# 非線形回帰モデルにおける予測情報量規準

九州大学大学院数理学府　金大柱
中央大学理工学部　小西貞則

## 1　はじめに

予測分布とは，将来新たに観測されるデータの予測を行うことを目的として，現在得られているデータとモデルのパラメータに関する事前分布に基づき，ベイズアプローチにより導出される将来新たに観測されるデータに関する分布 (以下，ベイズ型予測分布) である．予測分布は，回帰分析や判別分析など様々なモデルへの応用が試みられている．

Kitagawa (1997) では，分散共分散行列を既知とした下での正規線形モデルに対し，係数に自然共役な事前分布である正規分布を仮定することによって，ベイズ型予測分布が正規分布で与えられることを示した．さらに，情報量規準の観点に基づき，予測分布に基づくモデル評価基準の導出を行い，それを予測情報量規準 (Predictive Information Criterion; PIC) とした．しかし，モデルの分散を既知と仮定しているため理論的には予測分布や PIC の導出が正確に行われているが，一般には，モデルの分散が既知という場面は少なく，実用的な面での問題点が残っていると言える．

本研究では，基底展開法に基づく非線形回帰モデルを仮定し，予測情報量規準の導出を行った．誤差分散を未知パラメータとし，事前分布として逆ガンマ分布を仮定した場合，予測分布が t 分布となる (Denison et al. (2002), 中妻 (2003)) ため，PIC の導出は困難となるが，ラプラス近似 (Tierneay and Kadane (1986), Davison (1986), Kass and Raftery (1995)) を用いることによって PIC を解析的に導出し，基底関数の個数と調整パラメータを選択する手法について考察した．また，数値実験を通して，他のモデル評価基準との比較を行った．

## 2　基底展開法に基づく非線形回帰モデルにおける予測分布

事後分布の導出の際には，自然共役な事前分布を用いることにより，事後分布も事前分布と同じ形で解析的に表現することが出来る (小西・北川 (2004))．このような性質を用い，Bishop (2006) は，線形回帰モデルの分散を既知とした下で，回帰係数の事前分布として正規分布を仮定し事後分布の導出を解析的に行っている．さらに，分散共分散行列を既知とした線形回帰モデルに対し，このような事前分布を仮定すると，予測分布も解析的に求められることができ，正規分布となることを示した．

線形回帰モデルの回帰係数と分散を，それぞれ，未知のパラメータとして扱う場合の予測分布の導出は，Denison et al. (2002), 中妻 (2003) などで研究されている．このような場合には，回帰係数の事前分布に対して正規分布を仮定し，また，分散の事前分布として逆ガンマ分布を仮定することにより，これらの事前分布は自然共役な事前分布となっており，事後分布も，それぞれ，正規分布と逆ガンマ分布で与えられることが分かる．さらに，事前分布を上記のように仮定した場合のベイズ予測分布の導出を行い，予測分布がスチューデントの t 分布として，解析的に求められることを示した．

## 2.1 基底展開法に基づく非線形回帰モデル

1 次元目的変数 $Y$ と $p$ 次元説明変数ベクトル $\boldsymbol{X} = (X_1, X_2, \cdots, X_p)^T$ に関して観測された $n$ 組のデータを $\{(\boldsymbol{x}_i, y_i); i = 1, 2, \cdots, n\}$ とする．$\boldsymbol{x}_i$ に対して観測された $y_i$ は，真の値 $u(\boldsymbol{x}_i)$ に誤差変数 $\varepsilon_i$ が加わって

$$y_i = u(\boldsymbol{x}_i) + \varepsilon_i, \quad i = 1, 2, \cdots, n \tag{1}$$

で与えられたと仮定する．そこで，一般的に回帰モデルとは観測された $n$ 組のデータ $\{(\boldsymbol{x}_i, y_i, i = 1, 2, \cdots, n)\}$ に対して，

$$y_i = u(\boldsymbol{x}_i; \boldsymbol{w}) + \varepsilon_i, \quad i = 1, 2, \cdots, n \tag{2}$$

で与えられるものとする．この真の構造 $u(\boldsymbol{x})$ を $m$ 次元パラメータベクトル $\boldsymbol{w} = (w_1, w_2, \cdots, w_m)^T$ によって特徴づけられるモデル $u(\boldsymbol{x}; \boldsymbol{w})$ で近似することを考える．

基底展開法に基づく非線形回帰モデルは，回帰関数として，基底関数と呼ばれる既知の非線形関数 $b_j(\boldsymbol{x})$ $(j = 1, 2, \cdots, m)$ とパラメータベクトルの線形結合，

$$u(\boldsymbol{x}; \boldsymbol{w}) = w_0 + \boldsymbol{w}^T \boldsymbol{b}(\boldsymbol{x}) \tag{3}$$

を考える．ここで，$\boldsymbol{b}(\boldsymbol{x}) = (b_1(\boldsymbol{x}), b_2(\boldsymbol{x}), \cdots, b_m(\boldsymbol{x}))^T$ は $m$ 次元基底関数ベクトルとし，$\boldsymbol{w} = (w_1, w_2, \cdots, w_m)^T$ は未知の $m$ 次元パラメータベクトルとする．ここで，$w_0$ は切片である．

$$y_i = w_0 + \sum_{j=1}^{m} w_j b_j(\boldsymbol{x}_i) + \varepsilon_i = w_0 + \boldsymbol{w}^T \boldsymbol{b}(\boldsymbol{x}_i) + \varepsilon_i, \qquad i = 1, 2, \cdots, n \tag{4}$$

