

マルコフ連鎖モンテカルロ法による 高次元積分

お茶の水女子大学大学院 小 川 奈 美 子*
統計数理研究所 江 口 真 透

(1997 年 3 月 受付)

1. はじめに

歴史的にマルコフ連鎖モンテカルロ法 (MCMC 法) は統計物理の計算法として開発された。この方法は Metropolis et al. (1953) のメトロポリス法に始まり, Hastings (1970) によって一般化され, 統計の分野に導入された。この一般化されたメトロポリス法は Metropolis-Hastings 法と呼ばれる。それから Geman and Geman (1984) により Gibbs sampler と呼ばれる MCMC 法が画像復元問題に応用され, 注目を浴びることになる。さらに Gelfand and Smith (1990) がベイズ推論への応用の一般化を行ない, Gibbs sampler の様々な応用例を挙げている。今現在, MCMC 法はベイズ推論や尤度推定などに広く利用されている。

MCMC 法を利用する最大の利点は正規化されていない密度関数からのサンプリングが可能である, ということである。統計物理やベイズ推論に現れる分布は正規化定数が解析的に計算できないことが多い (伊庭 (1995))。これが, MCMC 法が広く利用されるゆえんである。多くの場合は期待値計算に使われており, この場合正規化定数の知識は不要である。しかし Thompson (1994) が述べているように, ベイズ推論におけるモデル比較のためのベイズ因子あるいは ABIC の計算, 欠測データを含む尤度の計算, そして混合分布を用いたインポートランス・サンプリングにおいて正規化定数が必要となる。そのような例として Geyer and Thompson (1992) は尤度曲線を近似し, 最尤推定量を求めるアルゴリズムを考えており, そのアルゴリズムの中で必要となる正規化定数の計算をインポートランス・サンプリングによって求めている。また Thompson (1994) ではインポートランス・サンプリングによる期待値計算問題を扱っている。推定の重要な鍵となるインポートランス・サンプリング関数として混合分布の密度関数を採用することを考えている。その際必要となる各々の分布の正規化定数を逆オートロジスティック回帰と呼ばれる方法を用いて求めている。

本研究では正規化定数を求めることを目的とする。この問題は $f(\mathbf{x}) = f(x_1, \dots, x_p)$ の p 次元区間 $[a, b]^p$ での定積分を求めることに帰着される。以後, $f(\mathbf{x})$ は下に有界であると仮定し, 一般性を失うことなく正の関数とする。定積分は

$$Z = \int_a^b \cdots \int_a^b f(\mathbf{x}) dx_1 \cdots dx_p$$

と表される。数値解析においては p 次元積分区間をメッシュに切って積分近似をするが, これは高次元になると計算が困難になる。実際には四次元以上になると積分近似は難しくなると言われている。そこで乱数を利用するモンテカルロ積分法がいくつか考え出されている。その多

* 現 富士総合研究所 研究開発 第 2 部: 〒101 東京都千代田区神田錦町 2-3.

くは統計物理の理論とテクニックによって導入されるものであり、MCMC法が有効に使われている。統計物理での自由エネルギーの推定法は一般の関数 $f(\mathbf{x})$ の積分にも利用できる。Ogata and Tanemura (1981, 1984, 1989)によって統計問題に導入された正規化定数を求める方法は自由エネルギー推定法の一つである熱力学的積分法に属する。本研究ではOgata-Tanemuraの方法について考察し、より良い積分近似値を求めるための方法とアルゴリズムを提案する。

この論文は次のような章立てになっている。2章では統計物理と絡めてモンテカルロ積分の紹介をし、3章では2章で紹介されるOgata-Tanemuraの方法について考察し、より良い積分近似を与えるための方法とアルゴリズムを提案したい。提案するアルゴリズムを用いたシミュレーションを4章で行なう。5, 6章において考察と今後の課題について述べる。

2. モンテカルロ積分法

本章ではモンテカルロ法を用いた積分法を紹介する。ほとんどのモンテカルロ積分法は統計物理において計算手法として開発されたマルコフ連鎖に基づくMCMC法が使われている(Neal (1993))。

以下で紹介する方法は統計物理における自由エネルギーの計算法として提案されるものである。自由エネルギーとは分配関数の逆数の対数であり、分配関数は系のエネルギー関数の高次元積分によって求められる。したがって一般の関数にある変換を施した量を系のエネルギーと見なすことにより、自由エネルギーの推定法はその関数の高次元積分を求める方法として応用できる。

ここでは統計物理における自由エネルギーとその推定法を簡単に紹介し、それらの方法を一般の関数の高次元積分法として説明する。

2.1 自由エネルギー推定

統計物理では気体や磁性体のように分子や原子などの特定の構成単位が多数集まった物理系を扱う。系の全ての実現可能な状態 \mathbf{x} はある一定のエネルギー $E(\mathbf{x})$ を持つ。孤立系すなわち外部とのエネルギーの交換がない系ではエネルギー E は固定される。この仮定は“固定されたエネルギー E を持つ状態は全て等しい出現確率を持ち、異なるエネルギーを持つ状態は起こり得ない”ことを意味する。この分布をマイクロ正準分布という。

次に系の占める空間とその中の粒子数は一定であるが、外部との間に相互作用がありエネルギーの交換が行なわれ、系と外部とを含めた大きな系がマイクロ正準分布にある場合を考える。温度が T で与えられたとき、この系が状態 \mathbf{x} である確率は、

$$P(\mathbf{x}) = \frac{1}{Z} \exp\left(-\frac{E(\mathbf{x})}{T}\right)$$

であり、この分布はギブス分布として知られている。ここで $Z = \int \exp(-E(\mathbf{x})/T) d\mathbf{x}$ は規格化因子で、分配関数と呼ばれる。これは定数のように見えるが、温度 T あるいは $E(\mathbf{x})$ の中に暗黙のうちに含まれている変数が変化すると考えれば関数である(荒木(1994))。

系の自由エネルギーは

$$F = -T \log Z$$

で定義される。統計物理の理論と統計推論とを関連付けるために、統計推論問題における確率変数の組を架空上の物理系の実現可能な状態と見なす。するとこれらの変数上のどんな確率分布も Z, T に対するエネルギー関数

$$E(\mathbf{x}) = -T \log P(\mathbf{x}) - T \log Z$$

に関してギブス分布であると考えることができる。エネルギーが無限大の状態は出現確率が0であると考えれば良い。ここで、高次元被積分関数 $f(\mathbf{x})$ をその積分値 Z で割ったものは密度関数であるから、温度 $T=1$ としてエネルギー関数を

$$(2.1) \quad E(\mathbf{x}) = -\log f(\mathbf{x})$$

とすれば、定積分 Z はこのエネルギー関数を持つ系の自由エネルギーと見なすことができる。

自由エネルギーを計算するには積分をギブス分布の期待値に書き直し統計量で推定するという方法が使われる。自由エネルギー推定法は

1. 同じ状態空間を持つ異なるエネルギー関数 $E_0(\mathbf{x})$, $E_1(\mathbf{x})$ を持つ二つの系,
2. 二つの異なる温度 T_0 , T_1 での同じ系,
3. エネルギー関数, 温度共に異なる二つの系,

などの自由エネルギーの差あるいはその指数を計算する方法である。エネルギー関数が異なる二つの系を考える場合、もし二つの系のうち一つの系の正確な自由エネルギーが既知であるならば $\log Z_1 - \log Z_0$ を知ることでもう一方の系の自由エネルギーを容易に計算することができる。この考え方は高次元積分に応用可能である。すなわち定積分 Z_0 が既知である関数 f_0 を採り、(2.1) をエネルギー関数と見なし自由エネルギー推定法により $\log Z - \log Z_0$ を求めることで $f(\mathbf{x})$ の定積分 Z を計算することができる。

2.2 インポートランス・サンプリング

最も簡単なモンテカルロ積分法はインポートランス・サンプリングを用いることである。インポートランス・サンプリングはサンプリングが比較的簡単な分布からサンプルを取り出し、そのサンプルに荷重を付けることでサンプリングが困難な分布からのサンプルを得る方法である。

インポートランス・サンプリング関数 $f_0(\mathbf{x})$ として正規化されている密度関数あるいは積分区間での積分値 Z_0 が既知である正規化されていない密度関数 $f(\mathbf{x})$ を採る。積分を期待値で書き表すと

$$\begin{aligned} \frac{Z}{Z_0} &= \int_a^b \frac{f(\mathbf{x})}{f_0(\mathbf{x})} \frac{f_0(\mathbf{x})}{Z_0} d\mathbf{x} \\ &= E_{f_0} \left[\frac{f(\mathbf{X})}{f_0(\mathbf{X})} \right] \end{aligned}$$

が成り立つ (正規化されている場合は $Z_0=1$)。これを利用して f_0 に従うサンプルを取り出し、期待値を標本平均で推定する。もし f_0 として正規化されていない分布を用いるのであれば、MCMC 法によりサンプル生成をする。インポートランス・サンプリング法は実装が簡単であるため広く利用されている。

例えば f_0 として p 次元一様分布を使うと、積分区間から独立な N 個の一様乱数 $\{\mathbf{x}_i\}$ ($i=1, \dots, N$) を取り、大数の法則から

$$\frac{1}{N} \sum_{i=1}^N f(\mathbf{x}_i) \rightarrow E[f(\mathbf{X})] \quad (N \rightarrow \infty)$$

であることを使い、

$$\hat{Z} = \frac{(b-a)^p}{N} \sum_{i=1}^N f(\mathbf{x}_i)$$

と推定することができる。インポートランス・サンプリング関数 f_0 からのサンプリングが容易である、または f_0 のサンプルがすでに利用できるとき、インポートランス・サンプリング法は有効

な方法である (Tierney (1994)). しかしこの \hat{Z} は次元 p が大きいとき, あるいは被積分関数 $f(\mathbf{x})$ の変動が大きいときは Z の推定値として実用的ではないことが知られている (Ogata (1989)). またこの推定法は f_0 の選択に大きく依存し, f に似通っている関数でなければ正しい結果は得られない (Gelfand and Smith (1990)).

2.3 熱力学的積分法

熱力学的積分法は二つの物理系の間にパスを取り, そのパスに沿ったいくつかの中間系で $\log Z$ の微分を求める方法で, 物理系の自由エネルギーを実験上で推定する方法に対応する.

まず p 次元区間 $[a, b]^p$ での定積分 Z_0 が既知である関数 $f_0(\mathbf{x})$ を見つける. この f_0 と f を 1 パラメータ $\sigma \in [0, 1]$ で結び $f(\mathbf{x}, \sigma)$ と定義する. ここで $f(\mathbf{x}, 0) = f_0(\mathbf{x})$, $f(\mathbf{x}, 1) = f(\mathbf{x})$ となるように結ぶ. このとき

$$P(\mathbf{x}, \sigma) = \frac{f(\mathbf{x}, \sigma)}{Z(\sigma)},$$

$$Z(\sigma) = \int_a^b \cdots \int_a^b f(\mathbf{x}, \sigma) d\mathbf{x}$$

とおけば, $\{P(\mathbf{x}, \sigma)\}$ は確率密度の 1 パラメータ族となる.

$f(\mathbf{x}, \sigma)$ が連続微分可能で正則条件を満たすとする. ただし正則条件とは次の三つの条件のことをさす.

1. ほとんどいたるところの $\mathbf{x} \in [a, b]^p$ で $f(\mathbf{x}) > 0$ である.
2. ほとんどいたるところの $\mathbf{x} \in [a, b]^p$ で $\partial f(\mathbf{x}, \sigma) / \partial \sigma$ が存在する.
3. 任意の $\sigma \in [0, 1]$ に対して $|\partial f(\mathbf{x}, \sigma) / \partial \sigma| \leq A(\mathbf{x})$ となる $[a, b]^p$ で積分可能な $A(\mathbf{x})$ が存在する.

すると $Z(\sigma)$ の対数微分は

$$\begin{aligned} \frac{\partial}{\partial \sigma} \log Z(\sigma) &= \frac{1}{Z(\sigma)} \frac{\partial}{\partial \sigma} \int f(\mathbf{x}, \sigma) d\mathbf{x} \\ &= \frac{1}{Z(\sigma)} \int \frac{\partial}{\partial \sigma} f(\mathbf{x}, \sigma) d\mathbf{x} \\ &= \int \frac{\partial}{\partial \sigma} \log f(\mathbf{x}, \sigma) \cdot P(\mathbf{x}, \sigma) d\mathbf{x} \\ &= E_\sigma \left[\frac{\partial}{\partial \sigma} \log f(\mathbf{x}, \sigma) \right] \end{aligned}$$

と書き直すことができる. ここで E_σ は $P(\mathbf{x}, \sigma)$ に関する期待値を表す. $\psi(\sigma) = (\partial / \partial \sigma) \log Z(\sigma)$, $\psi(\mathbf{x}, \sigma) = (\partial / \partial \sigma) \log f(\mathbf{x}, \sigma)$ とすれば,

$$(2.2) \quad \psi(\sigma) = E_\sigma[\psi(\mathbf{X}, \sigma)]$$

である. したがって,

$$(2.3) \quad \log Z - \log Z_0 = \int_0^1 \psi(\sigma) d\sigma$$

が成立する. このように p 次元積分を 1 次元積分に帰着して $\log Z$ の推定値を求めることができる.

