情報処理学会論文誌　数理モデル化と応用　Vol. 1　No. 1　17–26 (Sep. 2008)

外れ値データの発生を含む回帰モデルに対するベイズ予測アルゴリズム

須子統太[†1]　松嶋敏泰[†1]　平澤茂一[†2]

統計解析を行う際，得られたデータの中に外れ値が含まれることが多々ある．外れ値は少量であっても解析結果に大きく影響を与えることがあるため，従来から外れ値を含むデータに対する統計解析手法が数多く研究されている．従来，Box らにより線形回帰モデルに対し混合分布を用いて外れ値の発生をモデル化する研究が行われている．同様のモデルに対し様々な研究が行われているが，いずれも外れ値の検出やパラメータの推定を目的としている．そこで本研究では，外れ値データの発生を含む回帰モデルに対する予測法について扱う．まず，このモデルに対しベイズ基準のもとで最適な予測法を示す．しかし，この方法はデータ数に対し指数的に計算量が増大してしまう．そこで，EM アルゴリズムを用いて計算量を削減した近似アルゴリズムを提案し，シミュレーションにより有効性を検証する．

A Bayes Prediction Algorithm for Regression Models with Outliers

Tota Suko,[†1] Toshiyasu Matsushima[†1]
and Shigeichi Hirasawa[†2]

Outliers are often included in statistical data. A statistical analysis result is influenced from outliers. Therefore, there are many researches for a statistical analysis of data with outliers. Box modeled outliers using mixture distribution. There are many researches that aim parameter estimation or outlier detection about this model. In this paper, we treat prediction problem about this model. First, we present an optimal prediction method with reference to the Bayes criterion in this model. The computational complexity of this method grows exponentially with data size. Next, we propose an approximation algorithm reducing the computational complexity using EM algorithm, and evaluate this algorithm through some simulations.

1. はじめに

統計解析を行う際，得られたデータの中に外れ値が含まれることが多々ある．外れ値は転記ミスのようにデータに例外的な事象が含まれてしまう場合や，何らかの理由で別の母集団から発生したデータが混入してしまう場合など様々な要因によって発生する．外れ値は少量であっても解析結果に大きく影響を与えることがあるため，従来から外れ値を含むデータに対する解析手法が数多く研究されている．解析手法は主に，外れ値の発生に確率モデルを仮定する場合と仮定しない場合とに分けることができる[1)]．本研究では前者の外れ値の発生に確率モデルを仮定する場合を扱う．

従来，線形回帰モデルに対し混合分布を用いて外れ値の発生をモデル化する研究が行われている．Box らは，正常値の発生する分布と外れ値の発生する分布の混合分布を用いることで外れ値の発生をモデル化した[2)]．同様のモデルに対し様々な研究が行われているが，いずれの研究も外れ値の検出や混合分布のパラメータ推定を目的としている[3)–6)]．

他方，線形回帰モデルの主要な解析目的の 1 つに予測があげられる．予測問題は回帰分析だけではなく，パターン認識などデータ解析全般で扱われる基本的な問題の 1 つである．本研究では，外れ値データの発生を含む回帰モデルに対して，外れ値の検出ではなく予測を解析の目的とする．

また，統計的予測手法の 1 つにベイズ基準に基づく手法がある．ベイズ基準に基づく最適な予測法（以下，ベイズ最適な予測法と呼ぶ）については，従来から多くの研究が行われており，様々な有効性や性質が示されている[7)–10)]．本研究では外れ値データの発生を含む回帰モデルに対し，ベイズ基準に基づく最適な予測法を示す．しかしこれには，データ数に対して指数的に計算量が増えてしまうという問題点がある．

そこで，データ数が多い場合にも適用可能な計算量を削減した近似アルゴリズムを提案する．また，近似アルゴリズムの有効性についてシミュレーションにより評価を行う．

†1 早稲田大学基幹理工学部応用数理学科
Department of Applied Mathematics, School of Fundamental Science and Engineering, Waseda University
†2 早稲田大学創造理工学部経営システム工学科
Department of Industrial and Management Systems Engineering, School of Creative Science and Engineering, Waseda University

2. 混合分布による外れ値データのモデル化[2),3)]

i 番目のデータの説明変数を $\boldsymbol{x}_i = (1, x_{i1}, \ldots, x_{ip-1})^t$，目的変数を y_i とする．ただし，t は行列の転置を表すこととし，x_{ij}，$y_i \in \mathcal{R}$ とする．また，$\boldsymbol{\beta}_0$，σ_0^2 をそれぞれ正常値の従う p 次元回帰係数ベクトルおよび分散パラメータ，$\boldsymbol{\beta}_1$，σ_1^2 をそれぞれ外れ値の従う p 次元回帰係数ベクトルおよび分散パラメータとする．さらに，外れ値の発生する確率を α としたとき，y_i は確率 $1-\alpha$ で正規分布 $N(\boldsymbol{x}_i^t\boldsymbol{\beta}_0, \sigma_0^2)$ に従い発生し，確率 α で正規分布 $N(\boldsymbol{x}_i^t\boldsymbol{\beta}_1, \sigma_1^2)$ に従い発生すると仮定する．以降，正常値の分布のパラメータを $\boldsymbol{\theta}_0 = (\boldsymbol{\beta}_0, \sigma_0^2) \in \Theta_0$，外れ値の分布のパラメータを $\boldsymbol{\theta}_1 = (\boldsymbol{\beta}_1, \sigma_1^2) \in \Theta_1$ で表記し，全体の分布のパラメータを $\boldsymbol{\theta} = (\boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \sigma_0^2, \sigma_1^2) \in \Theta$ と表記する．このとき，y_i の確率分布を以下で定義する．

$$\begin{aligned} p(y_i|\boldsymbol{x}_i, \boldsymbol{\theta}) &= (1-\alpha)p_0(y_i|\boldsymbol{x}_i, \boldsymbol{\theta}_0) + \alpha p_1(y_i|\boldsymbol{x}_i, \boldsymbol{\theta}_1) \\ &= (1-\alpha)\frac{1}{\sqrt{2\pi\sigma_0^2}}\exp\left\{-\frac{1}{2\sigma_0^2}(y_i - \boldsymbol{x}_i^t\boldsymbol{\beta}_0)^2\right\} \\ &\quad + \alpha\frac{1}{\sqrt{2\pi\sigma_1^2}}\exp\left\{-\frac{1}{2\sigma_1^2}(y_i - \boldsymbol{x}_i^t\boldsymbol{\beta}_1)^2\right\}. \end{aligned} \tag{1}$$

また，$\boldsymbol{x}^n = (\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n)$，$y^n = (y_1, y_2, \cdots, y_n)$ とする．各データは独立に生起するものと仮定すると，y^n の確率分布は以下で表される．

$$p(y^n|\boldsymbol{x}^n, \boldsymbol{\theta}) = \prod_{i=1}^{n} p(y_i|\boldsymbol{x}_i, \boldsymbol{\theta}). \tag{2}$$