まず，誤差変数 $\varepsilon_i$ が互いに独立に平均 0，分散 $\sigma^2$ の正規分布 $N(0, \sigma^2)$ に従うとすると，基底展開法に基づく非線形回帰モデルは，

$$f(y_i|\boldsymbol{w}, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{\frac{1}{2}}} \exp\left[ -\frac{\left\{ y_i - \boldsymbol{w}^T \boldsymbol{b}(\boldsymbol{x}_i) \right\}^2}{2\sigma^2} \right] \tag{5}$$

で表わされる．ここで，推定すべきパラメータは $\boldsymbol{w}$ と $\sigma^2$ である．基底展開法に基づく非線形回帰モデルは，基底関数行列 $B$ を用い以下のように表すことができる．

$$y|\boldsymbol{w}, \sigma^2 \sim N(B\boldsymbol{w}, \sigma^2 I). \tag{6}$$

## 2.2 パラメータの事前分布

本研究では，回帰係数パラメータと分散パラメータの事前分布を，それぞれ，以下のようハイパーパラメータを持つ，正規分布，逆ガンマ分布とする．

$$\boldsymbol{w}|\sigma^2 \sim N(\boldsymbol{w}_0, \sigma^2 A^{-1}) \tag{7}$$

$$\sigma^2 \sim IG\left( \frac{\nu_0}{2}, \frac{\lambda_0}{2} \right). \tag{8}$$

### 2.3 事後分布の導出

基底展開法に基づく非線形回帰モデルに対し，回帰係数と分散のそれぞれのパラメータの事前分布が与えられた時，事後分布は次のように定義される．

$$\pi(\boldsymbol{w},\sigma^2|\boldsymbol{y}) = \frac{f(\boldsymbol{y}|\boldsymbol{w},\sigma^2)\pi(\boldsymbol{w},\sigma^2)}{\int\int f(\boldsymbol{y}|\boldsymbol{w},\sigma^2)\pi(\boldsymbol{w},\sigma^2)d\boldsymbol{w}d\sigma^2}. \tag{9}$$

また，尤度関数とパラメータの事前分布が (6)，(8) 式のように与えられた時，パラメータの事後分布は次のような計算により導出することが出来る．

(9) 式の分子は，

$$\begin{aligned}
& f(\boldsymbol{y}|\boldsymbol{w},\sigma^2)\pi(\boldsymbol{w},\sigma^2) \\
= \;& (2\pi)^{-\frac{m}{2}-\frac{n}{2}}(\sigma^2)^{-\frac{m}{2}-\frac{n}{2}-(\frac{\nu_0}{2}+1)}|A^{-1}|^{-\frac{1}{2}}\frac{\left(\frac{\lambda_0}{2}\right)^{\frac{\nu_0}{2}}}{\Gamma\left(\frac{\nu_0}{2}\right)} \\
& \times\exp\left[-\frac{1}{2\sigma^2}\left\{\lambda_0+(\boldsymbol{y}-B\hat{\boldsymbol{w}}_{\mathrm{MLE}})^T(\boldsymbol{y}-B\hat{\boldsymbol{w}}_{\mathrm{MLE}})\right.\right. \\
& \qquad \left.\left.+(\boldsymbol{w}_0-\hat{\boldsymbol{w}}_{\mathrm{MLE}})^T\left\{A^{-1}+(B^TB)^{-1}\right\}^{-1}(\boldsymbol{w}_0-\hat{\boldsymbol{w}}_{\mathrm{MLE}})\right\}\right] \\
& \times\exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{w}-\hat{\boldsymbol{w}}_n)^T\hat{A}_n^{-1}(\boldsymbol{w}-\hat{\boldsymbol{w}}_n)\right\}
\end{aligned} \tag{10}$$

として計算することが出来る．また，事後分布の分母は次のように表すことができる．

$$\begin{aligned}
& \int\int f(\boldsymbol{y}|\boldsymbol{w},\sigma^2)\pi(\boldsymbol{w},\sigma^2)d\boldsymbol{w}d\sigma^2 \\
= \;& (2\pi)^{-\frac{n}{2}}|A^{-1}|^{-\frac{1}{2}}\frac{\left(\frac{\lambda_0}{2}\right)^{\frac{\nu_0}{2}}}{\Gamma\left(\frac{\nu_0}{2}\right)}|\hat{A}_n|^{\frac{1}{2}}\times\frac{\Gamma\left(\frac{\hat{\nu}_n}{2}\right)}{\left(\frac{\hat{\lambda}_n}{2}\right)^{\frac{\hat{\nu}_n}{2}}}.
\end{aligned} \tag{11}$$

ここで，

$$\hat{\lambda}_n = \lambda_0+(\boldsymbol{y}-B\hat{\boldsymbol{w}}_{\mathrm{MLE}})^T(\boldsymbol{y}-B\hat{\boldsymbol{w}}_{\mathrm{MLE}})+(\boldsymbol{w}_0-\hat{\boldsymbol{w}}_{\mathrm{MLE}})^T\left\{A^{-1}+(B^TB)^{-1}\right\}^{-1}(\boldsymbol{w}_0-\hat{\boldsymbol{w}}_{\mathrm{MLE}}) \tag{12}$$

