi, Kerrie Mengersen, Christian Robert. (2015) Pre-processing for approximate Bayesian computation in image analysis. *Statistics and Computing*, Volume 25, Issue 1, pp 23-33
- Vikash K. Mansinghka, Tejas D. Kulkarni, Yura N. Perov, and Joshua B. Tenenbaum. (2013) Approximate Bayesian Image Interpretation using Generative Probabilistic Graphics Programs. *Advances in NIPS 2013*.