上記モデルは，パラメータ $\boldsymbol{\theta}$ に制約を置くことで，外れ値データの発生する構造を表現することができる．外れ値の発生する構造は，実際にデータをサンプリングする状況に応じて様々な場合が考えられる．Box らは，$\boldsymbol{\beta}_1 = \boldsymbol{\beta}_0$，$\sigma_1^2 = k^2\sigma_0^2$（$k$ は正の定数で既知），という制約を置き，外れ値は正常値よりも分散が大きいというモデルを表現している[2)]．それに対し，Abraham らは，$\sigma_1^2 = \sigma_0^2$ という制約を置くことで，平均値のみがずれるモデルを表現している[3)]．本研究では外れ値データが，正常値データとはまったく別の母集団から発生した場合を考え，$\boldsymbol{\theta}_0$ と $\boldsymbol{\theta}_1$ が独立であることを仮定する．これは，たとえば，本来解析対象としたいデータサンプルに，何かしらの要因で別の母集団から発生したデータサンプルが混ざってしまった場合などが考えられる．また，従来研究同様，外れ値の発生する確率 α は既知，パラメータ $\boldsymbol{\theta}$ は未知として扱う．

3. ベイズ最適な予測法

3.1 外れ値データの発生を含む回帰モデルにおける予測問題

本研究では予測問題として，説明変数と目的変数の n 個の組 $(\boldsymbol{x}^n, y^n)$ がデータとして得られたもとで，$\boldsymbol{x}_{n+1}$ が与えられたときの y_{n+1} の予測値 $\hat{y}_{n+1}$ を求める問題を扱う．

モデルとして外れ値データの発生を含む回帰モデルを仮定した場合，予測問題として大きく分けて 2 つの問題設定を考えることができる．1 つは，予測対象となる y_{n+1} が $p(y_{n+1}|\boldsymbol{x}_{n+1}, \boldsymbol{\theta})$ に従って発生すると仮定する問題と，もう 1 つは，$p_0(y_{n+1}|\boldsymbol{x}_{n+1}, \boldsymbol{\theta}_0)$ から発生すると仮定する問題である．前者の場合，予測対象となる y_{n+1} が外れ値の分布から発生する可能性があり，外れ値も予測する必要が出てくる．そのため，多くの場合では後者の，y_{n+1} は正常値の分布からのみ出現すると仮定することが考えられる．本研究では，どちらの問題設定においても同様の議論を行うことができるが，より一般的な式展開を示すため，特に断りのない限り前者の問題設定で議論を進める．

3.2 パラメトリックな確率モデルにおけるベイズ最適な予測[9)]

前述の外れ値データの発生を含む回帰モデルに限らず，$p(y_i|\boldsymbol{x}_i, \boldsymbol{\theta})$ のように，目的変数の出現構造がパラメトリックな確率モデルで表現される場合，ベイズ最適な予測法は一般に以下の手順で導出される[9)]．

一般に予測値は過去に得られたデータから導出される．そこで予測値をデータの関数として $\hat{y}_{n+1}(\boldsymbol{x}^n, y^n)$ と表す[⋆1]．次に，予測に対する損失関数を定義する．本研究では以下の平均二乗誤差損失を仮定する．

$$Loss(\hat{y}_{n+1}(\boldsymbol{x}^n, y^n), \boldsymbol{\theta}) = \int_{\mathcal{R}} (y_{n+1} - \hat{y}_{n+1}(\boldsymbol{x}^n, y^n))^2 p(y_{n+1}|\boldsymbol{x}_{n+1}, \boldsymbol{\theta}) dy_{n+1}. \tag{3}$$

損失関数は実際に得られたデータに依存した値をとるため，損失関数にデータの出現確率で期待値をとったものを危険関数として以下で定義する．

$$\begin{aligned} &Risk(\hat{y}_{n+1}, \boldsymbol{\theta}) \\ &= \int_{\mathcal{R}^n}\int_{\mathcal{R}^{n(p-1)}} Loss(\hat{y}_{n+1}(\boldsymbol{x}^n, y^n), \boldsymbol{\theta}) p(y^n|\boldsymbol{x}^n, \boldsymbol{\theta}) v(\boldsymbol{x}^n) dy^n\, d\boldsymbol{x}^n. \end{aligned} \tag{4}$$

ただし，$v(\cdot)$ は $\boldsymbol{x}^n$ の事前分布とする．本来であれば，この危険関数を最小にする $\hat{y}_{n+1}$ を求めたいのだが，パラメータによって危険関数を最小にする予測値は異なるため，任意の

⋆1 本来であれば $\hat{y}_{n+1}$ は $\boldsymbol{x}_{n+1}$ の関数になっているが，ここでは表記の簡略化のため省略する．

$\boldsymbol{\theta}$ すべてについて最小にする $\hat{y}_{n+1}$ は存在しない．そこで危険関数をパラメータの事前分布 $w(\boldsymbol{\theta})$ で期待値をとったベイズリスクを以下で定義する．

$$BR(\hat{y}_{n+1}) = \int_{\Theta} Risk(\hat{y}_{n+1}, \boldsymbol{\theta}) w(\boldsymbol{\theta}) d\boldsymbol{\theta}. \tag{5}$$

ベイズ最適な予測とはこのベイズリスクを最小にする予測を行うことであり，式 (3) の損失を仮定した場合次式で与えられる．

$$\hat{y}^*_{n+1} = \int_{\mathcal{R}^n} y_{n+1} \int_{\Theta} p(y_{n+1}|\boldsymbol{x}_{n+1}, \boldsymbol{\theta}) w(\boldsymbol{\theta}|\boldsymbol{x}^n, y^n) d\boldsymbol{\theta} dy_{n+1}. \tag{6}$$

このとき，$\int_{\Theta} p(y_{n+1}|\boldsymbol{x}_{n+1}, \boldsymbol{\theta}) w(\boldsymbol{\theta}|\boldsymbol{x}^n, y^n) d\boldsymbol{\theta}$，を予測分布と呼ぶ．つまり，ベイズ最適な予測は予測分布の期待値を求めることで得られる．

予測分布の計算にはパラメータ空間上での積分計算が必要となり，一般には解析的に解くことができない．そのため，積分計算を近似する手法として，MCMC 法[11] や変分ベイズ法[12] などが提案されている．また，y_i の分布が指数型の分布のとき，パラメータの事前分布に自然共役な事前分布が存在し，パラメータの事後分布 $w(\boldsymbol{\theta}|\boldsymbol{x}^n, y^n)$ が解析的に求まる．そのため予測分布も解析的に求まり，式 (6) も計算できることが知られている[7]．前述の外れ値データを含む回帰モデルは指数型の分布ではないので，自然共役な事前分布を構成できない．そのため式 (6) を直接，解析的に求めることはできない．

しかし，Box らによって提案されたパラメータの事後分布を求める手法[2] を応用することで，式 (6) のベイズ最適な予測値を解析的に求めることができる．

3.3　パラメータの事後分布の計算[2]

まず，次式で表される隠れ変数 z_i を導入する．

$$z_i = \begin{cases} 0 & y_i \text{ が正常値の分布から発生,} \\ 1 & y_i \text{ が外れ値の分布から発生.} \end{cases} \tag{7}$$

z_i は通常未知であるが，もし既知であった場合，y の分布は次式で表される．

$$p_Z(y_i|\boldsymbol{x}_i, \boldsymbol{\theta}, z_i) = \begin{cases} p_0(y_i|\boldsymbol{x}_i, \boldsymbol{\theta}_0) & z_i = 0, \\ p_1(y_i|\boldsymbol{x}_i, \boldsymbol{\theta}_1) & z_i = 1. \end{cases} \tag{8}$$