この一次元積分には $\psi(\sigma)$ が必要となるが, $\psi(\sigma)$ は正確には計算できない. そこで確率密度関数 $P(\mathbf{x}, \sigma)$ に従うサンプルをマルコフ連鎖モンテカルロ法 (MCMC 法) によるサンプリングを用いて取り出し, $\psi(\sigma)$ を推定することを考える. MCMC 法の特徴は正規化定数の知識なしに正

規格化されていない密度関数からのサンプリングを可能にする点である。すなわち $Z(\sigma)$ の知識なしに関数 $f(\mathbf{x}, \sigma)$ だけから $P(\mathbf{x}, \sigma)$ に従うサンプルが生成できる。こうして得られたサンプルから ψ を推定する。(2.3)の σ に関する一次元積分は積分区間 $[0, 1]$ を J 個に分割し、分割点 $\{\sigma_j\}$ ($0 = \sigma_0 < \sigma_1 < \dots < \sigma_{J-1} < \sigma_J = 1$) 上で ψ を推定して近似する。これより $\log Z$ の推定値 $\log \hat{Z}$ を得ることができる。

$\log Z$ を計算するためには、各分割点 σ_j で MCMC サンプリングによるデータが必要になる。しかし独立に各々のシミュレーションを実行するのは無駄である。隣合う分割点 σ_j と σ_{j+1} は近い系であるからパスの一端、すなわち $\sigma = 0$ または 1 から始めて σ_{j+1} のシミュレーションの初期状態として σ_j のシミュレーションの最終状態を使い次々ともう一端へ向かってシミュレートするのが良い。

これから分かるように熱力学的積分法は p 次元積分の対数を新しいパラメータを一つ加えて $(p+1)$ 次元で表し、そのうちの p 次元積分を p 個の確率ベクトルの期待値に表現して統計量で推定し、残りの加えたパラメータの一次元積分に帰着する方法である。 p 次元積分すなわち (2.2) の期待値の推定と、(2.3) の一次元積分の近似法の異なる方法が提案されている。

スロー・グロース法

スロー・グロース法は Bash et al. (1987) で提案されている方法である。(2.3)は一次元積分区間 $[0, 1]$ を十分大きな J 個に分割して、

$$\begin{aligned} \log Z - \log Z_0 &= \sum_{j=0}^{J-1} \int_{\sigma_j}^{\sigma_{j+1}} \mathbf{E}_{\sigma} \left[\frac{\partial}{\partial \sigma} \log f(\mathbf{X}, \sigma) \right] d\sigma \\ &\cong \sum_{j=0}^{J-1} (\sigma_{j+1} - \sigma_j) \mathbf{E}_{\sigma} \left[\frac{\partial}{\partial \sigma} \log f(\mathbf{X}, \sigma) \right]_{\sigma=\sigma_j} \\ &= \sum_{j=0}^{J-1} \mathbf{E}_{\sigma} \left[(\sigma_{j+1} - \sigma_j) \frac{\partial}{\partial \sigma} \log f(\mathbf{X}, \sigma) \right]_{\sigma=\sigma_j} \\ &\cong \sum_{j=0}^{J-1} \mathbf{E}_{\sigma} [\log f(\mathbf{X}, \sigma_{j+1}) - \log f(\mathbf{X}, \sigma_j)]_{\sigma=\sigma_j}. \end{aligned}$$

と近似できる。この方法は $[0, 1]$ の分割点を非常に大きく取るが、各分割点 σ_j で唯一つの MCMC サンプル $\mathbf{x}^{(\sigma_j)}$ を使って

$$\log \hat{Z} - \log Z_0 = \sum_{j=0}^{J-1} \{ \log f(\mathbf{x}^{(\sigma_j)}, \sigma_{j+1}) - \log f(\mathbf{x}^{(\sigma_j)}, \sigma_j) \}$$

と $\log Z$ を推定する。この方法は隣合う系 σ_j と σ_{j+1} が非常に近い系で σ_j での一回の推移の状態が平衡であるように、十分に分点の間隔を詰めなければならない。

Ogata-Tanemura の方法

この方法は統計物理とは別に空間点過程の統計問題における正規化定数を求める方法として提案された。後に Ogata (1989) はこれを一般の関数の高次元積分法として定式化している。それは期待値を標本平均で推定し、一次元積分を台形公式で近似する熱力学的積分法に相当する。

(2.2)の期待値は σ に対する MCMC サンプル $\{\mathbf{x}_i^{(\sigma)}\}$ ($i=1, \dots, I$) を使って、

$$(2.4) \quad \hat{\psi}(\sigma) = \frac{1}{I} \sum_{i=1}^I \psi(\mathbf{x}_i^{(\sigma)}, \sigma)$$

と推定する。これは $\{\mathbf{x}_i^{(\sigma)}\}$ の一貫性、不偏性をもつ $\psi(\sigma)$ の推定量である。さらに (2.3) の σ に関する積分に対して台形公式近似を使って近似する。 σ の積分区間 $[0, 1]$ を J 個の分割点 $\{\sigma_j\}$ において MCMC サンプルを生成させ、 $\log Z$ の推定値 $\log \hat{Z}$ を

$$S = \log \hat{Z} - \log Z_0 = \frac{1}{2} \sum_{j=1}^J \{ \hat{\psi}(\sigma_j) + \hat{\psi}(\sigma_{j-1}) \} (\sigma_j - \sigma_{j-1})$$

$$= \left(\frac{\sigma_1 - \sigma_0}{2} \right) \hat{\psi}(\sigma_0) + \sum_{j=1}^{J-1} \left(\frac{\sigma_{j+1} - \sigma_{j-1}}{2} \right) \hat{\psi}(\sigma_j) + \left(\frac{\sigma_J - \sigma_{J-1}}{2} \right) \hat{\psi}(\sigma_J)$$

で与える。

パスの選択

この方法に於ける $f_0(\mathbf{x})$ と $f(\mathbf{x})$ の結び方は重要である。結び方により積分の精度が異なる。Ogata (1989, 1995) では

1. $f(\mathbf{x}, \sigma) = f(\sigma \mathbf{x})$ (scaling parameter)
ただし、原点 $\mathbf{0} = (0, \dots, 0)$ が積分範囲に含まれているとき、 $f_0(\mathbf{x}) = f(\mathbf{0})$.
2. $f(\mathbf{x}, \sigma) = \{f(\mathbf{x})\}^\sigma \{f_0(\mathbf{x})\}^{1-\sigma}$ (reciprocal temperature)
3. $f(\mathbf{x}, \sigma) = f(\mathbf{x}_0 + \sigma(\mathbf{x} - \mathbf{x}_0))$
4. $f(\mathbf{x}, \sigma) = g(\mathbf{x}) \frac{f(\sigma \mathbf{x})}{g(\sigma \mathbf{x})}$

などの結び方を提案している。その他の結び方として、

$$f(\mathbf{x}, \sigma) = \sigma f(\mathbf{x}) + (1 - \sigma) f_0(\mathbf{x}) \quad (\text{mixture})$$

が考えられる。しかし正規分布の密度関数の正規化定数を求める数値実験では、被積分関数 $\psi(\sigma)$ がうまく推定できず、最終的に得られる結果は実用的なものではなかった。

また結ぶ関数 $f_0(\mathbf{x})$ の選び方でも得られる積分推定値は変わってくると思われる。しかし積分区間で積分値が分かっている関数という条件から f_0 として選ぶことができる関数は限られてしまう。

次章で Ogata-Tanemura の方法の考察と改善を試みる。本研究では結ぶ関数 $f_0(\mathbf{x})$ と結び方を固定して、等間隔の分点ではなく適切な分点を選ぶことにより、より良い $\log Z$ の推定値を得ることを考える。

2.4 Metropolis-Hastings 法

MCMC 法には Metropolis et al. (1953) によるメトロポリス法, Geman and Geman (1984) による Gibbs sampler がある。更に、Metropolis-Hastings 法は Hastings (1970) がメトロポリス法を統計分野に導入し一般化した MCMC 法を指す。

MCMC 法はマルコフ連鎖で表されるダイナミクスを利用して、与えられた確率分布からのサンプリングを可能にするアルゴリズムである。注目すべきことは、この確率分布は正規化されている必要がない、ということである。統計物理やベイズ統計に現れる分布は、正規化定数が解析的に計算できない場合があるためこの方法は広く利用されている。本研究では、正規化されていない確率密度関数 $P(\mathbf{x}, \sigma)$ を持つ分布からサンプルを得るためにこの方法を利用する。

この節では MCMC 法の一つである Metropolis-Hastings 法について説明する。

推移確率行列の作り方

$P = \{p_{ij}\}$ を状態 $0, 1, \dots, S$ をとるマルコフ連鎖の推移確率行列とする。時刻 t での確率過程の状態を $X(t)$ で表すとき、状態 i から状態 j への推移確率は

$$\text{Prob}(X(t+1)=j|X(t)=i)=p_{ij}, \quad (i, j=0, 1, \dots, S)$$

である。

平衡分布として $\pi=\{\pi_i\}$ を持つようなマルコフ連鎖を作るためには、推移確率 p_{ij} が詳細釣り合い条件

$$\pi_i p_{ij} = \pi_j p_{ji}, \quad (\forall i, j)$$

を満たすように定めればよい。これは全ての j に対し、

$$\sum_i \pi_i p_{ij} = \pi_j$$

となり π は P を持つマルコフ連鎖の平衡分布となることが分かる。

Hastings (1970) は P として

$$\begin{aligned} p_{ij} &= q_{ij} a_{ij}, & (i \neq j) \\ p_{ii} &= 1 - \sum_{j \neq i} p_{ij} \end{aligned}$$

を満たしているものを考えた。ここで $Q=\{q_{ij}\}$ は状態 $0, 1, \dots, S$ 上でのマルコフ連鎖の推移確率行列である。 a_{ij} は

$$a_{ij} = \frac{s_{ij}}{1 + \frac{\pi_i q_{ij}}{\pi_j q_{ji}}}$$

と定め、 s_{ij} は全ての i, j に対して、 $0 \leq a_{ij} \leq 1$ となるように選ばれた対称な関数 ($s_{ij}=s_{ji}$) とする。このように決められた p_{ij} は詳細釣り合い条件を満たす。よってこの推移確率に従って作り出されるマルコフ連鎖は平衡分布として π をもつことになる。この方法は Metropolis-Hastings 法と呼ばれる。アルゴリズムは任意の初期状態から出発して時刻 t における状態を $X(t)=i$ とするとき、以下の手順を繰り返すことで得られる。

1. 推移確率行列 Q の第 i 行で与えられる確率分布 $\{q_{ij}\}$ を用いて状態 j を選ぶ。
2. 確率 a_{ij} で採択し $X(t+1)=j$ とする、確率 $(1-a_{ij})$ で棄却し $X(t+1)=i$ とする。

平衡分布 P が d 次元分布であり、確率過程の状態 $X(t)=(X_1(t), \dots, X_d(t))$ が d 次元ベクトルである場合、次のような一単位時間の推移法を採ると一次元の場合と同様に簡単に実装される。

- (i) $X(t)$ の d 個の座標からランダムに選ばれた一つの座標だけを変える。
- (ii) 固定された一つの座標だけを変える。
- (iii) $X(t)$ の全ての座標を変える。

メトロポリス法

メトロポリス法はもともと、相互作用をしている分子から構成される物質の特性値を計算するために Metropolis et al. (1953) により導入された。このアルゴリズムは統計物理で広く利用されている。このメトロポリス法は Metropolis-Hastings 法の特別なケースである。 s_{ij} と q_{ij} を

$$s_{ij} = \begin{cases} 1 + \frac{\pi_i q_{ij}}{\pi_j q_{ji}} & \left(\frac{\pi_j q_{ji}}{\pi_i q_{ij}} \geq 1 \right), \\ 1 + \frac{\pi_j q_{ji}}{\pi_i q_{ij}} & \left(\frac{\pi_j q_{ji}}{\pi_i q_{ij}} \leq 1 \right) \end{cases}$$

$$q_{ij} = q_{ji}$$

と選ぶとき、メトロポリス法と言い、採択確率は

$$\alpha_{ij} = \min \left\{ \frac{\pi_j}{\pi_i}, 1 \right\}$$

となる。

3. 最適分割のための方法

3.1 パラメータ化

本研究では前章で紹介した熱力学的積分法の一つである Ogata-Tanemura の方法を改善することを考える。Ogata-Tanemura の方法によって推定される値は f と f_0 を結ぶパラメータの選択に依存することに注目する。実際、滑らかで単調増加である変換

$$t: \sigma \rightarrow \tau, \quad 0 \leq \tau \leq 1, \quad t(0)=0, \quad t(1)=1,$$

に対して、パラメータ化を

$$f^*(\mathbf{x}, \tau) = f(\mathbf{x}, t^{-1}(\tau)) = f(\mathbf{x}, \sigma)$$

とする(ここで $\sigma = t^{-1}(\tau)$)。このパラメータ化に従って、前章の Ogata-Tanemura の方法を適用する。このとき

$$Z^*(\tau) = \int_a^b \cdots \int_a^b f^*(\mathbf{x}, \tau) d\mathbf{x} = Z(\sigma)$$

$$\psi^*(\tau) = \frac{\partial}{\partial \tau} \log Z^*(\tau) = \frac{\psi(\sigma)}{t'(\sigma)}$$

$$\psi^*(\mathbf{x}, \tau) = \frac{\partial}{\partial \tau} \log f^*(\mathbf{x}, \tau) = \frac{\psi(\mathbf{x}, \sigma)}{t'(\sigma)}$$

