

ABC サンプラーの最近の発展について

小林 弦矢*

Recent Developments in Approximate Bayesian Computation Samplers

Genya Kobayashi*

尤度関数が解析的に求まらないモデルや尤度関数の計算が煩雑であるモデルに対しては、従来のマルコフ連鎖モンテカルロ法 (MCMC 法) などの適用によって事後分布を得ることは非常に困難であった。本稿は、困難な尤度関数の評価を回避して事後分布を近似する Approximate Bayesian Computation (ABC) という手法を取り扱う。ABC はモデルから発生させられたデータと観測されたデータが近くなるようなパラメータの値を事後分布からのサンプルとして採択する。今までに MCMC 法や逐次モンテカルロ法を基にしたアルゴリズムが提案されてきているが、本稿はそれらのアルゴリズムを紹介するとともに、尤度関数が得られない場合におけるモデル選択の方法などについても言及する。また、本稿で紹介されるアルゴリズムの特徴を単純なモデルに基づいた数値例によって示す。

When the likelihood is either analytically unavailable or computationally cumbersome to evaluate, it is extremely difficult to obtain the posterior distribution by applying the conventional Markov chain Monte Carlo (MCMC) methods. This paper introduces a method called approximate Bayesian computation (ABC) which avoids the direct evaluation of the likelihood. The ABC algorithm samples from the posterior distribution by finding the parameter values using which the data simulated from the model are close to the observed data. A number of ABC algorithms based on MCMC and sequential Monte Carlo have been proposed to date. This paper reviews those algorithms and also considers the further issues such as the method of model choice without likelihood. As numerical example, the algorithms considered in this paper are demonstrated using the simple model.

キーワード: Approximate Bayesian computation, 補助変数法, ベイズ推定, マルコフ連鎖モンテカルロ法, 逐次モンテカルロ法.

1. はじめに

ベイズ統計学では、未知パラメータに関する推測は事後分布に基づいて行なわれる。いま、未知パラメータを θ 、その事前分布を $\pi(\theta)$ と表すことにする。また、観察されたデータ $\mathbf{y} = (y_1, y_2, \dots, y_n)'$ の尤度関数を $f(\mathbf{y}|\theta)$ と表記することにする。このとき、パラメータ θ の事後分布は、ベイズの定理より

* 神戸大学経営学研究科博士課程後期課程: 〒 657-8501 神戸市灘区六甲台 2-1 (E-mail: g.kobayashi@kobe-u.ac.jp).

$$\pi(\boldsymbol{\theta}|\mathbf{y}) \propto f(\mathbf{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}) \quad (1.1)$$

で与えられる。

事後分布は単純な統計モデルに対しても複雑となり、解析的に事後分布を求めたり、あるいは関心のある値の推定値などを求めることが不可能である場合が多い。そのためベイズ統計学では、シミュレーションを利用した推定方法が用いられるようになってきた。その代表的な方法が、マルコフ連鎖モンテカルロ法 (以下、MCMC 法) である (例えば, Liu (2001), Gamerman and Lopes (2006) などを参照)。現在, MCMC 法は経済学や経営学に限らず幅広い分野において用いられており、最近では複雑な統計モデルに対するより高度な MCMC アルゴリズムの開発も行われている。

MCMC 法は事後分布からのサンプリングを行うための方法であり、適用するためには事前分布と尤度関数が評価できなければならない。しかし、統計モデルによっては、尤度関数を解析的に求めることができなかつたり、あるいは尤度関数の計算が煩雑であることがある。例えば、一般化ラムダ分布は、分位点関数によってその確率分布が定義され、尤度関数を明示的に求めることができない。また、安定分布も特性関数が明示的に示されているだけである。自己ロジスティックモデルでは、尤度関数の正規化定数がパラメータに依存しており、尤度関数を計算することが非常に難しい。このような場合、MCMC 法を用いて通常のベイズ分析を行うには困難が伴う。

尤度関数が得られないことによる計算上の問題点を克服するため、補助変数を導入することによって尤度関数の評価を回避する方法が提案されてきた。その1つが Approximate Bayesian Computation (ABC) と呼ばれる方法である。この方法は、Tavaré *et al.* (1997), Pitchard *et al.* (1999), Beaumont *et al.* (2002) などにより集団遺伝学の分野で提案され、いまでは遺伝子学、疫学、集団生物学などの分野でも広く用いられている。また ABC は、先に述べた一般化ラムダ分布を含む分位点分布 (Allingham *et al.* (2009), McVinish (2010)), 安定分布 (Peters *et al.* (2010a)), 自己ロジスティックモデル (Drovandi and Pettitt (2011)) の分析などにも応用されている。最近では、経済学や経営学においても適用され始めており、例えば、共和分 VAR モデル (Peters *et al.* (2010b)), 損害保険の保険給付支払い金準備金分析 (Peters *et al.* (2010c)), オペレーショナルリスク分析 (Peters and Sisson (2006)), オンラインポートフォリオ分析 (Jasra *et al.* (2010)) などが挙げられる。Csilléry *et al.* (2010), Sisson and Fan (2011), Marin *et al.* (2011) では ABC のサーベイがなされているので参考にされたい。

本稿は ABC の概要とこれまでに提案されてきたアルゴリズムの紹介を行うことで、経済学や経営学分野での ABC のさらなる応用への貢献となることを期待する。第2節ではまず

ABC の基本的なアルゴリズムについて説明を行う。また、ABC アルゴリズムを実行した後
に推定値を調整する方法やモデル選択の方法などについても言及する。これまでに数々の
ABC アルゴリズムが提案されてきた。第3節では、これらのアルゴリズムを MCMC 法に
基づくものと逐次モンテカルロ法 (例えば, Del Moral *et al.* (2006), Doucet *et al.* (2001),
Liu (2001) など) を参照) に基づくものとに区別して紹介する。最後に結論を第4節で述
べることにする。

2. ABC の基礎

2.1 棄却サンプリング

(1.1) で与えられる事後分布 $\pi(\boldsymbol{\theta}|\mathbf{y})$ から $\boldsymbol{\theta}$ をサンプリングしたいとする。ただし、尤度
関数 $f(\mathbf{y}|\boldsymbol{\theta})$ は評価が困難であるとする。このとき、尤度関数を直接評価することなく事後
分布からサンプリングを行う1つの方法は、モデルから別のデータ $\mathbf{z} = (z_1, z_2, \dots, z_n)'$ を
発生させ、 $\mathbf{y} = \mathbf{z}$ となるような $\boldsymbol{\theta}$ を事後分布からのサンプルとして採択するという方法で
ある。具体的なアルゴリズムは、以下で与えられる。

棄却サンプリング (1)

1. $\pi(\boldsymbol{\theta})$ から $\boldsymbol{\theta}$ を発生させる。
2. $f(\mathbf{z}|\boldsymbol{\theta})$ から \mathbf{z} を発生させる。
3. $\mathbf{y} = \mathbf{z}$ であれば $\boldsymbol{\theta}$ を採択し、 $\mathbf{y} \neq \mathbf{z}$ であれば1に戻る。

このアルゴリズムによって発生された $\boldsymbol{\theta}$ は、

$$p(\boldsymbol{\theta}) \propto \int f(\mathbf{z}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})I(\mathbf{y} = \mathbf{z})d\mathbf{z} = f(\mathbf{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}) \propto \pi(\boldsymbol{\theta}|\mathbf{y})$$

であることから、事後分布 $\pi(\boldsymbol{\theta}|\mathbf{y})$ からのサンプルとなっていることが容易に確認できる。
ここで、 $I(\cdot)$ は指標関数を表す。しかし、 \mathbf{y} が連続である場合には、 $\mathbf{y} = \mathbf{z}$ となる確率は0
であるので $\boldsymbol{\theta}$ が採択されないことになり、この方法は実用的ではない。また、 \mathbf{y} が離散で
あったとしても採択確率が小さくなり著しく非効率になる可能性がある。

そこで ABC では、 $\mathbf{y} = \mathbf{z}$ となるような $\boldsymbol{\theta}$ を探す代わりに、モデルから発生された \mathbf{z} が
観測されたデータ \mathbf{y} に近くなるような $\boldsymbol{\theta}$ を探すことを試みる。このとき、棄却サンプリ
ング (1) は次のアルゴリズムで置き換えられることになる。

棄却サンプリング (2)

1. $\pi(\boldsymbol{\theta})$ から $\boldsymbol{\theta}$ を発生させる。
2. $f(\mathbf{z}|\boldsymbol{\theta})$ から \mathbf{z} を発生させる。

3. $\rho(\mathbf{y}, \mathbf{z}) < \epsilon$ であれば θ を採択し、そうでなければ1に戻る。

ここで、 $\rho(\mathbf{y}, \mathbf{z})$ は \mathbf{y} と \mathbf{z} の距離を表す非負の関数であり、例えばユークリッド距離 $\rho(\mathbf{y}, \mathbf{z}) = \sqrt{\sum_{i=1}^n (y_i - z_i)^2}$ などがよく用いられる。また、 ϵ は分析者が指定する正の定数である。

棄却サンプリング(2)において、 $\epsilon \rightarrow \infty$ とすれば、 θ は必ず採択されることになり、 θ は事前分布 $\pi(\theta)$ からのサンプルとなる。逆に $\epsilon \rightarrow 0$ とすれば、棄却サンプリング(2)は棄却サンプリング(1)と同じとなり、 θ は事後分布から得られたサンプルとなる。したがって、小さい ϵ を選べば棄却サンプリング(2)によって発生される θ は、事後分布 $\pi(\theta|\mathbf{y})$ を近似する確率分布

$$p(\theta) = \int \pi_\epsilon(\theta, \mathbf{z}|\mathbf{y}) d\mathbf{z} \approx \pi(\theta|\mathbf{y})$$

からのサンプルとなっている。ここで、

$$\pi_\epsilon(\theta, \mathbf{z}|\mathbf{y}) \propto f(\mathbf{z}|\theta)\pi(\theta)I(\rho(\mathbf{y}, \mathbf{z}) < \epsilon)$$

である。

2.2 マルコフ連鎖モンテカルロ法

棄却サンプリングでは、データの情報を含まない事前分布から θ を発生させている。そのため、発生された θ は必ずしも事後密度が高い領域の値ではなく、1つの θ が採択されるまでに事前分布から多数の θ を発生させる必要があり非効率である。Marjoram *et al.* (2003)は、MCMC法を適用することによって効率性の改善を試みた。また、Wilkinson (2008)やWegmann *et al.* (2009)などによっても同様のアルゴリズムが提案されている。

彼らのアルゴリズムを説明するために、補助変数 \mathbf{z} を導入した拡張事後分布

$$\pi_{ABC}(\theta, \mathbf{z}|\mathbf{y}) \propto g_\epsilon(\mathbf{y}|\mathbf{z}, \theta)f(\mathbf{z}|\theta)\pi(\theta) \quad (2.1)$$

を考えることにする(Sisson and Fan (2011))。ここで、 $g_\epsilon(\mathbf{y}|\mathbf{z}, \theta) (\geq 0)$ はウェイト関数と呼ばれ、 ϵ をパラメータとして持つ。このウェイト関数は、 \mathbf{y} と \mathbf{z} の距離が近いときには大きな値を、距離が遠いときには小さな値を取り、 $\mathbf{y} = \mathbf{z}$ のときには定数となるように選ぶ必要がある。実際には

$$g_\epsilon(\mathbf{y}|\mathbf{z}, \theta) \propto I(\rho(\mathbf{y}, \mathbf{z}) < \epsilon) \quad (2.2)$$

や

$$g_\epsilon(\mathbf{y}|\mathbf{z}, \theta) \propto \exp\left\{-\frac{\rho(\mathbf{y}, \mathbf{z})}{\epsilon}\right\}$$