また，$z^n = (z_1, z_2, \cdots, z_n) \in Z^n$ とすると，

$$\begin{aligned} p_Z(y^n|\boldsymbol{x}^n, \boldsymbol{\theta}, z^n) &= \prod_{i \in \Gamma_0(z^n)} p_0(y_i|\boldsymbol{x}_i, \boldsymbol{\theta}_0) \prod_{j \in \Gamma_1(z^n)} p_1(y_j|\boldsymbol{x}_i, \boldsymbol{\theta}_1) \\ &= p_0(y^{\Gamma_0(z^n)}|\boldsymbol{x}^{\Gamma_0(z^n)}, \boldsymbol{\theta}_0) p_1(y^{\Gamma_1(z^n)}|\boldsymbol{x}^{\Gamma_1(z^n)}, \boldsymbol{\theta}_1), \end{aligned} \tag{9}$$

となる．ただし，$\Gamma_0(z^n) = \{i|z_i = 0, i = 1, 2, \cdots, n\}$，$y^{\Gamma_0(z^n)} = (y_i|i \in \Gamma_0(z^n))$，同様に $\Gamma_1(z^n) = \{i|z_i = 1, i = 1, 2, \cdots, n\}$，$y^{\Gamma_1(z^n)} = (y_i|i \in \Gamma_1(z^n))$，とする．また，$z_i$，$z^n$ の事前分布をそれぞれ以下で定義する．

$$q(z_i) = \begin{cases} 1 - \alpha & z_i = 0, \\ \alpha & z_i = 1, \end{cases} \tag{10}$$

$$q(z^n) = (1 - \alpha)^{|\Gamma_0(z^n)|} \alpha^{|\Gamma_1(z^n)|}. \tag{11}$$

z^n を用いることで，パラメータの事後分布 $w(\boldsymbol{\theta}|\boldsymbol{x}^n, y^n)$ は，z^n の事後確率 $q(z^n|\boldsymbol{x}^n, y^n)$ と，z^n が与えられたもとでの $\boldsymbol{\theta}$ の事後分布 $w_Z(\boldsymbol{\theta}|\boldsymbol{x}^n, y^n, z^n)$ を用いて次式で計算することができる．

$$w(\boldsymbol{\theta}|\boldsymbol{x}^n, y^n) = \sum_{z^n \in Z^n} q(z^n|\boldsymbol{x}^n, y^n) w_Z(\boldsymbol{\theta}|\boldsymbol{x}^n, y^n, z^n). \tag{12}$$

このことから，Box らは $\boldsymbol{\theta}$ の事前分布 $w(\boldsymbol{\theta})$ に，無情報事前分布を仮定することで，式 (12) が解析的に求まることを示した[2]．

本研究のパラメータ制約においても式 (12) が解析的に求まることを示す．今，$\boldsymbol{\theta}_0$ と $\boldsymbol{\theta}_1$ の独立性の仮定から次式が成り立つ．

$$w(\boldsymbol{\theta}) = w_0(\boldsymbol{\theta}_0) w_1(\boldsymbol{\theta}_1). \tag{13}$$

ただし，$w_0(\boldsymbol{\theta}_0)$，$w_1(\boldsymbol{\theta}_1)$ はそれぞれ，$\boldsymbol{\theta}_0$，$\boldsymbol{\theta}_1$ の事前分布とする．また，$\boldsymbol{\theta}$ と z^n は独立であることから，$w_Z(\boldsymbol{\theta}|z^n) = w(\boldsymbol{\theta})$，が成り立つ．以上とベイズの定理から，

$$\begin{aligned} w_Z(\boldsymbol{\theta}|\boldsymbol{x}^n, y^n, z^n) &= \frac{w_Z(\boldsymbol{\theta}|z^n) p_Z(y^n|\boldsymbol{x}^n, \boldsymbol{\theta}, z^n)}{\int_{\Theta} w_Z(\boldsymbol{\theta}|z^n) p_Z(y^n|\boldsymbol{x}^n, \boldsymbol{\theta}, z^n) d\boldsymbol{\theta}} \\ &= \frac{w_0(\boldsymbol{\theta}_0) p_0(y^{\Gamma_0(z^n)}|\boldsymbol{x}^{\Gamma_0(z^n)}, \boldsymbol{\theta}_0)}{\int_{\Theta_0} w_0(\boldsymbol{\theta}_0) p_0(y^{\Gamma_0(z^n)}|\boldsymbol{x}^{\Gamma_0(z^n)}, \boldsymbol{\theta}_0) d\boldsymbol{\theta}_0} \\ &\quad \times \frac{w_1(\boldsymbol{\theta}_1) p_1(y^{\Gamma_1(z^n)}|\boldsymbol{x}^{\Gamma_1(z^n)}, \boldsymbol{\theta}_1)}{\int_{\Theta_1} w_1(\boldsymbol{\theta}_1) p_1(y^{\Gamma_1(z^n)}|\boldsymbol{x}^{\Gamma_1(z^n)}, \boldsymbol{\theta}_1) d\boldsymbol{\theta}_1} \\ &= w_0(\boldsymbol{\theta}_0|\boldsymbol{x}^{\Gamma_0(z^n)}, y^{\Gamma_0(z^n)}) w_1(\boldsymbol{\theta}_1|\boldsymbol{x}^{\Gamma_1(z^n)}, y^{\Gamma_1(z^n)}), \end{aligned} \tag{14}$$

となる．今，$p_0(\cdot)$ と $p_1(\cdot)$ は正規分布であるため，$w_0(\boldsymbol{\theta}_0)$ と $w_1(\boldsymbol{\theta}_1)$ に正規分布に対する自然共役な事前分布を仮定すれば，事後分布 $w_0(\boldsymbol{\theta}_0|\boldsymbol{x}^{\Gamma_0(z^n)}, y^{\Gamma_0(z^n)})$，$w_1(\boldsymbol{\theta}_1|\boldsymbol{x}^{\Gamma_1(z^n)}, y^{\Gamma_1(z^n)})$ は解析的に求まる．以降，自然共役な事前分布として次式を仮定する．

$$w_0(\boldsymbol{\theta}_0) \propto (\sigma_0^2)^{-\frac{\nu_0'}{2}} \exp\{-\frac{1}{2}[\lambda_0' + (\boldsymbol{\beta}_0 - \boldsymbol{\beta}_0')^t C_0'(\boldsymbol{\beta}_0 - \boldsymbol{\beta}_0')]\}. \tag{15}$$

ここで，ν_0'，λ_0'，$\boldsymbol{\beta}_0'$，C_0' はハイパーパラメータとし，既知であるとする（$w_1(\boldsymbol{\theta}_1)$ についても同様の仮定を置く）．また，$q(z^n|\boldsymbol{x}^n, y^n)$ も同様に，

$$\begin{aligned} q(z^n|\boldsymbol{x}^n, y^n) &= \frac{q(z^n)\int_\Theta p_Z(y^n|\boldsymbol{x}^n, \boldsymbol{\theta}, z^n) w_Z(\boldsymbol{\theta}|z^n) d\boldsymbol{\theta}}{\sum_{z^n \in Z^n} q(z^n)\int_\Theta p_Z(y^n|\boldsymbol{x}^n, \boldsymbol{\theta}, z^n) w_Z(\boldsymbol{\theta}|z^n) d\boldsymbol{\theta}} \\ &\propto q(z^n) \int_{\Theta_0} p_0(y^{\Gamma_0(z^n)}|\boldsymbol{x}^{\Gamma_0(z^n)}, \boldsymbol{\theta}_0) w_0(\boldsymbol{\theta}_0) d\boldsymbol{\theta}_0 \\ &\quad \times \int_{\Theta_1} p_1(y^{\Gamma_1(z^n)}|\boldsymbol{x}^{\Gamma_1(z^n)}, \boldsymbol{\theta}_1) w_1(\boldsymbol{\theta}_1) d\boldsymbol{\theta}_1, \end{aligned} \tag{16}$$

