

## 異なる精度のデータを用いた予測法に関する考察

Study on prediction method from data with different precision

大阪大学 大学院情報科学研究科 土屋 耕治 (Koji Tsuchiya), 森田 浩 (Hiroshi Morita)  
Graduate School of Information Science and Technology,  
Osaka University

### 1 はじめに

データとは実験や調査, 観測等の試行の結果から得られるものであり, その採取および解析の目的は母集団の性質を知ることである. 同じ試行に対して必ずしも同じ結果が得られるとは限らない場合が多いため, 確率的な値として取り扱われる. また, 近年のデータ採取は技術の発達や手法の開発とともに多様化し, その精度も様々である. そのため, 目的に合わせて計画的に採取する必要があるとともに, その解析手法も考えなければならない.

データ解析のなかで予測に関する研究は多く見られる [1]. 一般に, 予測値の精度は得られたデータの精度とそのデータ数に依存している. しかし, この両者は相反する関係にある. 精度の低い器具や装置から得られたデータは誤差が大きく, 解析結果の信頼性は低下する. 一方で, 精度の高いデータは真値に近い値であると考えられるため信頼できる. さらにその数が多くあれば, 解析結果はより妥当なものとなる. しかし, 精度の高いデータを得るには, 精度の高い装置や整った環境といった多くの時間やコスト等を必要とする.

本研究ではトレードオフの関係にある精度の異なるデータを組み合わせた予測手法を提案する. 精度の低いデータからでは得られない予測精度を実現するとともに, 精度の高いデータのみを非効率的な予測を改善することでトレードオフの解消を図る. これにより, 従来よりも少ない時間とコストによる効率的な予測の可能性を探る.

### 2 モデル

本研究では精度の異なる標本を考え, 精度の差によってレベル化を行う. 実験や測定, 観測から得られた標本には不確実な要素が含まれているため, 標本にはばらつきが存在する. このばらつきの大きさを標本の精度として取り扱う. 真値を  $y$ , 標本  $Y$  とすると,  $Y = y + \epsilon$  と表される. ここで,  $\epsilon$  は誤差を表し, 正規分布  $N(0, \sigma^2)$  に従うため, 標本のばらつきを表す尺度として誤差分散  $\sigma^2$  を用いることができる. 精度の差を誤差分散の大きさで表し, 誤差分散の小さい標本を精度が高い, 誤差分散の大きい標本を精度が低いとする.

$t$  ( $t = 1, \dots, s$ ) 番目に精度の低い標本の組がレベル  $t$  に属するとし, 精度の高い標本がより高レベルに位置するように標本のレベル化を行う. レベル  $t$  に属する標本を  $Y_t(\mathbf{x})$ , その誤差を  $\epsilon_t$  とすると,

$$Y_t(\mathbf{x}) = y_0(\mathbf{x}) + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma_t^2) \quad (t = 1, \dots, s) \quad (1)$$

と表せるので、レベル化は誤差分散が

$$\sigma_1^2 > \sigma_2^2 > \dots > \sigma_s^2 \quad (t = 1, \dots, s) \quad (2)$$

を満たすように行う。最も精度の低い標本の組はレベル1、最も精度の高い標本の組はレベル $s$ となる。レベル $t$ の標本が必ずしもレベル $(t-1)$ 以下の標本よりも真値に近い値をとるとは限らないが、真値に近い値をとる確率が高い。そのため、高いレベルに属する標本ほど信頼できるものとして取り扱う。本研究では標本の誤差分散 $\sigma_t^2$ の値は未知であるが、異なる標本の組の誤差分散の大小関係を既知として、レベル化を行う。すなわち、式(2)が成り立つ条件下で予測値を求める。このレベル化により精度の異なる標本の組ごとに予測値を求めるとともに、精度の関係を1つの系列で見ることが可能になる。

このレベルの系列を時系列に見立てて予測値を求める。まず、低いレベルにおいて予測値を求める。この予測により精度の高いレベルの標本を加えることで新たな予測値を求める。精度の高い標本を多く得ることが困難な場合、高い予測精度を望むことができないが精度の低い標本の情報を用いることで予測精度を向上させる。標本数の少なさを下のレベルの情報で補う。高レベルに属する標本ほど信頼できる情報とし、レベル間の情報に対して重み付けを行う。低いレベルに属する標本を過去の時刻の標本と考えることで、時刻ごとに予測値を求めるとともに、時刻の新しい標本を加えていくことができる。このように低いレベルから高いレベルへと予測値の更新を行う。本研究では時系列解析などで用いられる手法である自己回帰モデルでレベル間の予測値の関係を考えていく [1, 2]。レベル $t$ における点 $\mathbf{x}$ の予測値 $Y_t(\mathbf{x})$ に対して、このモデルは

$$Y_1(\mathbf{x}) = \delta_1(\mathbf{x}) \quad (3)$$

$$Y_t(\mathbf{x}) = \rho_{t-1} \hat{Y}_{t-1}(\mathbf{x}) + \delta_t(\mathbf{x}) \quad (t = 2, \dots, s) \quad (4)$$