とした．よって，事後分布は次のように計算することが出来る．

$$\begin{aligned}
\pi(\boldsymbol{w},\sigma^2|\boldsymbol{y}) &= \frac{f(\boldsymbol{y}|\boldsymbol{w},\sigma^2)\pi(\boldsymbol{w},\sigma^2)}{\int\int f(\boldsymbol{y}|\boldsymbol{w},\sigma^2)\pi(\boldsymbol{w},\sigma^2)d\boldsymbol{w}d\sigma^2} \\
&= (2\pi)^{-\frac{m}{2}}|\sigma^2A^{-1}|^{-\frac{1}{2}}\exp\left[-\frac{1}{2}(\boldsymbol{w}-\hat{\boldsymbol{w}}_n)^T(\sigma^2\hat{A}_n)^{-1}(\boldsymbol{w}-\hat{\boldsymbol{w}}_n)\right] \\
&\quad\times\frac{\left(\frac{\hat{\lambda}_n}{2}\right)^{\frac{\hat{\nu}_n}{2}}}{\Gamma\left(\frac{\hat{\nu}_n}{2}\right)}(\sigma^2)^{-\left(\frac{n+\nu_0}{2}+1\right)}\exp\left[-\frac{\hat{\lambda}_n}{2\sigma^2}\right].
\end{aligned} \tag{13}$$

これより，回帰係数と分散の事後分布は，それぞれ，以下のようなパラメータを持つ，正規分布，逆ガンマ分布で与えられることが分かる．

$$\boldsymbol{w}|\sigma^2,\boldsymbol{y}\sim N(\hat{\boldsymbol{w}}_n,\sigma^2\hat{A}_n),\;\;\sigma^2|\boldsymbol{y}\sim IG\left(\frac{\hat{\nu}_n}{2},\frac{\hat{\lambda}_n}{2}\right). \tag{14}$$

### 2.4 予測分布の導出

予測分布は次のように計算することが出来る．

$$\begin{aligned} f(\boldsymbol{z}|\boldsymbol{y}) &= \int\int f(\boldsymbol{z}|\boldsymbol{w},\sigma^2)\pi(\boldsymbol{w},\sigma^2)d\boldsymbol{w}d\sigma^2 \\ &= \int \left\{ f(\boldsymbol{z}|\boldsymbol{w},\sigma^2)\pi(\boldsymbol{w}|\sigma^2,\boldsymbol{y}) \right\} \pi(\sigma^2|\boldsymbol{y})d\sigma^2 \\ &= \int f(\boldsymbol{z}|\sigma^2,\boldsymbol{y})\pi(\sigma^2|\boldsymbol{y})d\sigma^2. \end{aligned} \tag{15}$$

予測分布 $f(\boldsymbol{z}|\boldsymbol{y})$ を求めるために，まず，$f(\boldsymbol{z}|\sigma^2,\boldsymbol{y})$ を求める．

$$\begin{aligned} f(\boldsymbol{z}|\sigma^2,\boldsymbol{y}) &= (2\pi)^{-\frac{n}{2}-\frac{m}{2}}(\sigma^2)^{-\frac{n}{2}-\frac{m}{2}}|\hat{A}_n|^{\frac{1}{2}} \times (2\pi)^{\frac{m}{2}} \left| \left\{ \sigma^2(B^TB+\hat{A}_n^{-1}) \right\}^{-1} \right|^{\frac{1}{2}} \\ &\quad \times \exp\left[ -\frac{1}{2}(\boldsymbol{z}-B\hat{\boldsymbol{w}}_n)^T \left\{ \sigma^2(B\hat{A}_nB^T+I) \right\}^{-1} (\boldsymbol{z}-B\hat{\boldsymbol{w}}_n) \right] \end{aligned} \tag{16}$$

次に予測分布 $f(\boldsymbol{z}|\boldsymbol{y})$ を求める．予測分布 $f(\boldsymbol{z}|\boldsymbol{y})$ は (15) 式より，以下のように求めることができる．

$$\begin{aligned} f(\boldsymbol{z}|\boldsymbol{y}) &= \int f(\boldsymbol{z}|\sigma^2,\boldsymbol{y})\pi(\sigma^2|\boldsymbol{y})d\sigma^2 \\ &= \left\{ \Gamma\left(\frac{n+\hat{\nu}_n}{2}\right) / \Gamma\left(\frac{\hat{\nu}_n}{2}\right) (\pi\hat{\nu}_n)^{\frac{n}{2}} \right\} \times \left| \frac{\hat{\lambda}_n}{\hat{\nu}_n}(B\hat{A}_nB^T+I_n) \right|^{-\frac{1}{2}} \\ &\quad \times \left[ 1+\frac{1}{\hat{\nu}_n}(\boldsymbol{z}-B\hat{\boldsymbol{w}}_n)^T \left\{ \frac{\hat{\lambda}_n}{\hat{\nu}_n}(B\hat{A}_nB^T+I) \right\}^{-1} (\boldsymbol{z}-B\hat{\boldsymbol{w}}_n) \right]^{-\left(\frac{n+\hat{\nu}_n}{2}\right)}. \end{aligned} \tag{17}$$

(17) 式より，予測分布 $f(\boldsymbol{z}|\boldsymbol{y})$ は以下のようなパラメータを持つステューデントの t 分布に従うことがわかる．

$$\boldsymbol{z}|\boldsymbol{y} \sim St(\hat{\boldsymbol{\mu}}, \hat{\Sigma}^*, \hat{\nu}_n) \tag{18}$$