となる。これらを用いて $\psi^*(\tau)$ の推定値 $\hat{\psi}^*(\tau)$ は

$$\hat{\psi}^*(\tau) = \frac{1}{I} \sum_{i=1}^I \psi^*(\mathbf{x}_i^{(\tau)}, \tau) = \frac{1}{I} \sum_{i=1}^I \frac{\psi(\mathbf{x}_i^{(\sigma)}, \sigma)}{t'(\sigma)} = \frac{\hat{\psi}(\sigma)}{t'(\sigma)}$$

で与えられる。(2.3)は

$$\log Z - \log Z_0 = \int_0^1 \psi^*(\tau) d\tau$$

とも表現され、これを等分割の刻み $\{\tau_j = j/J\}$ で近似し、 Z の推定値を \hat{Z}^* とすると

$$S^* = \log \hat{Z}^* - \log Z_0 = \frac{1}{2J} \sum_{j=1}^J \left\{ \hat{\psi}^*\left(\frac{j}{J}\right) + \hat{\psi}^*\left(\frac{j-1}{J}\right) \right\}$$

が得られる。これは変換 $\tau_j = t(\sigma_j)$ を使って非等分割の分点で

$$S^* = \frac{1}{2} \sum_{j=1}^J \left\{ \frac{\hat{\psi}(\sigma_j)}{t'(\sigma_j)} + \frac{\hat{\psi}(\sigma_{j-1})}{t'(\sigma_{j-1})} \right\} \{t(\sigma_j) - t(\sigma_{j-1})\}$$

$$\cong \frac{1}{2} \sum_{j=1}^J \{ \hat{\psi}(\sigma_j) + \hat{\psi}(\sigma_{j-1}) \} (\sigma_j - \sigma_{j-1})$$

と近似される。このように $\log \hat{Z}^*$ は τ -パラメータで眺めると $[0, 1]$ 上の等分割に対して、 σ -パ

ラメータで眺めると非等分割 $0 = \sigma_0 < \sigma_1 < \dots < \sigma_{J-1} < \sigma_J = 1$ の分点に対して台形公式を実行した方法で求められた推定値であると近似的に見ることができる。刻み方のクラスを考えることは、この狭義の単調増加な変換 t 全体を考えることとほぼ一致する。したがって定積分を近似する最適な分割を見つけることは、置換積分で最適分割を施すような最適変換 t_{opt} を見つけることに帰着される。

3.2 最適変換

前章で導入された定積分の対数 $\log Z$ の推定値 $\log \hat{Z}^*$ に対して最適パラメータ $\tau = t(\sigma)$ を提案したい。良さの基準として推定された $\log Z$ の誤差分散 $\text{Var}(S)$ を考える。

$$\begin{aligned} \text{Var}(S) = & \left(\frac{\sigma_1 - \sigma_0}{2} \right)^2 \text{Var}(\hat{\psi}(\sigma_0)) + \sum_{j=1}^{J-1} \left(\frac{\sigma_{j+1} - \sigma_{j-1}}{2} \right)^2 \text{Var}(\hat{\psi}(\sigma_j)) \\ & + \left(\frac{\sigma_J - \sigma_{J-1}}{2} \right)^2 \text{Var}(\hat{\psi}(\sigma_J)). \end{aligned}$$

十分大きな J を仮定し、漸近評価することにより次の定理が得られる。

定理 1. 変換 t によるパラメータ化で求められる統計量 S^* の誤差分散 $\text{Var}(S^*)$ は漸近的に、

$$\text{Var}(S^*) = \frac{1}{2J} \int_0^1 \frac{v(\sigma)}{t'(\sigma)} d\sigma + o\left(\frac{1}{J}\right).$$

となる。ここで $v(\sigma) = \text{Var}(\hat{\psi}(\sigma))$ 。

証明. パラメータ τ の J -等分割を仮定して漸近評価する。 $\hat{\psi}^*(\tau_i)$ と $\hat{\psi}^*(\tau_j)$ ($i \neq j$) の独立性から

$$\begin{aligned} \text{Var}(S^*) &= \frac{1}{4J^2} \text{Var} \left[\sum_{j=1}^J \left\{ \hat{\psi}^* \left(\frac{j}{J} \right) + \hat{\psi}^* \left(\frac{j-1}{J} \right) \right\} \right] \\ &= \frac{1}{4J^2} \sum_{j=1}^J \left\{ \text{Var} \left[\hat{\psi}^* \left(\frac{j}{J} \right) \right] + \text{Var} \left[\hat{\psi}^* \left(\frac{j-1}{J} \right) \right] \right\} \\ &\approx \frac{1}{2J} \int_0^1 v^*(\tau) d\tau. \end{aligned}$$

ここで $v^*(\tau) = \text{Var}(\hat{\psi}^*(\tau))$ 。前節と同様に変換 t を用いて σ -パラメータで書き直すと、

$$\text{Var}(S^*) \approx \frac{1}{2J} \int_0^1 \frac{v(\sigma)}{t'(\sigma)} d\sigma,$$

となる。□

$[0, 1]$ 上の先の条件を満たすもので微分同型全体を \mathcal{T} と表す。漸近評価において \mathcal{T} 上で定義された汎関数

$$\phi(t) = \int_0^1 \frac{v(\sigma)}{t'(\sigma)} d\sigma$$

の最小化問題はシュワルツの不等式を使って次の解答を得る。

定理 2. 変換 $t_{\text{opt}}: \sigma \rightarrow \tau$ を

$$(3.1) \quad t_{\text{opt}}(\sigma) = \frac{\int_0^\sigma \sqrt{v(s)} ds}{\int_0^1 \sqrt{v(s)} ds},$$

と定めるとき、

$$\phi(t_{\text{opt}}) = \min\{\phi(t), t \in \mathcal{T}\}$$

を得る.

証明. シュワルツの不等式より

$$\phi(t) = \int_0^1 \frac{v(\sigma)}{t'(\sigma)} d\sigma \int_0^1 t'(\sigma) d\sigma \geq \left(\int_0^1 \sqrt{v(\sigma)} d\sigma \right)^2$$

である. 等号は

$$\sqrt{\frac{v(\sigma)}{t'(\sigma)}} \propto \sqrt{t'(\sigma)}$$

すなわち

$$t'(\sigma) = c^{-1} \sqrt{v(\sigma)}$$

のとき成立し, このとき最小値を取る.

$$\int_0^1 t'(\sigma) d\sigma = t(1) - t(0) = 1$$

より

$$c = \int_0^1 \sqrt{v(\sigma)} d\sigma$$

と定まる. 故に

$$t_{\text{opt}}(\sigma) = \frac{\int_0^\sigma \sqrt{v(s)} ds}{\int_0^1 \sqrt{v(s)} ds}$$

となる.

Remark. これは以下のように変分問題として考えることが出来る. クラス \mathcal{T} 上の t_{opt} を通る滑らかな曲線族を C とする. すなわち C は,

$$\{t_\varepsilon : |\varepsilon| < 1, (t_\varepsilon)_{\varepsilon=0} = t_{\text{opt}}\}$$

と書ける. このとき ϕ が

$$\text{第1変分: } \left(\frac{\partial}{\partial \varepsilon} \phi(t_\varepsilon) \right)_{\varepsilon=0} = 0,$$

$$\text{第2変分: } \left(\frac{\partial^2}{\partial \varepsilon^2} \phi(t_\varepsilon) \right)_{\varepsilon=0} > 0$$

を満たすことを示せば良い (Kobayashi and Nomizu (1969)). 実際 $\phi(t_\varepsilon)$ を ε で微分すると,

$$\begin{aligned} \frac{\partial}{\partial \varepsilon} \phi(t_\varepsilon) &= - \int_0^1 \frac{1}{t_\varepsilon'^2} \frac{\partial}{\partial \varepsilon} t_\varepsilon' v(\sigma) d\sigma, \\ \frac{\partial^2}{\partial \varepsilon^2} \phi(t_\varepsilon) &= \int_0^1 \left[-\frac{1}{t_\varepsilon'^2} \frac{\partial^2}{\partial \varepsilon^2} t_\varepsilon' + \frac{2}{t_\varepsilon'^3} \left\{ \frac{\partial}{\partial \varepsilon} t_\varepsilon' \right\}^2 \right] v(\sigma) d\sigma. \end{aligned}$$

となる. ここで上の二式に $\varepsilon=0$ を代入すると, t_{opt} の定義から

$$\begin{aligned}
\left(\frac{\partial}{\partial \varepsilon} \phi(t_\varepsilon)\right)_{\varepsilon=0} &= -\int_0^1 \frac{v(\sigma)}{(\sqrt{v(\sigma)/c})^2} \left(\frac{\partial}{\partial \varepsilon} t'_\varepsilon\right)_{\varepsilon=0} d\sigma \\
&= -c^2 \int_0^1 \left(\frac{\partial}{\partial \varepsilon} t'_\varepsilon(\sigma)\right)_{\varepsilon=0} d\sigma \\
\left(\frac{\partial^2}{\partial \varepsilon^2} \phi(t_\varepsilon)\right)_{\varepsilon=0} &= \int_0^1 \left[-\frac{v(\sigma)}{(\sqrt{v(\sigma)/c})^2} \left(\frac{\partial^2}{\partial \varepsilon^2} t'_\varepsilon\right)_{\varepsilon=0} + \frac{2v(\sigma)}{(\sqrt{v(\sigma)/c})^2} t'_0 \left\{ \left(\frac{\partial}{\partial \varepsilon} t'_\varepsilon\right)_{\varepsilon=0} \right\}^2 \right] d\sigma \\
&= -c^2 \int_0^1 \left(\frac{\partial^2}{\partial \varepsilon^2} t'_\varepsilon(\sigma)\right)_{\varepsilon=0} d\sigma + 2c^2 \int_0^1 \left\{ \left(\frac{\partial}{\partial \varepsilon} \log t'_\varepsilon(\sigma)\right)_{\varepsilon=0} \right\}^2 t'_0(\sigma) d\sigma
\end{aligned}$$

となる。これは

$$\int_0^1 t'_\varepsilon(\sigma) d\sigma = t_\varepsilon(1) - t_\varepsilon(0) = 1$$

となるので $\int (\partial^k / \partial \varepsilon^k) t'_\varepsilon(\sigma) d\sigma = 0$ ($k=1, 2$) であり, t'_ε は σ を確率変数とする密度関数と見なすことができる。第二変分の第二項に含まれる積分は

$$\int_0^1 \left\{ \left(\frac{\partial}{\partial \varepsilon} \log t'_\varepsilon(\sigma)\right)_{\varepsilon=0} \right\}^2 t'_0(\sigma) d\sigma = E \left[\left\{ \left(\frac{\partial}{\partial \varepsilon} \log t'_\varepsilon(\sigma)\right)_{\varepsilon=0} \right\}^2 \right]_{\varepsilon=0}$$

と書き直すことができ, これは ε をパラメータと見なしたときのフィッシャー情報量であるから正となる。したがって第一変分, 第二変分が示された。□

3.3 分散の推定

定理 2 から

$$t_{\text{opt}}(\sigma_j) = \frac{\int_0^{\sigma_j} \sqrt{v(s)} ds}{\int_0^1 \sqrt{v(s)} ds} = \frac{j}{J} \quad (= \tau_j), \quad (j=0, 1, \dots, J)$$

となる非等分割点 $\{\sigma_j\}$ を用いてより良い推定量 $\log \hat{Z}^*$ を求めることができる。

この最適変換は $v(\sigma)$ に基づいているが, $v(\sigma)$ は未知であるのでサンプルから推定することを考える。今, $\{\mathbf{x}_i\}$ ($i=1, \dots, I$) を確率密度関数 $P(\mathbf{x}, \sigma)$ からの MCMC サンプルとすると $v(\sigma)$ は近似的に,

$$\begin{aligned}
v(\sigma) &= \text{Var}[\hat{\psi}(\sigma)] = \text{Var} \left[\frac{1}{I} \sum_{i=1}^I \psi(\mathbf{x}_i, \sigma) \right] \\
&= \frac{1}{I^2} \sum_{m=-I+1}^{I-1} \sum_{i=1}^{I-|m|} \text{Cov}[\psi(\mathbf{x}_i, \sigma), \psi(\mathbf{x}_{i+|m|}, \sigma)] \\
&\cong \frac{1}{I} \text{Var}[\psi(\mathbf{X}, \sigma)] \left\{ \sum_{m=-M_0}^{M_0} \left(\frac{I-|m|}{I} \right) \rho_\sigma(|m|) \right\}
\end{aligned}$$

となる。ここで $\rho_\sigma(m)$ は自己相関,

$$\rho_\sigma(m) = \frac{\text{Cov}[\psi(\mathbf{x}_i, \sigma), \psi(\mathbf{x}_{i+m}, \sigma)]}{\sqrt{\text{Var}[\psi(\mathbf{x}_i, \sigma)]} \sqrt{\text{Var}[\psi(\mathbf{x}_{i+m}, \sigma)]}}, \quad (i=1, \dots, I-m)$$

である。タイムラグ m が十分大きいとき, 自己相関 $\rho_\sigma(m)$ は無視できるとして適切な M_0 を設定し, 全ての $|m| > M_0$ に対して $\rho_\sigma(|m|) = 0$ であると仮定する。そこで $\text{Var}[\psi(\mathbf{X}, \sigma)]$ の推定値を標本分散

$$\widehat{\text{Var}}[\psi(\mathbf{X}, \sigma)] = \frac{1}{I} \sum_{i=1}^I \{\psi(\mathbf{x}_i, \sigma) - \hat{\psi}(\sigma)\}^2$$

とし, 同様に標本自己相関

$$\hat{\rho}_\sigma(m) = \frac{\frac{1}{I-m} \sum_{i=1}^{I-m} (\psi(\mathbf{x}_i, \sigma) - \hat{\psi}(\sigma))(\psi(\mathbf{x}_{i+m}, \sigma) - \hat{\psi}(\sigma))}{\sqrt{\frac{1}{I-m} \sum_{i=1}^{I-m} \{\psi(\mathbf{x}_i, \sigma) - \hat{\psi}(\sigma)\}^2} \sqrt{\frac{1}{I-m} \sum_{i=1}^{I-m} \{\psi(\mathbf{x}_{i+m}, \sigma) - \hat{\psi}(\sigma)\}^2}}$$

で $\rho_\sigma(m)$ を推定する. このようにして $v(\sigma)$ は,

$$(3.2) \quad \hat{v}(\sigma) = \frac{1}{I} \widehat{\text{Var}}[\psi(\mathbf{X}, \sigma)] \left\{ \sum_{m=-M_0}^{M_0} \left(\frac{I-|m|}{I} \right) \hat{\rho}_\sigma(|m|) \right\}$$

と推定される. ただし実際には $\hat{\rho}_\sigma(m)$ の時系列の挙動を見て, $\hat{\rho}_\sigma(m)$ がそこから先はほとんど 0 となるような時刻を M_0 と決めて分散 $v(\sigma)$ を推定する.