と仮定できる。 $\rho_{t-1}$ は自己回帰係数でレベル間の値に対する重みを表し、 $\delta_t(\cdot)$ は $Y_{t-1}(\cdot), \dots, Y_1(\cdot)$ に独立な無相関ガウス過程とする。 $\mathbf{h}(\cdot) = (h_1(\cdot), \dots, h_\nu(\cdot))$ を回帰させる次元および次数をもつ回帰ベクトルとして、 $\delta_t(\cdot)$ の平均を

$$E\{\delta_t(\cdot)\} = \mathbf{h}(\cdot)^T \boldsymbol{\beta}_t \quad (5)$$

とする。また、2点 $\mathbf{x}, \mathbf{x}'$ の距離のみに依存する関数とし、

$$\text{Cov}\{\delta_t(\mathbf{x}), \delta_t(\mathbf{x}')\} = \sigma_t^2 c_t(\mathbf{x}, \mathbf{x}') \quad (6)$$

$$c_t(\mathbf{x}, \mathbf{x}') = \exp\left\{-\gamma_t(\mathbf{x} - \mathbf{x}')^T(\mathbf{x} - \mathbf{x}')\right\} \quad (7)$$

と仮定する。

$\boldsymbol{\beta}_t$ は回帰ベクトル $\mathbf{h}(\cdot)$ と同じ次元 $\nu$ を持つ平均ベクトルで、あらゆる値をとる。共分散関数(6), (7)より、 $\sigma_t^2$ はある点 $\mathbf{x}$ における $\delta_t(\mathbf{x})$ の分散 $\text{Var}\{\delta_t(\mathbf{x})\} = \sigma_t^2$ を表すパラメータである。また、 $\gamma_t$ は相関を表すパラメータで大きければ相関は弱くなり、小さければ相関は強くなる。 $\sigma_t^2, \gamma_t$ は正の値をとる。自己回帰モデルでレベル間の重みを表す $\rho_{t-1}$ は、高レベルに属する標本ほど信頼できる情報とするため、0から1の値をとるパラメータとする。

### 3 予測手法

表記として、レベル  $t$  における  $n_t$  個のサンプリング点  $\mathbf{x}_1^{(t)}, \dots, \mathbf{x}_{n_t}^{(t)}$  を  $\mathbf{d}_t$ , 標本を  $\mathbf{Y}_t^{n_t} = \mathbf{Y}_t(\mathbf{d}_t) = (Y_t(\mathbf{x}_1^{(t)}), \dots, Y_t(\mathbf{x}_{n_t}^{(t)}))^T$ , ある点  $\mathbf{x}$  における予測値を  $\hat{Y}_t(\mathbf{x})$  とする.

モデルに対する予測分布を考える. レベル化した標本に対して, 標本を加えるごとに予測値の更新を行えるアルゴリズムを導く. 最初に事前分布の構築を行う. ある点  $\mathbf{x}_0$  における予測値を  $Y_t(\mathbf{x}_0)$  を考えたとき, 予測値  $Y_t(\mathbf{x}_0)$  と標本  $\mathbf{Y}_t^{n_t}$  の同時分布  $(Y_t(\mathbf{x}_0), \mathbf{Y}_t^{n_t})$  を事前分布とする. この分布は式 (3), (4) におけるガウス過程  $\delta_t(\mathbf{x})$  から, 多変量正規分布で表される. 事前分布に対して, 標本が与えられたときの予測値の条件付分布  $[Y_t(\mathbf{x}_0) | \mathbf{Y}_t^{n_t}]$  が事後分布となる. 予測値として事後分布の平均, 事後平均を採用すると,  $\hat{Y}_t(\mathbf{x}_0) = E\{Y_t(\mathbf{x}_0) | \mathbf{Y}_t^{n_t}\}$  で与えられる [3].

レベル 1 における予測を考える. これは標本の精度が全て同一の場合に等しい. このとき, パラメータを与えたときの予測値と標本の条件付同時分布は

$$\begin{pmatrix} Y_1(\mathbf{x}_0) \\ \mathbf{Y}_1^{n_1} \end{pmatrix} \Big| \beta_1, \sigma_1^2, \gamma_1 \sim N_{n_1+1} \left[ \begin{pmatrix} \mathbf{h}(\mathbf{x}_0)^T \\ H(\mathbf{d}_1)^T \end{pmatrix} \beta_1, \begin{pmatrix} \sigma_1^2 & \mathbf{r}_1(\mathbf{x}_0, \mathbf{d}_1)^T \\ \mathbf{r}_1(\mathbf{x}_0, \mathbf{d}_1) & V_1(\mathbf{d}_1) \end{pmatrix} \right] \quad (8)$$

と表される. これがレベル 1 における事前分布である. ここで,  $H(\mathbf{d}_1)$  は回帰ベクトル  $\mathbf{h}(\cdot)$  からなる行列で,

$$H(\mathbf{d}_1) = (\mathbf{h}(\mathbf{x}_1^{(1)}), \dots, \mathbf{h}(\mathbf{x}_1^{(n_1)})) \quad (9)$$

となる. 回帰させる次元および次数が  $\nu$ , 標本数が  $n_1$  の場合,  $H(\mathbf{d}_1)$  は  $\nu \times n_1$  行列である.