などがよく用いられる。

Marjoram *et al.* (2003)らのアルゴリズムは、事後分布 $\pi(\theta|\mathbf{y})$ を

$$\pi_{ABC}(\theta|\mathbf{y}) = \int \pi_{ABC}(\theta, \mathbf{z}|\mathbf{y}) d\mathbf{z}$$

によって近似し、(2.1)からのサンプリングを以下に示すメトロポリス-ヘイスティングス法 (以下 MH 法) によって行う方法である。

MCMC-ABC

現在の状態を (θ, \mathbf{z}) とし、次のステップを繰り返し行う。

1. $q(\theta^*|\theta)$ から θ^* を発生させる。
2. $f(\mathbf{z}^*|\theta^*)$ から \mathbf{z}^* を発生させる。
3. (θ^*, \mathbf{z}^*) を確率

$$\alpha((\theta, \mathbf{z}), (\theta^*, \mathbf{z}^*)) = \min \left\{ 1, \frac{g_\epsilon(\mathbf{y}|\mathbf{z}^*, \theta^*)\pi(\theta^*)q(\theta|\theta^*)}{g_\epsilon(\mathbf{y}|\mathbf{z}, \theta)\pi(\theta)q(\theta^*|\theta)} \right\}$$

で採択する。

ここで、 $q(\theta^*|\theta)$ は θ の提案密度を表す。尤度関数が利用できないことから、通常はランダムウォーク連鎖が用いられる。また、MCMC-ABC の採択確率は、ウェイト関数、提案分布、そして事前分布のみに依存しており、計算することができない $f(\mathbf{z}|\theta)$ は含まれていないことに注意して欲しい。さらに、 $g_\epsilon(\mathbf{y}|\mathbf{z}, \theta) = I(\rho(\mathbf{y}, \mathbf{z}) < \epsilon)$ 、 $q(\theta^*|\theta) = \pi(\theta^*)$ であるときには、MCMC-ABC は棄却サンプリング (2) と等しくなり、MCMC-ABC が棄却サンプリング (2) の拡張となっていることが分かる。

数値例 数値例として、正規分布の平均が未知で分散が既知であるモデル $y_i \sim N(\theta, 1)$ ($i = 1, \dots, 500$) を考える。また、事前分布に $\theta \sim N(0, 100)$ を仮定することにする。このモデルでは、事後分布を明示的に得ることができるため、本稿ではこのモデルを繰り返し取り上げていく。

未知パラメータ θ の事後分布を MCMC-ABC によって近似する。ウェイト関数 g_ϵ には (2.2) を用い、距離関数は $\rho(\mathbf{y}, \mathbf{z}) = \sqrt{\sum_{i=1}^{500} (y_i - z_i)^2}$ を用いることにした。ここでは、 ϵ の値として 3.0, 2.5, 2.0 の 3 つの場合を考えることにする。最後に、アルゴリズムの初期値は $\theta = 0$ とし、標準偏差が 0.1 の正規分布を用いたランダムウォーク連鎖によって θ をサンプリングした。1 万個のサンプルを推定に用いた。

各 ϵ の値の下での MCMC-ABC による標本経路、自己相関関数、ヒストグラムが図 1 に示されている。ヒストグラム中の実線は真の事後密度を表す。図より、 ϵ の値が小さくなるにつれて事後分布がより正確に近似されていることが分かる。また、小さい ϵ を用いることによって、サンプルの自己相関が高くなっていくことも見て取れる。

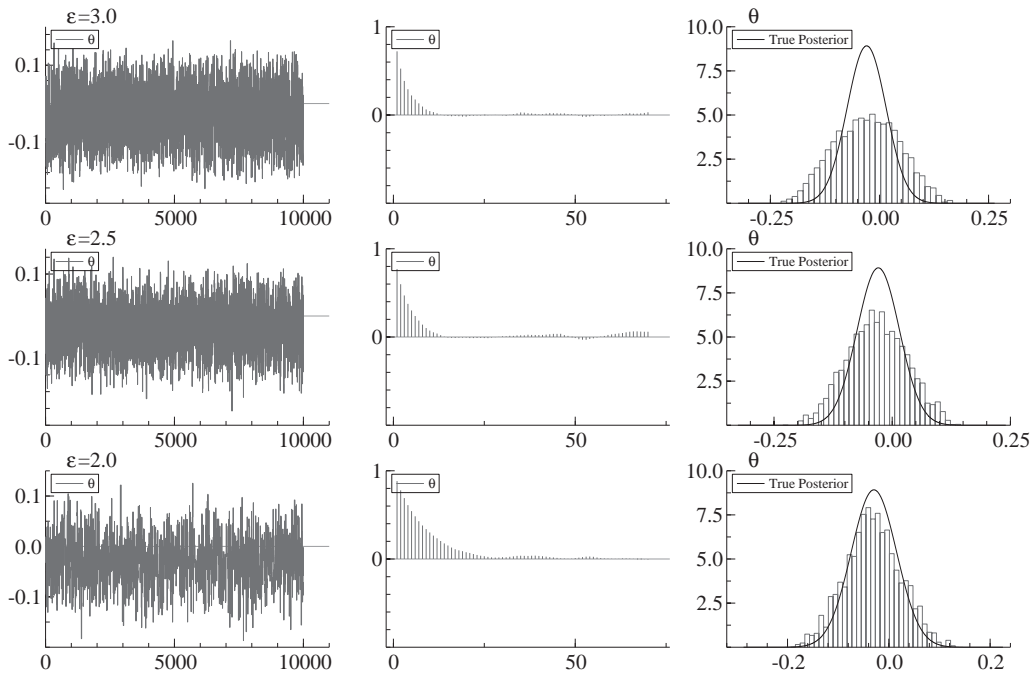


図1 MCMC-ABC の標本経路 (左), 自己相関関数 (中央), ヒストグラム (右).

2.3 要約統計量の使用

実際のデータ解析ではデータ数が大きいことがよくある. そのような場合, \mathbf{y} と \mathbf{z} の距離を表す $\rho(\mathbf{y}, \mathbf{z})$ を評価するには時間がかかり, 非効率になってしまう. そこで, \mathbf{y} と \mathbf{z} の距離を直接評価する代わりに, より次元の低い要約統計量 (summary statistics) $S(\cdot)$ を考え, $S(\mathbf{y})$ と $S(\mathbf{z})$ の距離を評価して分析を行うことがよくなされる. このとき, 拡張事後分布 (2.1) は,

$$\pi_{ABC}(\boldsymbol{\theta}, S(\mathbf{z})|S(\mathbf{y})) \propto g_{\epsilon}(S(\mathbf{y})|S(\mathbf{z}), \boldsymbol{\theta})f(\mathbf{z}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})$$

と書き直すことができ, ABC は $S(\mathbf{y})$ と $S(\mathbf{z})$ が近くなるようなパラメータ $\boldsymbol{\theta}$ を探すことになる.

要約統計量としては, データが $\boldsymbol{\theta}$ について保有している情報をすべて含んでいる十分統計量を使用することが望ましい. しかし, ABC を適用する場面ではモデルの尤度関数が求められないし, あるいはモデルが複雑であれば, 必ずしも十分統計量を見つけることができない. このような場合, なるべく十分統計量と同じ情報を持つような要約統計量を選択する必要がある, どの要約統計量を用いるかについては注意が必要である.

実際には, データのモーメントなど複数の要約統計量を用いることになる. 複数ある統計量の中から有用なものを選ぶために, Wegmann *et al.* (2009) は事前分布からサンプリングするパイロットランを行い, 部分最小自乗法 (partial least squares method) を用い

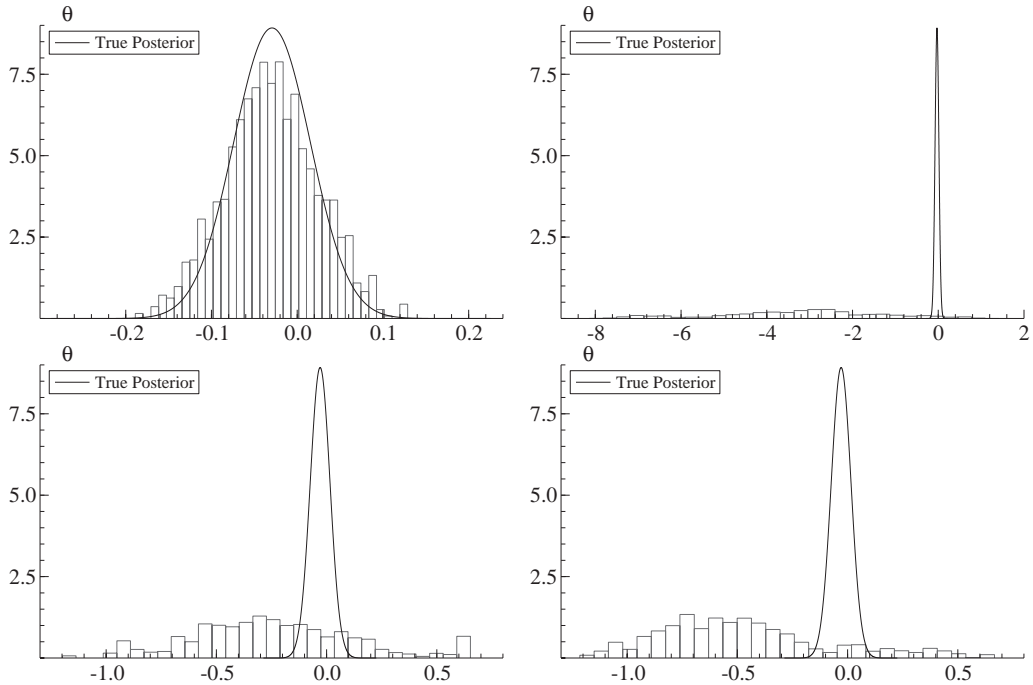


図2 様々な要約統計量を用いた MCMC-ABC によるヒストグラム：標本平均 (左上)，標本分散 (右上)，最大値 (左下)，最小値 (右下)。

て有効な要約統計量を選択している。Blum (2010) は、要約統計量の次元が大きくなるほど、次に説明する回帰調整を用いた事後推定値の平均二乗誤差が 0 に収束する速度が遅くなることを示している。Drovandi *et al.* (2011) では、要約統計量にはデータの複雑な関数を用いることでデータの情報をより多く受け継ぐことができるとして、最尤法などで比較的簡単に推定できるような補助的なモデルの推定値を要約統計量として採用することを提案している。Fearnhead and Prangle (2011) は二次損失関数の期待損失を最小化するために、事後分布の平均値を要約統計量として使用することを提案している。要約統計量である事後平均は、アルゴリズムのパイロットランにおいて次節で紹介する回帰調整の考え方を利用して推定される。また、Joyce and Marjoram (2008) や Nunes and Balding (2010) によっても要約統計量を選択するアルゴリズムが提案されている。

数値例 (続き) 前節での正規分布のモデルを再び考える。ここでは ABC において適切な要約統計量を使用する必要があるということを示すために、4つの異なる要約統計量を用いた結果を比較する。用いる要約統計量は標本平均、標本分散、最大値、そして最小値とした。ウェイト関数は

$$g_{\epsilon}(S(\mathbf{y})|S(\mathbf{z}), \theta) \propto I(\rho(S(\mathbf{y}), S(\mathbf{z})) < \epsilon)$$

を用い、 $\rho(S(\mathbf{y}), S(\mathbf{z})) = |S(\mathbf{y}) - S(\mathbf{z})|$ とした。また、 ϵ の値は0.03とし、前節と同じランダムウォーク連鎖によってサンプリングを行った。各要約統計量の下で得られたサンプルのヒストグラムが図2によって示されている。ヒストグラム中の実線は真の事後密度を表す。図2から明らかであるように、 θ の十分統計量である標本平均以外の要約統計量では事後分布の近似ができておらず、パラメータについての情報を持つ要約統計量を使用する必要があることが示されている。