3.4 アルゴリズム

これらを利用して我々は二段階アルゴリズムを提案する. 第一段階では等分割点 $\{j/J\}$ で $\psi(j/J)$ と標準偏差 $\sqrt{v(j/J)}$ を推定し, 変換 t_{opt} を使って最適分割を求める. 第二段階では最適分割点 $\{\bar{\sigma}_k\}$ で $\hat{\psi}(\bar{\sigma}_k)$ を求め, 台形公式近似で $\log \hat{Z}^*$ を求める. 以下にアルゴリズムを示す.

- (i) p 次元区間 $[a, b]^p$ での定積分 Z_0 が既知である関数 $f_0(\mathbf{x})$ を取る.
- (ii) 関数 $f(\mathbf{x})$ と $f_0(\mathbf{x})$ を 1 パラメータで結び, $\sigma=0$ のとき $f_0(\mathbf{x})$, $\sigma=1$ のとき $f(\mathbf{x})$ となるように関数族 $f(\mathbf{x}, \sigma)$ を定義する. すなわち

$$f(\mathbf{x}, \sigma), \sigma \in [0, 1], f(\mathbf{x}, 0) = f_0(\mathbf{x}), f(\mathbf{x}, 1) = f(\mathbf{x}).$$

- (iii) 区間 $[0, 1]$ を J -等分割して, 各分割点 j/J ($j=0, 1, \dots, J$) で確率密度関数 $P(\mathbf{x}, j/J)$ に従うサンプル $\{\mathbf{x}^{(j/J)}\}$ ($i=1, \dots, I$) を MCMC 法を用いて生成する (生成法は次節を参照).
- (iv) 統計量 $\hat{\psi}(j/J)$ と $\hat{v}(j/J)$ をそれぞれ (2.4), (3.2) を用いて計算する.
- (v) t_{opt} を計算し, K 個の最適刻み点 $\{\sigma_k\}$ ($k=1, \dots, K$) を求める.

$$\begin{aligned} \bar{\sigma}_k &= t_{\text{opt}}^{-1} \left(\frac{k}{K} \right) = \frac{j(k)}{J} + \frac{\frac{k}{K} - \gamma_{j(k)}}{J(\gamma_{j(k)+1} - \gamma_{j(k)})}, \\ \gamma_j &= \frac{\hat{\beta}_1 + \dots + \hat{\beta}_j}{\hat{\beta}_1 + \dots + \hat{\beta}_{J-1} + \hat{\beta}_J}, \\ \beta_j &= \int_{(j-1)/J}^{j/J} \sqrt{\hat{v}(\sigma)} d\sigma, \quad \hat{\beta}_j: \beta_j \text{ の積分近似値}, \\ j(k) &= \max \left\{ j \mid \gamma_j < \frac{k}{K} \right\} \end{aligned}$$

となる.

実際には (図 1):

- (a) $\sqrt{\hat{v}(j/J)}$ ($j=1, \dots, J$) を描く.
 - (b) この $\sqrt{\hat{v}(j/J)}$ を正規化されていない密度関数と考え, その分布関数を描く.
 - (c) 図 b の τ -軸を K -等分割して $t_{\text{opt}}(k/K) = \bar{\sigma}_k$ と決める.
- すなわち分割 $[\bar{\sigma}_{k-1}, \bar{\sigma}_k]$ での $\sqrt{\hat{v}(\sigma)}$ の面積が等しくなるように分割点 $\{\bar{\sigma}_k\}$ を決める.

- (vi) 上で求められた最適変換による分点 $\bar{\sigma}_k$ で確率密度関数 $P(\mathbf{x}, \bar{\sigma}_k)$ に従うサンプル $\{\mathbf{x}^{(\bar{\sigma}_k)}\}$ をステップ (iii) と同様に MCMC 法で発生させる.

- (vii) (2.4) から統計量 $\hat{\psi}(\bar{\sigma}_k)$ を求める.

- (viii) J -等分割点の集合 $\{j/J\}$ と最適変換による K 個の分点の集合 $\{\bar{\sigma}_k\}$ を合わせた $(J+K)$ 個の分点の集合を $\{\bar{\sigma}_k^*: k=0, 1, \dots, J+K\}$ とおく.

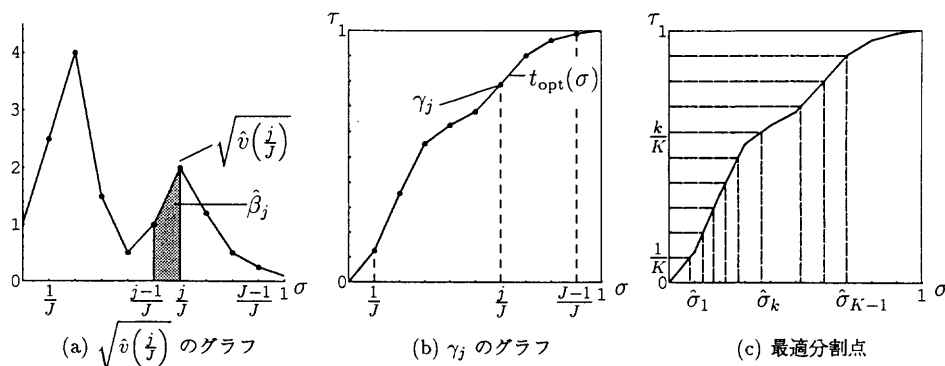


図1. 最適変換による分割.

この分点 $\hat{\sigma}_k^*$ における $\psi(\sigma)$ の推定値 $\{\hat{\psi}(\hat{\sigma}_k^*)\}$ の $[0, 1]$ 区間の面積を台形近似式(2.4)により求めて $\log \hat{Z}^*$ を推定する.

ここで提案した二段階アルゴリズムは各段階で一次元積分を行っており, 二通りの分割で $\hat{\psi}$ を計算することになる. 第一段階では最適変換 t_{opt} に必要とされる $\sqrt{\hat{v}(\sigma)}$ の積分値, すなわち $\sum_{j=0}^J \hat{\beta}_j$ を求めており, その際等分割点 $\{j/J\}$ での $\hat{\psi}(j/J)$ が計算される. 第二段階では ψ の積分近似のため, 最適分割点 $\{\hat{\sigma}_k\}$ において $\hat{\psi}(\hat{\sigma}_k)$ が計算される. ここで ψ の積分近似に対して,

1. 等分割での $\hat{\psi}(j/J)$ の情報を捨てて, 最適分割での $\hat{\psi}(\hat{\sigma}_k)$ のみを使う,

あるいは

2. 等分割と最適分割での $\hat{\psi}(j/J)$ と $\hat{\psi}(\hat{\sigma}_k)$ の両方の情報を使う

かが問題となる. 残念ながら現段階において理論的にどちらが良いのか結論を出すことはできていない. 分割点をたくさん取るとき, 各点で ψ がうまく推定できない場合には各点での誤差が積み重なり $\log Z$ の推定誤差が大きくなることも考えられるが, 分割点が多いほど一次元積分の近似が良くなることは明らかである. また試みた全てのシミュレーションでケース1とケース2による積分近似値を求めて検討してみたが, 多くの場合において分割点が多くなるケース2の方が良いという結果が出ている. そこでこのアルゴリズムのステップ(viii)では二つの分割点の集合を合わせて新しい分割点 $\{\hat{\sigma}_k^*\}$ による一次元積分近似を提案する.

次節で実際にプログラムで使ったMCMC法を示す.

3.5 MCMC サンプリング

この節では, 2.1節で述べたギブス分布 $P(\mathbf{x})$ を平衡分布として持つマルコフ連鎖の実現値を発生させるためのプログラムを説明する. 簡単のためにパラメータ σ をはずして表記し, 以下に実際のプログラム上での手続きを示す.

時刻 t における状態を p 次元ベクトル

$$\mathbf{x}(t) = (x_1(t), \dots, x_a(t), \dots, x_p(t))$$

で表すとする. メトロポリス法において変化させる座標はランダムに選ぶかわりに $1, 2, \dots, p$ と順番に取っても詳細釣り合い条件は満たされる. したがって各遷移において一つの座標を変え, 変える座標はランダムではなく順番に変えるという方法を使うことにする. 以下に示す手順を繰

り返し行なうことで、平衡分布として $P(\mathbf{x})$ をもつマルコフ連鎖を作り出すことができる。

(1) $\mathbf{y}=(y_1, \dots, y_p)=\mathbf{x}(t)$, $\alpha=1$ とする。

1. 次のステップの候補 $\mathbf{y}'=(y'_1, \dots, y'_\alpha, \dots, y'_p)$ を取る。この \mathbf{y}' は α 座標 y'_α が y_α からわずかに動いた値を取り、その他の座標は \mathbf{y} の座標と変わらないとする。すなわち

$$y'_\alpha = y_\alpha + \delta(1-2\xi),$$

$$y'_s = y_s \quad (s \neq \alpha).$$

ξ は $[0, 1]$ の一様乱数である。したがって y'_α は $[y_\alpha - \delta, y_\alpha + \delta]$ の一様乱数である。 δ は α 座標の値を変化させるときに許される最大変化量であり、制御パラメータと呼ばれる。 $\delta > 0$ であるとする。

2. $P(\mathbf{y})$ と $P(\mathbf{y}')$ の比を計算し、

$$r = \min \left\{ 1, \frac{P(\mathbf{y}')}{P(\mathbf{y})} \right\}$$

とする。

3. \mathbf{y} を確率 r で $\mathbf{y}:=\mathbf{y}'$, 確率 $1-r$ で $\mathbf{y}:=\mathbf{y}$ と更新する。
4. $\alpha < p$ ならば $\alpha=\alpha+1$ として 1 へ戻り次の座標を変化させる。 $\alpha=p$ ならば一単位時間が経過したとして (2) へ進む。

(2) 時刻 $t+1$ の状態を $\mathbf{x}(t+1)=\mathbf{y}$ とする。

ここで計算される量は

$$\frac{P(\mathbf{y}')}{P(\mathbf{y})} = \frac{f(\mathbf{y}')}{f(\mathbf{y})}$$

であり、正規化定数 Z はキャンセルされる。従って正規化定数の知識なしに正規化されていない密度関数から、その分布を平衡分布として持つようなマルコフ連鎖を作り出すことができる。このようにして得られたマルコフ連鎖 $\{\mathbf{x}(t)\}$ をサンプルとして利用する。始めのサンプルは状態が平衡になっていないと思われるので使わず捨てる。制御パラメータ δ は原則としてマルコフ連鎖が収束するような割合で調節しなければならない。 δ は経験的に“ステップ 4 で \mathbf{y}' が採用される割合 (acceptance ratio) が 50% 前後になるように決めるのが良い (Wood (1968))”と言われている。Acceptance ratio が大き過ぎると平衡状態に達するまでの時間が長くなることも注意しなければならない。

ステップ 2 とステップ 3 は、エネルギー関数を使えばエネルギーの差の符号判定で次のように更新させることができる。

- 2'. それぞれのエネルギー $U(\mathbf{y})$, $U(\mathbf{y}')$ を計算し、エネルギーの差 $\Delta = U(\mathbf{y}') - U(\mathbf{y})$ を求める。

- 3'. \mathbf{y} を次のように更新する。

$$\Delta \leq 0 \text{ ならば } \mathbf{y}:=\mathbf{y}'$$

$$\Delta > 0 \text{ ならば確率 } \exp(-\Delta) \text{ で } \mathbf{y}:=\mathbf{y}', \text{ 確率 } 1-\exp(-\Delta) \text{ で } \mathbf{y}:=\mathbf{y}$$

とする。

この 2' と 3' をステップ 2 とステップ 3 のかわりに用いても同様に平衡分布として P をもつマルコフ連鎖を作り出すことができる。本研究の数値実験ではエネルギー関数を用いたメトロポリスアルゴリズムを使用している。

ここで提案したアルゴリズムを使って数値実験を行なった結果を次章で示す。

4. シミュレーション

ここでは積分値が既知であるいくつかの関数を使ってシミュレーションを行ない、その結果を示し、最適分割による方法と等分割による方法を比較する。

4.1 指数関数

指数関数

$$f(\mathbf{x}) = e^{x_1} e^{x_2} \cdots e^{x_p}$$

の p 次元区間 $[0, 1]^p$ での定積分

$$Z = \int_0^1 \cdots \int_0^1 e^{x_1} \cdots e^{x_p} dx_1 \cdots dx_p$$

を考える。次元を $p=100$ とすると、真値は

$$\begin{aligned} Z &= (e-1)^p = 3.2318 \times 10^{23}, \\ \log Z &= p \log(e-1) = 54.132 \end{aligned}$$