$\mathbf{r}_1(\mathbf{x}_0, \mathbf{d}_1)$  は予測値  $Y_1(\mathbf{x}_0)$  と各標本  $Y_1(\mathbf{x}_1^{(1)}), \dots, Y_1(\mathbf{x}_1^{(n_1)})$  の共分散からなるベクトルで式 (6), (7) を用いて

$$\mathbf{r}_1(\mathbf{x}_0, \mathbf{d}_1) = (\sigma_1^2 c_1(\mathbf{x}_0, \mathbf{x}_1^{(1)}), \dots, \sigma_1^2 c_1(\mathbf{x}_0, \mathbf{x}_1^{(n_1)}))^T \quad (10)$$

と表すことができる. 同様に, 共分散行列  $V_1(\mathbf{d}_1)$  は各標本  $Y_1(\mathbf{x}_1^{(1)}), \dots, Y_1(\mathbf{x}_1^{(n_1)})$  間の共分散からなるので,  $V_1(\mathbf{d}_1)$  の  $(i, j)$  成分は  $\sigma_1^2 c_t(\mathbf{x}_i^{(t)}, \mathbf{x}_j^{(t)})$  となる.

事前分布 (8) に対する事後分布  $[Y_1(\mathbf{x}_0) | \mathbf{Y}_1^{n_1}, \beta_1, \sigma_1^2, \gamma_1]$  も正規分布となる. 事後平均  $E\{Y_1(\mathbf{x}_0) | \mathbf{Y}_1^{n_1}, \beta_1, \sigma_1^2, \gamma_1\}$  を予測値とするので, レベル 1 における予測値は

$$\hat{Y}_1(\mathbf{x}_0) = \mathbf{h}(\mathbf{x}_0)^T \beta_1 + \mathbf{r}_1(\mathbf{x}_0, \mathbf{d}_1)^T V_1(\mathbf{d}_1)^{-1} (\mathbf{Y}_1^{n_1} - H(\mathbf{d}_1)^T \beta_1) \quad (11)$$

となり, これは最良線形不偏予測値 (BLUP) である.

また, 事後分布  $[Y_1(\mathbf{x}_0), Y_1(\mathbf{x}_0') | \mathbf{Y}_1^{n_1}, \beta_1, \sigma_1^2, \gamma_1]$  の共分散は

$$C_1\{Y_1(\mathbf{x}_0), Y_1(\mathbf{x}_0')\} = \sigma_1^2 c_1(\mathbf{x}_0, \mathbf{x}_0') - \mathbf{r}_1(\mathbf{x}_0, \mathbf{d}_1)^T V_1(\mathbf{d}_1)^{-1} \mathbf{r}_1(\mathbf{x}_0', \mathbf{d}_1) \quad (12)$$

となる.

次に, レベル  $t (t = 2, \dots, s)$  における予測を考える. レベル  $t$  におけるサンプリング点は  $\mathbf{d}_t$  であり, これら点における下のレベルでの予測値は  $\hat{Y}_{t-1}(\mathbf{d}_t)$  と表すことができる. 標本  $\mathbf{Y}_t^{n_t}$  と下の

レベルの予測値  $\hat{Y}_{t-1}(\mathbf{d}_t)$  を用いてレベル  $t$  の予測を行う。このときレベル  $t-1$  に以下に属する全パラメータをまとめて  $\psi_{t-1}$  で表す。

レベル  $t$  以下の全パラメータ  $\psi_t$  とレベル  $t-1$  の予測値  $\hat{Y}_{t-1}(\mathbf{d}_t)$  を与えたときの予測値  $Y(\mathbf{x}_0)$  と標本  $Y_t^{n_t}$  の条件付同時分布は

$$\begin{pmatrix} Y_t(\mathbf{x}_0) \\ Y_t^{n_t} \end{pmatrix} \Big| \psi_t, \hat{Y}_{t-1}(\mathbf{d}_t) \sim N_{n_t+1} \left[ \begin{pmatrix} m_t(\mathbf{x}_0) \\ M_t(\mathbf{d}_t) \end{pmatrix}, \begin{pmatrix} \Sigma_t & \mathbf{r}_t(\mathbf{x}_0, \mathbf{d}_t)^T \\ \mathbf{r}_t(\mathbf{x}_0, \mathbf{d}_t) & V_t(\mathbf{d}_t) \end{pmatrix} \right] \quad (13)$$

と表すことができ、これがレベル  $t$  における事前分布である。

モデルの式 (4) より、事前分布 (13) の平均  $m_t(\mathbf{x}_0)$ ,  $M_t(\mathbf{d}_t)$  はそれぞれ

$$m_t(\mathbf{x}_0) = \rho_{t-1} \hat{Y}_{t-1}(\mathbf{x}_0) + \mathbf{h}(\mathbf{x}_0)^T \beta_t \quad (14)$$

$$M_t(\mathbf{d}_t) = \rho_{t-1} \hat{Y}_{t-1}(\mathbf{d}_t) + H(\mathbf{d}_t)^T \beta_t \quad (15)$$

と表すことができる。ここで、 $H(\mathbf{d}_t)$  は式 (9) と同様に回帰ベクトル  $\mathbf{h}(\cdot)$  からなる行列で、点  $\mathbf{d}_t$  における回帰を表している。

また、事前分布 (13) の分散共分散行列も式 (4) より導かれる。(1,1) ブロック  $\Sigma_t$  は

$$\Sigma_t = \rho_{t-1}^2 \Sigma_{t-1} + \sigma_t^2 \quad (16)$$

となる。(1,2) ブロックと (2,1) ブロック  $\mathbf{r}_t(\mathbf{x}_0, \mathbf{d}_t)$  は予測値  $Y_t(\mathbf{x}_0)$  と各標本  $Y_t(\mathbf{x}_i^{(1)}), \dots, Y_t(\mathbf{x}_i^{(n_t)})$  の共分散からなるベクトルで、第  $k$  成分は

$$[\mathbf{r}_t(\mathbf{x}_0, \mathbf{d}_t)]_k = \rho_{t-1}^2 C_{t-1} \{Y_{t-1}(\mathbf{x}_0), Y_{t-1}(\mathbf{x}_i^{(k)})\} + \sigma_t^2 c_t(\mathbf{x}_0, \mathbf{x}_i^{(k)}) \quad (17)$$

