

カーネルマルコフ連鎖モンテカルロ法による測定誤差モデル推定

Measurement Error Model Estimation by Kernel Markov Chain Monte Carlo Method

赤穂 昭太郎*
Shotaro Akaho

伊庭 幸人†
Yukito Iba

Abstract: Measurement error models are the statistical models in which random noise is added to input variables as well as output variables. We consider the estimation problem of the regression function in a reproducing kernel Hilbert space (RKHS) for the measurement error models. We apply Markov chain Monte Carlo approach to estimate the posterior of the function. To deal with the infinite dimensionality of RKHS, we introduce a trick to exchange the order of sampling of the hidden variable and the function.

Keywords: Bayesian framework, Markov chain Monte Carlo, reproducing kernel Hilbert space, Gaussian process, measurement error model

1 まえがき

入力値から出力値への関数を推定する回帰問題では、通常入力値の関数にノイズが加わったものが観測されていると仮定することが多い。ところが、位相応答曲線の推定 [1, 2] など、一般には入力値そのものにも観測ノイズが加わっている場合も多い。すなわち、

$$x = z + \epsilon_x, \quad y = f(z) + \epsilon_y \quad (1)$$

のように潜在変数 z を介して入力 x にも出力 y にもノイズがのる。このようなモデルは「測定誤差モデル」(Measurement Error Model) として知られている [3]。

関数 f のクラスとしては直線 [4] やスプライン関数 [5] の場合が研究されているが、本稿では再生核ヒルベルト空間 (RKHS: Reproducing Kernel Hilbert Space) 上の関数を考える。再生核ヒルベルト空間は、サポートベクトルマシンをはじめとするカーネル法で幅広く使われるようになった関数クラスであり、確率モデルである正規過程 (GP: Gaussian Process) と等価なものである [6, 7]。

再生核ヒルベルト空間はガウスカーネルなど無限個の

基底ベクトルをもつカーネル関数を用いれば無限次元の自由度をもつ関数空間となるので、あらかじめ関数クラスが規定できないような場合には柔軟な関数あてはめができる可能性がある。

無限自由度の関数によるあてはめを実現するためにはなんらかのトリックが必要である。出力にしかノイズが乗っていない通常の回帰で最適解 (MAP 推定値) のみが必要な場合は、観測点におけるカーネル関数だけを使って解析的に解が求められる (Representer 定理)。マルコフ連鎖モンテカルロ法 (MCMC) などを用いて事後分布からのサンプルを生成する場合には、サンプルそのものは無限自由度であって直接に扱えない。しかし、関数値を評価する点の集合があらかじめ与えられた場合には、それに対応する基底を加えることで、有限次元の範囲で処理することができる [8, 9]。

われわれの問題としている測定誤差モデルにおいては事後分布が閉じた形で求められないだけでなく、関数への実質的な入力には潜在変数 z となるため、この方法は使えず、カーネル関数の基底が無限個必要となるように思われる。本稿で提案する手法はその点をうまく解決して、測定誤差モデルに対してカーネル回帰の考え方を拡張するものである。提案手法の中心となるトリックはマルコフ連鎖モンテカルロ法を適用する際に潜在変数の次の値の候補の生成と関数のサンプリングの順序を交換することである。これによって、MCMC の各ステップでは有限次元の計算を行いつつ、全体としては無限自由度

*産業技術総合研究所脳神経情報研究部門, 305-8568 茨城県つくば市梅園 1-1-1 中央第 2, tel. 029-861-5549, e-mail s.akaho@aist.go.jp, The National Institute of Advanced Industry and Scientific Technology, Central 2, 1-1-1 Umezono, Tsukuba, Ibaraki 305-8568

†統計数理研究所, 〒 190-8562 東京都立川市緑町 10-3, tel. 050-5533-8500(代), e-mail iba@ism.ac.jp, Institute of Mathematical Statistics, 10-3 Midoricho, Tachikawa, Tokyo 190-8562