と展開される．式 (15) を仮定すると，式 (16) の $\int_{\Theta_0} p_0(y^{\Gamma_0(z^n)}|\boldsymbol{x}^{\Gamma_0(z^n)}, \boldsymbol{\theta}_0) w_0(\boldsymbol{\theta}_0) d\boldsymbol{\theta}_0$ は t 分布の尤度として解析的に計算することができる．$\int_{\Theta_1} p_1(y^{\Gamma_1(z^n)}|\boldsymbol{x}^{\Gamma_1(z^n)}, \boldsymbol{\theta}_1) w_1(\boldsymbol{\theta}_1) d\boldsymbol{\theta}_1$ についても同様である．以上より，$w_0(\boldsymbol{\theta}_0)$ と $w_1(\boldsymbol{\theta}_1)$ に通常の線形回帰モデルに対する自然共役な事前分布を仮定することで式 (12) が解析的に求まることが分かる．

3.4　ベイズ最適な予測法の導出

以上の事後分布の計算法を用いることで，式 (6) は以下のように解析的に求めることができる．式 (6) に式 (12) を代入し展開する．

$$\begin{aligned} &\hat{y}_{n+1}^* \\ &= \int_{\mathcal{R}^n} y_{n+1} \int_\Theta p(y_{n+1}|\boldsymbol{x}_{n+1}, \boldsymbol{\theta}) \sum_{z^n \in Z^n} q(z^n|\boldsymbol{x}^n, y^n) w_Z(\boldsymbol{\theta}|\boldsymbol{x}^n, y^n, z^n) d\boldsymbol{\theta} dy_{n+1} \\ &= \sum_{z^n \in Z^n} q(z^n|\boldsymbol{x}^n, y^n) \\ &\times \int_{\mathcal{R}^n} y_{n+1} \int_{\Theta_0} \int_{\Theta_1} \{(1-\alpha) p_0(y_{n+1}|\boldsymbol{x}_{n+1}, \boldsymbol{\theta}_0) + \alpha p_1(y_{n+1}|\boldsymbol{x}_{n+1}, \boldsymbol{\theta}_1)\} \\ &\times w_Z(\boldsymbol{\theta}|\boldsymbol{x}^n, y^n, z^n) d\boldsymbol{\theta}_0 d\boldsymbol{\theta}_1 dy_{n+1} \\ &= \sum_{z^n \in Z^n} q(z^n|\boldsymbol{x}^n, y^n) \\ &\times \left\{ (1-\alpha) \int_{\mathcal{R}^n} y_{n+1} \int_{\Theta_0} p_0(y_{n+1}|\boldsymbol{x}_{n+1}, \boldsymbol{\theta}_0) w_0(\boldsymbol{\theta}_0|\boldsymbol{x}^{\Gamma_0(z^n)}, y^{\Gamma_0(z^n)}) d\boldsymbol{\theta}_0 dy_{n+1} \right. \\ &\left. + \alpha \int_{\mathcal{R}^n} y_{n+1} \int_{\Theta_1} p_1(y_{n+1}|\boldsymbol{x}_{n+1}, \boldsymbol{\theta}_1) w_1(\boldsymbol{\theta}_1|\boldsymbol{x}^{\Gamma_1(z^n)}, y^{\Gamma_1(z^n)},) d\boldsymbol{\theta}_1 dy_{n+1} \right\}. \end{aligned} \tag{17}$$

今，式 (17) の波括弧内第 1 項，$\int_{\Theta_0} p_0(y_i|\boldsymbol{x}_i, \boldsymbol{\theta}_0) w_0(\boldsymbol{\theta}|\boldsymbol{x}^{\Gamma_0(z^n)}, y^{\Gamma_0(z^n)}) d\boldsymbol{\theta}_0$，は通常の線形回帰モデルにおける予測分布であり，前述の自然共役事前分布を仮定すると積分計算が解析的に求まり，予測分布は t 分布で表される[7]．同様に第 2 項も t 分布となる．以上より，z^n を用いることでベイズ最適な予測法は，t 分布の期待値の混合を z^n の事後確率で重み付けることで計算されることが分かる[⋆1]．

また同様に，y_{n+1} が正常値の分布から発生すると仮定した場合でも，

$$p(y_{n+1}|\boldsymbol{x}_{n+1}, \boldsymbol{\theta}) = p_0(y_{n+1}|\boldsymbol{x}_{n+1}, \boldsymbol{\theta}_0), \tag{18}$$

と書き換えることで，ベイズ最適な予測値が次式のように解析的に求まる．

$$\begin{aligned} &\hat{y}_{n+1}^* \\ &= \sum_{z^n \in Z^n} q(z^n|\boldsymbol{x}^n, y^n) \\ &\times \int_{\mathcal{R}^n} y_{n+1} \int_{\Theta_0} p_0(y_{n+1}|\boldsymbol{x}_{n+1}, \boldsymbol{\theta}_0) w_0(\boldsymbol{\theta}|\boldsymbol{x}^{\Gamma_0(z^n)}, y^{\Gamma_0(z^n)}) d\boldsymbol{\theta}_0 dy_{n+1}, \end{aligned} \tag{19}$$

以上より，ベイズ最適な予測値が解析的に求まることが分かった．

式 (17)，(19) では，$q(z^n|\boldsymbol{x}^n, y^n)$ での重み付け和計算部分が計算量の主要項となる．z^n のとりうる値は全部で $|Z^n| = 2^n$ 個あるため，n に対し指数オーダの計算量が必要となる．そこで次に，計算量を削減した近似予測アルゴリズムを提案する．

4. 計算量を削減した近似アルゴリズム

和計算を減らす方法として，重み付ける z^n の集合を，Z^n からそれより小さい何らかの集合に制限するという方法が考えられる．最も単純な方法としては，事後確率 $q(z^n|\boldsymbol{x}^n, y^n)$ が最大となる z^n のみを用いて近似する方法や，$q(z^n|\boldsymbol{x}^n, y^n)$ の高い順にいくつかの z^n だけを重み付けて近似する方法が考えられる．しかし，$q(z^n|\boldsymbol{x}^n, y^n)$ を求めるには，指数オーダの計算量が必要となる．そこで計算量を抑えて，$q(z^n|\boldsymbol{x}^n, y^n)$ が高い z^n の集合を近似的

⋆1 式 (12) が解析的に求まれば，ベイズ最適な予測値を解析的に求めることができる．そのため，Box らのパラメータ制約においても，同様にベイズ最適な予測値を解析的に求めることができるが，本稿では議論の煩雑を避けるため，パラメータの独立性を制約した場合のみ示す．