と表すことができる。ここで、 $C_{t-1}(\cdot, \cdot)$  はレベル  $t-1$  における事後分布の共分散である。

同様に (2,2) ブロック  $V_t(\mathbf{d}_t)$  は各標本  $Y_t(\mathbf{x}_i^{(1)}), \dots, Y_t(\mathbf{x}_i^{(n_t)})$  間の共分散からなる行列で  $(i, j)$  成分は

$$[V_t(\mathbf{d}_t)]_{ij} = \rho_{t-1}^2 C_{t-1} \{Y_{t-1}(\mathbf{x}_i^{(t)}), Y_{t-1}(\mathbf{x}_j^{(t)})\} + \sigma_t^2 c_t(\mathbf{x}_i^{(t)}, \mathbf{x}_j^{(t)}) \quad (18)$$

と表すことができる。

事前分布 (13) に対する事後分布  $[Y_t(\mathbf{x}_0) | \psi_t, \hat{Y}_{t-1}(\mathbf{d}_t)]$  の平均、すなわち事後平均を予測値とするので、予測値は  $\hat{Y}_t(\mathbf{x}_0) = E\{Y_t(\mathbf{x}_0) | \psi_t, \hat{Y}_{t-1}(\mathbf{d}_t)\}$  より、

$$\hat{Y}_t(\mathbf{x}_0) = m_t(\mathbf{x}_0) + \mathbf{r}_t(\mathbf{x}_0, \mathbf{d}_t)^T V_t(\mathbf{d}_t)^{-1} (Y_t^{n_t} - H(\mathbf{d}_t) M_t(\mathbf{d}_t)) \quad (19)$$

となる。また、事後分布  $[Y_t(\mathbf{x}_0), Y_t(\mathbf{x}_0') | \psi_t, \hat{Y}_{t-1}(\mathbf{d}_t)]$  の共分散は

$$C_t \{Y_t(\mathbf{x}_0), Y_t(\mathbf{x}_0')\} = C_{t-1} \{Y_{t-1}(\mathbf{x}_0), Y_{t-1}(\mathbf{x}_0')\} - \mathbf{r}_t(\mathbf{x}_0, \mathbf{d}_t)^T V_t(\mathbf{d}_t)^{-1} \mathbf{r}_t(\mathbf{x}_0', \mathbf{d}_t) \quad (20)$$

と表すことができる。 $\mathbf{x}_0 = \mathbf{x}_0'$  のとき、事後分散  $\text{Var}\{Y_t(\mathbf{x}_0) | \psi_t, \hat{Y}_{t-1}(\mathbf{d}_t)\}$  となる。この予測値と事後分散をレベル  $t+1$  の予測に用いる。

式(11)および式(19)で表される予測値に関して、サンプリング点  $\mathbf{x}_t^{(k)}$  における値を考えると、 $\mathbf{r}_t(\mathbf{x}_t^{(k)}, \mathbf{d}_t)$  は  $V_t(\mathbf{d}_t)$  の第  $k$  行と等しくなるので、 $\mathbf{r}_t(\mathbf{x}_t^{(k)}, \mathbf{d}_t)^T V_t(\mathbf{d}_t)^{-1} = \mathbf{e}_k$  となる。また、 $M_t(\mathbf{d}_t)$  の第  $k$  成分は  $m_t(\mathbf{x}_k)$  であることから、予測値は

$$\hat{Y}_t(\mathbf{x}_t^{(k)}) = m_t(\mathbf{x}_t^{(k)}) + \mathbf{e}_k(\mathbf{Y}_t^{n_t} - M_t(\mathbf{d}_t)) = Y_t(\mathbf{x}_t^{(k)}) \quad (21)$$

となる。すなわち、サンプリング点における予測値は標本と等しくなる。

#### 4 パラメータの推定

モデルに含まれる各パラメータを標本から推定する。パラメータ  $\beta_t$  は任意の値、 $\sigma_t^2$ 、 $\gamma_t$  は正の値、 $\rho_{t-1}$  は 0 から 1 の値を取る。それぞれ事前情報は一様分布を用いると、各パラメータの事前確率はそれぞれ

$$p(\beta_t) \propto 1, \quad p(\sigma_t^2) \propto \frac{1}{\sigma_t^2}, \quad p(\gamma_t) \propto \frac{1}{\gamma_t}, \quad p(\rho_{t-1}) = \begin{cases} 1, & 0 \leq \rho_{t-1} \leq 1 \\ 0, & \rho_{t-1} < 0, \rho_{t-1} > 1 \end{cases} \quad (22)$$

と表すことができる。また、互いに独立であることから

$$p(\beta_t, \sigma_t^2, \gamma_t, \rho_{t-1}) \propto p(\sigma_t^2) p(\gamma_t) p(\rho_{t-1}) \quad (23)$$