の空間をサンプルすることが可能になる。

2 生成モデル

入力 x , 出力 y , 潜在変数 z に関して以下の生成モデルを仮定する。

$$p(y | z) = N[f(z), \sigma_y^2], \quad p(x | z) = N[0, \sigma_x^2]. \quad (2)$$

ただし, $N[\mu, \sigma^2]$ は平均 μ , 分散 σ^2 の正規分布を表す。 z の事前分布は (計算可能であれば) 任意の確率分布でよいが, 本稿では一様分布を仮定する。

また, 本稿では簡単のため上記のハイパーパラメータ σ_x, σ_y は既知とする。 x にのるノイズと y にのるノイズはここでは独立であると仮定したが, 相関がある場合もアルゴリズムは容易に拡張できる。

一方, 関数 f の事前分布はカーネル関数 $k(z, z')$ を共分散関数とする平均 0 の正規過程

$$E[f(z)] = 0, \quad \text{Cov}[f(z), f(z')] = k(z, z') \quad (3)$$

とする。カーネル関数は正定値である必要があり, 例えばガウスカーネル

$$k(z, z') = \exp(-\beta(z - z')^2) \quad (4)$$

などがある。

以上の生成モデルに基づいて n 個の入力と出力の組 $\mathcal{D} = (x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ が与えられたときに f (や (x_i, y_i) に対応する潜在変数 z_i) の事後分布を求めることが問題である。なお, f, z_1, \dots, z_n の事後分布は

$$p(f, z_1, \dots, z_n | \mathcal{D}) \propto \exp\left(-\frac{(y_i - f(z_i))^2}{2\sigma_y^2} - \frac{(z_i - x_i)^2}{2\sigma_x^2} - \frac{1}{2}\|f\|^2\right) \quad (5)$$

と書ける。ただし, $\|f\|$ は $k(z, z')$ の定める再生核ヒルベルト空間のノルムである。

3 マルコフ連鎖モンテカルロ法

マルコフ連鎖モンテカルロ法 [10] は, 事後分布を解析的に求めることが困難な場合にマルコフ連鎖を用いたシミュレーションによってモンテカルロサンプリングを行う手法である。マルコフ連鎖は, その定常分布が目的の事後分布となるように設計され, Metropolis-Hastings アルゴリズムや Gibbs サンプラなどのアルゴリズムが知られている。

我々の問題では, \mathcal{D} が与えられたときに潜在変数 z_1, z_2, \dots, z_n と f に関してシミュレーションを行う必要があるが, 同時に複数の確率変数に関する乱数を生成するの

は困難であるため, ほかの変数を固定して 1 変数ずつ更新する手法を用いる。すなわちまず, $z = (z_1, z_2, \dots, z_n)$ をランダムに初期化した上で, $i = 1, 2, \dots, n$ について

1. z を固定して f をサンプリングする
2. f と z_{-i} を固定して z_i をサンプリングする。

という 2 ステップで交互にサンプリングを行うことを一つのサイクルとし, サイクルを繰り返すことによって f のサンプル集合を得る。

なお, 以下の節で見ると, f のサンプリングは Gibbs サンプラと同様に目標分布の事後分布で生成させることができるが, z_i の方は複雑な分布となるので Metropolis-Hastings 法を用いる。

3.1 f のサンプリングのためのトリック

z を固定して f をサンプリングする際, f は一般に無限次元となる。単純なカーネル回帰では, 入力点が固定であるため Representer 定理によってその固定された入力点だけの関数として書くことができる。ただし, 測定誤差モデルでは入力点が潜在変数となっており, サンプリングによって時々刻々と変動するため, 入力点の可能性が無限に増えてしまう。例えば, f の値を, ある潜在変数 z_i に対して生成した $f(z_i)$ の値は, z_i が別の値 z'_i に変動してしまうともはや z'_i の関数値とみなすことはできない。

