

# 変分ベイズ法に基づく AR モデルによる混合予測について

## A Note on Mixture Prediction in Autoregressive Model based Variable Bayesian Approach

倉持佳生\*  
Yoshitaka Kuramochi

後藤正幸†  
Masayuki Goto

平澤茂一\*  
Shigeichi Hirasawa

**Abstract**— The AR model is useful to predict time series data. When we predict the Bayes optimal values, the posterior probability can not calculate easily, because it is integrated on the parameter space. So a conventional method for the calculation uses Laplace Approximation. The other method which does not use approximation is MonteCarlo method. But it is not practical use because of the plenty of computational complexity.

W.D.Penny et. al. have propose algorithm which selects the optimal model with Variational Bayes method with smaller computational complexity than MonteCarlo method and more precise than Laplace Approximation. In this paper, we propose the Bayes optimal prediction using mixture models with Variational Bayes method.

**Keywords**— Time series data, Autoregressive model, Bayes decision theory, Mixture model, Variational Bayes

### 1 はじめに

時間の経過に伴って不規則に変動する時系列データにおいて、過去のデータ系列から将来のデータを予測するにはパワースペクトルによる方法と線形予測による方法がある。

パワースペクトルによる方法は、時系列を構成する調和成分の強さ、周波数の空間での分布をとらえ、不規則な変動の特性を知ることによって予測を行う。一方、線形予測による方法では過去の観測値に適当な係数を乗じて加え合わせ、予測値を得る [9]。

AR モデル (自己回帰モデル) は線形予測の考え方から定義されているモデルであり、このモデルは自己回帰係数  $\phi_1, \phi_2, \dots, \phi_k$  と誤差分散  $\sigma_\varepsilon^2$  の 2 つのパラメータから構成されている。予測に適したモデル式を得るにはこれらのパラメータの決定とモデル次数  $k$  の決定という 2 つの問題があり、前者の問題に対しては最小 2 乗法、ユール・ウォーカー法 [9]、J.W.Miller の手法 [7] などが解法として用いられる。また、AR モデルにおいては候補となるモデルは次数により決定されるので、後者の次数決定問題はモデル選択問題として考えることができる。この問題は AIC や BIC、あるいは各モデルの事後確率を計算し最大になるようなモデルを選択する方法が提案されている [9]。そこで、W.D.Penny らは事後確率の算出に変分ベイズ法を用い、事後確率を最大化するモデル次数を一

意に選択して予測を行う方法を提案している [2]。

一方、ベイズ決定理論により予測を行う場合、全モデルを事後確率で重み付け平均をとった混合モデルで予測するのがベイズ最適である。しかし、事後確率の算出にはパラメータ空間での積分操作を行わなければならないために計算が非常に困難となる。そこで、事後確率の算出にラプラス近似を用いたうえで、全モデルの混合をとり、時系列データの予測を行う近似解法が提案されている [1]。

本研究では、W.D.Penny らと同様、モンテカルロ法よりも少ない計算量ながら、ラプラス近似よりも推定精度が良いと近年注目されている変分ベイズ法を用いる。これにより事後確率を求めた上で、全モデルの混合をすることにより、時系列データのベイズ最適な混合予測を精度良く計算する方法を提案する。

### 2 本研究の設定

#### 2.1 AR モデル

$k$  次の AR モデルは (1) 式で与えられる。

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_k x_{t-k} + \varepsilon_t \quad (1)$$

ここで、 $x_t \in \mathcal{R}$  は時点  $t$  の時系列データ (状態データ)、 $\{\phi_j\}$  は自己回帰係数、 $k$  はモデルの次数、 $\varepsilon_t$  は時点  $t$  における予測誤差を表す。ただし、 $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$  の白色雑音 (White Noise) である。なお、本研究では問題を単純にするために分散  $\sigma_\varepsilon^2$  は既知とする。 $(\beta = \frac{1}{\sigma_\varepsilon^2})$  とした精度  $\beta$  を今後利用する。