となる。

レベル 1 に属するパラメータ  $\beta_1$ 、 $\sigma_1^2$ 、 $\gamma_1$  の同時事前確率は  $p(\beta_1, \sigma_1^2, \gamma_1) \propto p(\sigma_1^2) p(\gamma_1)$  となる。この確率に対して、標本  $\mathbf{Y}_1^{n_1}$  を与えたときの事後確率は

$$p(\beta_1, \sigma_1^2, \gamma_1 \mid \mathbf{Y}_1^{n_1}) \propto p(\mathbf{Y}_1^{n_1} \mid \beta_1, \sigma_1^2, \gamma_1) p(\sigma_1^2) p(\gamma_1) \quad (24)$$

となる。ここで、パラメータを与えたときの標本  $\mathbf{Y}_1^{n_1}$  の分布は  $\mathbf{Y}_1^{n_1} \mid \beta_1, \sigma_1^2, \gamma_1 \sim N_{n_1}[H(\mathbf{d}_1)\beta_1, V_1(\mathbf{d}_1)]$  となるため、式(24)は、

$$p(\beta_1, \sigma_1^2, \gamma_1 \mid \mathbf{Y}_1^{n_1}) \propto (\sigma_1^2 \gamma_1)^{-1} |V_1(\mathbf{d}_1)|^{-\frac{1}{2}} \times \exp\left\{-\frac{1}{2}(\mathbf{Y}_1^{n_1} - H(\mathbf{d}_1)\beta_1)^T V_1(\mathbf{d}_1)^{-1} (\mathbf{Y}_1^{n_1} - H(\mathbf{d}_1)\beta_1)\right\} \quad (25)$$

となる。

また、回帰パラメータ  $\beta_1$  の事前確率は  $p(\beta_1) \propto 1$  であるから、事後分布は  $\beta_1 \mid \sigma_1^2, \gamma_1, \mathbf{Y}_1^{n_1} \sim N_\nu[\hat{\beta}_1, B_1(\mathbf{d}_1)]$  となり、平均  $\hat{\beta}_1$  と分散  $B_1(\mathbf{d}_1)$  はそれぞれ

$$\hat{\beta}_1 = B_1(\mathbf{d}_1) H(\mathbf{d}_1)^T V(\mathbf{d}_1)^{-1} \mathbf{Y}_1^{n_1} \quad (26)$$

$$B_1(\mathbf{d}_1) = (H(\mathbf{d}_1)^T V(\mathbf{d}_1)^{-1} H(\mathbf{d}_1))^{-1} \quad (27)$$

となる。 $\beta_1$  の推定量を事後分布の平均  $\hat{\beta}_1$  とする。式(25)を  $\beta_1$  で積分すると、

$$p(\sigma_1^2, \gamma_1 \mid \mathbf{Y}_1^{n_1}) \propto (\gamma_1 \sigma_1^2)^{-1} |V_1(\mathbf{d}_1)|^{-\frac{1}{2}} |B_1(\mathbf{d}_1)|^{\frac{1}{2}} \times \exp\left\{-\frac{1}{2}(\mathbf{Y}_1^{n_1} - H(\mathbf{d}_1)\hat{\beta}_1)^T V_1(\mathbf{d}_1)^{-1} (\mathbf{Y}_1^{n_1} - H(\mathbf{d}_1)\hat{\beta}_1)\right\} \quad (28)$$

となる.

分散パラメータ  $\sigma_1^2$  を分離するために  $V_1(\mathbf{d}_1) = \sigma_1^2 V_1^*(\mathbf{d}_1)$ ,  $B_1(\mathbf{d}_1) = \sigma_1^2 B_1^*(\mathbf{d}_1)$  とおくと, 式 (28) は,

$$p(\sigma_1^2, \gamma_1 | \mathbf{Y}_1^{n_1}) \propto \gamma_1^{-1} (\sigma_1^2)^{-\frac{n_1-\nu}{2}+1} |V_1^*(\mathbf{d}_1)|^{-\frac{1}{2}} |B_1^*(\mathbf{d}_1)|^{\frac{1}{2}} \exp\left\{-\frac{(n_1-\nu)S^2}{2\sigma_1^2}\right\} \quad (29)$$

と書き直すことができる. ここで,

$$S^2 = \frac{1}{n_1 - \nu - 2} (\mathbf{Y}_1^{n_1} - H(\mathbf{d}_1)\hat{\beta}_1)^T V_1^*(\mathbf{d}_1)^{-1} (\mathbf{Y}_1^{n_1} - H(\mathbf{d}_1)\hat{\beta}_1) \quad (30)$$

である.

$\sigma_1^2$  の事前確率が  $p(\sigma_1^2) \propto 1/\sigma_1^2$  で与えられたときの事後分布は自由度  $n_1 - \nu$ , 尺度  $S^2$  の逆カイ 2 乗分布  $\sigma_1^2 | \rho_1, \mathbf{Y}_1^{n_1} \sim \chi^{-2}(n_1 - \nu, S^2)$  となる. この事後分布の平均と分散はそれぞれ

$$\hat{\sigma}_1^2 = E\{\sigma_1^2 | \rho_1, \mathbf{Y}_1^{n_1}\} = \frac{(n_1 - \nu)S^2}{n_1 - \nu - 2} \quad (31)$$

$$\text{Var}\{\sigma_1^2 | \rho_1, \mathbf{Y}_1^{n_1}\} = \frac{(n_1 - \nu)^2 S^4}{(n_1 - \nu - 2)^2 (n_1 - \nu - 4)} \quad (32)$$

である. 事後分布の平均  $\hat{\sigma}_1^2$  を  $\sigma_1^2$  の推定量とする. 式 (29) を  $\sigma_1^2$  で積分すると,

$$p(\gamma_1 | \mathbf{Y}_1^{n_1}) \propto \gamma_1^{-1} |V_1^*(\mathbf{d}_1)|^{-\frac{1}{2}} |B_1^*(\mathbf{d}_1)|^{\frac{1}{2}} (S^2)^{\frac{n_1-\nu}{2}} \quad (33)$$

となる.