ただし, ほかの変数を固定して z_i だけを変動させるとき, それを採択するかどうかを決めるのに必要なのは後で見ると, $f(z_i)$ と $f(z'_i)$ の二つの値だけである。そこで, z_i のほかに, z_i が次に変動する候補値 z'_i をあらかじめある確率分布によって生成しておき, $f(z_i)$ と $f(z'_i)$ をランダムに生成し, その後で z'_i を採択するかどうかを決めるというトリックを使う。これは一見 z'_i と f の生成する順番が逆になってように見えるが, f のサンプリングは入力点とは独立して行うことができるため, このようなテクニックが使えるのである。

これにより (一般に無限自由度の) 操作:

「関数 f を事後分布から生成し, $f(z_1), \dots, f(z_n), f(z'_i)$ という値を評価すること」

が, 有限次元の操作:

「 z'_i の生成を f よりも先に行うことによって, $f(z_1), \dots, f(z_n), f(z'_i)$ の事後分布からサンプルを生成すること」

に還元できることを意味している。

z'_i の提案分布はエルゴード的であれば任意でよいが、本稿では x_i の近くに z_i があることを想定し、

$$p(z'_i) \propto \exp\left(-\frac{1}{2}(z'_i - x_i)^2\right) \quad (6)$$

に従って z'_i を生成する。

事後分布に従って f を生成する具体的な方法は 3.3 節で述べる。

3.2 z_i の採否

z_i が次にどこに動くかという候補点 z'_i は f のサンプリングを行う前に決めていたが、 $f(z_i), f(z'_i)$ を生成した後これを採用するかどうか決める手続きが必要となる。本稿では (6) 式のように z'_i を z_i とは独立に生成したので Metropolis-Hastings 法による採択確率は

$$\alpha = \exp\left(-\frac{(y_i - f(z'_i))^2 - (y_i - f(z_i))^2}{2\sigma_y^2}\right) \quad (7)$$

となり、 $\min(1, \alpha)$ の確率で z'_i を採択すればよい。

3.3 f のサンプリングの詳細

ここでは f のサンプリングを行う数値計算の詳細について述べる。

関数 f の統計量を計算するためには、サンプリングした f 自身を保存する必要があるが、先にも書いたように f は有限個の点でのみ評価可能な関数である。そこで本稿では、実軸上にグリッド点 u_1, u_2, \dots, u_m という有限個の点を追加し、これらの点で f の値をサンプリングして統計量を計算する。このように有限個グリッド点での値の評価を行っても、サンプリングしている関数全体は無限次元のままであることを注意する。

最終的に、与えられたデータ点と新規に生成した候補点、それから関数値の評価のためのグリッド点を合わせた $n + 1 + m$ 個の点

$$\mathbf{z} = (z_1, \dots, z_n, z'_i, u_1, \dots, u_m)^T \quad (8)$$

における f の値が出てきた。それらをまとめて

$$\mathbf{f} = (f(z_1), \dots, f(z_n), f(z'_i), f(u_1), \dots, f(u_m))^T \quad (9)$$

とおくと、 \mathbf{z} を固定した元での f の事後分布は

$$p(\mathbf{f} | \mathbf{z}, \mathcal{D}) \propto \exp\left(-\frac{1}{2\sigma_y^2}(\mathbf{y}_+ - \mathbf{f})^T J(\mathbf{y}_+ - \mathbf{f}) - \frac{1}{2}\mathbf{f}^T K^{-1}\mathbf{f}\right) \quad (10)$$

という多変量正規分布となる。ただし、

$$\mathbf{y}_+ = (\mathbf{y}^T, 0, \dots, 0)^T, \quad \mathbf{y} = (y_1, \dots, y_n)^T, \quad (11)$$

$$J = \begin{pmatrix} I_n & O \\ O & O \end{pmatrix} \quad (12)$$

であり、 K は \mathbf{z} に対するグラム行列で、 $n, 1, m$ 個の成分のブロックに分割すると

$$K = \begin{pmatrix} K_1 & \mathbf{k}_2 & K_3 \\ \mathbf{k}_2^T & k_3 & \mathbf{k}_4^T \\ K_3^T & \mathbf{k}_4 & K_5 \end{pmatrix} \quad (13)$$