である。

サンプル数

被積分関数 $\psi(\sigma)$ を推定するために使う MCMC サンプルの総数は 10000 個とする。各点での MCMC サンプルは初めの半数を平衡状態に達していないとして捨て、残りの半数を平衡状態におけるサンプルとして推定に使用する。

分割点の数

まず妥当な分割点の数を知るために、一次元積分近似に使う全サンプル数を 10000 個に固定し、等分割による積分近似の実験を試みた。(分割点数) \times (各 σ_j でのサンプル数) = 10000 を、

- 10×1000 ,
- 20×500 ,
- 40×250 ,
- 50×200 ,
- 80×125 ,
- 100×100 ,

とし、それぞれ 10 回の実験を行なった。表 1 に 10 回の実験での推定値を平均した値を示す。

この表から分割点が 50 個のときが最も真値に近いことがわかる。分割点数が少なくても正しい結果は得られないが、分割点数を増やしてゆくと真値からのずれが大きくなるが見えてとれる。この例は被積分関数の推定値 $\hat{\psi}(\sigma)$ がほぼ直線になっているので分割点を多くとる必要はなく、少ない分割点による折れ線で十分 ψ の近似ができる。分割点を多く取り各点でのサン

表 1. $\log Z$ の推定値 (真値 54.132)。

サンプル数	$\log \hat{Z}$	サンプル数	$\log \hat{Z}$
10×2000	54.135	50×400	54.132
20×1000	54.142	80×250	54.136
40×500	54.140	100×200	54.144

プルを少なくすると被積分関数の変動を捉えることはできるが、各 σ で $\hat{\psi}$ を正しく推定することができないため、最終的な $\log Z$ の推定値も誤差が大きくなると思われる。各点でのサンプル数が少な過ぎるとマルコフ連鎖が平衡状態に到達しないことも考えられる。また表1から分割点数が少な過ぎても一次元積分が十分に近似されず、正しく推定できないことも窺える。したがって全サンプル数を同じ数だけ取るのであれば、分割点は少な過ぎても正しい結果は得られないが、分割点数は多く取らずに各点での推定のための MCMC サンプルを多くする方が良いことが分かる。その結果、分割点数は 50 が妥当であると判断した。

分割方法

分割点の数を 50 で固定し四通りの分割方法、

- 50 等分割、
- 10 等分割と 40 最適分割、
- 20 等分割と 30 最適分割、
- 40 等分割と 10 最適分割、

で同じサンプル数 ($50 \times 200 = 10000$) を使ったシミュレーションを行ない、推定値を比較する。

関数

結ぶ関数は $f_0(\mathbf{x})=1$ とし、 $f(\mathbf{x})$ と $f_0(\mathbf{x})$ は scaling parameter (この場合は scaling parameter と reciprocal temperature は同じパスになる) で結ぶと、定義する関数は次のようになる。

$$\begin{aligned} f(\mathbf{x}, \sigma) &= \exp \left\{ \sigma \sum_{i=1}^p x_i \right\}, \\ U(\mathbf{x}, \sigma) &= -\sigma \sum_{i=1}^p x_i, & \psi(\mathbf{x}, \sigma) &= \sum_{i=1}^p x_i, \\ Z(\sigma) &= \frac{(e^\sigma - 1)^p}{\sigma^p}, & \psi(\sigma) &= p \left\{ \frac{(\sigma - 1)e^\sigma + 1}{\sigma(e^\sigma - 1)} \right\}. \end{aligned}$$

シミュレーション結果

分割点の数を 50 と判断し、異なる分割法でそれぞれ 10 回ずつ行なったシミュレーションの $\log Z$ の推定値を表2に示す。また図2, 図3にはそれぞれ 50 等分割, 20 等分割と 30 最適分割

表2. $\log Z$ の推定値 (真値 54.132, 分散 $\times 10^{-3}$).

No.	50 等分割		10+40 分割		20+30 分割		40+10 分割	
	$\log \hat{Z}$	$\text{Var}(S)$	$\log \hat{Z}^*$	$\text{Var}(S^*)$	$\log \hat{Z}^*$	$\text{Var}(S^*)$	$\log \hat{Z}^*$	$\text{Var}(S^*)$
1	54.154	1.1179	54.158	1.1737	54.133	1.2107	54.149	1.2801
2	54.138	1.1789	54.151	1.2180	54.110	1.2029	54.145	1.2203
3	54.142	1.1128	54.130	1.1238	54.159	1.1467	54.135	1.1905
4	54.078	1.1802	54.058	1.1631	54.090	1.1121	54.111	1.2142
5	54.134	1.2258	54.144	1.2218	54.138	1.2479	54.132	1.1550
6	54.157	1.1095	54.168	1.1816	54.178	1.1762	54.141	1.2702
7	54.152	1.1215	54.125	1.1398	54.121	1.1708	54.125	1.1710
8	54.126	1.1180	54.135	1.1593	54.118	1.1795	54.134	1.2011
9	54.122	1.1543	54.124	1.1674	54.116	1.1425	54.152	1.2453
10	54.151	1.1414	54.125	1.3006	54.167	1.1775	54.137	1.2005

点による被積分関数の近似 $\hat{\psi}$ のグラフを示す。

等分割のみの求積と等分割と最適分割とを合わせた分割による求積ではほとんど $\log Z$ の推定に差異はないが、値を見る限りでは等分割の方が良い推定量を出していると言わざるを得ない。加えて最適分割を含む求積の方が分散が大きくなっている。また図から最適変換による分割は等分割とほぼ一致し、 $[0, 1]$ 区間を偏ることなく均等に刻んでいることが分かる。この例のように変動が少なく、直線で十分近似できるような被積分関数であるならば等分割点で求積をするのが良い。アルゴリズムの複雑さ、プログラミングにかかる労力から考えても、等分割点による求積の方がはるかに優れている。

次に被積分関数の変動が大きい例を示す。

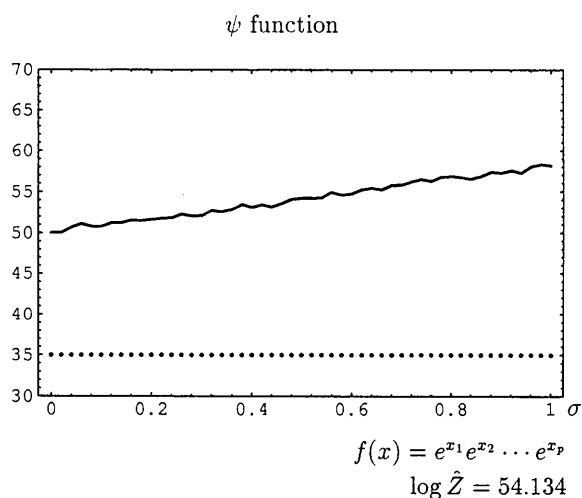


図 2. 50 等分割点による被積分関数の近似。

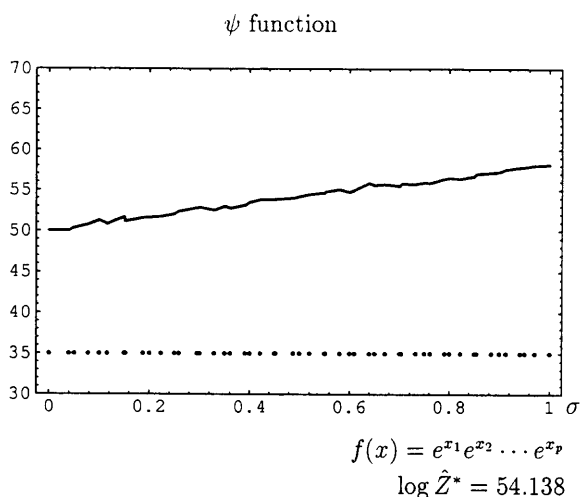


図 3. 20 等分と 30 最適分割による積分近似。

4.2 正規分布

多変量正規分布の正規化定数を求める。すなわち

$$f(\mathbf{x}) = \exp\left\{-\frac{1}{2}\mathbf{x}^T B \mathbf{x}\right\}$$

の p 次元無限領域 $(-\infty, \infty)$ の定積分

$$Z = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}\mathbf{x}^T B \mathbf{x}\right\} dx_1 \cdots dx_p,$$

を考える。 B は三重対角行列で、

$$B = \begin{bmatrix} 1 & -.5 & & & 0 \\ -.5 & 1 & \ddots & & \\ & \ddots & \ddots & 1 & -.5 \\ 0 & & & -.5 & 1 \end{bmatrix}$$

とする。次元を $p=100$ とすると真値は

$$Z = (2\pi)^{\frac{p}{2}} \{\det(B)\}^{-\frac{1}{2}} = 2^p \pi^{\frac{p}{2}} (p+1)^{\frac{1}{2}} = 9.1760 \times 10^{55},$$

$$\log Z = p \log 2 + \frac{p}{2} \log \pi + \frac{1}{2} \log(p+1) = 124.244$$

である。

この積分を計算する方法の一つは、無限領域からいくつかの有界な領域へ置き換えることであるが、ここでは十分大きな範囲（実際には $[-20.0, 20.0]^p$ ）を取ることで積分を近似する。

結ぶ関数 $f_0(\mathbf{x})$ として、二つの関数 $f_0(\mathbf{x})=1$, $f_0(\mathbf{x})=\exp\{-(\mathbf{x}^T \mathbf{x})/2\}$ を採り積分近似の精度を比較する。

サンプル数

被積分関数 $\psi(\sigma)$ を推定するために各 σ_j に対して MCMC サンプルは 1000 個取る。初めの 500 個のサンプルは平衡状態に達していないとして捨てる。残りの 500 個を平衡分布に従うサンプルとして利用する。

分割方法

分点の個数は 100 個に固定する。分割方法を次の四通り、

- 100 個の等分割,
- 20 個の等分割と 80 個の最適分割,
- 30 個の等分割と 70 個の最適分割,
- 40 個の等分割と 60 個の最適分割,

で同じサンプル数 ($100 \times 500 = 5 \times 10^4$ 個) を使ったシミュレーションを行ない、推定した値を比較する。

関数 ($f_0(\mathbf{x})=1$ の場合)

結ぶ関数として $f_0(\mathbf{x})=1$ を採り scaling parameter で結ぶ。ここでは等分割による求積と最適刻みによる求積の比較をするために、意図的に $\sigma = \xi^{1/\alpha}$ という変換を施し、 α を変えて定積分を試みる。これは $\alpha=2$ のとき reciprocal temperature のパスと一致する。定義する関数は次のようになる。

$$\begin{aligned}
 f_a(\mathbf{x}, \xi) &= \exp \left\{ -\frac{1}{2} (\xi^{\frac{1}{a}} \mathbf{x})^T B (\xi^{\frac{1}{a}} \mathbf{x}) \right\}, \\
 U_a(\mathbf{x}, \xi) &= \frac{1}{2} \xi^{\frac{2}{a}} \mathbf{x}^T B \mathbf{x}, \\
 \psi_a(\mathbf{x}, \xi) &= -\frac{1}{a} \xi^{\frac{2}{a}-1} \mathbf{x}^T B \mathbf{x}, \\
 Z_a(\xi) &= (2\pi)^{\frac{p}{2}} (\det B)^{-\frac{1}{2}} \xi^{-\frac{p}{a}}, \\
 \psi_a(\xi) &= -\frac{p}{a\xi}.
 \end{aligned}$$

シミュレーション結果

表3にシミュレーションの結果を示す。また $\alpha=1, 2$ の場合の $\hat{\psi}$ のグラフを図4~10に示す。この被積分関数 $\psi(\xi)$ は $\xi=0$ で $-\infty$ になる。したがって $\xi=0$ の付近で大きな変動をする。しかし $\alpha \leq 2$ ならば関数 $\psi(\mathbf{x}, \xi)$ は $\psi(\mathbf{x}, 0)=0$ であり無限大にならないため実用に耐え得る推定値を出していると言える。 $\alpha > 2$ になると $\psi(\mathbf{x}, 0)$ も無限大になるので、 $\hat{\psi}(0)$ は補外して求めなければならない。 $\alpha \leq 2$ のとき $\hat{\psi}(\xi)$ は $\xi=0$ で0を取り、 $\xi=0$ の付近で大きく減少してその後急激な上昇をする。等分割ではその上昇を捉えることが出来ず、 $\log Z$ を小さく推定してしまっている。最適刻みでは上昇する区間を細かく刻み、上昇を捉えることができる。 $\alpha > 2$ のとき $\hat{\psi}(\xi)$ は $\xi=0$ で大きな負の値を取り、その後一気に増加する。等分割では $\alpha \leq 2$ の場合と同様に、急激な上昇を捉えることが出来ず過小推定になっている。最適刻みでは増加傾向を捉えているが、 $\psi(0)=-\infty$ を大きく推定してしまい結果的に大きく推定している。

関数 ($f_0(\mathbf{x})=\exp\{-\mathbf{x}^T \mathbf{x}/2\}$ の場合)