#### 2.2 パラメータ

自己回帰係数をまとめて、ベクトル表現で

$$\theta^k = (\phi_1, \phi_2, \dots, \phi_k)^T$$

と記述する。 $\phi_i, i = 1, 2, \dots, k$  は互いに独立な正規分布に従って発生し、 $\theta^k$  の精度を  $\alpha$  と定義する。すなわち、 $\sigma_{\phi_i}^2 = \frac{1}{\alpha}, i = 1, 2, \dots, k$  である。なお、 $\alpha \sim \Gamma(\alpha; b, c)$  である。 $b, c$  はそれぞれ  $\Gamma$  分布のパラメータである。

また、(1) 式におけるモデルの次数  $k$  は未知とする。ここで扱うモデルとは次数の異なる AR モデルのことであり、たとえば次数  $l$  の AR モデルを  $m_l$  と表す。今後はモデル  $m$  における次数を  $k_m$  と置き、 $\theta^{k_m} = (\phi_1, \phi_2, \dots, \phi_{k_m})^T$  とする。

#### 2.3 ベイズ最適な予測法

モデルを固定した予測法ではなく、事後確率による混合モデルを用いて予測するベイズ決定理論による予測モデルは以下のように定められる。

\* 〒 169-8555 東京都新宿区大久保 3-4-1 早稲田大学理工学部経営システム工学科 Dept. of Industrial & Management System Engineering, Waseda Univ. 3-4-1 Okubo Shinjuku Tokyo 169-8555, Japan

† 〒 113-8656 東京都文京区本郷 7-3-1 東京大学大学院工学系研究科環境海洋工学専攻 Dept. of Environmental & Ocean Engineering, The Univ. of Tokyo. 7-3-1 Hongo Bunkyo Tokyo 113-8656,

予測値を表す決定関数を  $d(\cdot)$  とし、2乗誤差損失を表す損失関数を  $L(\cdot, \cdot)$  と表すと (2) 式のようになる。

$$L(x_n, d(x^{n-1})) = (x_n - d(x^{n-1}))^2 \quad (2)$$

また、 $x^{n-1}$  が与えられたもとの条件付リスク関数を  $R(\cdot)$  とすると (3) 式のようになる。

$$R(\theta^{k_m}, m | x^{n-1}) = \int_{x_n} L(x_n, d(x^{n-1})) p(x_n | x^{n-1}, \theta^{k_m}, m) dx_n \quad (3)$$

条件付リスク関数を最小にするような決定関数を定めなければならないが、決定されたモデルまたはパラメータによってリスク関数を最小にする決定関数は異なるために、任意の  $\theta^{k_m}, m$  すべてについて (3) 式を最小にする決定関数は存在しない。そこで、リスク関数を事後確率  $p(\theta^{k_m}, m | x^{n-1})$  で平均化した条件付ベイズリスク  $BR$  を考え、それを最小化する。

$$BR = \int_{x_n} L(x_n, d(x^{n-1})) \sum_m \int_{\theta^{k_m}} p(x_n | x^{n-1}, \theta^{k_m}, m) \times p(\theta^{k_m}, m | x^{n-1}) d\theta^{k_m} dx_n \quad (4)$$

ここで、事後混合分布として  $p_{mix}(x_n | x^{n-1})$  を用いると (4) 式は次式のように書き換えられる。

$$BR = \int_{x_n} L(x_n, d(x^{n-1})) p_{mix}(x_n | x^{n-1}) dx_n \quad (5)$$

$$p_{mix}(x_n | x^{n-1}) = \sum_m \int_{\theta^{k_m}} p(x_n | x^{n-1}, \theta^{k_m}, m) \times p(\theta^{k_m} | x^{n-1}, m) p(m | x^{n-1}) d\theta^{k_m} \quad (6)$$