となる。このうち z_i を z'_i に更新する場合に新たに計算が必要なのは $\mathbf{k}_2, k_3, \mathbf{k}_4$ の $n+1+m$ 個の成分だけである。 K_5 は最初に一度だけ計算しておけばよく、 K_1, K_3 も最初に一度計算して、 z'_i が採択された場合には、 $\mathbf{k}_2, k_3, \mathbf{k}_4$ を z_i に相当する成分 (i 行と i 列) に置き換えるだけでよい。

さて、これらの値をもとに f を実際にサンプリングするのだが、そのために平均 μ 、分散共分散行列 V を求めよう。まず平均は、上記の事後分布からも導出できるが、より計算効率がよい方法は以下の事実を利用したものである。

命題 1 \mathbf{z} を固定したときの f の事後分布の平均 (=正規分布の場合 MAP 解) は、 $(z_1, y_1), \dots, (z_n, y_n)$ に対する正則化付き最小二乗カーネル回帰の解 $f(\mathbf{z}) = \sum_{i=1}^n \alpha_i k(z_i, \mathbf{z})$ に \mathbf{z} の各要素を代入してできるベクトルに等しく、

$$\mu = \begin{pmatrix} K_1 \\ \mathbf{k}_2^T \\ K_3^T \end{pmatrix} (K_1 + \sigma_y^2 I_n)^{-1} \mathbf{y} \quad (14)$$

で与えられる。

これは Representer 定理からも理解できる。

次に分散共分散行列は (10) 式の 2 次形式を取りだして -2 倍して逆行列を取ればよく、

$$V = \left(K^{-1} + \frac{1}{\sigma_y^2} J\right)^{-1} \quad (15)$$

となる。ただし通常 K は 0 に近い固有値が多いため計算結果が不安定になりやすい。そのため K の Cholesky 分解を $K = L^T L$ としたときに、

$$V = L^T \left(I_{n+1+m} + \frac{1}{\sigma_y^2} L J L^T\right)^{-1} L \quad (16)$$

と変形したのを使った方が数値的に安定であり、さらに L を K の不完全 Cholesky 分解で近似すれば計算量も減らすことができる。

結果として、平均 μ と分散共分散行列 V を計算し、 $N[\mu, V]$ に従う多次元正規乱数を発生して f のサンプルとすればよい。

3.4 アルゴリズムのまとめ

以上説明してきたサンプリングアルゴリズムをまとめると以下ようになる。

1. 初期化: z_1, \dots, z_n を生成する ((6) 式) . また , (13) 式のグラム行列 K のうち , K_1, K_3, K_5 を計算する .
2. $i = 1, 2, \dots, n$ について以下を実行する
 - (a) z'_i を生成する ((6) 式)
 - (b) グラム行列 K の k_2, k_3, k_4 を計算する .
 - (c) μ, V を計算する ((14), (16) 式) .
 - (d) $N[\mu, V]$ に従って f を生成する .
 - (e) (7) 式の確率で z'_i を採択し , 採択されればグラム行列を更新し z_i を z'_i で置き換える .
3. ステップ 2 のサイクルを反復し , グリッド点における f を生成する .

4 数値実験

人工的に生成したデータを用いて数値実験を行った . データはまず $(-6, 6)$ 区間に一様分布で z を $n = 20$ 点生成し , 各 z_i から分散 σ_x^2 のノイズを加えて x_i を生成した . 一方 y_i は関数

$$f(z_i) = \exp(-z_i^2/5) \sin(z_i) \quad (17)$$

に分散 σ_y^2 のノイズを加えて生成した . カーネル関数にはいずれも $\beta = 1$ のガウスカーネルを用いた . 関数を評価するグリッド点は $(-6, 6)$ の間に 50 個等間隔に置いた .