ここで,

$$\begin{aligned} \hat{\boldsymbol{\mu}} &= B\hat{\boldsymbol{w}}_n \\ \hat{\Sigma}^* &= \frac{\hat{\lambda}_n}{\hat{\nu}_n}(B\hat{A}_nB^T+I_n) \\ \hat{A}_n &= (B^TB+A)^{-1} \\ \hat{\nu}_n &= n+\nu_0 \\ \hat{\lambda}_n &= \lambda_0 + (\boldsymbol{y}-B\hat{\boldsymbol{w}}_{\mathrm{MLE}})^T(\boldsymbol{y}-B\hat{\boldsymbol{w}}_{\mathrm{MLE}}) + (\boldsymbol{w}_0-\hat{\boldsymbol{w}}_{\mathrm{MLE}})^T \left\{ A^{-1}+(B^TB)^{-1} \right\}^{-1} (\boldsymbol{w}_0-\hat{\boldsymbol{w}}_{\mathrm{MLE}}) \\ \hat{\boldsymbol{w}}_n &= (B^TB+A^{-1})^{-1}(B^TB\hat{\boldsymbol{w}}_{\mathrm{MLE}}+A\boldsymbol{w}_0) \\ \hat{\boldsymbol{w}}_{\mathrm{MLE}} &= (B^TB)^{-1}B^T\boldsymbol{y} \end{aligned} \tag{19}$$

とする．

## 3 予測分布に対するラプラス近似

ラプラス近似の目的は，連続変数の集合上に定義される確率密度分布に対しガウス分布による近似を見出すことである．ベイズ予測分布は以下のように表すことができる．

$$h(\boldsymbol{z}|\boldsymbol{y}) = \frac{\int \exp[n\left\{q(\boldsymbol{w},\sigma^2|\boldsymbol{y}) + n^{-1}\log f(\boldsymbol{z}|\boldsymbol{w},\sigma^2)\right\}]d\boldsymbol{w}d\sigma^2}{\int \exp\left\{nq(\boldsymbol{w},\sigma^2|\boldsymbol{y})\right\} d\boldsymbol{w}d\sigma^2}. \tag{20}$$

小西・北川 (2004) によると，予測分布に対するラプラス近似を用いることにより，上式の $q(\boldsymbol{w},\sigma^2|\boldsymbol{y})$ のモードを用いることにより，予測分布の正規分布による近似を得ることができる．$q(\boldsymbol{w},\sigma^2|\boldsymbol{y})$ の $\boldsymbol{w}$ と $\sigma^2$ に関するモードをそれぞれ

$$\hat{\boldsymbol{w}}_{mode} = \arg\max_{\boldsymbol{w}} q(\boldsymbol{w},\sigma^2|\boldsymbol{y}), \quad \hat{\sigma}^2_{mode} = \arg\max_{\sigma^2} q(\boldsymbol{w},\sigma^2|\boldsymbol{y}) \tag{21}$$

とする．ここで

$$\begin{aligned} q(\boldsymbol{w},\sigma^2|\boldsymbol{y}) &= \frac{1}{n}\log\Bigg[(2\pi)^{-\frac{m}{2}-\frac{n}{2}}(\sigma^2)^{-\frac{m}{2}-\frac{n}{2}-(\frac{\nu_0}{2}+1)}|A^{-1}|^{-\frac{1}{2}}\frac{\left(\frac{\lambda_0}{2}\right)^{\frac{\nu_0}{2}}}{\Gamma\left(\frac{\nu_0}{2}\right)} \\ &\quad\times\exp\Bigg[-\frac{1}{2\sigma^2}\Big\{\lambda_0+(\boldsymbol{y}-B\hat{\boldsymbol{w}}_{\mathrm{MLE}})^T(\boldsymbol{y}-B\hat{\boldsymbol{w}}_{\mathrm{MLE}}) \\ &\qquad +(\boldsymbol{w}_0-\hat{\boldsymbol{w}}_{\mathrm{MLE}})^T\left\{A^{-1}+(B^TB)^{-1}\right\}^{-1}(\boldsymbol{w}_0-\hat{\boldsymbol{w}}_{\mathrm{MLE}})\Big\}\Bigg] \\ &\quad\times\exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{w}-\hat{\boldsymbol{w}}_n)^T\hat{A}_n^{-1}(\boldsymbol{w}-\hat{\boldsymbol{w}}_n)\right\}\Bigg] \end{aligned} \tag{22}$$

となる．したがって，

$$\arg\max_{\boldsymbol{w}} q(\boldsymbol{w},\sigma^2|\boldsymbol{y}) = \hat{\boldsymbol{w}}_n = (B^TB+A^{-1})^{-1}(B^TB\hat{\boldsymbol{w}}_{\mathrm{MLE}}+A\boldsymbol{w}_0) \tag{23}$$

次に，

$$\hat{\sigma}^2_{mode} = \arg\max_{\sigma^2} q(\boldsymbol{w},\sigma^2|\boldsymbol{y}) \tag{24}$$

を求めるため，$q(\boldsymbol{w},\sigma^2|\boldsymbol{y})$ の $\sigma^2$ に関する 1 階微分を計算する．$\sigma^2$ に依存しない項を $C$ とおくと，

$$\begin{aligned} q(\boldsymbol{w},\sigma^2|\boldsymbol{y}) &= -\frac{1}{n}\left(\frac{(m+n+\nu_0)}{2}+1\right)\log\sigma^2+C \\ &\quad+\frac{1}{n}\Bigg[-\frac{1}{2\sigma^2}\Big\{\lambda_0+(\boldsymbol{y}-B\hat{\boldsymbol{w}}_{\mathrm{MLE}})^T(\boldsymbol{y}-B\hat{\boldsymbol{w}}_{\mathrm{MLE}}) \\ &\qquad +(\boldsymbol{w}_0-\hat{\boldsymbol{w}}_{\mathrm{MLE}})^T\left\{A^{-1}+(B^TB)^{-1}\right\}^{-1}(\boldsymbol{w}_0-\hat{\boldsymbol{w}}_{\mathrm{MLE}})+(\boldsymbol{w}-\hat{\boldsymbol{w}}_n)^T\hat{A}_n^{-1}(\boldsymbol{w}-\hat{\boldsymbol{w}}_n)\Big\}\Bigg] \end{aligned} \tag{25}$$