パラメータ  $\beta_1, \sigma_1^2$  の推定量はそれぞれの事後平均としたが, 式 (33) を  $\gamma_1$  で積分計算することは容易ではない. そこで  $\gamma_1$  の事後平均を求める代わりに, 事後確率 (33) を最大にする  $\gamma_1$  を探索する. すなわち,

$$L(\gamma_1) = \log \gamma_1 + \frac{1}{2} \{ \log |V_1^*(\mathbf{d}_1)| - \log |B_1^*(\mathbf{d}_1)| + (n_1 - \nu) \log S^2 \} \quad (34)$$

とおくと,  $\gamma_1$  の推定量は  $\hat{\gamma}_1 = \min[\log L(\gamma_1)] (\gamma_1 > 0)$  となる.

次に, レベル  $t (t = 2, \dots, s)$  におけるパラメータの推定を行う. このとき, レベル  $t-1$  以下に属する全パラメータの推定量をまとめて  $\hat{\psi}_{t-1}$  とおく. サンプル点  $\mathbf{d}_t$  におけるレベル  $t-1$  の予測値  $\hat{Y}_{t-1}(\mathbf{d}_t)$  とパラメータ  $\hat{\psi}_{t-1}$  を与えられたときのレベル  $t$  の標本  $\mathbf{Y}_t^{n_t}$  の分布を考えると, 式 (15), (18) を用いて

$$\mathbf{Y}_t^{n_t} | \beta_t, \sigma_t^2, \gamma_t, \rho_{t-1}, \hat{Y}_{t-1}(\mathbf{d}_t), \hat{\psi}_{t-1} \sim N_{n_t} [M(\mathbf{d}_t), V_t(\mathbf{d}_t)] \quad (35)$$

と表すことができる.

表記として, レベル  $t$  におけるパラメータの推定を行うとき,  $\hat{Y}_{t-1}(\mathbf{d}_t)$ ,  $\hat{\psi}_{t-1}$  は常に与えられているため,  $\hat{Y}_{t-1}(\mathbf{d}_t)$ ,  $\hat{\psi}_{t-1}$  を以下省略する.

標本  $\mathbf{Y}_t^{n_t}$  を与えたときのパラメータ  $\beta_t, \sigma_t^2, \gamma_t, \rho_{t-1}$  の同時確率は

$$p(\beta_t, \sigma_t^2, \gamma_t, \rho_{t-1} | \mathbf{Y}_t^{n_t}) \propto p(\mathbf{Y}_t^{n_t} | \beta_t, \sigma_t^2, \gamma_t, \rho_{t-1}) p(\beta_t, \sigma_t^2, \gamma_t, \rho_{t-1}) \quad (36)$$

となり、パラメータの同時事前確率 (23) と標本  $Y_t^{n_t}$  の分布 (35) から、式 (36) は

$$p(\beta_t, \sigma_t^2, \gamma_t, \rho_{t-1} \mid Y_t^{n_t}) \propto p(\sigma_t^2)p(\gamma_t)p(\rho_{t-1})|V_t(\mathbf{d}_t)|^{-\frac{1}{2}} \\ \times \exp\left\{-\frac{1}{2}(\mathbf{Y}_t^{n_t} - M_t(\mathbf{d}_t))^T V_t(\mathbf{d}_t)^{-1}(\mathbf{Y}_t^{n_t} - M_t(\mathbf{d}_t))\right\} \quad (37)$$

となる。

回帰パラメータ  $\beta_t$  の事前確率は  $p(\beta_t) \propto 1$  であるから、事後分布は  $\beta_t \mid \sigma_t^2, \gamma_t, \rho_{t-1}, Y_t^{n_t} \sim N_\nu[\hat{\beta}_t, B_t(\mathbf{d}_t)]$  となる。ここで、平均  $\hat{\beta}_t$  と分散  $B_t(\mathbf{d}_t)$  はそれぞれ

$$\hat{\beta}_t = B_t(\mathbf{d}_t)H(\mathbf{d}_t)^T V(\mathbf{d}_t)^{-1}(\mathbf{Y}_t^{n_t} - \rho_{t-1}\hat{Y}_{t-1}(\mathbf{d}_t)) \quad (38)$$

$$B_t(\mathbf{d}_t) = (H(\mathbf{d}_t)^T V(\mathbf{d}_t)^{-1}H(\mathbf{d}_t))^{-1} \quad (39)$$

である。事後平均の  $\hat{\beta}_t$  を  $\beta$  の推定量とする。

式 (37) を  $\beta_t$  で積分すると、

$$p(\sigma_t^2, \gamma_t, \rho_{t-1} \mid Y_t^{n_t}) \propto p(\rho_{t-1})(\gamma_t \sigma_t^2)^{-1} |V_t(\mathbf{d}_t)|^{-\frac{1}{2}} |B_t(\mathbf{d}_t)|^{\frac{1}{2}} \exp\left\{-\frac{1}{2}T(\mathbf{d}_t)\right\} \quad (40)$$

