応用力学論文集 Vol.7,pp.145-153 (2004年8月) 土木学会

赤池ベイズ情報量規準による事前分布の選択を考慮した逆解析に関する基礎的研究

On selection of prior distribution in inverse analyses by Akaike Bayesian Information Criterion

本城勇介* ・ Budhi SETIAWAN **
Yusuke HONJO and Budihi SETIAWAN

*正会員 Ph.D. 岐阜大学教授 工学部 (〒 501-1193 岐阜市柳戸 1-1)
**岐阜大学 工学部 大学院生 (同上)

The first author has proposed to use Akaike Bayesian Information criterion (ABIC) to adjust relative weight between objective and subjective information in inverse analysis in order to overcome the problem of ill-posedness. The method has been applied to various civil engineering problems in the past and found to be very effective. However, the reason for this effectiveness was not necessarily clearly explained. In this study, an attempt is made to explain the behavior of ABIC from the viewpoint of information entropy. It is found that ABIC chooses the estimates of parameters that maximize the reduction of information entropy from the entropy given at the beginning of the analysis. This fact actually extends the use of ABIC to wider selection of the prior information, i.e. not only choice of prior variance but also alternative prior means can be examined using ABIC as a criterion. The findings are not explained theoretically, but also illustrated using a simple numerical example.

Key Words : *Inverse analysis, regularization procedure, ABIC, extended Bayesian method, Information entropy*

1. はじめに

工学分野で出会う, 多くの逆問題は, 不適切性を有する. 本城 (1989, 1995)[1),2)]はこの不適切性を克服するための一つの方法として, 拡張ベイズ法と呼ばれる方法により, 事前情報を適当な強度で観測データに重ね合わせることによりこの問題を解決することを提案している. この方法では赤池ベイズ情報量基準 (以下 ABIC と呼ぶ) を用いることにより, 事前情報と, 観測データにより与えられる情報の適当な融合を試みている. そして, そのような融合がなぜうまく機能するかと言う点については, 他の方法との比較において説明が試みられている (本城・酒向・菊池, 1999)[3)].

ところで元々ABIC は赤池が, フィッシャー以来の伝統的な統計的推論と, ベーズ推定の融合を試みると言う野心的な構想の元に提案した理論的な枠組み (Akaike,1982; 赤池,1980,1989)[4),5),6)] を利用したものであり, AIC が提案しているモデル間の選択の問題を, 事前分布の選択の問題に拡張したものである.

ベイズ推定では, 対象とする事象の確率分布のパラメータの確率分布は, これらパラメータの事前分布と, 観測データのデータ分布関数よりベイズの定理を介して得られる条件付確率分布である事後分布として得られる. それでは, 事前分布として何を想定すれば, 得られた事後分布が客観的といえるであろうか. これがBayes 以来繰り返された議論である. この議論を貫く客観性の根拠を, 「エントロピーの推定量である対数尤度に見出そう」 (赤池, 1980)[5)] というのが, 赤池の基本的な立場である. この結果, 赤池の提案する情報量基準を用いれば, ベイズ推定は「データから特定の目的に適した情報を抽出するための道具立てであり」, 「情報を適切に利用する仕組みを与えるにすぎないこと」が理解される (赤池, 1989)[6)]. この立場からすると, 事前分布の選択の問題は, 単に観測情報と事前情報の不確実性の程度の調整に留まらず, いろいろな代替的な事前分布の選択に応用できることになる.

しかしながら, 赤池の上記の問題に対する理論的説明は, かなり抽象的であり, これを理解することは容易ではない. 本研究では, この説明をかなり一般性を持った逆解析問題の定式化を基に, 具体的に展開し, この方法が有効に機能する理由を, 情報エントロピーの立場から説明しようと試みたものである. その結果, 赤池の情報量基準は, 情報エントロピーにより表現されるデータ分布のあいまいさと事前分布のあいまいさの総和と, 事後分布のあいまいさの差, つまりエントロピーの減少量を最大にするように, 事前分布を選択する, という解釈を行うことができることを見出した. さらに本研究では, 非常に簡単な線形逆問題を例として, 上記のメカニズムを数値的に例示している.

読者は, 本論文の結果より, 赤池の情報量基準が, 従来からAIC として知られるモデル選択の問題ばかりでなく, これを含んでさらにベイズ推定の事前分布の選択

¶ Dedicated to the memory of Prof. Michihiro KITAHARA

にも利用できることを理解されるであろう．さらに赤池の提案している枠組みに従えば，伝統的な統計的推定も，ベイズ推定も本質的に同じ枠組みで議論できることを認識されるであろう．

2. 拡張ベイズ法による逆解析

2.1 離散型逆問題の定式化

離散型の線形逆問題は，最終的に一つの連立方程式を解くことに帰着する (グロエッチェ, 1996)[7]．

$$y = X\theta + \varepsilon \tag{1}$$

ここに，y は n 次元観測値ベクトル，X は与えられた $n \times m$ 観測行列，θ は推定しようとする m 次元モデルパラメータベクトル，ϵ は n 次元誤差ベクトルである．