に求める方法を考える．

EM アルゴリズムは，混合数の少ない混合正規分布のパラメータ推定において，非常に良い推定精度を示すことが知られている[13]．そこで，EM アルゴリズムより求められたパラメータの推定値 $\hat{\boldsymbol{\theta}}$ を利用し，$q(z_i|\boldsymbol{x}_i, y_i, \hat{\boldsymbol{\theta}})$ という分布を計算する．これは，パラメータの推定値が与えられたもとで，i 番目のデータが外れ値であるか，正常値であるかを表した分布になっており，良いパラメータの推定値が与えられれば，$q(z^n|\boldsymbol{x}^n, y^n)$ の高い z^n を見つける指標になると考えられる．また，$q(z_i|\boldsymbol{x}_i, y_i, \hat{\boldsymbol{\theta}})$ は，EM アルゴリズムの反復計算の中で計算され，簡単に求めることができる．提案近似アルゴリズムを以下に示す．

近似アルゴリズム

step1: EM アルゴリズムを用いて $q(z_i|\boldsymbol{x}_i, y_i, \hat{\boldsymbol{\theta}})$ を求める．

step1-1: パラメータの初期値 $\boldsymbol{\theta}^{(0)} = (\boldsymbol{\beta}_0^{(0)}, \sigma_0^{2(0)}, \boldsymbol{\beta}_1^{(0)}, \sigma_1^{2(0)})$ を設定する．

step1-2: 以下の E-step と M-step を $\boldsymbol{\theta}^{(l)}$ が収束するまで繰り返す．

E-step: l 回目の反復において，次式を計算する．

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(l-1)}) = \sum_{i=1}^{n} \sum_{z_i \in \{0,1\}} q(z_i|\boldsymbol{x}_i, y_i, \boldsymbol{\theta}^{(l-1)}) \log p_Z(y_i|\boldsymbol{x}_i, \boldsymbol{\theta}, z_i). \tag{20}$$

ここで，

$$q(z_i|\boldsymbol{x}_i, y_i, \boldsymbol{\theta}^{(l-1)}) = \frac{q(z_i) p_Z(y_i|\boldsymbol{x}_i, \boldsymbol{\theta}^{(l-1)}, z_i)}{\sum_{z_i \in \{0,1\}} q(z_i) p_Z(y_i|\boldsymbol{x}_i, \boldsymbol{\theta}^{(l-1)}, z_i)}, \tag{21}$$

である．

M-step: $q(z_i|\boldsymbol{x}_i, y_i, \boldsymbol{\theta}^{(l-1)})$ を重みとした，重み付き最小二乗法を用いて $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(l-1)})$ の最大化を行い，

$$\boldsymbol{\theta}^{(l)} = \arg\max_{\boldsymbol{\theta} \in \Theta} Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(l-1)}), \tag{22}$$

とする．

step1-3: 収束した $\boldsymbol{\theta}^{(l)}$ を $\hat{\boldsymbol{\theta}}$ とし，式 (21) から $q(z_i|\boldsymbol{x}_i, y_i, \hat{\boldsymbol{\theta}})$ をすべての $i = 1, 2, \cdots, n$ について求める．

step2: $\hat{z}_i$ を $i = 1, 2, \cdots, n$ について次式で求め，

$$\hat{z}_i = \begin{cases} 0 & q(z_i = 0|\boldsymbol{x}_i, y_i, \hat{\boldsymbol{\theta}}) > 0.5, \\ 1 & otherwise, \end{cases} \tag{23}$$

確信度 $r_i = |0.5 - q(\hat{z}_i|\boldsymbol{x}_i, y_i, \hat{\boldsymbol{\theta}})|$ を計算する．

step3: $i = 1, 2, \cdots, n$ について，r_i が何番目に小さいかを表す関数を $\eta(r_i)$ とし，$\Omega^A = \{i|\eta(r_i) \leq A, i = 1, 2, \cdots, n\}$ とする．このとき，以下の集合を求める．

$$\tilde{Z}^n(A) = \{(\tilde{z}_1, \tilde{z}_2, \cdots, \tilde{z}_n)|\tilde{z}_i \in \tilde{Z}_i(A), i = 1, 2, \cdots, n\}. \tag{24}$$

ただし，

$$\tilde{Z}_i(A) = \begin{cases} \{0, 1\} & i \in \Omega^A, \\ \{\hat{z}_i\} & otherwise, \end{cases} \tag{25}$$

とする．

step4: 次式で予測値を計算．

$$\begin{aligned} &\tilde{y}_{n+1} \\ &= \sum_{\tilde{z}^n \in \tilde{Z}^n(A)} \tilde{q}(\tilde{z}^n|\boldsymbol{x}^n, y^n) \\ &\times \left\{ (1-\alpha) \int_{\mathcal{R}^n} y_{n+1} \int_{\Theta_0} p_0(y_{n+1}|\boldsymbol{x}_{n+1}, \boldsymbol{\theta}_0) w_0(\boldsymbol{\theta}_0|\boldsymbol{x}^{\Gamma_0(\tilde{z}^n)}, y^{\Gamma_0(\tilde{z}^n)}) d\boldsymbol{\theta}_0 dy_{n+1} \right. \\ &+ \left. \alpha \int_{\mathcal{R}^n} y_{n+1} \int_{\Theta_1} p_1(y_{n+1}|\boldsymbol{x}_{n+1}, \boldsymbol{\theta}_1) w_1(\boldsymbol{\theta}_1|\boldsymbol{x}^{\Gamma_1(\tilde{z}^n)}, y^{\Gamma_1(\tilde{z}^n)}) d\boldsymbol{\theta}_1 dy_{n+1} \right\}. \end{aligned} \tag{26}$$

ただし，

$$\tilde{q}(\tilde{z}^n|\boldsymbol{x}^n, y^n) = \frac{q(\tilde{z}^n) \int_\Theta p_Z(y^n|\boldsymbol{x}^n, \boldsymbol{\theta}, \tilde{z}^n) w_Z(\boldsymbol{\theta}|\tilde{z}^n) d\boldsymbol{\theta}}{\sum_{\tilde{z}^n \in \tilde{Z}^n(A)} q(\tilde{z}^n) \int_\Theta p_Z(y^n|\boldsymbol{x}^n, \boldsymbol{\theta}, \tilde{z}^n) w_Z(\boldsymbol{\theta}|\tilde{z}^n) d\boldsymbol{\theta}}, \tag{27}$$

である．

□

上記近似アルゴリズムは $\tilde{Z}^n(A)$ を求めるところのみに EM アルゴリズムを用いている．具体的には，step3 において，確信度の低い A 個の z_i に対しては $\hat{z}_i$ の値が誤っている可能性が高いと考え，正常値となる場合，外れ値となる場合の両方のパターンを重み付ける集合に入れ，それ以外の z_i に対しては $\hat{z}_i$ の値に固定している．また，予測値自体は式 (26) の

ように，$\hat{\boldsymbol{\theta}}$ を使わず計算している．そのため，ベイズ最適な予測値との違いは，重み付ける z^n の集合を Z^n から $\tilde{Z}^n(A)$ に制限している部分のみである．