と計算することができる．よって，(25) 式の $\sigma^2$ に関する 1 階微分を計算すると，

$$\begin{aligned} \frac{\partial q(\boldsymbol{w},\sigma^2|\boldsymbol{y})}{\partial\sigma^2} &= -\frac{1}{n}\left\{\frac{(m+n+\nu_0)}{2}+1\right\}\frac{1}{\sigma^2}+\frac{1}{2n(\sigma^2)^2}\Big\{\lambda_0+(\boldsymbol{y}-B\hat{\boldsymbol{w}}_{\mathrm{MLE}})^T(\boldsymbol{y}-B\hat{\boldsymbol{w}}_{\mathrm{MLE}}) \\ &\qquad +(\boldsymbol{w}_0-\hat{\boldsymbol{w}}_{\mathrm{MLE}})^T\left\{A^{-1}+(B^TB)^{-1}\right\}^{-1}(\boldsymbol{w}_0-\hat{\boldsymbol{w}}_{\mathrm{MLE}})+(\boldsymbol{w}-\hat{\boldsymbol{w}}_n)^T\hat{A}_n^{-1}(\boldsymbol{w}-\hat{\boldsymbol{w}}_n)\Big\} \end{aligned}$$

となる．したがって，$q(\boldsymbol{w},\sigma^2|\boldsymbol{y})$ の $\sigma^2$ に対するモードは以下のように計算することができる．

$$\begin{aligned} &\frac{\partial q(\boldsymbol{w},\sigma^2|\boldsymbol{y})}{\partial\sigma^2}=0 \\ &\Leftrightarrow \hat{\sigma}^2_{mode}=\frac{1}{m+n+\nu_0+2}\Big\{\lambda_0+(\boldsymbol{y}-B\hat{\boldsymbol{w}}_{\mathrm{MLE}})^T(\boldsymbol{y}-B\hat{\boldsymbol{w}}_{\mathrm{MLE}}) \\ &+(\boldsymbol{w}_0-\hat{\boldsymbol{w}}_{\mathrm{MLE}})^T\left\{A^{-1}+(B^TB)^{-1}\right\}^{-1}(\boldsymbol{w}_0-\hat{\boldsymbol{w}}_{\mathrm{MLE}})+(\boldsymbol{w}-\hat{\boldsymbol{w}}_n)^T\hat{A}_n^{-1}(\boldsymbol{w}-\hat{\boldsymbol{w}}_n)\Big\}. \end{aligned} \tag{26}$$

これは，以下の式と同等である．

$$\arg\max_{\sigma^2} q(\boldsymbol{w},\sigma^2|\boldsymbol{y}) \quad = \quad \frac{1}{m+n+\nu_0+2}\left\{\lambda_0+(\hat{\boldsymbol{w}}_n-\boldsymbol{w}_0)^T A(\hat{\boldsymbol{w}}_n-\boldsymbol{w}_0)+(\boldsymbol{y}-B\hat{\boldsymbol{w}}_n)^T(\boldsymbol{y}-B\hat{\boldsymbol{w}}_n)\right\}.$$

よって，ラプラス近似を用いることにより，予測分布は，

$$\begin{aligned} h(\boldsymbol{z}|\boldsymbol{y},\hat{\boldsymbol{w}}_{mode},\hat{\sigma}^2_{mode}) &= f(\boldsymbol{z}|\hat{\boldsymbol{w}}_{mode},\hat{\sigma}^2_{mode})(1+O_p(n^{-1})) \quad (27)\\ &= N(B\hat{\boldsymbol{w}}_{mode},\hat{\sigma}^2_{mode}I)(1+O_p(n^{-1})) \quad (28)\end{aligned}$$

と近似することが出来る．

# 4 基底展開法に基づく非線形回帰モデルにおける予測情報量規準 PIC の導出

ベイズアプローチに基づく予測情報量規準 PIC (Kitagawa, 1997) は予測分布とデータを生成した真の分布との K-L 情報量の推定量として導かれた．そのとき，予測情報量規準 PIC は，以下のように定義される (Kitagawa (1997))．

$$\mathrm{PIC} = -2\log h(\boldsymbol{z}|\boldsymbol{y},\hat{\boldsymbol{w}}_{mode},\hat{\sigma}^2_{mode}) + B_p(q(\cdot),\hat{\boldsymbol{w}}_{mode},\hat{\sigma}^2_{mode}) \quad (29)$$

## 4.1 バイアス補正項

ベイズ型予測分布に対する対数尤度関数 $\log h(\boldsymbol{z}|\boldsymbol{y},\hat{\boldsymbol{w}}_{mode},\hat{\sigma}^2_{mode})$ は，

$$\log h(\boldsymbol{z}|\boldsymbol{y},\hat{\boldsymbol{w}}_{mode},\hat{\sigma}^2_{mode}) = -\frac{n}{2}\log(2\pi)-\frac{1}{2}\log|\hat{\sigma}^2_{mode}I| - \frac{1}{2\hat{\sigma}^2_{mode}}(\boldsymbol{z}-B\hat{\boldsymbol{w}}_{mode})^T(\boldsymbol{z}-B\hat{\boldsymbol{w}}_{mode}) \quad (30)$$