ベイズリスクを最小化するために (5) 式を  $d(x^{n-1})$  で微分すると

$$\frac{\partial}{\partial d(x^{n-1})} BR = \int_{x_n} (-2x_n + 2d(x^{n-1})) p_{mix}(x_n | x^{n-1}) dx_n \quad (7)$$

$$\frac{\partial}{\partial d(x^{n-1})} BR = 0 \text{ として解くと,}$$

$$d(x^{n-1}) \int_{x_n} p_{mix}(x_n | x^{n-1}) dx_n = \int_{x_n} x_n p_{mix}(x_n | x^{n-1}) dx_n \quad (8)$$

$\int_{x_n} p_{mix}(x_n | x^{n-1}) dx_n = 1$  であるので次式のようになる [8].

$$d(x^{n-1}) = \int_{x_n} x_n p_{mix}(x_n | x^{n-1}) dx_n \quad (9)$$

実際には (9) 式を用いて計算することによってベイズ最適な予測値が得られる。

### 3 ラプラス近似による事後確率 [1]

$$p(m | x^{n-1}) = \frac{p(x^{n-1} | m) p(m)}{\sum_{m'} p(x^{n-1} | m') p(m')} \quad (10)$$

事後確率は (10) 式で求めることができるが、その際、尤度をパラメータ上で積分することが必要となる。この計算は一般に非常に数値計算が難しいため様々な近似が用いられているが、ここではラプラス法による事後確率の近似を述べる [6].

実数値関数  $g(\theta)$  を  $\theta_0$  の周りでテーラー展開する。ただし、 $\theta$  は  $k$  次元ベクトルである。

$$g(\theta) \approx g(\theta_0) + (\theta - \theta_0)^T (\nabla g_{\theta_0}) + \frac{1}{2} (\theta - \theta_0)^T (\nabla^2 g_{\theta_0}) (\theta - \theta_0) \quad (11)$$

ただし、 $(\nabla g_{\theta_0})$  と  $(\nabla^2 g_{\theta_0})$  はそれぞれ  $g(\theta)$  の gradient, Hessian である。

ここで、 $\theta_0$  を  $g(\theta)$  の最頻値とすると、 $(\nabla g_{\theta_0}) = 0$  となり、

$$\begin{aligned} \int_{\Theta} \exp[g(\theta)] d\theta &\approx \int_{\Theta} \exp \left[ g(\theta_0) + \frac{1}{2} (\theta - \theta_0)^T (\nabla^2 g_{\theta_0}) (\theta - \theta_0) \right] d\theta \\ &= \exp[g(\theta_0)] \int_{\Theta} \exp \left[ \frac{1}{2} (\theta - \theta_0)^T (\nabla^2 g_{\theta_0}) (\theta - \theta_0) \right] d\theta \\ &= \exp[g(\theta_0)] |\nabla^2 g_{\theta_0}|^{-1/2} (2\pi)^{k/2} \int_{\Theta} \varphi(\theta) d\theta \quad (12) \end{aligned}$$

ただし、 $\varphi(\theta)$  は平均  $\theta_0$ 、分散共分散行列  $[-\nabla^2 g_{\theta_0}]^{-1}$  の  $k$  次元の多変量正規密度である。

(12) 式を用いて事後確率を求める。 $\theta^{k_m}$  を次数  $k_m$  の AR モデルのパラメータとすると  $\theta^{k_m} = (\phi_1, \phi_2, \dots, \phi_{k_m})^T$  となり、

$$g(\theta^{k_m}) = \log[p(x^{n-1} | \theta^{k_m}, m) p(\theta^{k_m} | m)] \quad (13)$$

とすると、

$$p(x^{n-1} | m) \approx (2\pi)^{k_m/2} |\nabla^2 g_{\theta^{k_m}}|^{1/2} p(x^{n-1} | \theta^{k_m}, m) \times p(\theta^{k_m} | m) \int_{(-1,1)^{k_m}} \varphi(\theta^{k_m}) d\theta^{k_m} \quad (14)$$