2.4 回帰調整

ABCによって得られる推定値はあくまで近似である。そこで、パラメータ推定の精度を改善するため、Beaumont *et al.* (2002)はABCアルゴリズムを用いて事後分布からサンプリングした後に回帰分析を行うことを提案している。本稿では、この方法を回帰調整(regression adjustment)とよぶことにする。

いま、ABCアルゴリズムによって θ (ここではスカラーとする)のサンプルが T 個得られたとし、これを $\theta^{(t)}$ ($t = 1, \dots, T$)と表す。このとき、回帰式

$$\theta^{(t)} = \alpha + (S(\mathbf{z}^{(t)}) - S(\mathbf{y}))' \beta + \omega^{(t)} \quad (t = 1, \dots, T) \quad (2.3)$$

を考える。ここで $\omega^{(t)}$ は互いに独立で、平均がゼロ、同一の分散を持つ確率変数である。サンプルの調整は、回帰式(2.3)の最小二乗推定値 $(\hat{\alpha}, \hat{\beta})'$ を用いて

$$\theta^{*(t)} = \theta^{(t)} - (S(\mathbf{z}^{(t)}) - S(\mathbf{y}))' \hat{\beta}$$

によって行なわれる。

もし、 $\theta^{(t)}$ と $S(\mathbf{z}^{(t)}) - S(\mathbf{y})$ の真の関係が(2.3)のように線形のとときには、 $\theta^{*(t)}$ は $\pi(\theta|S(\mathbf{y}))$ からの正確なサンプルであり、最小二乗推定値 $\hat{\alpha}$ は事後平均 $E[\theta|S(\mathbf{y})]$ の推定値となる。しかし、回帰式の線形性は特にABCの枠組みでは制約的であるので、Beaumont *et al.* (2002)では目的関数

$$\sum_{t=1}^T (\theta^{(t)} - \alpha - [S(\mathbf{z}^{(t)}) - S(\mathbf{y})]' \beta)^2 K \left(\frac{\rho(\mathbf{y}, \mathbf{z}^{(t)})}{\epsilon} \right)$$

を最小化する局所線形回帰を提案している。ここで、 $K(\cdot)$ は適当なカーネル関数を表す。さらに、Blum and François (2010)ではノンパラメトリック回帰モデルを推定することを提案し、Blum (2010)は回帰調整による事後分布のバイアスと分散を分析している。

数値例(続き) 再び第2.2節での正規分布のモデルを考える。ここではMCMC-ABCによって得られたサンプルに対して回帰調整を適用した例を示す。要約統計量として標本平均を用い、ウェイト関数と距離関数は前節と同じものを用いた。ただし、 ϵ の値は0.1とした。第2.2節と同じランダムウォーク連鎖によってサンプリングを行った。MCMC-ABC

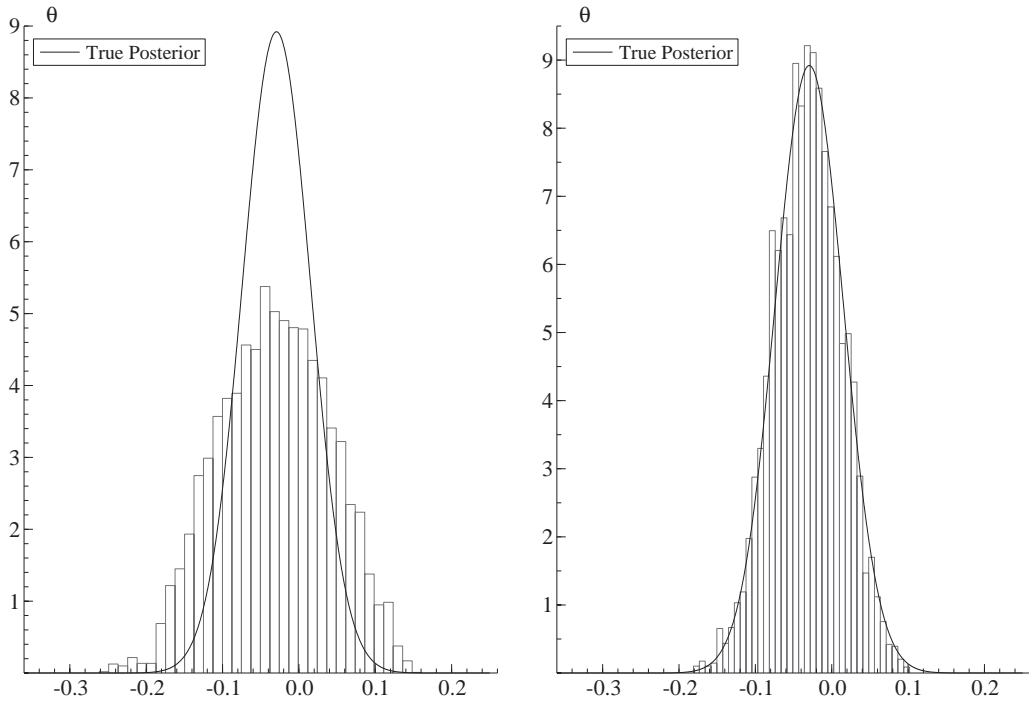


図3 回帰調整前(左)と回帰調整後(右)のヒストグラム。

によって得られた1万個のサンプルを(2.3)の回帰式によって回帰調整を行った。図3によって回帰調整前と回帰調整後のヒストグラムが示されている。ヒストグラム中の実線は真の事後密度を表す。図より、回帰調整後に事後分布の近似が改善されていることが分かる。

2.5 モデル選択

複数のモデルの中からモデルを選ぶとき、ベイズ統計学では各モデルに対する事前分布 $\pi(\mathcal{M} = m)$ ($m = 1, \dots, M$) を仮定し、モデルの事後確率

$$\pi(\mathcal{M} = m | \mathbf{y}) = \frac{\pi(\mathcal{M} = m) f_m(\mathbf{y})}{\sum_{i=1}^M \pi(\mathcal{M} = i) f_i(\mathbf{y})}$$

に基づいてモデル選択を行う。ここで、 $f_m(\mathbf{y})$ はモデル m の周辺尤度を表し、

$$f_m(\mathbf{y}) = \int \pi_m(\boldsymbol{\theta}_m) f_m(\mathbf{y} | \boldsymbol{\theta}_m) d\boldsymbol{\theta}_m$$

によって定義される。また、 $\boldsymbol{\theta}_m$ 、 $\pi_m(\boldsymbol{\theta}_m)$ 、 $f_m(\mathbf{y} | \boldsymbol{\theta}_m)$ はそれぞれモデル m のパラメータ、事前密度、尤度関数を表す。2つのモデル ($m = 1, 2$) を比較するときには、ベイズファクター

$$B_{12} = \frac{f_1(\mathbf{y})}{f_2(\mathbf{y})}$$

が重要な数量となる。

尤度関数の評価ができる場合には、様々な方法によって周辺尤度を求めることができる (Han and Carlin (2001) を参照). しかし、尤度関数の評価が困難である場合、これまでの方法は適用が困難となってしまう. そこで, Cornuet *et al.* (2008) や Grelaud *et al.* (2009) は, ABC に基づいた次の棄却サンプリングを提案している.

ABC-MC

次のステップを繰り返し行う.

1. モデルの事前分布 $\pi(\mathcal{M} = m)$ から \mathcal{M} を発生させる.
2. $\pi_{\mathcal{M}}(\boldsymbol{\theta}_{\mathcal{M}})$ から $\boldsymbol{\theta}_{\mathcal{M}}$ を発生させる.
3. $f_{\mathcal{M}}(\mathbf{z}|\boldsymbol{\theta}_{\mathcal{M}})$ から \mathbf{z} を発生させる.
4. もし $\rho(S(\mathbf{y}), S(\mathbf{z})) < \epsilon$ ならば \mathcal{M} と $\boldsymbol{\theta}$ を採択し, そうでなければ 1 に戻る.

この棄却サンプリングによって得られたサンプルを $\mathcal{M}^{(t)}$ ($t = 1, \dots, T$) と表せば, モデルの事後確率 $\pi(\mathcal{M} = m|\mathbf{y})$ は,

$$\frac{1}{T} \sum_{t=1}^T I(\mathcal{M}^{(t)} = m)$$

によって推定することができる. また, ベイズファクターは,

$$\hat{B}_{12} = \frac{\pi(\mathcal{M} = 2) \sum_{t=1}^T I(\mathcal{M}^{(t)} = 1)}{\pi(\mathcal{M} = 1) \sum_{t=1}^T I(\mathcal{M}^{(t)} = 2)} \quad (2.4)$$

により推定される. 最近になり, Cornuet *et al.* (2010), Leuenberger and Wegmann (2010) は前節の回帰調整を応用した方法を, Didelot *et al.* (2011) は MCMC 法を応用した方法を, そして Toni *et al.* (2009) や Toni and Stumpf (2010) は第 3.3 節で紹介される逐次モンテカルロ法を応用した方法を提案している.

一方で, Didelot *et al.* (2011) や Robert *et al.* (2011) らは, ABC-MC に基づくモデル選択が適用可能であるのはギブス確率場 (Grelaud *et al.* (2009)) などのごく一部のモデルに限られており, 一般的にはうまく機能しないと主張している. Robert *et al.* (2011) にならい, 2つのモデルを比較する状況を考える. ここで, $\boldsymbol{\eta}(\mathbf{y})$ が $\boldsymbol{\theta}_1$ と $\boldsymbol{\theta}_2$ 両方のパラメータの十分統計量であるとする. さらに, 十分統計量 $\boldsymbol{\eta}(\mathbf{y})$ の密度関数を $f_m^n(\boldsymbol{\eta}(\mathbf{y})|\boldsymbol{\theta}_m)$ とすれば, 尤度関数は $f_m(\mathbf{y}|\boldsymbol{\theta}_m) = g_m(\mathbf{y})f_m^n(\boldsymbol{\eta}(\mathbf{y})|\boldsymbol{\theta}_m)$ のように分解できる. したがって, ベイ

ズファクターは

$$\begin{aligned}
 B_{12} &= \frac{\int \pi_1(\boldsymbol{\theta}_1) f_1(\mathbf{y}|\boldsymbol{\theta}_1) d\boldsymbol{\theta}_1}{\int \pi_2(\boldsymbol{\theta}_2) f_2(\mathbf{y}|\boldsymbol{\theta}_2) d\boldsymbol{\theta}_2} \\
 &= \frac{g_1(\mathbf{y}) \int \pi(\boldsymbol{\theta}_1) f_1^\eta(\boldsymbol{\eta}(\mathbf{y})|\boldsymbol{\theta}_1) d\boldsymbol{\theta}_1}{g_2(\mathbf{y}) \int \pi(\boldsymbol{\theta}_2) f_2^\eta(\boldsymbol{\eta}(\mathbf{y})|\boldsymbol{\theta}_2) d\boldsymbol{\theta}_2} \\
 &= \frac{g_1(\mathbf{y})}{g_2(\mathbf{y})} B_{12}^\eta
 \end{aligned} \tag{2.5}$$

と表すことができる。ここで、 B_{12}^η は $\boldsymbol{\eta}(\mathbf{y})$ の値だけが観察されたときのベイズファクターであり、推定量 (2.4) は要約統計量を用いてこれを近似している。よって (2.5) から、ABC によるベイズファクターは $g_1(\mathbf{y}) = g_2(\mathbf{y})$ でない限り、比 $g_1(\mathbf{y})/g_2(\mathbf{y})$ の分だけ本来のベイズファクター B_{12} と異なることになる。そして、この比は一般的に近似することが不可能である。これにより Robert *et al.* (2011) は、要約統計量を用いて近似的にベイズファクターを求め、これに基づいてモデル選択を行うことはできないと主張している。代わりに Ratmann *et al.* (2009) や Drovandi *et al.* (2011) のように、 $\rho(S(\mathbf{y}), S(\mathbf{z}))$ の分布を確認するようなモデル診断方法が ABC の枠組みにおいては有用であるとしている。さらに、Didelot *et al.* (2011) は指数型分布族の場合にはすべてのモデルの十分統計量を用いることを提案している。