結ぶ関数として $f_0(\mathbf{x})=\exp\{-\mathbf{x}^T \mathbf{x}/2\}$ を採り、reciprocal temperature で結ぶ。ここでも同様に、等分割による求積と最適刻みによる求積の比較のため、意図的に $\sigma=\xi^{2/\alpha}$ という変換を施し、 α を変えて定積分を試みる。すると定義する関数は次のようになる。

$$f_a(\mathbf{x}, \xi) = \exp \left\{ -\frac{1}{2} \mathbf{x}^T (\xi^{\frac{2}{a}} B + (1 - \xi^{\frac{2}{a}}) I) \mathbf{x} \right\},$$

表3. $\log Z$ の推定値 (真値 124.24)。

α	100 等分割		20+80 分割		30+70 分割		40+60 分割	
	$\log \tilde{Z}$	$\text{Var}(S)$	$\log \tilde{Z}^*$	$\text{Var}(S^*)$	$\log \tilde{Z}^*$	$\text{Var}(S^*)$	$\log \tilde{Z}^*$	$\text{Var}(S^*)$
0.1	124.10	2.0730	124.88	0.8286	125.09	0.8427	123.96	0.9420
0.2	126.79	1.1868	123.46	0.7251	121.36	0.6911	123.33	0.7033
0.3	123.55	1.0460	122.24	0.7064	124.68	0.6637	124.27	0.7360
0.5	124.53	0.8043	121.95	0.6831	123.97	0.6398	123.04	0.6245
1.0	121.87	0.9126	124.57	0.7700	125.59	0.6270	124.96	0.7253
2.0	114.35	3.2342	122.94	0.7453	122.24	0.8331	124.26	0.9073
3.0	-176.21	1.0084	141.15	1.2842	134.24	1.1644	134.00	1.4711
4.0	-486.26	0.7348	172.77	1.2541	171.17	1.2096	163.44	1.7221

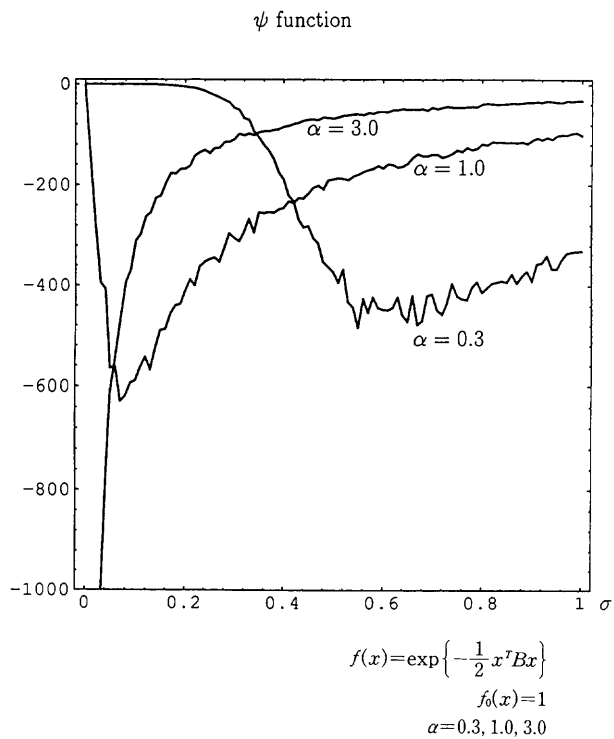
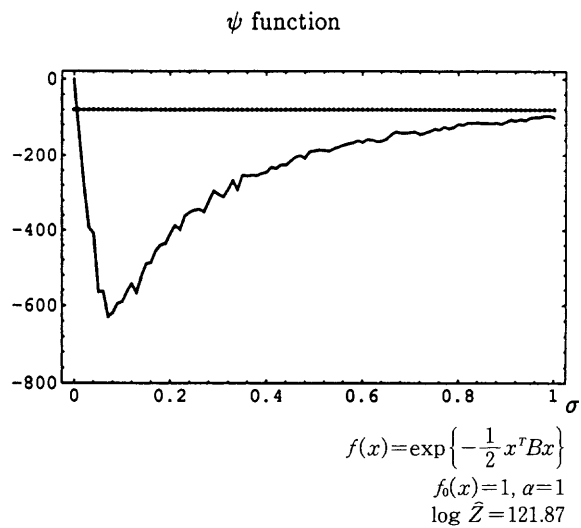
図4. ψ 関数の比較.

図5. 100等分割による積分近似.

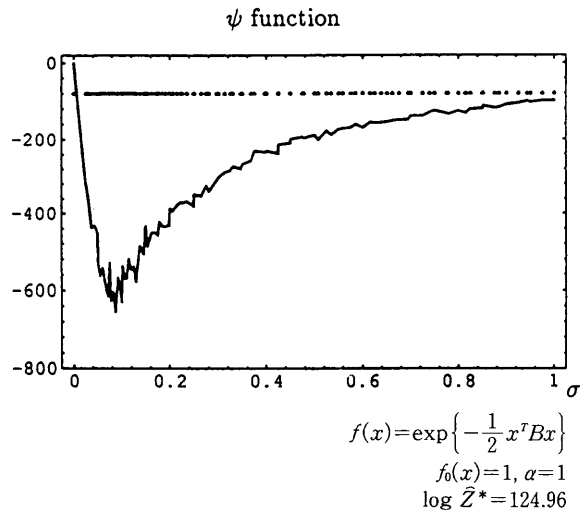


図 6. 40 等分と 60 最適分割による積分近似.

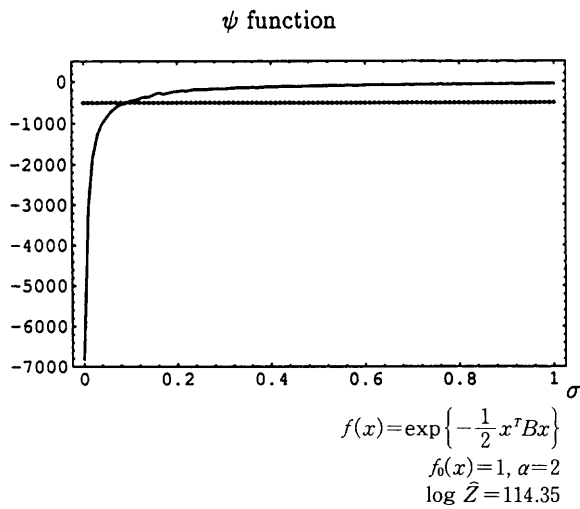


図 7. 100 等分割による積分近似.

$$U_a(x, \xi) = \frac{1}{2} x^T (\xi^{\frac{2}{a}} B + (1 - \xi^{\frac{2}{a}}) I) x,$$

$$\psi_a(x, \xi) = \frac{1}{\alpha} \xi^{\frac{2}{a}-1} x^T (I - B) x,$$

$$\psi_a(\xi) = -\frac{1}{\alpha} \xi^{\frac{2}{a}-1} \text{tr}[(I - B)(\xi^{\frac{2}{a}} B + (1 - \xi^{\frac{2}{a}}) I)^{-1}].$$

シミュレーション結果

表 4 にシミュレーションの結果を示す. また $\alpha=1$ の場合の $\hat{\psi}$ のグラフを図 11~15 に示す. この被積分関数 $\psi(\xi)$ は始めは 0 に近い値をとり, 徐々に増加し $\xi=1$ で急激に大きな値を取

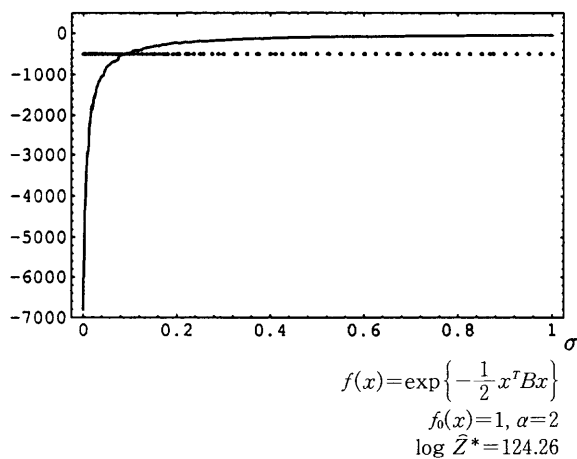
ψ function

図8. 40等分と60最適分割による積分近似.

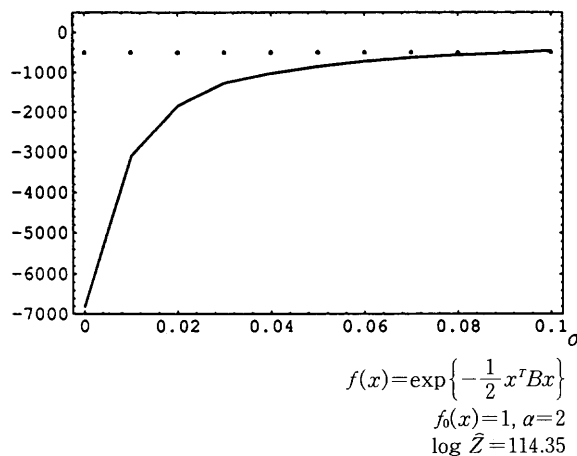
 ψ function (区間 [0.0,0.1] を拡大)

図9. 100等分割による積分近似.

る. 等分割では $\xi=1$ の付近で急激な増加を捉えることが出来ずに大きめの推定をしている. 最適刻みではこの急上昇を良く捉えてよい推定値を出していると言える.

4.3 ランダム行列の固有値分布

ランダム行列の固有値の同時密度関数の正規化定数を求める. すなわち

$$f(\mathbf{x}) = \exp\left\{-\frac{1}{2}\beta \sum_{j=1}^p x_j^2\right\} \prod_{j < k} |x_j - x_k|^\beta$$

の無限領域 $(-\infty, \infty)$ の定積分

$$Z = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}\beta \sum_{j=1}^p x_j^2\right\} \prod_{j < k} |x_j - x_k|^\beta dx_1 \cdots dx_p$$

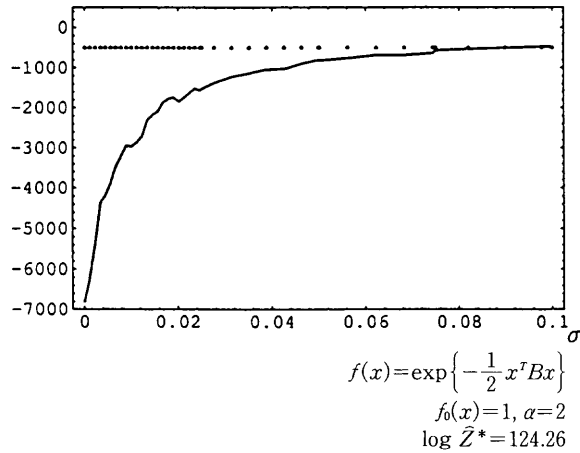
ψ function (区間 [0.0,0.1] を拡大)

図 10. 40 等分と 60 最適分割による積分近似.

表 4. $\log Z$ の推定値 (真値 124.24). ここで例えば $e+3$ は 10^3 を表す. 以下, 同様の約束をする. 一般には $e+n$ は 10^n と約束する.

α	100 等分割		20+80 分割		30+70 分割		40+60 分割	
	$\log \bar{Z}$	$\text{Var}(S)$	$\log \bar{Z}^*$	$\text{Var}(S^*)$	$\log \bar{Z}^*$	$\text{Var}(S^*)$	$\log \bar{Z}^*$	$\text{Var}(S^*)$
0.1	156.99	2.0390e+3	123.76	4.7156e-1	123.64	1.8028e-1	123.47	5.5437e-1
0.2	136.49	2.2569e+2	125.54	4.4094e-1	124.25	1.9535e-1	124.66	4.7321e-1
0.3	126.77	1.0977	122.34	1.5059e-1	124.11	2.7941e-1	123.20	1.7478e-1
0.5	125.85	2.7566	124.38	1.4672e-1	123.97	2.4821e-1	123.96	1.7460e-1
1.0	126.51	2.3254	123.38	7.2518e-2	124.47	8.9014e-2	123.67	1.2798e-1
2.0	125.74	4.4145e-1	123.84	1.0505e-1	123.95	1.1157e-1	124.40	9.3675e-2
3.0	123.20	3.0348e-1	122.95	7.3866e-1	123.71	7.5163e-2	124.11	1.3289e-1
4.0	123.96	4.4745e-1	123.37	9.0122e-2	123.96	1.2151e-1	122.39	7.3236e-2

を考える. 真値は β の関数として

$$\begin{aligned}
 Z(\beta) &= (2\pi)^{\frac{p}{2}} \beta^{-\frac{p}{2} - \frac{\beta}{4} p(p-1)} \frac{\prod_{j=1}^p \Gamma\left(1 + \frac{1}{2} \beta j\right)}{\left(\Gamma\left(1 + \frac{1}{2} \beta\right)\right)^p}, \\
 \log Z(\beta) &= \frac{p}{2} \log(2\pi) - \left(\frac{p}{2} + \frac{\beta}{4} p(p-1)\right) \log \beta \\
 &\quad + \sum_{j=1}^p \log\left(\Gamma\left(1 + \frac{1}{2} \beta j\right)\right) - p \log\left(\Gamma\left(1 + \frac{1}{2} \beta\right)\right)
 \end{aligned}$$

となる.