$\sigma_x^2 = \sigma_y^2 = 0.1^2$ のときの実行例 (図 1) を示す . 提案手法と比較する手法として , 入力へのノイズを仮定しない通常のカーネル最小二乗回帰によるあてはめを行った .

提案手法の結果は繰り返しのサイクルを 100 サイクル行った単純期待値であり , burn-in 等の処理は行っていない . ノイズレベルが比較的小さいのでカーネル回帰との違いがわかりにくい (提案手法の実線とカーネル回帰の破線が重なっている部分が多い) が , この図では左端のデータがない部分でカーネル回帰のあてはまりが悪い . そこで , フィッティングを定量的に評価するために , データを生成した曲線との区間 $(-6, 6)$ での二乗距離 (グリッド点における関数値の二乗平均) の変化をプロットしたのが図 2 である . この例では二乗距離で 25 % 程度の改善が見られる .

これらのデータはあくまでワンショットのあてはめなので , 乱数の種を変えて 50 回行ったときの提案手法と

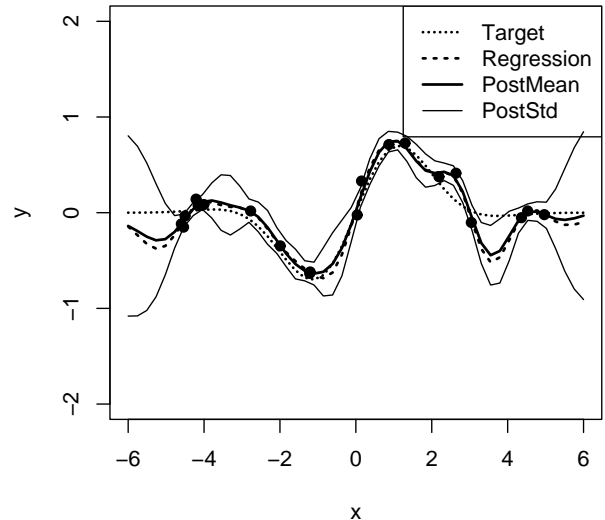


図 1: あてはめの例 . $\sigma_x^2 = \sigma_y^2 = 0.1^2$. 提案手法はモンテカルロ法によるグリッド点 $(\{u_i\}_{i=1}^m)$ における $f(u_i)$ の期待値と標準偏差 . 点線 (Target): データを生成した f , 破線 (Regression): 通常のカーネル回帰 , 実線 (PostMean): 提案手法 (期待値) , 細い実線 (PostStd): エラーバー (期待値 \pm 標準偏差)

通常のカーネル回帰との比較を行った . 結果は図 3 に示したとおり , ほとんどの場合でカーネル回帰の結果を上回っている ..

次に , ノイズレベルを $\sigma_x^2 = \sigma_y^2 = 0.3^2$ とかなり大きくした場合にどうなるかについて同様の実験をやった結果が図 4, 5, 6 である . 図 3 と図 6 を比べてみるとノイズレベルが高い方が提案手法による改善が大きい様子がわかる .

5 おわりに

測定誤差モデルにおいて , 再生核ヒルベルト空間上の関数の事後分布を生成するようなマルコフ連鎖モンテカルロ法を提案した . サンプリングの順序を入れ替えるトリックによって無限次元空間のサンプリングを有限次元のサンプリングに変換することができた .

測定誤差モデルは非常に推定が困難なモデルとされている [4] . 今後は理論的あるいは実験的に収束性や近似精度の分析が必要である . また , 位相応答曲線の推定など実データを用いた有効性の検証も今後の課題である .

謝辞 本研究は科研費 18500184 の助成を受けたものである .