となり，ベイズ型予測分布 $h(\boldsymbol{z}|\boldsymbol{y},\hat{\boldsymbol{w}}_{mode},\hat{\sigma}^2_{mode})$ に対する対数尤度と平均対数尤度の差の期待値は，以下のように計算できる．

$$\begin{aligned} & B_p(q(\cdot),\hat{\boldsymbol{w}}_{mode},\hat{\sigma}^2_{mode})\\ =\ & E_{q(\boldsymbol{y})}\left[\log h(\boldsymbol{y}|\boldsymbol{y},\hat{\boldsymbol{w}}_{mode},\hat{\sigma}^2_{mode}) - E_{q(\boldsymbol{z})}\left[\log h(\boldsymbol{z}|\boldsymbol{y},\hat{\boldsymbol{w}}_{mode},\hat{\sigma}^2_{mode})\right]\right]\\ =\ & -\frac{1}{2\hat{\sigma}^2_{mode}}\mathrm{tr}\left\{E_{q(\boldsymbol{y})}\left[(\boldsymbol{y}-B\hat{\boldsymbol{w}}_{mode})(\boldsymbol{y}-B\hat{\boldsymbol{w}}_{mode})^T - E_{q(\boldsymbol{z})}\left[(\boldsymbol{z}-B\hat{\boldsymbol{w}}_{mode})(\boldsymbol{z}-B\hat{\boldsymbol{w}}_{mode})^T\right]\right]\right\} \end{aligned} \quad (31)$$

## 4.2 真の分布に対する仮定

バイアス補正項を導出するため，真の分布に対し，次のような仮定を置く．真の分布 $q(\boldsymbol{z})$ がある $\boldsymbol{w}^*\in R^m$，$\sigma^{2*}\in R^1$ によって $q(\boldsymbol{z})=f(\boldsymbol{z}|\boldsymbol{w}^*,\sigma^{2*})$ で与えられると仮定する．この時，$\boldsymbol{z}|\boldsymbol{w}^*,\sigma^{2*}\sim N(B\boldsymbol{w}^*,\sigma^{2*}I)$ とする．

### 4.3 予測情報量規準 PIC

また，パラメータの事前分布の分散共分散行列は $A = (n\lambda)I_m$ とする．この時，ベイズ予測分布に対する平均対数尤度は以下のように計算出来る．

$$E_{q(\boldsymbol{z})}\left\{(\boldsymbol{z}-B\hat{\boldsymbol{w}}_{mode})(\boldsymbol{z}-B\hat{\boldsymbol{w}}_{mode})^T\right\} = E_{f(\boldsymbol{z}|\boldsymbol{w}^*,\sigma^{2*})}\left\{(\boldsymbol{z}-B\boldsymbol{w}^*)(\boldsymbol{z}-B\boldsymbol{w}^*)^T\right\} + (B\boldsymbol{w}^*-B\hat{\boldsymbol{w}}_{mode})(B\boldsymbol{w}^*-B\hat{\boldsymbol{w}}_{mode})^T \quad (32)$$

ここで，表記 $\triangle\boldsymbol{w} \equiv \boldsymbol{w}^* - \boldsymbol{w}_0$，$\alpha = n\lambda$ とすると，

$$\begin{aligned} B\boldsymbol{w}^* - B\hat{\boldsymbol{w}}_{mode} &= (BB^T+\alpha I_n)^{-1}BB^T(B\boldsymbol{w}^*-\boldsymbol{y}) + \alpha(BB^T+\alpha I_n)^{-1}B\triangle\boldsymbol{w} \quad (33)\\ \boldsymbol{y} - B\hat{\boldsymbol{w}}_{mode} &= \alpha(B^TB+\alpha I_m)^{-1}(\boldsymbol{y}-B\boldsymbol{w}^*) + \alpha(B^TB+\alpha I_m)^{-1}B\triangle\boldsymbol{w} \end{aligned}$$

と計算することができる．したがって，バイアス補正項は以下のように計算することができる．

$$\begin{aligned} & E_{q(\boldsymbol{y})}\left[E_{q(\boldsymbol{z})}\left[(\boldsymbol{z}-B\hat{\boldsymbol{w}}_{mode})(\boldsymbol{z}-B\hat{\boldsymbol{w}}_{mode})^T\right] - (\boldsymbol{y}-B\hat{\boldsymbol{w}}_{mode})(\boldsymbol{y}-B\hat{\boldsymbol{w}}_{mode})^T\right] \\ =\ & E_{f(\boldsymbol{y}|\boldsymbol{w}^*,\sigma^{2*})}\left[E_{f(\boldsymbol{z}|\boldsymbol{w}^*,\sigma^{2*})}\left[(\boldsymbol{z}-B\hat{\boldsymbol{w}}_{mode})(\boldsymbol{z}-B\hat{\boldsymbol{w}}_{mode})^T\right] - (\boldsymbol{y}-B\hat{\boldsymbol{w}}_{mode})(\boldsymbol{y}-B\hat{\boldsymbol{w}}_{mode})^T\right] \\ =\ & 2\sigma^{2*}(BB^T+\alpha I_n)^{-1}(BB^T) \quad (34) \end{aligned}$$