3. ABC サンプラーの最近の発展

現在 ABC アルゴリズムの発展には、MCMC 法に基づくものと逐次モンテカルロ法に基づくものと 2 つの大きな流れがある。この節ではまず MCMC-ABC が抱えている問題点を指摘する。その後、この問題点を克服するために提案されている MCMC 法によるアルゴリズムの紹介から行っていく。

3.1 MCMC-ABC の問題点

ABC-MCMC は比較的単純な MH 法であるが、いくつか問題点があることが Wegmann *et al.* (2009), Drovandi and Pettitt (2011), Kobayashi and Kozumi (2011) などによって報告されている。

1 つ目の問題点は、初期値の選択に関するものである。もし、パラメータ $\boldsymbol{\theta}$ の初期値が事後密度の高い領域の値ではない場合、その下で発生された \mathbf{z} は観測されたデータ \mathbf{y} とは離れた値となってしまうであろう。すると、ウェイト関数 $g_\epsilon(\mathbf{y}|\mathbf{z}, \boldsymbol{\theta})$ の値が小さく、つまり採択確率が小さくなり、次の状態へなかなか移ることができないことになる。特にウェイト関数として $I(\rho(\mathbf{y}, \mathbf{z}) < \epsilon)$ を用いたときには、 $\rho(\mathbf{y}, \mathbf{z}) < \epsilon$ を満たす \mathbf{z} が得られないため、採択確率がゼロとなってしまう。Allingham *et al.* (2009) が行った数値実験では、新たな状態へ移るのに最大で 52 万回もの繰り返しを行っている。したがって、MCMC-ABC においては、通常の MCMC に比べて初期値を注意深く選択しなければならない。

次に、ABC では尤度関数が利用できないため、適切な提案分布を構築することが難しい。そのため、先に述べたように、 θ のサンプリングにはランダムウォーク連鎖がよく用いられている。しかし、ランダムウォーク連鎖から発生されたサンプルは非常に高い自己相関を持ち、定常分布への収束が遅いことが知られており、MCMC-ABC も同じ問題を有することになる。

最後に、ABC においてより精度の高い事後分布の近似を望むのであれば、より小さな ϵ を使用しなければならない。しかし、MCMC-ABC の採択確率は ϵ に依存しており、この値を小さくすれば採択確率は低くなってしまふ。つまり、MCMC-ABC では事後分布の近似の精度と採択確率のトレードオフが存在しており、効率的なサンプリングを行うためには採択確率を改善する必要がある。

数値例 (続き) 上で述べた MCMC-ABC の問題を例示するために、第 2.2 節でのモデルを引き続き考える。要約統計量として標本平均を用い、ウェイト関数と距離関数は第 2.3 節のものを用いた。第 2.2 節と同じランダムウォーク連鎖によってサンプリングを行った。

まず、MCMC-ABC の初期値問題について考える。アルゴリズムを 50 回走らせ、そのうち 100 回の繰り返しの間に初期値から動くことができた回数を調べることで、初期値の選択に対する敏感さを評価する。初期値は標準正規分布から毎回発生させた。ここで $\epsilon = 0.03$ の下でアルゴリズムを 50 回走らせたところ、100 回の繰り返しの間に初期値から動くことができた回数は 12 回であった。

次に、初期値を $\theta = 0$ にし、MCMC-ABC によって 1 万個のサンプルを得た。このときの採択確率は 19.2% であった。図 4 はサンプルの自己相関関数を示している。非常に単純なモデルへの適用にもかかわらず、MCMC-ABC の自己相関がなかなか消えないことが見て取れる。

3.2 MCMC-ABC の発展

3.2.1 ϵ の段階的な調整

Sisson and Fan (2011) は、アルゴリズムの初期段階では大きめの ϵ の値を用い、サンプリングが進むにつれて ϵ の値を小さくしていく方法を提案している。彼らの方法では、まず $\epsilon_0 > \epsilon_1 > \dots > \epsilon_T = \epsilon$ の値を設定する。そして、 ϵ_0 からアルゴリズムを始め、バーンイン期間に $\epsilon_1 \rightarrow \epsilon_2 \rightarrow \dots$ と値を小さくしていき、バーンイン期間が終わると ϵ の下でサンプリングを行う。この方法は、最初に大きめの ϵ を設定することで採択確率を上げ、パラメータの値を更新されやすくし、定常分布への収束が促されることが期待できる。

Sisson and Fan (2011) が提案した方法では、 ϵ_t を小さくする速度が速すぎるとアルゴリズムがうまく機能しないが、一方で遅すぎると計算量が大きくなってしまふので、 ϵ_t の設定

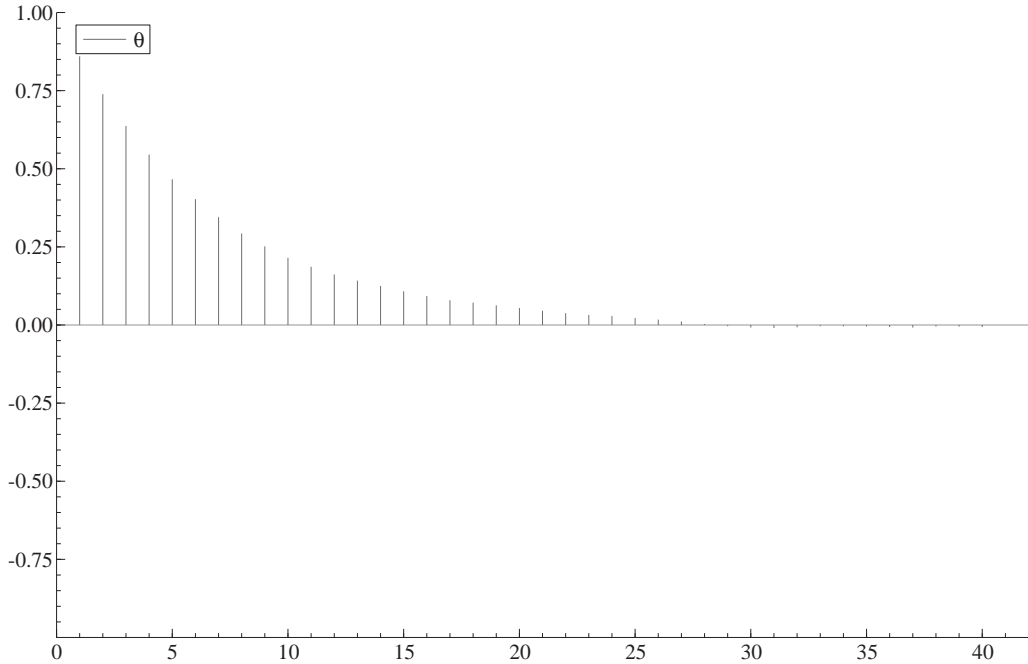


図4 MCMC-ABC の自己相関関数.

には注意が必要である. Sisson and Fan (2011) は, $\epsilon_t = \max\{\epsilon_0 - ct, \epsilon\}$ とし, ϵ_T へ収束する速度を定数 $c (> 0)$ によってコントロールする方法と, $\epsilon^* = \max\{\epsilon, \min\{\rho(\mathbf{y}, \mathbf{z}^*), \epsilon_{t-1}\}\}$ とし, 候補 (θ^*, \mathbf{z}^*) が採択されれば $\epsilon_t = \epsilon^*$, そうでなければ $\epsilon_t = \epsilon_{t-1}$ とする方法を提案している.

3.2.2 状態空間拡張による方法

Bortot *et al.* (2007) は, MCMC-ABC で考えられている (θ, \mathbf{z}) の推移の代わりに, 状態空間を拡張して $(\theta, \mathbf{z}, \epsilon)$ の推移を行うアルゴリズムを提案している. 同様のアルゴリズムがモデル不確実性の観点から Ratmann *et al.* (2009) によっても提案されている. このアルゴリズムは, ϵ に対する事前分布 $\pi(\epsilon)$ を導入し, (2.1) を拡張した

$$\pi(\theta, \mathbf{z}, \epsilon | \mathbf{y}) \propto g_\epsilon(\mathbf{y} | \mathbf{z}, \theta) f(\mathbf{z} | \theta) \pi(\theta) \pi(\epsilon)$$

からのサンプリングを行う方法である.

Augmented MCMC-ABC

現在の状態を $(\theta, \mathbf{z}, \epsilon)$ とし, 次のステップを繰り返し行う.

1. $q(\mathbf{z}^*, \epsilon^* | \theta, \epsilon)$ から (θ^*, ϵ^*) を発生させる.
2. $f(\mathbf{z}^* | \theta^*)$ から \mathbf{z}^* を発生させる.

3. $(\theta^*, \mathbf{z}^*, \epsilon^*)$ を確率

$$\alpha((\theta, \mathbf{z}, \epsilon), (\theta^*, \mathbf{z}^*, \epsilon^*)) = \min \left\{ 1, \frac{g_{\epsilon^*}(\mathbf{y}|\mathbf{z}^*, \theta^*)\pi(\theta^*)\pi(\epsilon^*)q(\theta, \epsilon|\theta^*, \epsilon^*)}{g_{\epsilon}(\mathbf{y}|\mathbf{z}, \theta)\pi(\theta)\pi(\epsilon)q(\theta^*, \epsilon^*|\theta, \epsilon)} \right\}$$

で採択する.

近似の精度は ϵ の事前分布のパラメータによってコントロールされ, ϵ の事前分布には指数分布などが用いられる. このアルゴリズムは, 時折大きな ϵ の値を発生させることによってパラメータが更新しやすくなることを目指したものである. また, ϵ の初期値に大きな値を設定しておくことで悪い初期値からでも容易に動くことが容易になり, MCMC-ABC の初期値選択の問題が改善されることが期待できる. ところが, 追加的に ϵ のサンプリングを行うため, 事後分布からのサンプルの自己相関はさらに高くなってしまう.

大きな ϵ の値とともに真の事後分布のもとの事後密度が低い領域の点も採択され, このアルゴリズムによる事後分布の近似の精度が悪くなることが懸念される. Bortot *et al.* (2007) は, 閾値 ϵ_T を導入して $\epsilon^* < \epsilon_T$ となるような $(\theta^*, \mathbf{z}^*, \epsilon^*)$ のみを採択することで, 大きすぎる ϵ^* を採択することによって事後分布の近似の精度が悪くなることを防ぐことも提案している.

数値例 (続き) 引き続き正規分布のモデルを考え, 第 3.1 節で述べた問題を Augmented MCMC-ABC がどのように改善するかを調べることにする. ここでは ϵ の事前分布として平均が 0.03 の指数分布を仮定し, ϵ のサンプリングは標準偏差が 0.05 の正規分布を用いたランダムウォーク連鎖によって行った.

まず, Augmented MCMC-ABC を 50 回走らせ, 100 回の繰り返しの間に初期値から動くことができた回数を調べた. パラメータ θ の初期値は標準正規分布から毎回発生させ, ϵ の初期値は 1.0 とした. 100 回の繰り返しの間にアルゴリズムが初期値から動くことができた回数は, MCMC-ABC が 12 回であったのに対し, Augmented MCMC-ABC は 34 回であり, MCMC-ABC の初期値の問題について大きな改善が見られた.