この密度関数は β が 1, 2, 4 のとき重要な意味を持つ. $\beta=1$ のときは次のような標準実対称ガウス型ランダム行列, すなわち $X=(x_{ij})_{i,j=1}^p$ は, 各要素について $x_{ij}=x_{ji}$ であり, $\{x_{ij}: 1 \leq i \leq$

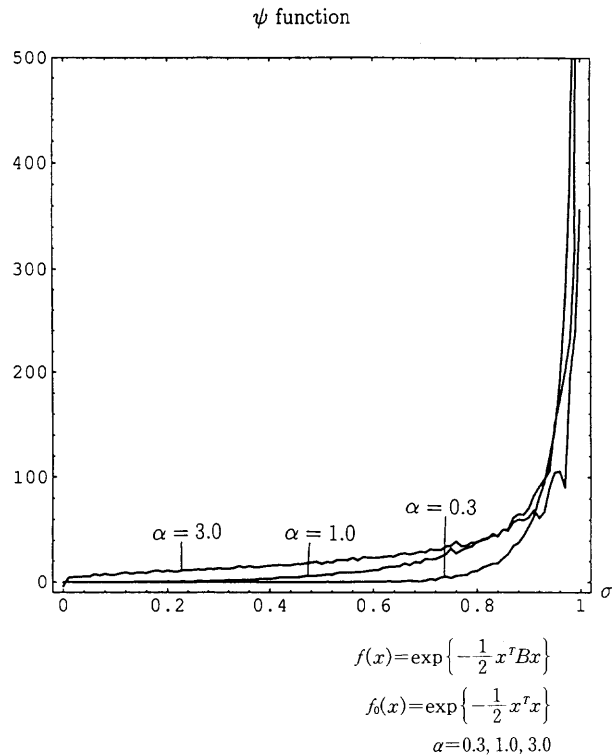


図 11. ψ 関数の比較.

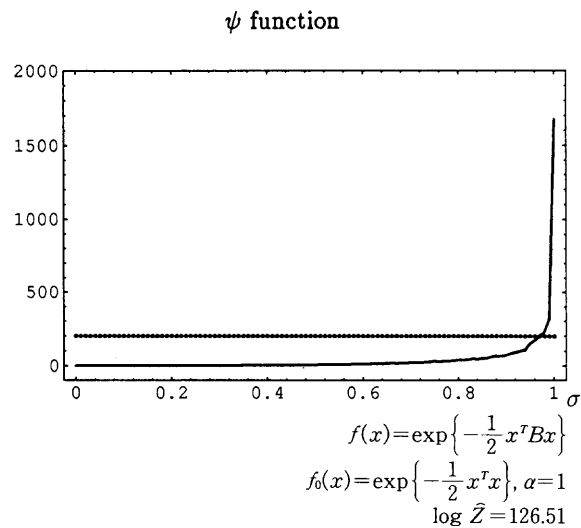


図 12. 100 等分割による積分近似.

$j \leq p\}$ は独立に正規分布に従う確率変数で $x_{ij} \sim N(0, 1/2p)$ ($1 \leq i \leq j \leq p$), $x_{ii} \sim N(0, 1/p)$ ($1 \leq i \leq p$) であるランダム行列に関連している。また $\beta=2, \beta=4$ のときはそれぞれ行列の各要素を複素数, 四元数に拡張した同様のランダム行列に関連する。ここで取り挙げる関数は, このような

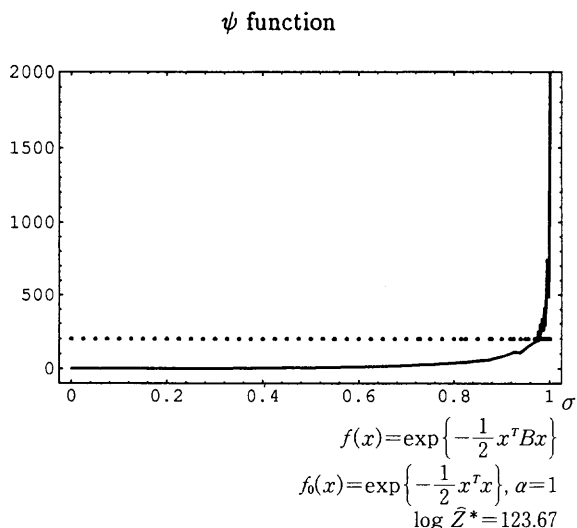


図 13. 40 等分と 60 最適分割による積分近似.

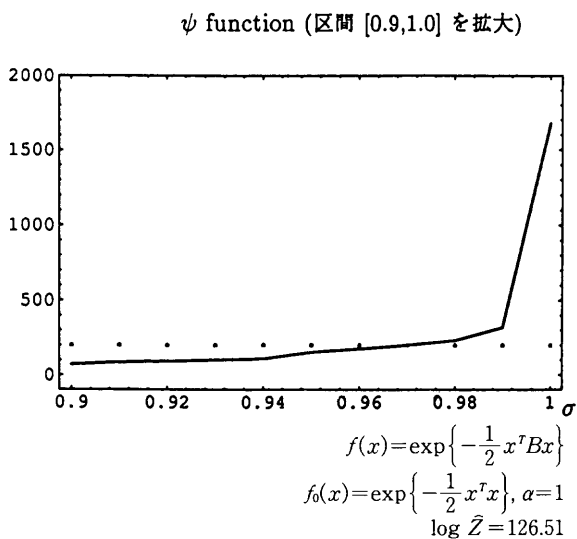


図 14. 100 等分割による積分近似.

ランダム行列に対する p 個の固有値の同時密度関数となる.

ここでは $\beta=1, 2, 4$ の場合についてシミュレーションを行なう. 次元を $p=100$ としたときそれぞれの真値は,

$$\begin{aligned} Z(1) &= 8.7092 \times 10^{2799}, & \log Z(1) &= 6.4471 \times 10^3, \\ Z(2) &= 1.5520 \times 10^{5475}, & \log Z(2) &= 1.2607 \times 10^4, \\ Z(4) &= 2.8441 \times 10^{10836}, & \log Z(4) &= 2.4952 \times 10^4 \end{aligned}$$

となる.

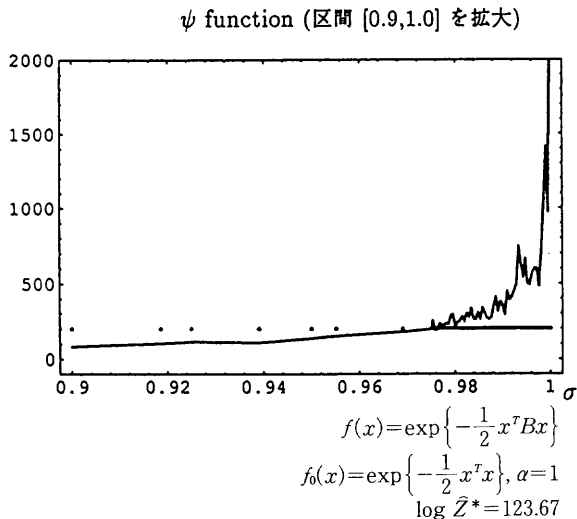


図 15. 40 等分と 60 最適分割による積分近似.

分割方法

分点の個数は 100 個に固定する. 分割を次の四通りの方法,

- 100 個の等分割,
- 20 個の等分割と 80 個の最適分割,
- 30 個の等分割と 70 個の最適分割,
- 40 個の等分割と 60 個の最適分割,

で同じサンプル数 ($100 \times 500 = 5 \times 10^4$ 個) を使ったシミュレーションを行ない, 推定した値を比較する.

関 数

結ぶ関数 $f_0(x)=1$ として reciprocal temperature で結ぶ. 先の例と同様に, 等分割による求積と最適刻みによる求積の比較のため意図的に $\sigma = \xi^{1/\alpha}$ という変換を施し, α を変えて定積分を試みる. すると定義する関数は次のようになる.

$$\begin{aligned}
 f_\alpha(x, \xi) &= \exp\left\{-\frac{1}{2}\xi^{\frac{1}{\alpha}}\beta \sum_{j=1}^p x_j^2\right\} \prod_{j < k} |x_j - x_k|^{\xi^{1/\alpha}\beta}, \\
 U_\alpha(x, \xi) &= \frac{1}{2}\xi^{\frac{1}{\alpha}}\beta \sum_{j=1}^p x_j^2 - \xi^{\frac{1}{\alpha}}\beta \sum_{j < k} \log|x_j - x_k|, \\
 \psi_\alpha(x, \xi) &= -\frac{1}{2\alpha}\xi^{\frac{1}{\alpha}-1}\beta \sum_{j=1}^p x_j^2 + \frac{1}{\alpha}\xi^{\frac{1}{\alpha}-1}\beta \log|x_j - x_k|.
 \end{aligned}$$

シミュレーション結果

$\beta=1, 2, 4$ の場合の $\log Z$ の推定値をそれぞれ表 5, 表 6, 表 7 に示す. また $\beta=1, \alpha=1$ の場合の $\hat{\psi}$ のグラフを図 16~19 に示す. いずれの場合も 100 等分割による方法と最適分割による推定値の差はほとんどない. ただし自然な結び方である $\alpha=1$ のとき等分割による方法とうまく推定することができていないが, 最適分割による方法は比較的うまく推定している. これは

表 5. $\log Z$ の推定値, $\beta=1$ (真値 6447.10).

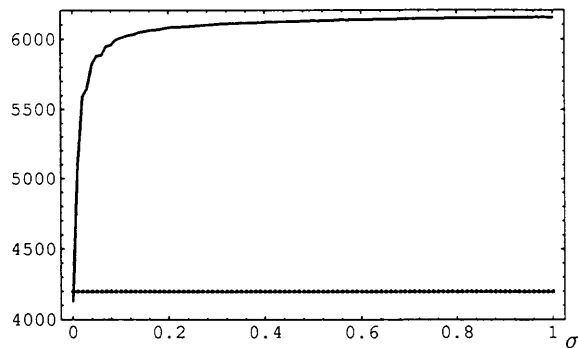
α	100 等分割		20+80 分割		30+70 分割		40+60 分割	
	$\log \hat{Z}$	$\text{Var}(S)$	$\log \hat{Z}^*$	$\text{Var}(S^*)$	$\log \hat{Z}^*$	$\text{Var}(S^*)$	$\log \hat{Z}^*$	$\text{Var}(S^*)$
0.1	6451.38	3.4886e-1	6453.20	1.7055-e1	6452.75	1.5510e-1	6453.70	1.6175e-1
0.2	6447.68	2.7793e-1	6449.69	1.3828-e1	6452.25	1.5289e-1	6450.54	1.4960e-1
0.3	6447.66	2.0030e-1	6448.81	1.8065-e1	6448.73	1.6344e-1	6448.31	1.4536e-1
0.5	6447.30	3.4567e-1	6447.15	1.2832-e1	6447.55	1.6795e-1	6447.66	1.7041e-1
1.0	6445.46	1.1429e-1	6446.84	1.9716-e1	6448.09	2.3954e-1	6446.77	1.9046e-1

表 6. $\log Z$ の推定値, $\beta=2$ (真値 12607.1).

α	100 等分割		20+80 分割		30+70 分割		40+60 分割	
	$\log \hat{Z}$	$\text{Var}(S)$	$\log \hat{Z}^*$	$\text{Var}(S^*)$	$\log \hat{Z}^*$	$\text{Var}(S^*)$	$\log \hat{Z}^*$	$\text{Var}(S^*)$
0.1	12616.9	3.3130e-1	12613.0	2.1980e-1	12617.4	1.8376e-1	12620.6	2.0293e-1
0.2	12608.6	2.9167e-1	12614.7	1.6919e-1	12612.7	1.7807e-1	12612.9	2.1558e-1
0.3	12608.4	2.2528e-1	12610.5	1.8780e-1	12610.2	1.6458e-1	12609.9	2.6767e-1
0.5	12607.6	3.8797e-1	12607.2	1.5596e-1	12607.8	2.0172e-1	12607.4	2.2005e-1
1.0	12601.6	2.3442	12607.9	2.7377e-1	12607.1	2.8864e-1	12606.0	2.6476e-1

表 7. $\log Z$ の推定値, $\beta=4$ (真値 24951.9).

α	100 等分割		20+80 分割		30+70 分割		40+60 分割	
	$\log \hat{Z}$	$\text{Var}(S)$	$\log \hat{Z}^*$	$\text{Var}(S^*)$	$\log \hat{Z}^*$	$\text{Var}(S^*)$	$\log \hat{Z}^*$	$\text{Var}(S^*)$
0.1	24970.6	4.4132e-1	24966.3	2.3189e-1	24969.4	2.5341e-1	24970.2	2.8049e-1
0.2	24956.5	3.1404e-1	24962.5	2.2205e-1	24961.5	2.4445e-1	24963.9	2.3299e-1
0.3	24953.6	3.5416e-1	24957.4	2.2649e-1	24957.7	2.0552e-1	24956.9	2.4043e-1
0.5	24951.9	6.2637e-1	24951.3	2.5738e-1	24951.8	2.3550e-1	24952.7	2.3346e-1
1.0	24934.8	7.9803	24951.8	3.1677e-1	24952.1	3.4573e-1	24951.7	3.9115e-1

 ψ function

$$f(x) = \exp\{-\tfrac{1}{2}\beta \sum x_j^2\} \prod |x_j - x_k|^\beta$$

$$\log \hat{Z} = 6445.46$$

図 16. 100 等分割による積分近似.

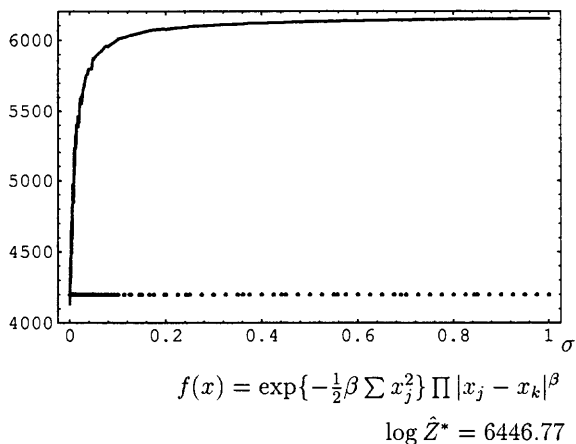
ψ function

図 17. 40 等分と 60 最適分割による積分近似.