また，$|\tilde{Z}^n(A)| = 2^A$ となるので，近似アルゴリズムでは 2^A 個の z^n を重み付けていることになる．そのため，仮に $A = n$ とすると，近似アルゴリズムによる予測値とベイズ最適な予測値は一致するという特徴がある．

y_{n+1} が正常値の分布から発生すると仮定した場合には，式 (26) を次式に置き代えることで，同様の近似アルゴリズムを得ることができる．

$$\begin{aligned}\tilde{y}_{n+1} &= \sum_{\tilde{z}^n \in \tilde{Z}^n(A)} \tilde{q}(\tilde{z}^n|\boldsymbol{x}^n, y^n) \\ &\times \int_{\mathcal{R}^n} y_{n+1} \int_{\Theta_0} p_0(y_{n+1}|\boldsymbol{x}_{n+1}, \boldsymbol{\theta}_0) w_0(\boldsymbol{\theta}_0|\boldsymbol{x}^{\Gamma_0(\tilde{z}^n)}, y^{\Gamma_0(\tilde{z}^n)}) d\boldsymbol{\theta}_0 dy_{n+1}. \end{aligned} \tag{28}$$

5. シミュレーションによる評価

近似アルゴリズムの性能をシミュレーションにより評価する．以下，y_{n+1} が正常値の分布から発生すると仮定した場合についてシミュレーションを行う．

5.1 実験 1：z^n の重み付け数の変化による評価（データ数が少ない場合）

データ数が少ない場合には，事後確率 $q(z^n|\boldsymbol{x}^n, y^n)$ を正確に計算することができる．そこで，正確な $q(z^n|\boldsymbol{x}^n, y^n)$ が高い順に z^n をいくつか重み付ける場合と比較する．$q(z^n|\boldsymbol{x}^n, y^n)$ が高い順に重み付ける z^n の数を $1, 2, \cdots, 2^n$ と増やしていった場合と，近似アルゴリズムにより $A = 0, 1, \cdots, n$ と $\tilde{Z}^n(A)$ の集合を大きくしていった場合の予測誤差について調べた．

データは人工データを用いた．データ数 $n = 13$ とし，1 回の実験で 1 組の $\boldsymbol{\theta}$ を式 (15) の自然共役事前分布に従いランダムに発生させ，そのもとで 1 組の学習データ $(\boldsymbol{x}^n, y^n)$ と $(\boldsymbol{x}_{n+1}, y_{n+1})$ を発生させた．各実験で予測値の二乗誤差を測り，30000 回の実験について平均を求めた．実験結果を**図 1**，**図 2** に示す．縦軸に平均二乗誤差，横軸に重み付ける z^n の個数をとりプロットした．図 1，図 2 の近似アルゴリズムについてのプロットは，左から $A = 0, 1, \cdots, 13$ と A の値を増やしたときの予測誤差を点で示している．近似アルゴリズムと，事後確率の高い順に重み付けていった場合については，重み付け数 2^{13} の点は，ベイズ最適な予測値と等しい．

図 1，図 2 の結果から正確な事後確率の高い順番で重み付けを行う場合，重み付け数が非常に少ない段階で平均二乗誤差がベイズ最適な予測とほぼ同じ値になっていることが分

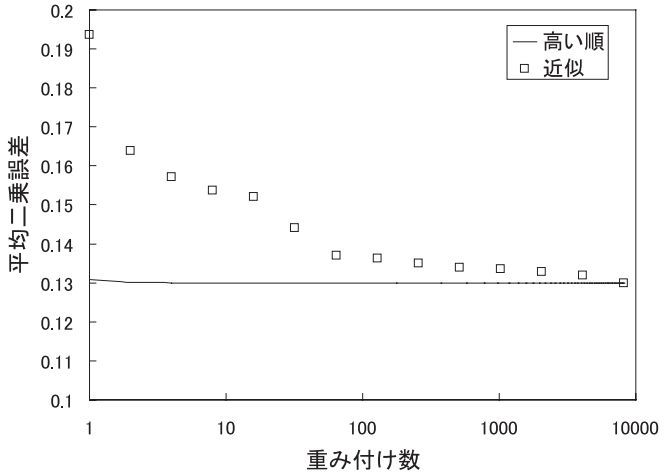

図 1　実験 1：重み付け数の変化による平均二乗誤差（$n = 13$，$\alpha = 0.1$）
Fig. 1　Experiment 1: mean square error when changing a weighting number ($n = 13, \alpha = 0.1$).

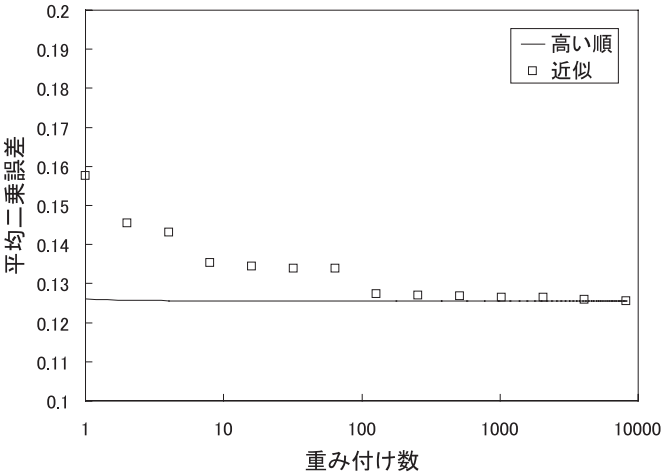

図 2　実験 1：重み付け数の変化による平均二乗誤差（$n = 13$，$\alpha = 0.05$）
Fig. 2　Experiment 1: mean square error when changing a weighting number ($n = 13, \alpha = 0.05$).

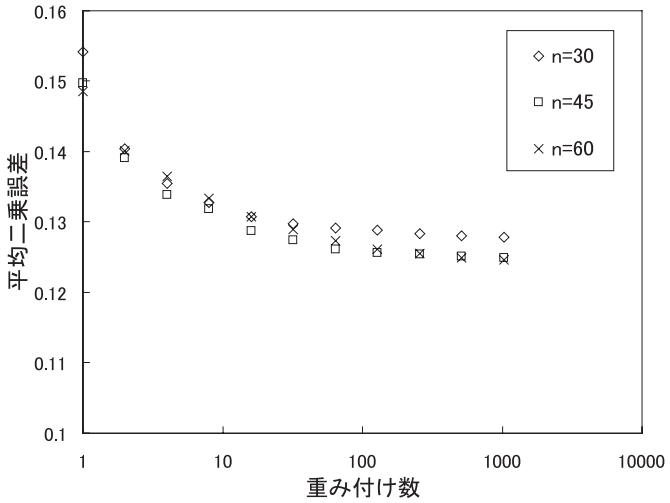

図 3　実験 2：重み付け数の変化による平均二乗誤差（$n = 30$，45，60，$\alpha = 0.1$）
Fig. 3　Experiment 2: mean square error when changing a weighting number ($n = 30, 45, 60$, $\alpha = 0.1$).

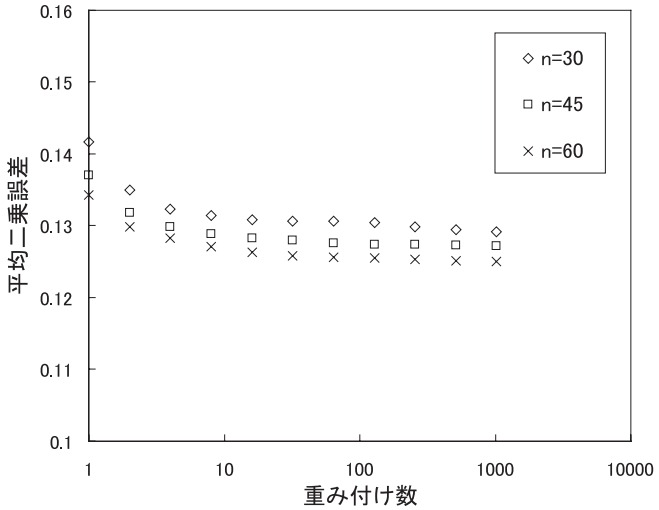

図 4　実験 2：重み付け数の変化による平均二乗誤差（$n = 30$，45，60，$\alpha = 0.05$）
Fig. 4　Experiment 2: mean square error when changing a weighting number ($n = 30, 45, 60$, $\alpha = 0.05$).