次に, 初期値を $(\theta, \epsilon) = (0, 0.2)$ とし, 1 万個のサンプルを得た. 採択確率は 18.2% であった. また, ϵ の事後平均は 0.063, 95% 信用区間は (0.007, 0.177) であり, 比較的大きな ϵ の値も採択されていることが分かる. 図 5 は Augmented MCMC-ABC による θ と ϵ の標本経路, 自己相関関数, ヒストグラムを示している. 図より, Augmented MCMC-ABC による自己相関は MCMC-ABC よりもさらに高くなっていることが分かる.

3.2.3 パラレルテンパリング

効率性を改善するための手法として, Geyer (1991) によって提案されたパラレルテンパリング (parallel tempering, PT) を MCMC-ABC に応用することが考えられる. PT は, m 個の異なった定常分布 $\pi_l(\theta, \mathbf{z}|\mathbf{y})$ ($l = 1, \dots, m$) を持ったマルコフ連鎖を並行して走ら

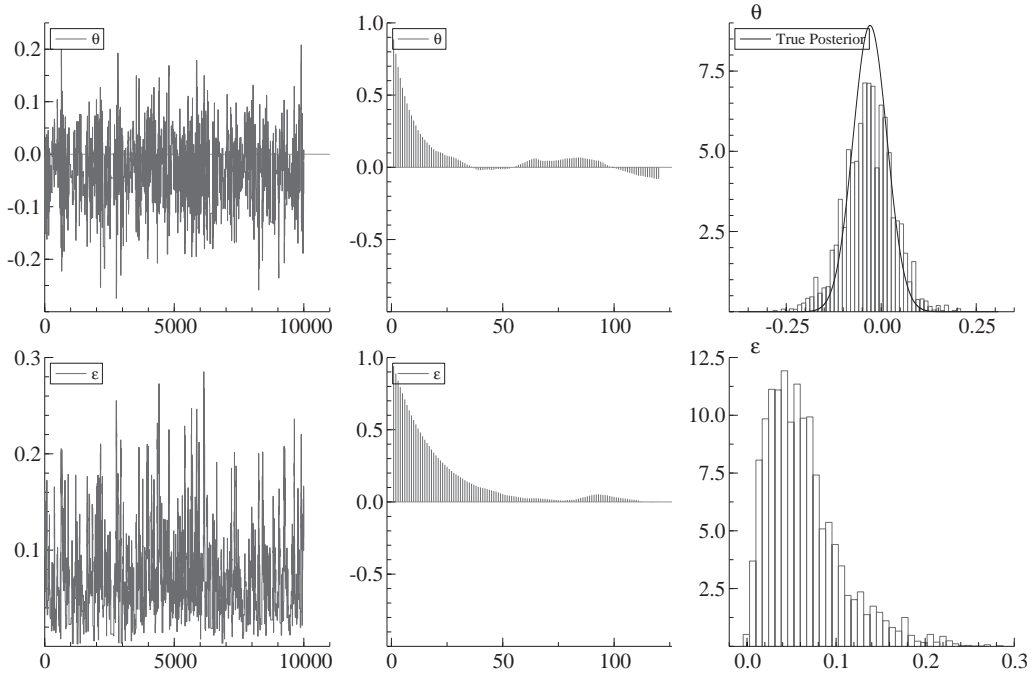


図5 Augmented MCMC-ABC による θ (上段) と ϵ (下段) の標本経路 (左), 自己相関関数 (中央), ヒストグラム (右).

せる方法であり, 通常, 各定常分布は $\pi_l(\boldsymbol{\theta}, \mathbf{z}|\mathbf{y}) = \pi(\boldsymbol{\theta}, \mathbf{z}|\mathbf{y})^{1/T_l}$ の形をとっている. ここで, $1 = < T_1 < T_2 < \dots < T_m$ であり, T_1 を持つ分布 $\pi_1(\boldsymbol{\theta}, \mathbf{z}|\mathbf{y})$ が本来関心のある分布である. 大きな T_l の値を持つ分布は小さな T_l の値を持つ分布よりもその形状がフラットなので, パラメータは状態空間をより自由に動くことができる. そこで, PT は隣り合う 2 つの連鎖の状態を交換することによって, サンプルの混合を改善することを意図したアルゴリズムである.

いま, 各定常分布を

$$\pi_l(\boldsymbol{\theta}, \mathbf{z}|\mathbf{y}) \propto f_{\epsilon_{(l)}}(\mathbf{y}|\mathbf{z}_{(l)}, \boldsymbol{\theta}_{(l)})f(\mathbf{z}_{(l)}|\boldsymbol{\theta}_{(l)})\pi(\boldsymbol{\theta}_{(l)})$$

と表すことにする. ここで, $\boldsymbol{\theta}_{(l)}, \mathbf{z}_{(l)}, \epsilon_{(l)}$ はそれぞれ l 番目の連鎖の $\boldsymbol{\theta}, \mathbf{z}, \epsilon$ の値である. また, $\epsilon_{(l)} = \epsilon T_l$ とする. このとき, PT を用いたアルゴリズムは以下ようになる.

PT-ABC

次のステップを繰り返す.

1. パラレルステップ: m 個の連鎖についてそれぞれ MCMC-ABC を適用する. このときの状態を $(\boldsymbol{\theta}_{(l)}, \mathbf{z}_{(l)})$ ($l = 1, \dots, m$) とする.
2. スワップステップ: 隣り合う 2 つの連鎖 l と $l+1$ を無作為に選び, 2 つの連鎖の状

態 $(\boldsymbol{\theta}_{(l)}, \mathbf{z}_{(l)})$ と $(\boldsymbol{\theta}_{(l+1)}, \mathbf{z}_{(l+1)})$ を確率

$$\min \left\{ 1, \frac{f_{\epsilon_{(l)}}(\mathbf{y}|\mathbf{z}_{(l+1)}, \boldsymbol{\theta}_{(l+1)})f_{\epsilon_{(l+1)}}(\mathbf{y}|\mathbf{z}_{(l)}, \boldsymbol{\theta}_{(l)})}{f_{\epsilon_{(l)}}(\mathbf{y}|\mathbf{z}_{(l)}, \boldsymbol{\theta}_{(l)})f_{\epsilon_{(l+1)}}(\mathbf{y}|\mathbf{z}_{(l+1)}, \boldsymbol{\theta}_{(l+1)})} \right\}$$

で交換する.

大きな $\epsilon_{(l)}$ の値を持つ連鎖においては \mathbf{y} と $\mathbf{z}_{(l)}^*$ の距離が遠くても, 提案された $(\boldsymbol{\theta}_{(l)}^*, \mathbf{z}_{(l)}^*)$ が採択されやすい. また, スワップステップによって隣り合う連鎖の状態が交換されるため, $\epsilon_{(l)}$ の小さい連鎖において事後密度の低い領域から抜け出すことができるようになり, 混合が改善されることが期待できる. しかし, PT-ABC では $l=1$ と $l=2$ の交換の採択確率を改善することが難しく (Kobayashi and Kozumi (2011)), アルゴリズムがうまく機能するためには連鎖の数 m と $\epsilon_{(l)}$ の設定に注意を払わなければならない.

数値例 (続き) PT-ABC が MCMC-ABC の問題をどのように改善するかを調べる. ここでは $m=3$ 個のマルコフ連鎖を用意し, 各連鎖について $\epsilon_{(l)} = 0.03 \times 1.5^{l-1}$, $l=1, \dots, 3$ のように設定した.

まず, 第 3.1 節と同様に 100 回の繰り返しの間に $l=1$ が初期値から動くことができた回数を調べる. 初期値は標準正規分布から毎回発生させ, すべての連鎖について同じものを用いた. 50 回アルゴリズムを走らせたところ, 初期値から動くことができた回数は 15 回であり, MCMC-ABC よりも初期値に対する敏感さが若干改善されていた.

次に, 各連鎖の初期値を $\theta_{(l)} = 0$ とし 1 万個のサンプルを得た. 連鎖 $l=1$ の採択確率は 36.6% と MCMC-ABC に対する改善が見られた. 図 6 は PT-ABC による標本経路, 自己相関関数, ヒストグラムを示している. PT-ABC による自己相関は MCMC-ABC に比べて速くゼロに近づいていき, より効率的にサンプリングを行うことができるということが分かる.

3.2.4 複数提案アルゴリズム

状態空間をより効率的に探索するためには, 提案分布から候補を 1 つだけ発生させるよりも, 一度に複数の候補を提案し, その中から一番よいものを選ぶ方がより効果的であろう. こうしたアルゴリズムは, 以前より Liu *et al.* (2000), Qi and Liu (2001), Pandolfi *et al.* (2010) などによって開発されているが, Kobayashi and Kozumi (2011) はこれらのアルゴリズムを一般化した generalized multiple-point Metropolis (GMPM) アルゴリズムを提案している. 以下に GMPM を ABC の枠組みにおいて適用した場合を紹介する.

GMPM-ABC

現在の状態を $(\boldsymbol{\theta}, \mathbf{z})$ とし, 次のステップを繰り返し行う.

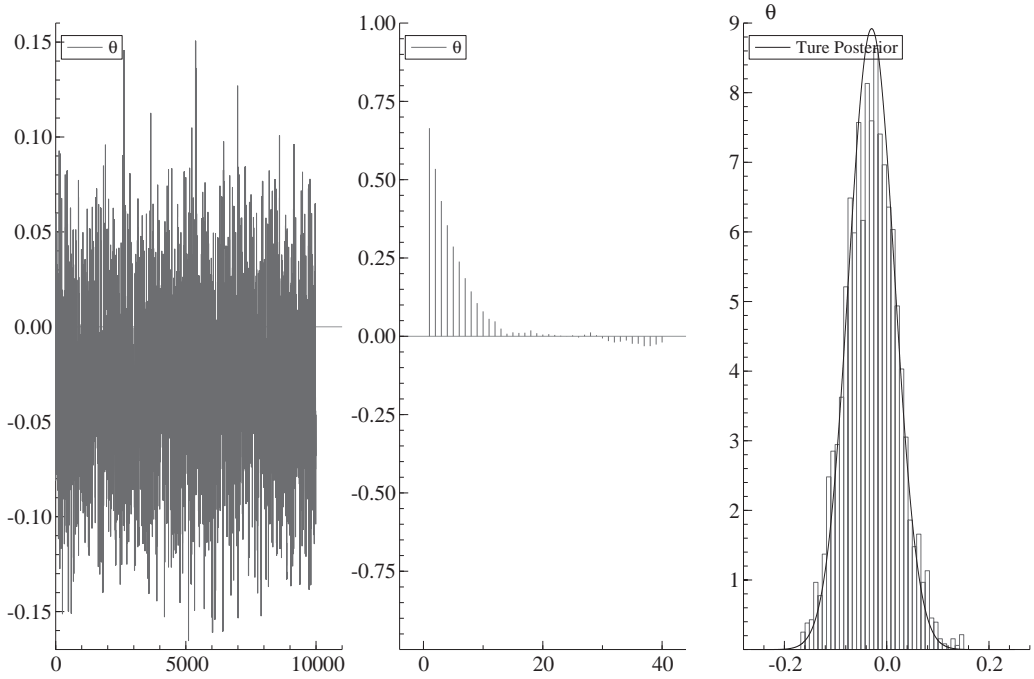


図6 PT-ABCによる標本経路(左), 自己相関関数(中央), ヒストグラム(右).