となる。ただし、

$$T(\mathbf{d}_t) = (\mathbf{Y}_t^{n_t} - H(\mathbf{d}_t)\hat{\beta}_t)^T V_t(\mathbf{d}_t)^{-1}(\mathbf{Y}_t^{n_t} - H(\mathbf{d}_t)\hat{\beta}_t) \quad (41)$$

である。

パラメータ  $\beta_t$  の推定量は事後平均としたが、式 (40) において、 $\sigma_t^2, \gamma_t, \rho_{t-1}$  で積分を実行することは困難である。そこで、それぞれの事後平均を求める代わりに、事後確率 (40) を最大にする  $(\sigma_t^2, \gamma_t, \rho_{t-1})^T$  を探索する。すなわち、

$$L_t(\sigma_t^2, \gamma_t, \rho_{t-1}) = \log \sigma_t^2 + \log \gamma_t + \frac{1}{2} \{ \log |V_t(\mathbf{d}_t)| - \log |B_t(\mathbf{d}_t)| + T(\mathbf{d}_t) \} \quad (42)$$

とおくと、 $\sigma_t^2, \gamma_t, \rho_{t-1}$  の推定量は  $(\hat{\sigma}_t^2, \hat{\gamma}_t, \hat{\rho}_{t-1}) = \min [L_t(\sigma_t^2, \gamma_t, \rho_{t-1})]$  ( $\sigma_t^2 > 0, \gamma_t > 0, 0 \leq \rho_{t-1} \leq 1$ ) となる。

## 5 数値実験

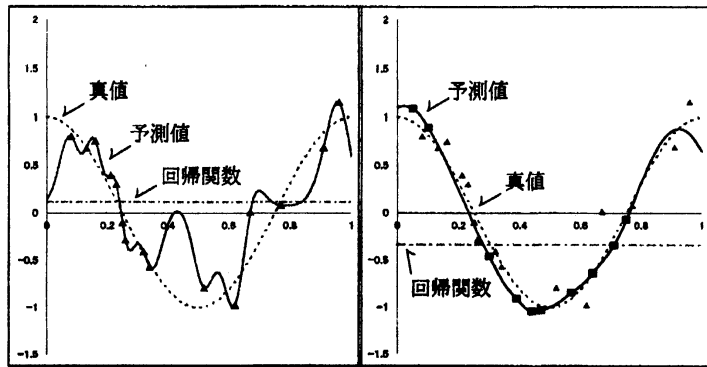
精度の異なる標本の組に提案手法を適用した数値実験を行い、その予測精度および効率性を評価する。関数  $f(x) = \cos(2\pi x)$  ( $0 \leq x \leq 1$ ) に対して、2組の異なる精度の標本を無作為に選び、予測を行う。

レベル1の標本を15点、レベル2の標本を10点抽出した場合の予測値の変化を調べる。誤差分散をそれぞれ  $\sigma_1^2 = 0.2^2, \sigma_2^2 = 0.1^2$  とするが、数値実験を行う上では未知の値とする。予測に用いる回帰関数を0次、1次、2次と変えた場合、すなわち、 $\mathbf{h}(\mathbf{x})^T \beta = \beta_0, \mathbf{h}(\mathbf{x})^T \beta = \beta_0 + \beta_1 x, \mathbf{h}(\mathbf{x})^T \beta$

$= \beta_0 + \beta_1x + \beta_2x^2$  にした場合の予測値をそれぞれ図 1, 図 2, 図 3 に示す. 各図において, (a) はレベル 1 における予測, (b) はレベル 2 における予測を示し, レベル 1 の標本  $Y_1^{15}$  を ■, レベル 2 の標本  $Y_2^{10}$  を ▲ で表わす.

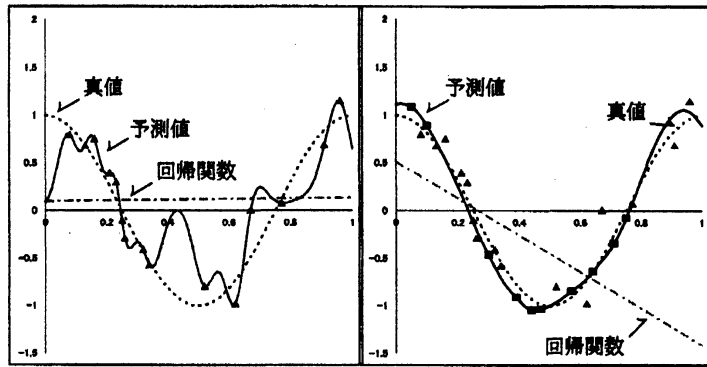
レベル 1, 2 ともに予測値はサンプリング点を必ず通過する. レベル 1 の予測値はサンプリング点がない付近では回帰関数の影響を受けている. これは式 (6), (7) より, 2 点間の距離が大きくなると相関関係が弱くなるため, 標本の影響より回帰の影響が大きくなる. また, レベル 2 の予測値は回帰関数に関わらず, 真値に近い曲線を描いている. 特に  $x \geq 0.8$  の範囲でレベル 2 の標本はないが, 回帰関数の影響よりもレベル 1 における予測値の影響を大きく受けている. そのため, レベル 2 では回帰関数の次数に依らず, 真値に近い曲線を描いている.