が得られる。ただし、 $\theta^{k_m}$  は  $g(\theta^{k_m})$  を最大にする  $\theta^{k_m}$  の値である。

### 4 変分ベイズ法を用いたベイズ最適な時系列予測

ベイズ最適な予測を近似するために、従来ではラプラス法などが用いられていたが、本研究では変分ベイズを用いて事後確率を計算した上で、近似的にベイズ最適な予測を行う方法を提案する。

本研究において  $t$  時点のデータ  $x_t$  とパラメータベクトル  $\theta^{k_m}$  は正規分布に従っているという仮定から、

$$p(x^{n-1} | \theta^{k_m}, \beta, m) = \frac{1}{C} \exp \left( -\beta \sum_{t=1}^{n-1} x_t \theta^{k_m} \right) \quad (15)$$

$$p(\theta^{k_m} | \alpha, m) = \frac{1}{C} \exp \left( -\alpha \frac{1}{2} \theta^{k_m T} \theta^{k_m} \right) \quad (16)$$

ただし、 $\alpha$  は  $\theta^{k_m}$  の精度パラメータ、 $\beta$  は  $x^{n-1}$  の精度パラメータを表す。また、

$$E_D(\theta^{k_m}) = \frac{1}{2}(\mathbf{Y} - \mathbf{X}\theta^{k_m})^T(\mathbf{Y} - \mathbf{X}\theta^{k_m}) \quad (17)$$

ここで、 $\mathbf{Y}$  は列ベクトル、 $\mathbf{X}$  は行列を表し、 $\theta^{k_m}$  はパラメータの列ベクトルを表す。

$$\mathbf{Y} = (x_{k+1}, x_{k+2}, \dots, x_n)^T \quad (18)$$

$$\mathbf{X} = \begin{pmatrix} x_k & x_{k-1} & \dots & x_1 \\ x_{k+1} & x_k & \dots & x_2 \\ \vdots & \vdots & \ddots & \vdots \\ x_{n-1} & x_{n-2} & \dots & x_k \end{pmatrix} \quad (19)$$

$$\theta^{k_m} = (\phi_1, \phi_2, \dots, \phi_{k_m})^T \quad (20)$$

また、パラメータの事後確率は (21) 式で表される。

$$p(\theta^{k_m}|x^{n-1}, \alpha, m) = \frac{p(x^{n-1}|\theta^{k_m}, \alpha, m)p(\theta^{k_m}|\alpha, m)}{p(x^{n-1}|\alpha, m)} \quad (21)$$

今、新たな分布  $Q$  を導入して対数尤度を次式のように書き表す。

$$\log p(x^{n-1}|m) = \log \int Q(\theta^{k_m}, \alpha) \times \frac{p(x^{n-1}|\theta^{k_m}, m)p(\theta^{k_m}|\alpha, m)p(\alpha|m)}{Q(\theta^{k_m}, \alpha)} d\theta^{k_m} d\alpha \quad (22)$$

ここで、 $Q$  は変分事後分布と呼ばれる。新たに、

$$F(Q) = - \int Q(\theta^{k_m}, \alpha) \log \frac{p(x^{n-1}|\theta^{k_m}, m)p(\theta^{k_m}|\alpha, m)p(\alpha|m)}{Q(\theta^{k_m}, \alpha)} d\theta^{k_m} d\alpha \quad (23)$$

を導入すると、Jensen の不等式から

$$F(Q) \geq - \log p(x^{n-1}|m) \quad (24)$$

つまり、 $F(Q)$  を最小にするような  $Q$  を見つけることによって尤度の近似解を求めることができる。このように、 $-\log p(x^{n-1}|m)$  は積分操作が含まれるので、これを厳密に解くことをあきらめ、 $Q$  により  $F(Q)$  を最小化し、 $-\log p(x^{n-1}|m)$  の近似解を得ることを考えるのが変分ベイズ法である [3]。