1. 密度関数が

$$q(\boldsymbol{\theta}_{[1:k]}^* | \boldsymbol{\theta}) = q(\boldsymbol{\theta}_1^* | \boldsymbol{\theta}) q(\boldsymbol{\theta}_2^* | \boldsymbol{\theta}, \boldsymbol{\theta}_1^*) \times \cdots \times q(\boldsymbol{\theta}_k^* | \boldsymbol{\theta}, \boldsymbol{\theta}_{[1:k-1]}^*) \quad (3.1)$$

によって与えられる結合提案分布から $\boldsymbol{\theta}_j^*$ ($j = 1, \dots, k$) を発生させる. さらに, \mathbf{z}_j^* ($j = 1, \dots, k$) を $f(\mathbf{z}_j^* | \boldsymbol{\theta}_j^*)$ から発生させる. ここで $\mathbf{x}_{[1:j]} = (\mathbf{x}_1, \dots, \mathbf{x}_j)$ である.

2. $\tilde{w}((\boldsymbol{\theta}_{[j:1]}^*, \mathbf{z}_{[j:1]}^*), (\boldsymbol{\theta}, \mathbf{z}))$ ($j = 1, \dots, n$) を計算する. ここで, \tilde{w} は正の関数を表し, また $\mathbf{x}_{[j:1]} = (\mathbf{x}_j, \dots, \mathbf{x}_1)$ である..

3. $\{(\boldsymbol{\theta}_1^*, \mathbf{z}_1^*), \dots, (\boldsymbol{\theta}_k^*, \mathbf{z}_k^*)\}$ の中から確率

$$p_l^* = \frac{\tilde{w}((\boldsymbol{\theta}_{[l:1]}^*, \mathbf{z}_{[l:1]}^*), (\boldsymbol{\theta}, \mathbf{z}))}{\sum_{j=1}^k \tilde{w}((\boldsymbol{\theta}_{[j:1]}^*, \mathbf{z}_{[j:1]}^*), (\boldsymbol{\theta}, \mathbf{z}))}$$

で $(\boldsymbol{\theta}_l^*, \mathbf{z}_l^*)$ を選び, $(\boldsymbol{\theta}^*, \mathbf{z}^*) = (\boldsymbol{\theta}_l^*, \mathbf{z}_l^*)$ とする.

4. 参照セット $\{(\boldsymbol{\theta}_1^{**}, \mathbf{z}_1^{**}), (\boldsymbol{\theta}_2^{**}, \mathbf{z}_2^{**}), \dots, (\boldsymbol{\theta}_k^{**}, \mathbf{z}_k^{**})\}$ を構築する. ここで, $j = 1, \dots, l-1$ について $\boldsymbol{\theta}_j^{**} = \boldsymbol{\theta}_{l-j}^*$, $\mathbf{z}_j^{**} = \mathbf{z}_{l-j}^*$ とし, また $\boldsymbol{\theta}_l^{**} = \boldsymbol{\theta}$, $\mathbf{z}_l^{**} = \mathbf{z}$ とする. $j = l+1, \dots, k$ については $\boldsymbol{\theta}_j^{**}$ を $q(\boldsymbol{\theta}_j^{**} | \boldsymbol{\theta}^*, \boldsymbol{\theta}_{[1:j-1]}^{**})$ から発生させ, \mathbf{z}_j^{**} を $f(\mathbf{z}_j^{**} | \boldsymbol{\theta}_j^{**})$ から発生させる.

5. $\tilde{w}((\boldsymbol{\theta}_{[j:1]}^{**}, \mathbf{z}_{[j:1]}^{**}), (\boldsymbol{\theta}^*, \mathbf{z}^*))$ ($j = 1, \dots, k$) を計算し,

$$p_l^{**} = \frac{\tilde{w}((\boldsymbol{\theta}_{[l:1]}^{**}, \mathbf{z}_{[l:1]}^{**}), (\boldsymbol{\theta}^*, \mathbf{z}^*))}{\sum_{j=1}^k \tilde{w}((\boldsymbol{\theta}_{[l:1]}^{**}, \mathbf{z}_{[l:1]}^{**}), (\boldsymbol{\theta}^*, \mathbf{z}^*))}$$

を計算する.

6. $(\boldsymbol{\theta}_k^*, \mathbf{z}_k^*)$ を確率

$$\min \left\{ 1, \frac{g_\epsilon(\mathbf{y}|\mathbf{z}^*, \boldsymbol{\theta}^*)\pi(\boldsymbol{\theta}^*)q(\boldsymbol{\theta}_{[1:l]}^{**}|\boldsymbol{\theta}^*)p_l^{**}}{g_\epsilon(\mathbf{y}|\mathbf{z}^{**}, \boldsymbol{\theta}^{**})\pi(\boldsymbol{\theta}^{**})q(\boldsymbol{\theta}_{[1:l]}^*|\boldsymbol{\theta}^{**})p_l^*} \right\} \quad (3.2)$$

で採択する.

Kobayashi and Kozumi (2011) は, このアルゴリズムによって生成されるマルコフ連鎖が詳細釣り合い条件を満たすことを示している.

このアルゴリズムは, それまでに提案されたパラメータの値に依存して新しいパラメータの値を提案することにより, 現在の点から遠い点を提案することが可能である. よって, MCMC-ABC よりも事後密度が低い領域から抜け出すことが容易になり, また効率性も改善されることが期待される. 提案分布については, 例えばランダムウォーク連鎖

$$q(\boldsymbol{\theta}_{[1:k]}^*|\boldsymbol{\theta}) = q(\boldsymbol{\theta}_1^*|\boldsymbol{\theta})q(\boldsymbol{\theta}_2^*|\boldsymbol{\theta}_1^*) \times \dots \times q(\boldsymbol{\theta}_k^*|\boldsymbol{\theta}_{k-1}^*) \quad (3.3)$$

を採用することができる. また, 提案される k 個の点の一部を現在の値の周りに独立に提案することも可能である.

正の関数 \tilde{w} の選択は任意であるが, ABC では $\rho(\mathbf{y}, \mathbf{z})$ の値が小さくなるような \mathbf{z} が望ましいので, 例えば

$$\tilde{w}((\boldsymbol{\theta}_{[j:1]}^*, \mathbf{z}_{[j:1]}^*), (\boldsymbol{\theta}, \mathbf{z})) = \exp\{-c\rho(S(\mathbf{y}), S(\mathbf{z}_j^*))\} \quad (3.4)$$

といったものを用いることが考えられる. ここで, $c (> 0)$ は分析者によって設定する定数であり, アルゴリズムの効率性や採択確率を高めるように設定される.

数値例 (続き) 正規分布のモデルを GMPM-ABC の場合についても考える. ここでは, 標準偏差が 0.1 の正規分布を用いたランダムウォーク連鎖 (3.3) によって $k = 5$ 個の点を提案する GMPM-ABC アルゴリズムを用いることにする. ウェイト関数は (3.4) を使い, c の値は 100 とした.

まず, GMPM-ABC が初期値の問題について MCMC-ABC のに対してどのように改善するかを調べる. アルゴリズムを 50 回走らせ, θ の初期値は毎回標準正規分布から発生させた. 100 回の繰り返しの間に GMPM-ABC アルゴリズムが初期値から動くことができた回数は 29 回であり, MCMC-ABC に対して大きな改善が見られた.

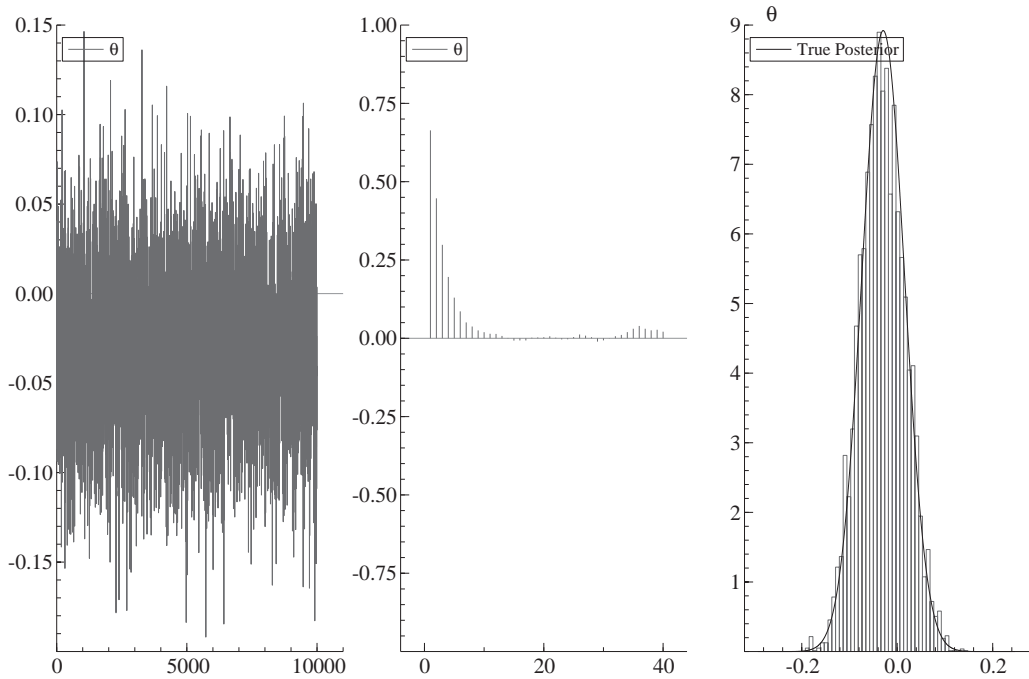


図7 GMPM-ABCによる標本経路(左), 自己相関関数(中央), ヒストグラム(右).

次に初期値を $\theta = 0$ として GMPM-ABC を走らせ, 1 万個のサンプルを得た, 採択確率は 45.0% であり, 採択確率においても MCMC-ABC に対しての改善が見られた. 図7は GMPM-ABC による標本経路, 自己相関関数およびヒストグラムを示している. GMPM-ABC の自己相関は PT-ABC よりも速く消え, より効率的なアルゴリズムとなっていることが分かる.

3.3 逐次モンテカルロ法

3.3.1 SMC-ABC

ABC アルゴリズムの発展のもうひとつの流れとして, 逐次モンテカルロ法 (以下, SMC 法) に基づいたものが挙げられる. SMC 法は目標の分布 f からのサンプリングを行うために, サンプリングを行う分布 $f_1, f_2, \dots, f_T = f$ の列を考える. 初期分布 f_1 からのサンプリングは容易に行えるものとする. SMC 法は粒子 $X_i^{(t)}$ ($i = 1, \dots, N, t = 1, \dots, T$) に対し, マルコフ連鎖による推移と f_t ($t = 1, \dots, T$) に対するウェイト付けを繰り返し行うことによって, $\{X_1^{(T)}, \dots, X_N^{(T)}\}$ を f からのサンプルとして得ることができる.

尤度関数が得られない場合には, SMC 法は MCMC-ABC と同様に拡張事後分布 (2.1) からのサンプリングを行う. 第2節で述べたように, 大きな ϵ の値のもとでは提案された値が拡張事後分布からのサンプルとして採択されやすい. よって SMC 法は, 第 3.2.1 節のように $\epsilon_1 > \epsilon_2 > \dots > \epsilon_T = \epsilon$ の値を設定し $\epsilon_1 \rightarrow \epsilon_2 \dots$ と値を小さくしていくことで, 目標

の分布である (2.1) にサンプリングを行う分布を近づけていく. SMC 法に基づいた ABC アルゴリズムの適用は, 特に複雑なモデルに対して有用であるとされている.

Sisson *et al.* (2007) は, Del Moral *et al.* (2006) の SMC 法を尤度関数が得られない場合に適用できるように拡張し, ABC の枠組みの中ではじめて逐次型のアルゴリズムを提案した. しかし, Beaumont *et al.* (2009) は Sisson *et al.* (2007) の方法では事後分布の近似にバイアスが生じることを指摘し, 集団モンテカルロ法 (Cappé *et al.* (2004)) に基づくアルゴリズムを提案した. その後, Sisson *et al.* (2009) により当初のアルゴリズムの修正が行われた. 次に示す Sisson *et al.* (2009) のアルゴリズムと Beaumont *et al.* (2009) のアルゴリズムとは基本的に同一である.