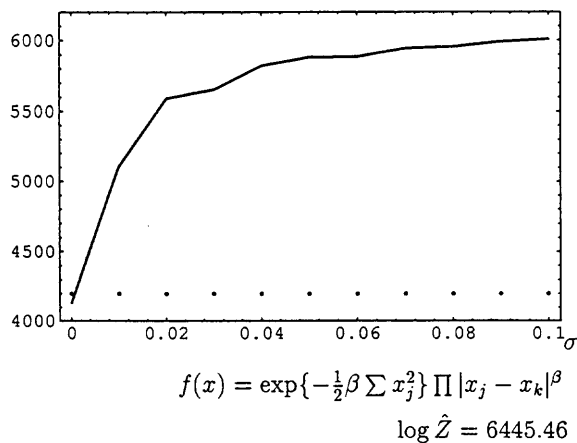
 ψ function (区間 $[0.0, 0.1]$ を拡大)

図 18. 100 等分割による積分近似.

$\xi=0$ の付近での被積分関数の大きな変動を等分割点では捉えることができないと思われる。

4.4 シミュレーションの考察

提案したアルゴリズムを実行するには、一次元積分のための分割点の数、被積分関数推定のための各分割点でのサンプル数を予め設定しなければならない。このアルゴリズムでは二段階で一次元積分を行なっている。第一段階では最適変換 $t_{\text{opt}}(\sigma)$ に標準偏差 $\sqrt{v(\sigma)}$ の積分が含まれているため一次元積分が必要である。第二段階では $\log Z$ の推定のために $\psi(\sigma)$ の積分を行なわなければならない。第一段階での積分はこの積分値を求めることが本来の目的ではないので、厳密である必要はないが、第二段階での積分は目的の $\log Z$ を近似するための積分であるから

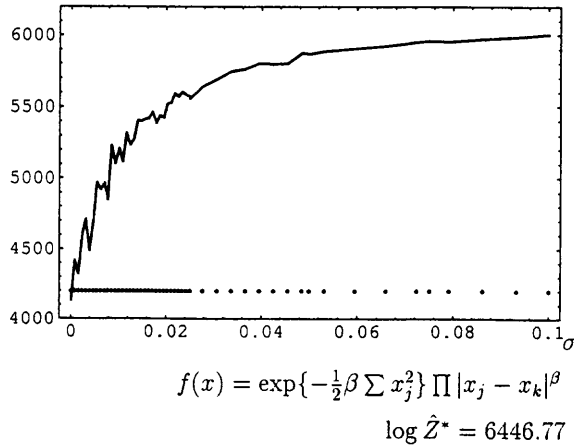
ψ function (区間 $[0.0, 0.1]$ を拡大)

図 19. 40 等分と 60 最適分割による積分近似.

正確でなければならない。したがって第一段階の積分より第二段階の積分に対して分割点をたくさん取り、密に刻んで求積しなければならないことは明らかであるが、それぞれの積分にどれだけの分割点が必要であるのかははっきりした答えは見つかっていない。積分を正確に近似するためには、分割点は被積分関数 $\psi(\sigma)$ の変動を捉えることができるだけの十分な数を用意しなければならない。第一段階の求積において、すなわち $\sqrt{v(\sigma)}$ の積分近似において求められた $\hat{\psi}(\sigma)$ をプロットし、おおまかな形状を見て $\psi(\sigma)$ の一次元積分のための最適な分割点の数を決める、というのが一つの手である。サンプル数に関しては、刻み数が十分に大きいのであれば期待値を推定するためのサンプルは多量に取る必要はないと言われている。しかし正確に $\psi(\sigma)$ を推定するためには多くのサンプルを要すると思われる。

またサンプル生成に使う MCMC 法には制御しなければならないパラメータがある。連続状態空間の場合、次元 k 個の確率変数を順々に推移させ全ての要素を推移させたとき、その状態を一単位時間後の状態としマルコフ連鎖を作り出す。この一つの要素を推移させるときに許される最大変化量である制御パラメータを決めなければならない。さらにマルコフ連鎖が平衡状態に達しているかどうかのチェックも必要になり、チェックした上で平衡状態に達するまでのサンプルは捨てて、平衡状態におけるマルコフ連鎖をサンプルとして使用する。MCMC サンプルを利用する場合にはこれらのことを十分注意しなければ正しい結果を得ることはできない。

5. 考察と問題点

ここで提案したアルゴリズムは関数 f_0 と f_0, f のパスを固定し $f(x, \sigma)$ を定義したとき、一次元積分の被積分関数 $\psi(\sigma)$ の変動を良く捉えることができる最適な分割点で一次元積分をする。これは経験的に一次元積分の被積分関数 $\psi(\sigma)$ の絶対値と $\psi(\sigma)$ の分散 $\sqrt{v(\sigma)}$ は正の相関があることがわかっており (Ogata (1989)), このアルゴリズムは結果的に $\psi(\sigma)$ の変動が大きい区間では細かな分割をし、変動が少なく直線で十分近似できる区間では大きな分割をすることになる。したがって $\psi(\sigma)$ が大きく変動するような $f(x, \sigma)$ しか採ることができないとき、このアルゴリズムは威力を発揮すると言える。しかし変動が少なくほぼ直線で近似できるような関数を $\psi(\sigma)$ として取ることができるのであれば、アルゴリズムの複雑さやプログラミングにか

かる手間などから考えても等分割による積分近似をするほうが良い。

密度関数の1パラメータ族 $\{P(\mathbf{x}, \sigma)\}$ の \mathbf{x} をデータ, σ をパラメータとすれば $P(\mathbf{x}, \sigma)$ は σ が与えられたときの \mathbf{x} の条件付き分布の密度関数と見ることもできる。また, ここで考えた変換 $t(\sigma)$ はパラメータ σ の分布すなわち事前分布の分布関数と見なすことができる。これより $\log Z - \log Z_0$ は \mathbf{x}, σ の同時分布に関する $\psi(\mathbf{x}, \sigma)/t'(\sigma)$ の期待値である。

$$\log Z - \log Z_0 = \int_0^1 \int \left(\frac{\psi(\mathbf{x}, \sigma)}{t'(\sigma)} \right) P(\mathbf{x}, \sigma) t'(\sigma) d\mathbf{x} d\sigma.$$

このように見ると, 最適刻みのための変換 $t_{\text{opt}}(\sigma)$ は最適な事前分布を選択していると考えられることもできる。

6. 今後の課題

ここで提案した最適分割による積分が良い近似を与えるかどうかを調べるためには, ここに例として挙げた関数だけでなく様々な関数で数値実験を行わなければならない。ただ積分推定値を比較するだけでなく, 最適分割のためのアルゴリズムの複雑さ, プログラミングにかかる労力なども考慮に入れて, 等分割による積分近似との比較・検討が必要である。

Ogata-Tanemura の積分法は結ぶ関数 f_0 として何を選ぶか, また f と f_0 のパスをどのように選ぶかにより積分値の精度が変わってくる。 f_0 は積分区間で積分値が分かっている関数を取らなければならないことから, 関数 f_0 の候補は限られてしまい, 最適な f_0 を考えることは困難である。したがって f と f_0 を結ぶパスとして前述のパス以外にどのようなパスがあるか, またどのパスが積分近似に最適であるかを考察し, 最適分割, 最適パスによってより正確に積分を近似することを考えたい。

このプログラムは一次元積分近似をする際, 各分割点 σ_i で独立に実行することが出来る。したがって Ogata (1989) が述べているように, 手続きの並列化が可能である。より高速で正確な演算をするためには並列化プログラミングが望まれる。

また統計問題への応用についても考えたい。主に正規化定数が必要となるベイズ解析に使うことができるだろう。ベイズ推論においては統計物理の理論が応用されることが多く, その場合状態は k 個の未知パラメータの組に対応する物理系によって表現される。未知パラメータの k 次元積分により求められる正規化定数は重要な意味を持つ。この場合正規化定数はデータを得る確率であり, 周辺尤度である。正規化定数を求めることで予測確率の計算が可能になる。ベイズ推論における正規化定数推定法はすでにラプラス近似による方法や, 事後分布や事後分布と事前分布の混合分布を用いたインポートランス・サンプリングによる方法などが考え出されている (Chib (1995), Raftery (1994), Kass and Raftery (1995))。ここで提案した積分推定法がこの正規化定数を求める方法として有効かどうかを検討したい。また異なる二つのベイズモデルの正規化定数の比はベイズ因子と, ベイズモデル選択の指標となる。二つのモデルの正規化定数を Z_1, Z_0 とすれば, ベイズ因子は Z_1/Z_0 であり, この対数はここで紹介した方法における $S = \log Z_1 - \log Z_0$ に相当する。したがって積分値が既知である f_0 を採る必要はなく, 正規化していないそれぞれのモデルの事後分布を f_1, f_0 として容易に推定することができる。これまでの研究においては Z_1, Z_0 をそれぞれ正規化定数推定法で計算してベイズ因子を計算する方法が主である。直接ベイズ因子が計算できるこの方法を使ってベイズ因子を計算し, 正規化定数推定による計算法と比較することが今後の課題として挙げられる。

参 考 文 献

- 荒木不二洋 (1994). 『統計物理の数理』, 岩波講座 応用数学, 岩波書店, 東京.
- Bash, P. A., Singh, U. C., Langridge, R. and Kollman, P. A. (1987). Free energy calculations by computer simulation, *Science*, **236**, 564-568.
- Chib, S. (1995). Marginal likelihood from the Gibbs output, *J. Amer. Statist. Assoc.*, **90**, 1313-1321.
- Gelfand, A. E. and Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities, *J. Amer. Statist. Assoc.*, **85**, 398-409.
- Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **6**, 721-741.
- Geyer, C. J. and Thompson, E. A. (1992). Constrained Monte Carlo maximum likelihood for dependent data (with discussion), *J. Roy. Statist. Soc. Ser. B*, **54**, 657-699.
- Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications, *Biometrika*, **57**, 97-109.
- 伊庭幸人 (1995). マルコフ連鎖モンテカルロ法とその統計学への応用, Research Memo., No. 570, The Institute of Statistical Mathematics, Tokyo.
- Kass, R. E. and Raftery, A. E. (1995). Bayes factors, *J. Amer. Statist. Assoc.*, **90**, 773-795.
- Kobayashi, S. and Nomizu, K. (1969). *Foundations of Differential Geometry*, Vol. 2, Wiley, New York.
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. and Teller, E. (1953). Equation of state calculations by fast computing machines, *J. Chem. Phys.*, **21**, 1087-1092.
- Neal, R. M. (1993). Probabilistic inference using Markov chain Monte Carlo methods, Tech. Report, CRG-TR-93-1, Department of Computer Science, University of Toronto, Canada.
- Ogata, Y. (1989). A Monte Carlo method for high dimensional integration, *Numer. Math.*, **55**, 137-157.
- Ogata, Y. (1995). Markov chain Monte Carlo integration by annealing and its application to the likelihood computation of Bayesian models, *Bull. Internat. Statist. Inst.*, **56**, Book 4, 1873-1891.
- Ogata, Y. and Tanemura, M. (1981). Estimation of interaction potentials of spatial point patterns through the maximum likelihood procedure, *Ann. Inst. Statist. Math.*, **33**, 315-338.
- Ogata, Y. and Tanemura, M. (1984). Likelihood analysis of spatial point patterns, *J. Roy. Statist. Soc. Ser. B*, **46**, 496-518.
- Ogata, Y. and Tanemura, M. (1989). Likelihood estimation of soft-core interaction potentials for Gibbsian point patterns, *Ann. Inst. Statist. Math.*, **41**, 583-600.
- Raftery, A. E. (1994). Hypothesis testing and model selection via posterior simulation (preprint).
- Thompson, E. A. (1994). Monte carlo likelihood in genetic mapping, *Statist. Sci.*, **9**(3), 355-366.
- Tierney, L. (1994). Markov chains for exploring posterior distributions, *Ann. Statist.*, **22**, 1701-1762.
- Wood, W. W. (1968). *Monte Carlo Studies of Simple Liquid Models* (eds. H. N. V. Termperley, J. S. Rowlinson and G. S. Rushbrooke), Physics of Simple Liquids, Chapter 5, 115-230, North-Holland, Amsterdam.

The High Dimensional Integration with Markov Chain Monte Carlo Method

Namiko Ogawa

(Graduate School of Information Sciences, Ochanomizu University)*

Shinto Eguchi

(The Institute of Statistical Mathematics)

This paper is concerned with numerical integration of functions defined on high dimensional spaces. Recently, methods of Markov chain Monte Carlo (MCMC) have been proposed in statistical literature; they have been established for the problem in statistical physics. This paper is focused to obtain a normalizing constant by the use of free energy method. The method leads the high dimensional integration to 1-dimensional one by MCMC sampling. Among all the segmentations an optimal is uniquely obtained in the sense that the mean square error of the estimated integral is minimal. We implement the improved method and investigate some Gaussian examples in the comparison with the method proposed by Ogata and Tanemura (1981, *Ann. Inst. Statist. Math.*, **33**, 315-338).

Key words: Markov chain Monte Carlo method, high-dimensional integration, Metropolis-Hastings algorithm, importance sampling.

* Now at Fuji Research Institute Corporation.