したがって，バイアス補正項の計算は以下のようになる．

$$\begin{aligned} & E_{q(\boldsymbol{y})}\left[\log h(\boldsymbol{y}|\boldsymbol{y},\hat{\boldsymbol{w}}_{mode},\hat{\sigma}^2_{mode}) - E_{q(\boldsymbol{z})}\left[\log h(\boldsymbol{z}|\boldsymbol{y},\hat{\boldsymbol{w}}_{mode},\hat{\sigma}^2_{mode})\right]\right] \\ =\ & -\frac{1}{2\hat{\sigma}^2}\mathrm{tr}\left[E_{q(\boldsymbol{y})}\left[(\boldsymbol{y}-B\hat{\boldsymbol{w}}_{mode})(\boldsymbol{y}-B\hat{\boldsymbol{w}}_{mode})^T - E_{q(\boldsymbol{z})}\left[(\boldsymbol{z}-B\hat{\boldsymbol{w}}_{mode})(\boldsymbol{z}-B\hat{\boldsymbol{w}}_{mode})^T\right]\right]\right] \\ =\ & -\left(\frac{\sigma^{2*}}{\hat{\sigma}^2_{mode}}\right)\mathrm{tr}\left[B(B^TB+n\lambda I_m)^{-1}B^T\right] \quad (35) \end{aligned}$$

よって，ベイズ型予測分布のモデル評価基準である，予測情報量規準 PIC は以下のように計算することが出来る．

$$\begin{aligned} \mathrm{PIC} &= n\log(2\pi) + n\log(\hat{\sigma}^2_{mode}) + \frac{1}{\hat{\sigma}^2_{mode}}(\boldsymbol{y}-B\hat{\boldsymbol{w}}_{mode})^T(\boldsymbol{y}-B\hat{\boldsymbol{w}}_{mode}) \\ &\quad +2\left(\frac{\sigma^{2*}}{\hat{\sigma}^2_{mode}}\right)\mathrm{tr}\left[B(B^TB+n\lambda I_n)^{-1}B^T\right]. \quad (36) \end{aligned}$$

### 4.4 数値実験

本節では，シミュレーションによって，提案する予測情報量規準 PIC と従来の一般化情報量規準 (GIC) とベイズ型一般化情報量規準 (GBIC) および Eilers and Marx (1996) による方法 (mAIC) を，基底関数との個数と平滑化パラメータおよび事前分布のハイパーパラメータの選択を通して比較検討する．PIC に含まれる真の分散 $\sigma^{2*}$ に関しては，最尤推定量を代入し数値実験を行った．ここでは，選択した基底関数の個数と平滑化パラメータの平均と標準偏差，推定曲線と将来の新たなデータとの予測二乗誤差に基づく比較を行う．データは，$y_\alpha = w(x_\alpha) + \varepsilon_\alpha$ に基づいて発生させ，真の曲線 $w(x)$ に対しては次の関数を仮定した．

$$w(x) = 1 - 48x + 218x^2 - 315x^3 + 145x^4. \quad (37)$$

$x$ の定義域は $[0,1]$ とし，データは等間隔に取った．誤差項 $\varepsilon_\alpha$ に対して正規分布 $\varepsilon_\alpha \sim N(0,\sigma^2)$, $\sigma = 0.2, 0.5, 1$ を仮定し，標本数 $n$ は $n = 20, 50, 100$ の 3 つの設定で実験を行った．表 $1 \sim 3$ は 500 回のシミュレーションに対する結果である．mean は，基底関数の個数と平滑化パラメータの平均，SD はそれらの標準偏差を表し，PSE は予測値 $\hat{y}$ と新たに観測されたデータ $z$ との予測二乗誤差

$$\mathrm{PSE} = \frac{1}{n}\sum_{\alpha=1}^{n}\{\hat{y}_\alpha - z_\alpha\}^2 \tag{38}$$

である．基底関数の個数は $4 \sim 15$ の範囲で調べ，平滑化パラメータの値は，$\lambda = 10^{-1}$ から $\lambda = 10^{-8}$ を 100 等分した $\lambda_1, \cdots, \lambda_{100}$ を調べた．

表 1: $n = 20$

| **model selection** | **number of basis function mean − SD** | **hyper parameter of prior distribution mean − SD** | **predictive squared error mean − SD** |
|---|---|---|---|
| $(\tau = 0.2)$ | | | |
| mAIC | 12.064 – 3.545 | 404.001 – 1029.419($\times 10^{-5}$) | 1.341 – 0.483 |
| GIC | 12.536 – 3.205 | 91.001 – 433.306 ($\times 10^{-5}$) | 1.361 – 0.486 |
| GBIC | 8.996 – 1.396 | 1149.001 – 684.277($\times 10^{-5}$) | 1.470 – 0.534 |
| PIC | 11.864 – 0.615 | 3167.001 – 92.672 ($\times 10^{-5}$) | 2.333 – 0.588 |
| $(\tau = 0.5)$ | | | |
| mAIC | 12.044 – 3.711 | 350.801 – 705.816 ($\times 10^{-5}$) | 8.254 – 2.981 |
| GIC | 12.3 – 3.471 | 104.401 – 375.210 ($\times 10^{-5}$) | 8.423 – 2.476 |
| GBIC | 8.828 – 0.976 | 1603.401 – 498.503 ($\times 10^{-5}$) | 7.900 – 2.437 |
| PIC | 12.992 – 1.548 | 2858.001 – 200.390 ($\times 10^{-5}$) | 7.788 – 2.437 |
| $(\tau = 1)$ | | | |
| mAIC | 12.208 – 3.532 | 299.401 – 565.171 ($\times 10^{-5}$) | 32.680 – 11.805 |
| GIC | 12.368 – 3.422 | 114.201 – 338.478 ($\times 10^{-5}$) | 33.377 – 12.132 |
| GBIC | 8.688 – 1.250 | 3280.001 – 2596.629 ($\times 10^{-5}$) | 28.620 – 9.182 |
| PIC | 13.81 – 1.815 | 2357.601 – 278.563 ($\times 10^{-5}$) | 26.754 – 8.709 |