SMC-ABC

θ の初期分布を $f_1(\theta)$, $\epsilon_1 > \epsilon_2 > \dots > \epsilon_T$, 粒子を $X_i^{(t)} = (\theta_i^{(t)}, \mathbf{z}_i^{(t)})$ ($i = 1, \dots, N, t = 1, \dots, T$) とし, 次のステップを繰り返し行う.

1. $i = 1, \dots, N$ に対して

(a) $t = 1$ のとき, $\theta^{(1)}$ を $f_1(\theta)$ から発生させ, ウェイトを $w_i^{(1)} = 1/N$ とする.

$t > 1$ のとき, 粒子 $\{X_1^{(t-1)}, \dots, X_N^{(t-1)}\}$ からウェイト $\{w_1^{(t-1)}, \dots, w_N^{(t-1)}\}$ に基づいて $X_k^{(t)}$ をサンプリングする.

(b) θ^* を $q_t(\theta^* | \theta_k^{(t)})$ から発生させる.

(c) \mathbf{z}^* を $f(\mathbf{z}^* | \theta^*)$ から発生させる. もし $\rho(S(\mathbf{y}), S(\mathbf{z}^*)) \geq \epsilon_t$ ならば (a) に戻る.

(d) $\theta_i^{(t)} = \theta^*$ とし,

$$w_i^{(t)} \propto \begin{cases} \pi(\theta_i^{(t)})/f_1(\theta_i^{(t)}), & t = 1 \text{ のとき} \\ \pi(\theta_i^{(t)})/\sum_{j=1}^N w_j^{(t-1)} q_t(\theta_i^{(t)} | \theta_j^{(t-1)}), & t > 1 \text{ のとき} \end{cases} \quad (3.5)$$

とする.

2. ウェイトが $\sum_{i=1}^N w_i^{(t)} = 1$ となるように基準化する.

3. もし $ESS = \left[\sum_{i=1}^N (w_i^{(t)})^2 \right]^{-1} < E$ ならばウェイトに基づいて粒子をリサンプリングし, すべての粒子のウェイトを $1/N$ にする. ここで, E ($0 \leq E \leq 1$) は分析者が指定する定数である.

4. $t \leftarrow t + 1$ とする.

アルゴリズムを走らせた後の粒子 $\{X_1^{(T)}, \dots, X_N^{(T)}\}$ は事後分布からのサンプルであるので, ある関数 h の事後分布に関する期待値は

$$\sum_{i=1}^N h(X_i^{(T)}) w_i^{(T)} \quad (3.6)$$

によって推定することができる。

Beaumont *et al.* (2009) は提案分布に現在の θ_i の値を中心とした正規分布を使用している。また、適応的に提案分布を調整していくために、その分散には $t-1$ 期の粒子の分散の 2 倍を用いることを提案している。

第 3.1 節で述べたように、MCMC-ABC においては、悪い初期値や事後密度の低い領域にある点から次の状態へ映ることが難しかった。一方で SMC-ABC では、そのような領域にある粒子のウェイトはゼロとなり、サンプルとしての代表性を失うことになる。そして、リサンプリングステップにおいてウェイトが正の粒子に置き換わることにより、事後密度の低い領域にある粒子を取り除くことができる。このように、事後分布からのサンプルとしての候補をたくさん提案しておいてその中からよいものだけを残しておくという点で、第 3.2.4 節で紹介した GMPM アルゴリズムによる方法は SMC-ABC の考え方と近いところがある。また、SMC-ABC によって得られた粒子は事後分布からの独立なサンプルであるので、MCMC-ABC でのようなサンプルにおける高い自己相関の問題は存在しない。

SMC-ABC では ϵ_t ($t = 1, \dots, T$) を事前に設定する必要があったが、第 3.2.1 節で述べたようにその選び方には注意が必要である。また、粒子のウェイトを評価するとき、すべての粒子について $q_t(\theta^{(t)} | \theta_j^{(t-1)})$ を評価しなければならない。これによって計算量が $O(N^2)$ となってしまう非効率的である。次節では ϵ_t を適応的に決定し、計算量が $O(N)$ である SMC アルゴリズムを取り上げる。

3.3.2 適応的 SMC-ABC

Del Moral *et al.* (2011) や Drovandi and Pettitt (2011) は、 ϵ_t をアルゴリズムの中で適応的に決定するような方法を提案している。

まず、Del Moral *et al.* (2011) はウェイトが正である粒子の割合

$$PA(\{w_i^{(t)}\}, \epsilon_t) = \frac{1}{N} \sum_{i=1}^N I(w_i^{(t)} > 0)$$

をコントロールすることに着目し、

$$PA(\{w_i^{(t)}\}, \epsilon_t) = \alpha PA(\{w_i^{(t-1)}\}, \epsilon_{t-1}) \quad (3.7)$$

を満たすような ϵ_t を 2 分法によって決定する。ここで、 α ($0 < \alpha < 1$) は分析者によって指定される値である。もし 1 に近い α を用いれば、 ϵ_t が徐々に目標の ϵ に近づいていくので計算時間が増えることになるが、 N が小さいときにはアルゴリズムの精度を改善することができる。一方で、Drovandi and Pettitt (2011) では、単純に $\{\rho(S(\mathbf{y}), S(\mathbf{z}_i^{(t-1)}))\}$ ($i = 1, \dots, N$) の α 分位点を ϵ_t として選んでいる。

最後にこれらのアルゴリズムは次のようにまとめられる。

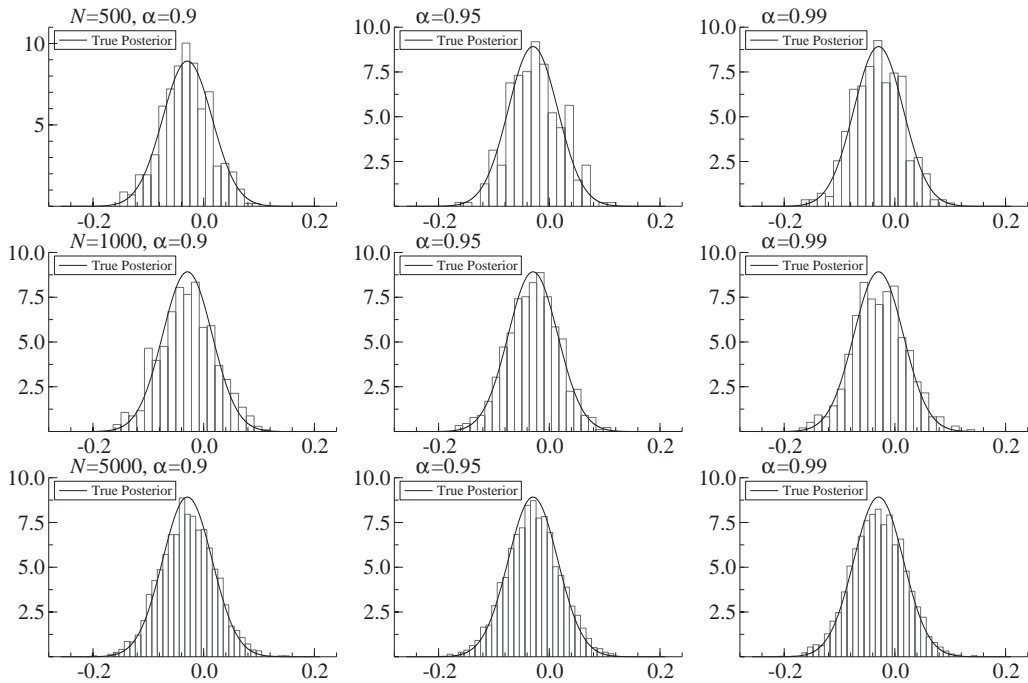


図8 Adaptive SMC-ABC による事後分布の近似.

Adaptive SMC-ABC

次のステップを繰り返し行う.

1. もし $\epsilon_t = \epsilon$ ならばアルゴリズムを終了する.

そうでないなら $PA(\{w_i^{(t)}\}, \epsilon_t) = \alpha PA(\{w_i^{(t-1)}\}, \epsilon_{t-1})$ を ϵ_t について解くか, $\{\rho(S(\mathbf{y}), S(\mathbf{z}_i^{(t-1)}))\}$ の α 分位点によって ϵ_t を決定する.

2. もし $ESS < E$ ならば粒子をリサンプリングし, $w_i^{(t)} = 1/N$ ($i = 1, \dots, N$) とする.

3. $i = 1, \dots, N$ について

(a) $\theta_i^{(t)}$ を $q_t(\theta_i^{(t)} | \theta_i^{(t-1)})$ から発生させる.

(b) $\mathbf{z}_i^{(t)}$ を $f(\mathbf{z}_i^{(t)} | \theta_i^{(t)})$ から発生させる.

(c) ウェイトを

$$w_i^{(t)} \propto w_i^{(t-1)} \frac{I(\rho(\mathbf{y}, \mathbf{z}_i^{(t)}) < \epsilon_t)}{I(\rho(\mathbf{y}, \mathbf{z}_i^{(t-1)}) < \epsilon_{t-1})} \tag{3.8}$$

とする.

4. $t \leftarrow t + 1$ とする.

Adaptive SMC-ABC のウェイト (3.8) を評価するには $t-1$ 期と t 期の ABC のウェイト関数 (第 2.2 節を参照) を評価するだけでよい。これにより, Adaptive SMC-ABC の計算量は $O(N)$ となっていることが分かる。

数値例 (続き) Del Moral *et al.* (2011) の Adaptive SMC-ABC を用いて第 2.2 節の正規分布のモデルを考えることにする。アルゴリズムの目標の ϵ_t の値は 0.01 に設定した。粒子の数 N は 500, 1000, 5000 の 3 つの場合を考え, (3.7) の α の値はそれぞれの場合において 0.9, 0.95, 0.99 を用いることにした。粒子の初期値は事前分布から互いに独立に発生させた。

図 8 は各 N および α の値の下で Adaptive SMC-ABC によるヒストグラムを示している。図中の実線は真の事後密度を表す。図より, 同じ α の下で粒子の数を増やすことで近似の精度が改善されることが分かる。また, $N = 500$ と 1000 の場合のように, 粒子の数が少ないときでも 1 に近い α の値を用いることで, サンプルングを行う分布がよりゆっくり目標の分布に近づいていくので, 近似の精度が改善されることが分かる。

4. おわりに

本稿では, モデルの尤度関数が得られない場合にパラメータの事後分布を求める際に用いられる ABC という方法についての紹介を行った。そして MCMC 法や SMC 法を基にした様々なバージョンの ABC アルゴリズムを紹介した。また, 単純な数値例を用いてそれぞれの ABC の手法がどのように働くかも示した。

ABC は未だに多くの課題をかかえているといえる。例えば, 第 2.5 節で紹介したように, モデル選択が特別なモデルを除いてベイズファクターに基づいて行うことができないことが挙げられ, 代替的なモデルの診断方法を考える必要がある。本稿で取り上げた ABC アルゴリズムは基本的にどのようなモデルに対しても適用可能であるが, より効率的なサンプルングを行うためにはモデルにより即したアルゴリズムの開発を行う必要がある。モデルが単純であるときは Neal (2010) の効率的な棄却サンプルングを用いることができるが, その適用はパラメータの次元が低い場合に限られてしまう。よって, 複雑なモデルに対してもより効率的に ABC を適用するために, モデルに即した MCMC 法および SMC 法の開発が必要とされる。また, Peters *et al.* (2010b) や Peters *et al.* (2010d) のような通常の MCMC 法と ABC とのハイブリッド型のサンプルング方法も興味深く, 今後 ABC アルゴリズムの様々な方向への発展が期待される。