かる．そのため，事後確率の大きい z^n を見つけることができれば，すべての z^n の重み付けをしなくてもベイズ最適に近い予測が可能であると考えられる．また，近似アルゴリズムは A がある一定の値になると平均二乗誤差が収束し，ベイズ最適な予測に近づいている．このことから，収束した時点の A を用いることができれば，計算量を削減したもとで精度の高い予測が可能であるといえる．

5.2　実験 2：z^n の重み付け数の変化による評価（データ数が多い場合）

近似アルゴリズムについて，データ数が多い場合に実験 1 と同様の傾向が見られるのかを調べた．

実験 1 と同様の方法で発生させたデータに対し，データ数 $n = 30$，45，60 としたときの近似アルゴリズムの予測誤差を，実験 1 と同様にプロットした．実験結果を**図 3**，**図 4** に示す．

図 3，図 4 の結果から，データ数の違いによらず，ほぼ同じ A で平均二乗誤差が収束していることが分かる．このことから，A はデータ数 n によらず決めることができ，n が増えた場合でも，計算量を増やさずに精度の高い予測ができると考えられる．また，α の値が小さいときの方が，A が小さい段階で近似アルゴリズムとベイズ最適な予測の差がなくなり，近似アルゴリズムの効果が高いことが分かる．外れ値がデータに含まれる割合は，データに対して非常に小さい割合であると考えられるため，今回のような問題設定では，提案した近似アルゴリズムが有効であるといえる．

5.3　実験 3：データ数の変化による評価

近似アルゴリズムについて，データ数が変化した場合の予測誤差の変化を調べた．

実験 1 と同様の方法で発生させたデータに対し，$A = 0$，2，6，8，10 と固定したもとで，縦軸に平均二乗誤差，横軸にデータ数をとりプロットした．また比較のため，外れ値がすべて正確に分かっていたもとでのベイズ予測，EM アルゴリズムによって求めた推定値 $\hat{\boldsymbol{\theta}}$ を用いて，$\hat{y} = \boldsymbol{x}_{n+1}^t \hat{\boldsymbol{\beta}}_0$ と予測した場合，MCMC 法の 1 つであるギブスサンプラ法による予測[⋆1]，変分ベイズ法による予測についても実験を行った[⋆2]．実験結果を**図 5**，**図 6** に示す．

⋆1 今回の実験では提案近似アルゴリズムにおける $A = 10$ の場合と計算量を同等にするため，ギブスサンプラ法におけるサンプリング数は 2^{10} 個で固定した．
⋆2 ギブスサンプラ法と変分ベイズ法のシミュレーションについては，文献 14) を参考にした．

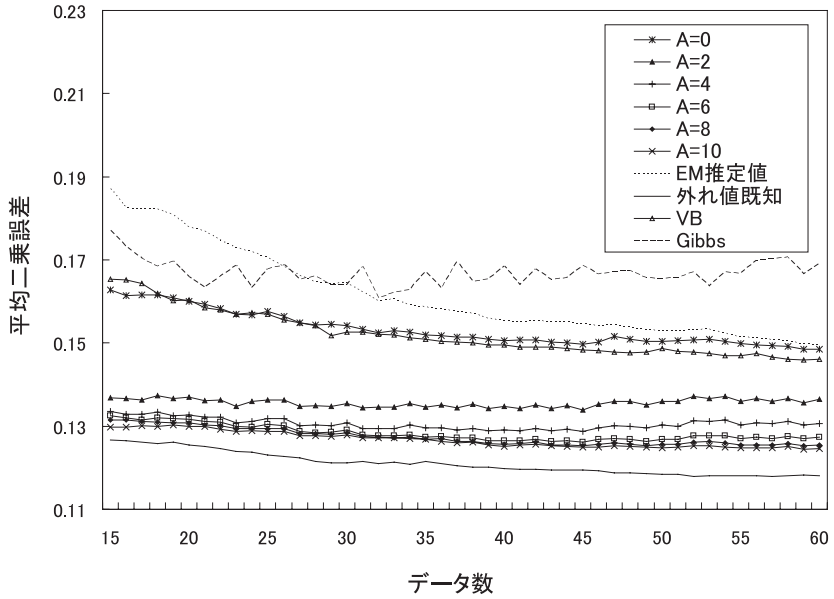

図 5　実験 3：データ数の変化による平均二乗誤差（$\alpha = 0.1$）
Fig. 5　Experiment 3: mean square error when changing a data number ($\alpha = 0.1$).

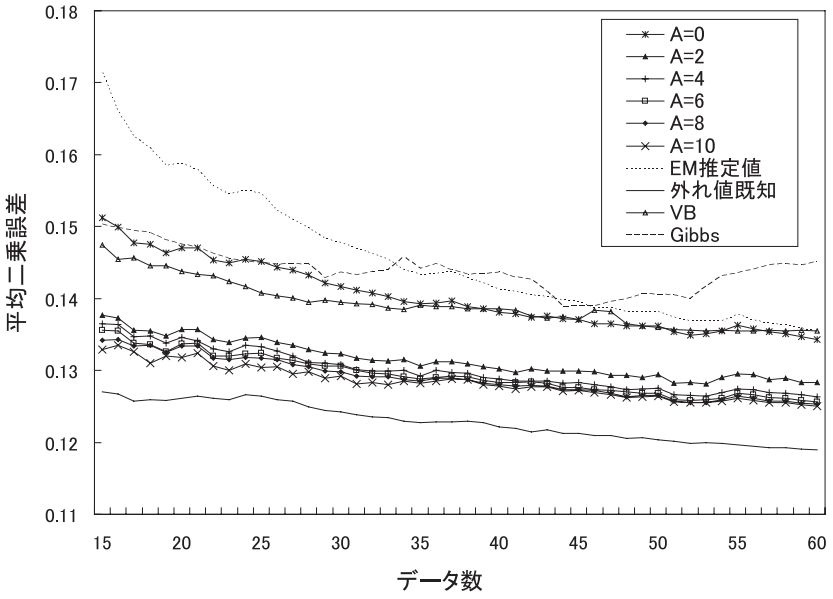

図 6　実験 3：データ数の変化による平均二乗誤差（$\alpha = 0.05$）
Fig. 6　Experiment 3: mean square error when changing a data number ($\alpha = 0.05$).