なお、 $Q$  は通常、次式のように各未知変量ごとに分解した形を仮定する。各分布族は任意でよい [4]。

$$Q(\theta^{k_m}, \alpha) = Q(\theta^{k_m})Q(\alpha) \quad (25)$$

最適な  $Q(\theta^{k_m})$  と  $Q(\alpha)$  を同時に計算することは出来ないで、片方に分布を固定しておき、 $F$  を最小にするような分布を求めることを交互に行う。その結果、それぞれの最適な分布は以下ようになる [3]。

$$Q(\theta^{k_m}) = p(\theta^{k_m}|x^{n-1}, \bar{\alpha}) = \text{Normal}(\theta_{MP|\bar{\alpha}}^{k_m}, \Sigma) \quad (26)$$

$$Q(\alpha) = \Gamma(\alpha; b', c') \quad (27)$$

ここで  $\Sigma$  は  $\theta^{k_m}$  の分散共分散行列を表し、また、

$$1/b' = 1/b_\alpha + \frac{1}{2}\theta_{MP|\bar{\alpha}}^{k_m T} \theta_{MP|\bar{\alpha}}^{k_m} + \frac{1}{2}\text{Trace}\Sigma \quad (28)$$

$$c' = k/2 + c_\alpha$$

$$\theta_{MP|\bar{\alpha}}^{k_m} = \arg \max_{\theta^{k_m}} p(\theta^{k_m}|x^{n-1}, \alpha_{MP}, m) \quad (29)$$

ただし、 $\alpha_{MP}$  は  $\alpha$  のモードを表す。

ベイズ最適な予測値を得るための (9) 式を以下のように変形する。

$$d(x^{n-1}) = \sum_m p(m|x^{n-1}) \int_{\theta^{k_m}, \alpha, x_n} x_n p(x_n|\theta^{k_m}, \alpha, x^{n-1}, m) \times f(\theta^{k_m}, \alpha|m, x^{n-1}) d\theta^{k_m} d\alpha dx_n \quad (30)$$

ここで、(30) 式を計算するために以下の2つの手順を踏む。

1.  $p(m|x^{n-1})$  の計算
2.  $\int_{\theta^{k_m}, \alpha, x_n} x_n p(x_n|\theta^{k_m}, \alpha, x^{n-1}, m) \times f(\theta^{k_m}, \alpha|m, x^{n-1}) d\theta^{k_m} d\alpha dx_n$  の計算

手順1はベイズの定理 (10) 式によって求められる。また、(10) 式中のモデルに関する尤度  $p(x^{n-1}|m)$  は変分ベイズ法により、(23) 式の  $F$  を最小化することで近似することができる。

(23) 式は次のように書き換えることができる。

$$F = - \int d\theta^{k_m} Q(\theta^{k_m}) \left[ \int d\alpha Q(\alpha) \{ \log p(\theta^{k_m}|\alpha, m) + \log p(x^{n-1}|\theta^{k_m}, m) - \log Q(\theta^{k_m}) \} \right] + const$$

$$= \int d\theta^{k_m} Q(\theta^{k_m}) \left[ \int d\alpha Q(\alpha) \left\{ \bar{\alpha} \frac{1}{2} \theta^{k_m T} \theta^{k_m} + \beta E_D(\theta^{k_m}) + \log Q(\theta^{k_m}) \right\} \right] + const \quad (31)$$

$\bar{\alpha} = \int Q \log(Q/P)$  より、 $F$  は

$$F = \int d\theta^{k_m} Q(\theta^{k_m}) \left\{ \bar{\alpha} \frac{1}{2} \theta^{k_m T} \theta^{k_m} + \beta E_D(\theta^{k_m}) + \log Q(\theta^{k_m}) \right\} + const \quad (32)$$

実際に  $F$  の値を計算するときは (32) 式に代入して値を求める。

一方で、手順2は以下のように式を変形して求める。

$$\int_{\theta^{k_m}, \alpha} \int_{x_n} x_n p(x_n|\theta^{k_m}, \alpha, x^{n-1}, m) \times f(\theta^{k_m}, \alpha|m, x^{n-1}) d\theta^{k_m} d\alpha dx_n$$

$$= \int_{\theta^{k_m}, \alpha} \hat{x}_n(\theta^{k_m}, \alpha, x^{n-1}, m) f(\theta^{k_m}, \alpha|m, x^{n-1}) d\theta^{k_m} d\alpha \quad (33)$$