表 2: $n = 50$

| **model selection** | **number of basis function mean − SD** | **hyper parameter of prior distribution mean − SD** | **predictive squared error mean − SD** |
|---|---|---|---|
| $(\tau = 0.2)$ | | | |
| mAIC | 7.852 – 2.723 | 56.801 –302.120($\times 10^{-5}$) | 2.359 – 0.537 |
| GIC | 8.4 – 2.904 | 30.001 – 188.603 ($\times 10^{-5}$) | 2.374 – 0.537 |
| GBIC | 9.676 – 1.596 | 328.801 – 266.970($\times 10^{-5}$) | 2.493 – 0.541 |
| PIC | 12.112 – 0.469 | 1802.801 – 34.562 ($\times 10^{-5}$) | 3.902 – 0.723 |
| $(\tau = 0.5)$ | | | |
| mAIC | 7.326 – 3.032 | 91.201 – 263.859 ($\times 10^{-5}$) | 14.506 – 3.259 |
| GIC | 7.826 – 3.218 | 62.001 –217.648 ($\times 10^{-5}$) | 14.642 – 3.286 |
| GBIC | 10.034 – 1.392 | 659.201 – 126.520 ($\times 10^{-5}$) | 15.248 – 3.118 |
| PIC | 12.804 – 1.335 | 1530.401 – 84.440 ($\times 10^{-5}$) | 15.591 – 3.152 |
| $(\tau = 1)$ | | | |
| mAIC | 7.852 – 3.334 | 158.001 – 289.696 ($\times 10^{-5}$) | 57.688 – 13.013 |
| GIC | 7.94 – 3.315 | 93.601 – 240.238 ($\times 10^{-5}$) | 58.280 – 13.208 |
| GBIC | 9.34 – 1.593 | 682.401 –244.724 ($\times 10^{-5}$) | 57.462 – 11.763 |
| PIC | 12.316 – 2.340 | 1091.201 – 125.035 ($\times 10^{-5}$) | 56.861 – 11.529 |

表 3: $n = 100$

| **model selection** | **number of basis function mean − SD** | **hyper parameter of prior distribution mean − SD** | **predictive squared error mean − SD** |
|---|---|---|---|
| $(\tau = 0.2)$ | | | |
| mAIC | 7.638 – 2.039 | 5.401 – 41.885($\times 10^{-5}$) | 4.390 – 0.679 |
| GIC | 8.046 – 2.377 | 3.201 – 40.904 ($\times 10^{-5}$) | 4.409 – 0.686 |
| GBIC | 10.146 – 1.334 | 158.401 – 102.813($\times 10^{-5}$) | 4.555 – 0.697 |
| PIC | 12.826 – 0.522 | 1098.801–10.899 ($\times 10^{-5}$) | 6.265 – 0.8857 |
| $(\tau = 0.5)$ | | | |
| mAIC | 6.934–2.691 | 34.401– 121.020 ($\times 10^{-5}$) | 27.299 – 4.239 |
| GIC | 7.238 – 2.807 | 23.401 – 98.039 ($\times 10^{-5}$) | 27.363 – 4.215 |
| GBIC | 10.438 – 1.512 | 337.801 – 75.912 ($\times 10^{-5}$) | 28.501 – 4.257 |
| PIC | 13.322 – 1.238 | 870.401 – 47.837 ($\times 10^{-5}$) | 28.955 – 4.316 |
| $(\tau = 1)$ | | | |
| mAIC | 6.95 – 2.896 | 57.601 – 136.233 ($\times 10^{-5}$) | 108.240 – 16.684 |
| GIC | 7.134 – 2.916 | 42.801 – 118.305 ($\times 10^{-5}$) | 108.543 – 16.726 |
| GBIC | 9.582 – 1.694 | 329.001 – 77.918 ($\times 10^{-5}$) | 110.175 – 16.503 |
| PIC | 11.854 – 2.327 | 582.001 – 56.590 ($\times 10^{-5}$) | 109.693 – 16.402 |

## 参考文献


[1] Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.

[2] Davison, A. C. (1986). Approximate predictive likelihood. *Biometrika*, **73**, 323–32.

[3] Denison, D. G. T., Holmes, C. C., Mallick, B. K. and Smith A. F. M. (2002). *Bayesian Methods for Nonlinear Classification and Regression*. Wiley.

[4] Eilers, P. and Marx, B. (1996). Flexible smoothing with B-splines and penalties (with discussion). *Statistical Science*, **11**, 89–121.

[5] Hastie, T. and Tibshirani, R. (1990). *Generalized Additive models*. Chapman and Hall.

[6] Kitagawa, G. (1997). Information criteria for the predictive evaluation of bayesian models. *Communications in Statistics-Theory and Methods*, **26**, 2223–2246.

[7] Konishi, S., Ando, T. and Imoto, S. (2004). Bayesian information criteria and smoothing parameter selection in radial basis function networks. *Biometrika*, **91**, 27–43.

[8] 小西貞則，北川源四朗 (2004)． 情報量規準．朝倉書店.

[9] Kullback, S. and Leibler, R. A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, **22**, 79–86.

[10] 中妻照雄 (2003). ファイナンスのための MCMC 法によるベイズ分析, 三菱経済研究所

[11] Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, **6**, 461–464.

[12] Tierney, L and and Kadane, J. B. (1986). Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association*, **81**, 82–86.