また，**図 7** に近似アルゴリズム中に用いられる EM アルゴリズムの平均反復回数をデータ数ごとに示した．

図 5，図 6 より，近似アルゴリズムは $A = 2$ 以上において，他の手法よりも予測精度が大幅に向上していることが分かる．また，ギブスサンプラ法はサンプル数が少ないため不安定な挙動を示している．

6. 考　　察

実験の結果からも，提案近似アルゴリズムが良い近似性能を持っていることが分かったが，それ以外の長所として，$A = n$ と設定することでベイズ最適に一致する点があげられる．変分ベイズ法の場合，いくら計算量をかけてもベイズ最適には一致しない．また，ギブスサンプラ法がベイズ最適に一致するには無限の計算量が必要となる．それに対し提案近似アルゴリズムは，計算量を多くかけることができる場合にはベイズ最適な予測が可能であり，A を変えることで計算量と予測精度のトレードオフを簡単に調整することができる．これは実用上有効な性質であると考えられる．

また，近似アルゴリズムの計算量について考察する．近似アルゴリズムは A を固定した場合，重み付ける z^n の個数は n によらず一定となる．そのため step1 以外にかかる計算量はデータ数の線形オーダとなる．また，step1 における EM アルゴリズムの反復回数は，図 7 から，α によって違いが出てくるものの，データ数に対しては，線形オーダより少ない増え方しかしていないことが分かる．EM アルゴリズムの 1 回の反復にかかる計算量はデータ数の線形オーダとなるので，近似アルゴリズム全体としてもデータ数の線形オーダ程度の計算量になると考えられる．

7. ま と め

本研究では，外れ値データの発生を含む回帰モデルに対し，ベイズ最適な予測法を示した．また，ベイズ最適な予測法はデータ数が増えると計算量が指数的に増えてしまうため，

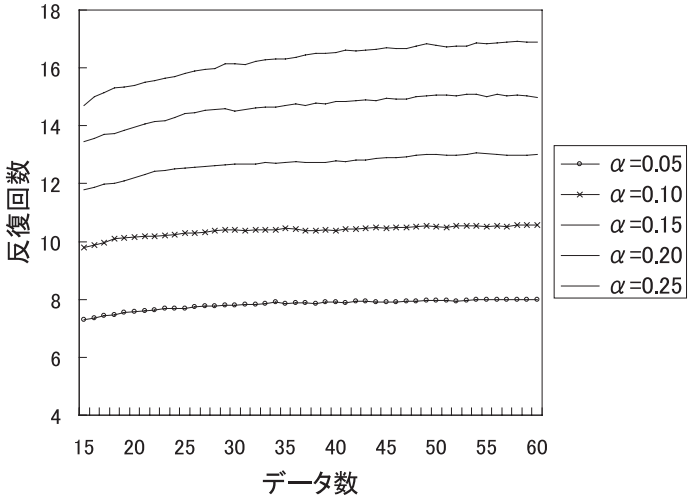

図 7　実験 3：データ数の変化による EM 反復回数
Fig. 7　Experiment 3: repetition of EM algorithm when changing a data number ($\alpha = 0.05$).

EM アルゴリズムを用いた近似アルゴリズムを提案し，その性能をシミュレーションによって示した．シミュレーションの結果，近似アルゴリズムは計算量を削減しながら，十分に精度の高い予測が可能であることが分かった．

今回提案した近似アルゴリズムは，EM アルゴリズムを用いることで近似的に事後確率の高い z^n の集合を求めた．しかし，変分ベイズ法などその他の手法を用いても，同様に z^n の集合を求めることができると考えられる．このような，他の z^n の集合を求める手法との比較検討については今後の課題としたい．

謝辞　本研究に関して貴重なご意見をいただきました，松嶋研究室，平澤研究室の方々に深く感謝いたします．本研究の一部は，日本学術振興会科学研究費基盤（C）（No.18560391）の援助による．

参 考 文 献

1) Barnett, V.: *Outliers in statistical data*, John Wiley&Sons (1994).
2) Box, G.E.P. and Tiao, G.C.: A Baysian approach to some outlier problems, *Biometrika*, Vol.55, No.1, pp.119–129 (1968).
3) Abraham, B. and Box, G.E.P.: Linear Models and Squrious Observations, *Applied Statistics*, Vol.27, No.2, pp.131–138 (1978).
4) 北川源四郎：異常値解析ベイズモデル，数理科学，No.213, pp.62–66 (1981).
5) Pena, D. and Guttman, I.: Comparing probabilitic methods for outlier detection in linear models, *Biometrika*, Vol.80, No.3, pp.603–610 (1993).
6) Hoeting, J., Raftry, A.E. and Madigan, D.: A Method for simultaneous Variable Selection and Outlier Identification in Linear Regression, *Computational Statistics and Data Analysis*, Vol.22, No.3, pp.251–270 (1989).
7) Bernardo, J.M. and Smith, A.F.M.: *Bayesian theory*, John Wiley & Sons (1994).
8) Berger, J.: *Statistical Decision Theory and Bayesian Analysis*, Springer (1993).
9) 松嶋敏泰：帰納・演繹推論と予測—決定理論による学習モデル，第 1 回情報論的学習理論ワークショップ予稿集，pp.1–8 (1998).
10) Komaki, F.: On asymptotic properties of predictive distributions, *Biometrika*, Vol.83, No.2, pp.299–313 (1996).
11) Hastings, W.: Monte Carlo Sampling Methods Using Markov Chains and Their Applications, *Biometrika*, Vol.57, pp.97–109 (1970).
12) Waterhouse, S.R., Mackay, D. and Robinson, A.J.: Bayesian Methods for Mixtures of Experts, *Advances in Neural Information Processing Systems*, Vol.8, pp.351–357 (1996).
13) Dempster, A.P., Laird, N.M. and Rubin, D.B.: Maximum likelihood from incomplete data via the EM algorithm, *J. of Royal Statistical Society Series B*, Vol.39, pp.1–38 (1977).
14) 上田修功：ベイズ学習のアルゴリズム—高次元積分の近似手法，人工知能学会誌，Vol.19, No.6, pp.656–663 (2004).

(平成 19 年 11 月 22 日受付)
(平成 20 年 1 月 10 日再受付)
(平成 20 年 1 月 28 日採録)

須子　統太

平成 13 年早稲田大学理工学部経営システム工学科卒業．平成 15 年同大学大学院修士課程修了．同大学院博士課程入学．情報源符号化および情報理論とその応用に関する研究に従事．

松嶋　敏泰（正会員）

昭和 53 年早稲田大学理工学部工業経営学科卒業．昭和 55 年同大学大学院修士課程修了．同年日本電気（株）入社．昭和 61 年早稲田大学大学院理工学研究科博士後期課程入学．平成元年横浜商科大学講師．平成 3 年同大学助教授．平成 4 年早稲田大学理工学部工業経営学科（現在経営システム工学科）助教授．平成 9 年同大学教授．現在に至る．知識情報処理および情報理論とその応用に関する研究に従事．工学博士．平成 13 年ハワイ大学客員研究員．IEEE，情報理論とその応用学会，人工知能学会，OR 学会，日本経営工学会等各会員．

平澤　茂一（正会員）

昭和 36 年早稲田大学理工学部数学科卒業．昭和 38 年同電気通信学科卒業．同年三菱電機入社．昭和 56 年早稲田大学理工学部工業経営学科（現在経営システム工学科）教授．現在に至る．情報理論とその応用，データ伝送方式，ならびに計算機応用システムの開発等の研究に従事．工学博士．昭和 54 年 UCLA 計算機科学科客員研究員．昭和 60 年ハンガリー科学アカデミー，昭和 61 年伊トリエステ大学客員研究員．平成 5 年電子情報通信学会小林記念特別賞，業績賞受賞．IEEE Fellow，情報理論とその応用学会，人工知能学会，OR 学会，日本経営工学会等各会員．