ABC による経済・経営データの分析は第 1 節で挙げたように現在まだ限られているが, 経済学や経営学は ABC の重要な応用先であると考えられる。まず, 経済時系列データ分析における Peters *et al.* (2010b) や Jasra *et al.* (2010) による研究は, 時系列データ分析に

においてよく用いられる VAR モデルや状態空間モデルの推定に ABC を応用しており、ABC の有用性について重要な結果を残している。一方で、マイクロデータの分析に対しても ABC の応用が進んでいくことと思われる。以前、ABC の枠組みで回帰モデルを考える場合には、要約統計量が使用できない上に $\rho(\mathbf{y}, \mathbf{z})$ が小さくなるように \mathbf{z} を発生させるのが難しかったため、パラメータの推定が非常に困難であった。この問題は McVinish (2010) のアルゴリズムを用いることで克服され、ABC によって回帰モデルを推定することが可能になった。また、マイクロデータ分析では個人の異質性を考慮した階層モデルが用いられるが、尤度関数が求まらない場合の階層モデルの推定には Bazin *et al.* (2010) の提案している方法が応用可能であると考えられる。このようにごく最近になって、従来より用いられてきた重要なモデルへの ABC の応用を可能にするような研究結果が増えてきていると言える。今後、経済学や経営学の領域において、ABC が尤度関数の得られないモデルについて推測を行うためのアプローチとして有用なものとなることが期待される。

謝辞

本稿に対して本誌レフリーより大変貴重で丁寧なコメントを頂いた。ここに記して感謝の意を申し上げる。本研究は日本学術振興会特別研究員奨励費 (DC1) の助成を受けたものである。本稿の計算結果は Ox バージョン 6.20 (Doornik (2007)) によって得られたものである。

参考文献

- Allingham, D., King, R. A. R. and Mengersen, K. L. (2009). Bayesian estimation of quantile distributions, *Stat. Comput.*, **19**, 189–201.
- Bazin, E., Dawson, K. J. and Beaumont, M. A. (2010). Likelihood-free inference of population structure and local adaptation in a Bayesian hierarchical model, *Genetics*, **185**, 587–602.
- Beaumont, M. A., Zhang, W. and Balding, D. J. (2002). Approximate Bayesian computation in population genetics, *Genetics*, **162**, 2025–2035.
- Beaumont, M. A., Cornuet, J.-M., Marin, J. and Robert, C. P. (2009). Adaptive approximate Bayesian computation, *Biometrika*, **96**, 983–990.
- Blum, M. G. B. (2010). Approximate Bayesian Computation: a nonparametric perspective, *J. Am. Stat. Assoc.*, **105**, 1178–1187.
- Blum, M. G. B. and François, O. (2010). Non-linear regression models for Approximate Bayesian Computation, *Stat. Comput.*, **20**, 63–73.
- Bortot, P., Coles, S. G. and Sisson, S. A. (2007). Inference for stereological extremes, *J. Am. Stat. Assoc.*, **102**, 84–92.
- Cappé, O., Guillin, A., Marin, J. M. and Robert, C. P. (2004). Population Monte Carlo, *J. Comput. Graph. Stat.*, **13**, 907–929.
- Cornuet, J., Santos, F., Beaumont, M. A., Robert, C. P., Marin, J., Balding, D. J., Guillemaud, T. and Estoup, A. (2008). Inferring population history with DIY ABC: a user-friendly approach to approximate Bayesian computation, *Bioinformatics*, **24**, 2713–2719.
- Cornuet, J., Ravigné, V. and Estoup, A. (2010). Inference on population history and model checking using DNA sequence and microsatellite data with the software DIYABC (v1.0), *Bioinformatics*, DOI:10.1186/1471-2105-11-401.

- Csilléry, K., Blum, M. G. B., Gaggiotti, O. E. and François, O. (2010). Approximate Bayesian Computation (ABC) in practice, *Trends Ecol. Evol.*, **25**, 410–418.
- Del Moral, P., Doucet, A. and Jasra, A. (2006). Sequential Monte Carlo samplers, *J. R. Stat. Soc. Ser. B*, **68**, 411–436.
- Del Moral, P., Doucet, A. and Jasra, A. (2011). An adaptive sequential Monte Carlo method for Approximate Bayesian Computation, *Stat. Comput.*, DOI: 10.1007/s11222-011-9271-y.
- Didelot, X., Everitt, R. G., Johansen, A. M. and Lawson, D. J. (2011). Likelihood-free estimation of model evidence, *Bayesian Analysis*, **6**, 49–76.
- Doornik, J. A. (2007). *Object-Oriented Matrix Programming Using Ox* (3rd ed.), Timberlake Consultants Press and Oxford, London.
- Doucet, A., Freitas N. and Gordon, N. (2001). *Sequential Monte Carlo Methods in Practice*, New York, Springer.
- Drovandi, C. C. and Pettitt, A. N. (2011). Estimation of parameters for macroparasite population evolution using approximate Bayesian computation, *Biometrics*, **67**, 225–223.
- Drovandi, C. C., Pettitt, A. N. and Faddy, M. J. (2011). Approximate Bayesian computation using indirect inference, *J. R. Stat. Soc. Ser. C*, **60**, 317–337.
- Fearnhead, P. and Prangle, D. (2011). Constructing summary statistics for approximate Bayesian computation: semi-automatic ABC. arXiv:1004.1112v2.
- Gamerman, D. and Lopes, H. S. (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference* (2nd ed.), Chapman & Hall/CRC, Boca Raton.
- Geyer, C. J. (1991). Markov chain Monte Carlo maximum likelihood, *Computing Science and Statistics: The 23rd Symposium on the Inference Interface Foundation* (eds. E. Keramigas), Fairfax, pp. 156–163.
- Grelaud, A., Robert, C. P., Rodlphé, F. and Marin, J. (2009). ABC likelihood free methods for model choice in Gibbs random fields, *Bayesian Analysis*, **4**, 317–336.
- Han, C. and Carlin, B. P. (2001). Markov chain Monte Carlo methods for computing Bayes factors: a comparative review, *J. Am. Stat. Assoc.*, **96**, 1122–1132.
- Jasra, A., Singh, S. S. and Martin, J. S. (2010). Filtering via approximate Bayesian computation, *Stat. Comput.*, DOI: 10.1007/s11222-010-9185-0.
- Joyce, P. and Marjoram, P. (2008). Approximately sufficient statistics and Bayesian computation, *Statistical Applications in Genetics and Molecular Biology*, **7**, DOI: 10.2202/1544-6115.1389.
- Kobayashi, G. and Kozumi, H. (2011). Generalized multiple-point Metropolis algorithms for approximate Bayesian computation, Technical Report, Kobe University.
- Leuenberger, C. and Wegmann, D. (2010). Bayesian computation and model selection without likelihoods, *Genetics*, **184**, 243–252.
- Liu, J. S. (2001). *Monte Carlo Strategies in Scientific Computing*, Springer, New York.
- Liu, J. S., Liang, F. and Wong, W. H. (2000). The use of multiple-try method and local optimization in Metropolis sampling, *J. Am. Stat. Assoc.*, **94**, 121–134.
- Marjoram, P., Molitor, J., Plagnol, V. and Tavaré, S. (2003). Markov chain Monte Carlo without likelihoods, *Proc. Natl. Acad. Sci.*, **100**, 15324–15328.
- Marin, J., Pudlo, P. and Robert, C. P. (2011). Approximate Bayesian computational methods, arXiv: 1101.0955v2.
- McVinish, R. (2010). Improving ABC for quantile distributions, *Stat. Comput.*, DOI: 10.1007/s11222-010-9209-9.
- Neal, P. (2010). Efficient likelihood-free Bayesian computation for household epidemics, *Stat. Comput.*, DOI: 10.1007/s11222-010-9216-x.
- Nunes, M. A. and Balding, D. J. (2010). On optimal selection of summary statistics for approximate Bayesian computation, *Statistical Applications in Genetics and Molecular Biology*, **9**, DOI: 10.2202/1544-6115.1576.
- Pandolfi, S., Bartolucci, F. and Friel, N. (2010). A generalization of the Multiple-try Metropolis algorithm for Bayesian estimation and model selection, in Proceedings of the 13th International Conference on Artificial Intelligence and Statistics, pp. 581–588.

- Peters, G. W. and Sisson, S. A. (2006). Bayesian inference, Monte Carlo sampling and operational risk, *Journal of Operational Risk*, **1**, 27–50.
- Peters, G. W., Sisson, S. A. and Fan, Y. (2010a). Likelihood-free Bayesian inference for α -stable models, *Comput. Stat. Data Anal.*, DOI:10.1016/j.csda.2010.10.004.
- Peters, G. W., Kannan, B. B., Lasscock, B. and Mellen, C. (2010b). Bayesian cointegrated vector autoregression models incorporating α -stable noise for inter-day price movements via approximate Bayesian computation, arXiv:1008.0149v1.
- Peters, G. W., Wüthrich, M. V. and Shevchenko, P. V. (2010c). Chain ladder method: Bayesian bootstrap versus classical bootstrap, *Insur., Math. Econ.*, **4**, 36–51.
- Perets, G. W., Nevat, I., Sisson, S. A., Fan, Y. and Yuan, J. (2010d). Bayesian symbol detection in wireless relay networks via likelihood-free inference, *IEEE Transactions on Signal Processing*, **58**, 5206–5218.
- Pitchard, J. K., Seielstad, M. T., Perez-Lezaun, A. and Feldman, M. W. (1999). Population growth of human Y chromosomes: A study of Y chromosome microsatellites, *Mol. Biol. Evol.*, **16**, 1791–1798.
- Qi, Z. S. and Liu, J. S. (2001). Multipoint Metropolis method with application to hybrid Monte Carlo, *J. Comput. Phys.*, **172**, 827–840.
- Ratmann, O., Andrieu, C., Wiuf, C. and Richardson, S. (2009). Model criticism based on likelihood-free inference, with an application to protein network evolution, *Proc. Natl. Acad. Sci.*, **106**, 10576–10581.
- Robert, C. P., Cornuet, J., Marin, J. and Pillai, N. (2011). Lack of confidence in ABC model choice, arXiv:1101.5909v4.
- Sisson, S. A. and Fan, Y. (2011). Likelihood-free Markov chain Monte Carlo, *Handbook of Markov Chain Monte Carlo* (eds. S. P. Brooks, A. Gelman, G. Jones and X. L. Meng), Chapman & Hall/CRC, Boca Raton.
- Sisson, S. A., Fan, Y. and Tanaka, M. M. (2007). Sequential Monte Carlo without likelihoods, *Proc. Natl. Acad. Sci.*, **104**, 1760–1765.
- Sisson, S. A., Fan, Y. and Tanaka, M. M. (2009). Correction for Sisson *et al.* sequential Monte Carlo without likelihoods, *Proc. Natl. Acad. Sci.*, **106**, 16889.
- Tavaré, S., Balding, D. J., Griffiths, R. C. and Donnelly, P. (1997). Inferring coalescence times from DNA sequence data, *Genetics*, **145**, 505–518.
- Toni, T. and Stumpf, M. P. H. (2010). Simulation-based model selection for dynamical systems in systems and population biology, *Bioinformatics*, **26**, 104–110.
- Toni, T., Welch, D., Strelkowa, N., Ipsen, A. and Stumpf, P. H. (2009). Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems, *Journal of the Royal Society Interface*, **6**, 187–202.
- Wegmann, D., Leuenberger, C. and Excoffier, L. (2009). Efficient Approximate Bayesian Computation coupled with Markov chain Monte Carlo without likelihood, *Genetics*, **182**, 1207–1218.
- Wilkinson, R. D. (2008). Approximate Bayesian computation (ABC) gives exact results under the assumption of model error, arXiv:0811.3355v1.