ここで、 $\hat{x}_n(\theta^{k_m}, \alpha, x^{n-1}, m)$  は  $x_n$  の平均を表し、(34) 式で表される。

$$\hat{x}_n(\theta^{k_m}, \alpha, x^{n-1}, m) = E[x_n] = \phi_1 E[x_{n-1}] + \dots + \phi_{k_m} E[x_{n-k_m}] + E[\varepsilon_t] = \phi_1 x_{n-1} + \phi_2 x_{n-2} + \dots + \phi_{k_m} x_{n-k_m} \quad (34)$$

$(E[x_i] = x_i, i = 1, 2, \dots, k_m), E[\varepsilon_t] = 0$  より)  
 また,  $f(\theta^{k_m}, \alpha | m, x^{n-1}) = Q(\theta^{k_m})Q(\alpha)$  とすると, (26) 式と (27) 式から (33) 式は次式のようになる.

$$= \phi_{(MP|\bar{\alpha})_1} x_{n-1} + \phi_{(MP|\bar{\alpha})_2} x_{n-2} + \dots + \phi_{(MP|\bar{\alpha})_{k_m}} x_{n-k_m} \quad (35)$$

ただし, (35) 式における回帰係数  $\phi_{(MP|\bar{\alpha})}$  は (29) 式により定まる  $\theta_{MP|\bar{\alpha}}^{k_m} = (\phi_{(MP|\bar{\alpha})_1}, \phi_{(MP|\bar{\alpha})_2}, \dots, \phi_{(MP|\bar{\alpha})_{k_m}})$  である. 今,  $\theta^{k_m}$  は (26) 式で示したように正規分布で表現できることから,  $\theta_{MP|\bar{\alpha}}^{k_m}$  は (21) 式で表される  $\theta^{k_m}$  の分布の平均値に等しい. ここで, (21) 式の分子を考えると, (15) 式と (16) 式から指数部は

$$\begin{aligned} & -\frac{\beta}{2}(\mathbf{Y} - \mathbf{X}\theta^{k_m})^T(\mathbf{Y} - \mathbf{X}\theta^{k_m}) + \alpha\theta^{k_m T}\theta^{k_m} \\ & = -\frac{1}{2}(\beta\mathbf{Y}^T\mathbf{Y} - \beta\mathbf{Y}^T\mathbf{X}\theta^{k_m} - \beta\theta^{k_m T}\mathbf{X}^T\mathbf{Y} \\ & \quad + \theta^{k_m T}(\beta\mathbf{X}^T\mathbf{X} + \alpha\mathbf{I})\theta^{k_m}) \quad (36) \end{aligned}$$

$\mathbf{I}$  は単位行列を表す. 一方, 平均  $\mu$ , 分散共分散行列  $\Psi$  の多次元正規分布  $Y$  の分布密度関数の指数部分は

$$-\frac{1}{2}(\mathbf{Y} - \mu)^T\Psi(\mathbf{Y} - \mu) \quad (37)$$

(36) 式と (37) 式から,

$$\theta_{MP|\bar{\alpha}}^{k_m} = (\beta\mathbf{X}^T\mathbf{X} + \alpha\mathbf{I})^{-1}\beta\mathbf{X}^T\mathbf{Y} \quad (38)$$

同様に分散共分散行列も以下のように求められる.

$$\Sigma^{-1} = \beta\mathbf{X}^T\mathbf{X} \quad (39)$$

(38) 式における  $\alpha$  は (28) 式より,

$$\alpha = \bar{\alpha} = b'c' = \frac{\frac{k_m}{2} + c}{\frac{1}{b} + \frac{1}{2}\theta_{MP|\bar{\alpha}}^{k_m T}\theta_{MP|\bar{\alpha}}^{k_m} + \frac{1}{2}\text{Trace}\Sigma} \quad (40)$$