提案手法の予測精度を評価する. 予測は区間  $[0, 1]$  から等間隔に選んだ点  $\{0, 0.01, 0.02, \dots, 1\}$  の計 101 点で行う. しかし, 区間  $[0, 1]$  の両端における予測はサンプリング点の有無,  $x^2$  の回帰係数の値に大きく影響するため, 区間  $[0.2, 0.8]$  の点における予測値を用いる. 予測精度の評価は  $SE = \sum (\hat{Y}(x_k) - f(x_k))^2$  で行う. ここで,  $\hat{Y}(x_k)$ ,  $f(x_k)$  はそれぞれ予測値と真値である. 2 つの異なる精度を用いた場合, レベル 2 における予測値が  $\hat{Y}(x_k)$  となる. サンプリング点を無作為に選び, 50 回の予測を行う. その結果を SE の平均を表 1 に示す. また, 1 つの精度のみを用いた予測, すなわち, 式 (11) で得られる予測と比較するとともに, の重み付最小 2 乗法 (WLS) による予測を行う. 誤差分散を既知とし,



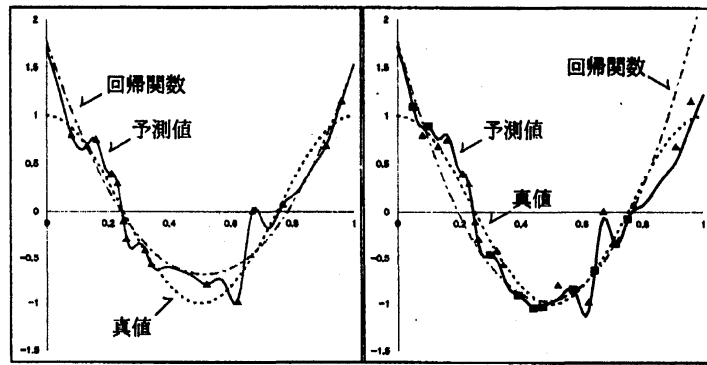
(a)  $Y_1$  を用いた 0 次回帰 (b)  $Y_2$  を加えた 0 次回帰

図 1:  $Y_1^n, Y_2^n$  を用いた 0 次回帰による予測



(a)  $Y_1$  を用いた 1 次回帰 (b)  $Y_2$  を加えた 1 次回帰

図 2:  $Y_1^n, Y_2^n$  を用いた 1 次回帰による予測



(a)  $Y_1$  を用いた 2 次回帰 (b)  $Y_2$  を加えた 2 次回帰

図 3:  $Y_1^n, Y_2^n$  を用いた 2 次回帰による予測



重みを誤差分散の逆数で与える。

表 1 より、回帰関数がいずれの場合も異なる精度を組合せた予測は精度が高い。2 次回帰の場合を見ると、 $Y_1$  を 15 点、 $Y_2$  を 10 点を組合せて行った予測の SE は 1.47 となり、 $Y_2$  のみを 17 点用いた予測よりも良い予測が行えていることがわかる。 $Y_2$  をさらに 7 点得ることが困難な場合、精度の異なる標本を組合せた予測が効率的である。

表 1: 予測結果

予測手法	標本数		0 次	1 次	2 次
	$Y_1$	$Y_2$			
異なる精度	15	10	2.91	2.16	1.47
単一精度	28	-	6.17	6.66	1.59
	-	17	9.52	9.03	1.48
WLS	15	10	26.9	25.8	1.94

また、 $Y_1$  のみを 28 点用いて行った予測の SE は 1.59 であり、 $Y_1$  と  $Y_2$  を組合せた予測は  $Y_1$  のみで行う予測よりも予測精度が高い場合がある。WLS と比較すると同じ標本数であるが、提案手法のほうが予測精度は非常に高いものが得られている。

## 6 考察・展望

精度の高いの標本を多く得ることが困難な場合は、精度の低いの標本と組み合わせることで予測精度を高められることが確認できた。また、しかし、高精度の標本を加えれば必ずしも予測精度が良くなるわけではない。誤差が大きい標本が含まれた場合は予測精度の低下が考えられる。

また、式 (21) から、予測値はサンプリング点では必ず標本の値と等しくなる。そのため、標本数が増えると標本を全て通過するようになるため、標本間の相関関係を適切に定めること、適切なパラメータを求めることが困難になり、標本数の増加が予測精度の低下を招く場合がある。本研究では共分散関数を式 (6)、(7) のように取ったが、適切な共分散関数を考えるとともにロバストな推定を行う方法を考えなければならない。さらに、標本の精度がどの程度異なっているかを考慮に入れた予測、誤差分散に対する事前情報が活用できる場合に適切な推定方法とモデルが必要である。

## 参考文献

- [1] M. C. Kennedy and A. O'Hagan : Predicting and output from a complex computer code when fast approximations are available, *Biometrika*, **87**, 1-13 (2000).
- [2] M. C. Kennedy and A. O'Hagan : Bayesian calibration of computer models, *Journal of the Royal Statistical Society, Series B*, **63**, 425-464 (2001).
- [3] Thomas J.Santner , Brian J. Williams and William I.Notz : The Design and Analysis of Computer Experiments, Springer(2003).
- [4] David Hirst, Geir Storvik, and Anne Randi Syversveen: A hierarchical modelling approach to combining environmental data at different scales, *Journal of the Royal Statistical Society, Series C*, **52**, 377-390 (2003).