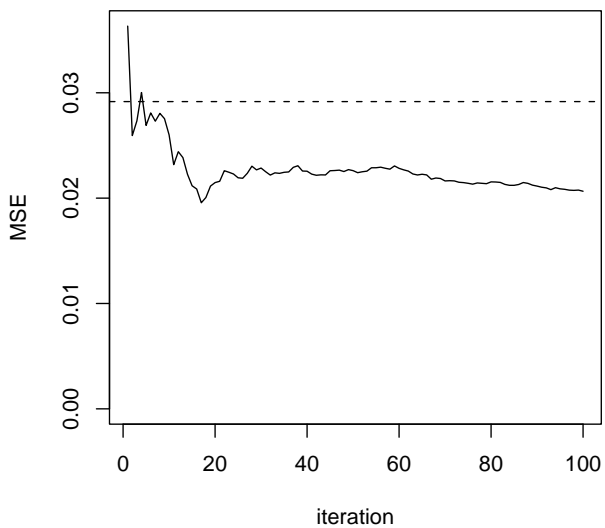


図 2: 二乗誤差の推移 ($\sigma_x^2 = \sigma_y^2 = 0.3^2$) . 破線は通常のカネル回帰による結果

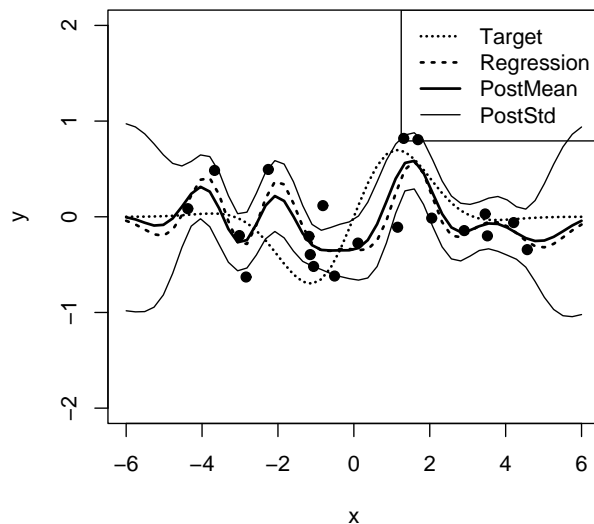


図 4: あてはめの例 ($\sigma_x^2 = \sigma_y^2 = 0.3^2$)

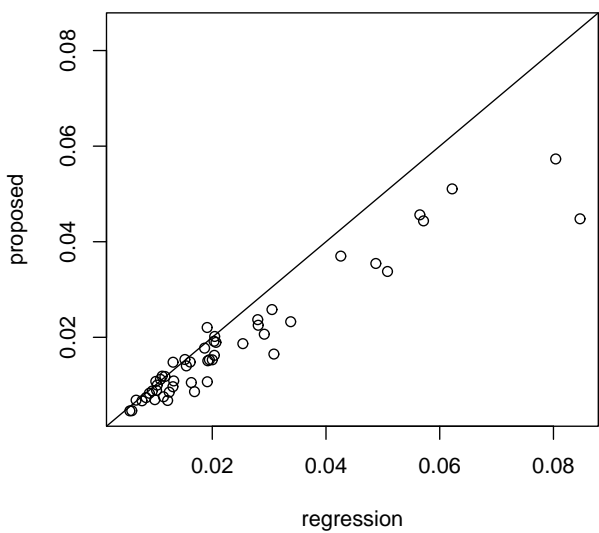


図 3: カネル回帰と提案手法との比較 (50 回, $\sigma_x^2 = \sigma_y^2 = 0.1^2$)