ただし,  $b, c$  は  $\Gamma$  分布のパラメータを表し,  $k_m$  はモデル  $m$  の次元数を表す.  $\Sigma$  は分布  $p(\theta^{k_m} | x^{n-1}, \alpha_{MP}, m)$  の分散共分散行列を表し, (39) 式で表される. これら  $\theta_{MP|\bar{\alpha}}^{k_m}$  および  $\bar{\alpha}$  は (38) 式と (40) 式を交互に解くことによって得られる.

## 5 考察

変分ベイズ法はラプラス近似より精度がよく, モンテカルロ法よりも計算が煩雑ではない手法として知られている. その根拠として, 下に示すようにラプラス近似は未知パラメータの同時確率を単一の正規分布で近似している ((41) 式) のに対して, 変分ベイズ法では同時確率を分解した形に仮定し, それぞれが任意の分布に従う ((42) 式) と仮定できることによる [5].

$$Q(\theta^{k_m}, \alpha) \sim \text{Normal} \quad (41)$$

$$\begin{aligned} Q(\theta^{k_m}, \alpha) &= Q(\theta^{k_m})Q(\alpha) \\ Q(\theta^{k_m}) &\sim \text{Normal} \\ Q(\alpha) &\sim \Gamma \end{aligned} \quad (42)$$

## 6 まとめ

本稿では時系列データの予測という観点から, ベイズ決定理論に基づき, モデルの次数を一意に決定することなく, モデルの混合をとって予測値を得るという手法を示した. そのためには事後確率の算出が必須であるが, 問題となるのがパラメータ空間で積分操作を行うということである. この解決法としてラプラス近似を用いたもの [6] やモンテカルロシミュレーションを使った方法などがあった. しかし, ラプラス近似は計算が容易である代わりに精度の点で問題がある. また, モンテカルロ法は近似を使わないという点では評価されるが, 計算量が膨大なために実用的ではない. そこで, 本稿は近年注目されている変分ベイズ法を用いた.

今後はシミュレーションを行ってラプラス近似を使用した場合より良い精度で予測できていることを実証することを考えている.

謝辞: 著者の一人倉持は, 本研究を進めるにあたり, 日頃よりご討論, ご助言を頂いている早稲田大学 平澤研究室各氏に心から深く感謝致します.

なお, 本研究の一部は 2001 年度早稲田大学特定課題研究助成費 (課題番号 2001A-566) の助成によるものです.

## 参考文献

- [1] 倉持佳生, 後藤正幸, 平澤茂一, "観測雑音を考慮した AR モデルによるベイズ最適な予測法," 信学技報, NLP2000-17, 2000.
- [2] W.D.Penny and S.J.Roberts, "Bayesian Methods For Autoregressive Models"
- [3] David J.C.MacKay, "Ensemble Learning and Evidence Maximization"
- [4] 上田修功, "ベイズ学習法の最前線," 情報処理, Vol.42, No.1, 2001.
- [5] 上田修功, "最良モデル探索のための変分ベイズ学習," 人工知能学会論文誌, Vol.16, No.2, 2001.
- [6] Nhu D.Le, Adrian E.Raftery, and R.Douglas Martin, "Robust Bayesian Model Selection for Autoregressive Processes With Additive Outliers," Journal of the American Statistical Association, Vol.91, No.433, 1996.
- [7] James W.Miller, "Exact Maximum Likelihood Estimation In Autoregressive Processes," Journal of Time Series Analysis, Vol.16, No.6, 1995.
- [8] 鈴木友彦, 後藤正幸, 俵信彦, "線形回帰モデルのベイズ最適な予測法に関する研究," 日本経営工学会論文誌, 2000.
- [9] 尾崎統, 時系列論, 日本放送出版協会, 1988.