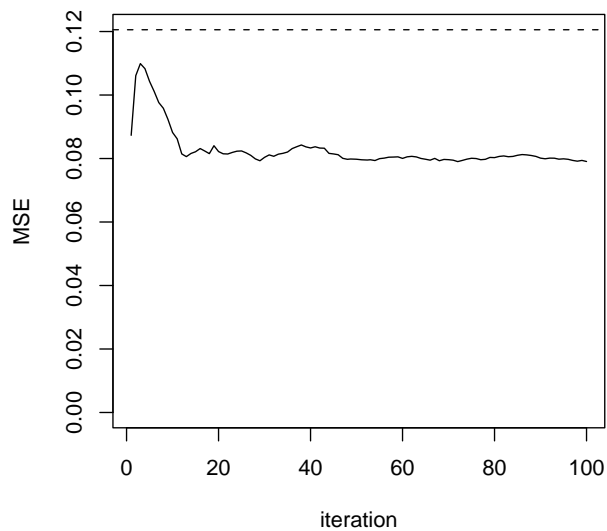


図 5: 二乗誤差の推移 ($\sigma_x^2 = \sigma_y^2 = 0.3^2$)

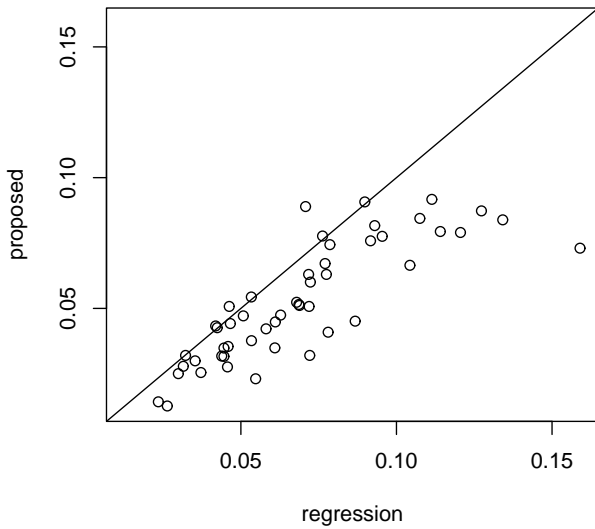


図 6: カーネル回帰と提案手法との比較 (50 回, $\sigma_x^2 = \sigma_y^2 = 0.3^2$)

参考文献

- [1] K. Nakae, Y. Iba, and T. Aoyagi. Statistical estimation of phase response curves and application in neural science. In *Int. Soc. for Bayesian Analysis 9th World Meeting*, 2008.
- [2] 中江, 伊庭, 青柳. レプリカ交換モンテカルロ法による位相応答曲線の統計的推定. 第 11 回情報論的学習理論ワークショップ (IBIS), 2008.
- [3] P. J. Bickel, C. A. J. Klaassen, Y. Ritov, and J. A. Wellner. *Efficient and Adaptive Estimation for Semiparametric Models*. Springer-Verlag, 1993.
- [4] 甘利, 川鍋. 線形関係の推定 - 最小 2 乗法は最良であるのか? *応用数理*, Vol. 6, No. 2, pp. 96–109, 1996.
- [5] S. M. Berry, R. J. Carroll, and D. Ruppert. Bayesian Smoothing and Regression Splines for Measurement Error Problems. *J. of the American Statistical Association*, Vol. 97, No. 457, pp. 160–169, 2002.
- [6] 赤穂. *カーネル多変量解析 — 非線形データ解析の新しい展開*. 確率の情報と科学. 岩波書店, 2008.
- [7] C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. MIT Press, 2006.

- [8] A. J. Smola, S. V. N. Vishwanathan, and T. Hofmann. Kernel methods for missing variables. In *Proc. of 10th Int. Workshop on AI& Statistics 2005*, pp. 325–334, 2005.
- [9] 藤木, 赤穂. 入力空間での計量に基づいた核主成分分析. *信学技報 PRMU2008-121*, Vol. 108, No. 327, pp. 69–74, 2008.
- [10] 伊庭, 種村, 大森, 和合, 佐藤, 高橋. *計算統計 II. 統計科学のフロンティア 6*. 岩波書店, 2005.