通常の逆問題では，観測値の数は，推定しようとするパラメータの数よりも多く，典型的には最小二乗法により解が求められる．

$$\begin{aligned} min. \qquad J(\theta) &= \varepsilon^T \varepsilon \\ &= (y - X\theta)^T (y - X\theta) \end{aligned} \tag{2}$$

この式は，解析的に解くことができ，

$$\hat{\theta} = (X^T X)^{-1} X^T y \tag{3}$$

このとき逆行列を求めようとする行列 $X^T X$ の det が 0(換言すると，行列 $X^T X$ にランク落ちがある)，あるいは 0 に近い状態のとき，この逆行列の計算は不可能となり，解を求められなくなるし，またたとえ解が求まったとしても，観測値のわずかな変化が，パラメータ推定値に極端な影響をあたえるような不安定な結果を与える．このような状態を一般に不適切性 (ill-posedness) と言う．

不適切性の克服が逆解析では重要である．この問題の解決の一つに，事前情報の導入がある．これを定式化するのが，拡張ベイズ法であり，これを次に説明する．

2.2 拡張ベイズ法の定式化

(1) 基本モデル

拡張ベイズ法によるパラメータ θ の推定過程は，次のように定式化される．まず，次の 2 つのモデルを考える．

観測モデル： 観測方程式により与えられ，n 個の観測値 y を得ることにより，これと n 行 m 列の観測行列 X により関係付けられた m 個のパラメータ θ を推定する．

$$y = X\theta + \varepsilon \tag{4}$$

ここに $\varepsilon \sim N(0, {\sigma_\varepsilon}^2 V_\varepsilon)$．$V_\varepsilon$ は，誤差間の相関行列であり，${\sigma_\varepsilon}^2$ は，分散の程度を表すスカラー量である．

従って，データ分布は次のような多変量正規分布として与えられる．

$$\begin{aligned} f(y|\theta) &= \frac{1}{(2\pi)^{n/2} |{\sigma_\varepsilon}^2 V_\varepsilon|} \\ & exp\left\{ -\frac{1}{2} (y - X\theta)^T \frac{1}{{\sigma_\varepsilon}^2} {V_\varepsilon}^{-1} (y - X\theta) \right\} \end{aligned} \tag{5}$$

事前情報モデル： パラメータ θ に関する情報としては，観測値 y の他に，これに関する次の式で表現されるような事前情報が存在する．

$$\theta = \theta^* + \delta \tag{6}$$

ここに $\delta \sim N(0, {\sigma_\theta}^2 V_\theta)$．$V_\theta$ は，事前平均値間の相関行列で，${\sigma_\theta}^2$ はそれらの分散の程度を表す．

従って，θ の事前分布は，次の多変量正規分布となる．

$$\begin{aligned} f(\theta|\sigma_\theta, \theta^*) &= \frac{1}{(2\pi)^{m/2} |{\sigma_\theta}^2 V_\theta|^{1/2}} \\ & exp\left\{ -\frac{1}{2} (\theta - \theta^*)^T \frac{1}{{\sigma_\theta}^2} {V_\theta}^{-1} (\theta - \theta^*) \right\} \end{aligned} \tag{7}$$

ここでデータ分布と事前分布の相対的な重みを表すパラメータ λ^2（超パラメータ）を定義する．

$$\lambda^2 = \frac{{\sigma_\epsilon}^2}{{\sigma_\theta}^2} \tag{8}$$

式 (8) からわかるように，λ^2 が小さいほど事前分布の分散は相対的に大きくなり，従って，推定におけるデータの重みが増すことになる．

(2) パラメータのベイズ推定

以上より，θ の事後分布は，ベイズの定理により次のようになる．

$$\begin{aligned} f(\theta|\lambda^2, \theta^*, {\sigma_\varepsilon}^2, y) &\propto f(y|\theta, {\sigma_\varepsilon}^2) f(\theta|\theta^*, \lambda^2) \\ &= \frac{\lambda^{2m}}{(2\pi {\sigma_\varepsilon}^2)^{(m+n)}} |V_\varepsilon|^{-\frac{n}{2}} |V_\theta|^{-\frac{m}{2}} \\ & \exp\left[-\frac{1}{2{\sigma_\varepsilon}^2} \left\{ (y - X\theta)^T {V_\varepsilon}^{-1} (y - X\theta) \right.\right. \\ & \left.\left. + \lambda^2 (\theta - \theta^*)^T {V_\theta}^{-1} (\theta - \theta^*) \right\} \right] \end{aligned} \tag{9}$$

θ のベイズ推定量は，式 (9) を最大化することにより得られるから，最大化に関係のない定数項を省略すると，次に示す関数を最小化することになる．

$$\begin{aligned} J(\theta) &= (y - X\theta)^T {V_\varepsilon}^{-1} (y - X\theta) \\ & + \lambda^2 (\theta - \theta^*)^T {V_\theta}^{-1} (\theta - \theta^*) \end{aligned} \tag{10}$$

従って，θ のベイズ推定量 $\hat{\theta}$ は，次のように求められる．

$$\hat{\theta} = \theta^* + P X^T {V_\varepsilon}^{-1} X (y - X\theta^*) \tag{11}$$

ここに，

$$P = (X^T {V_\varepsilon}^{-1} X + \lambda^2 {V_\theta}^{-1})^{-1} \tag{12}$$

P はまた，ベイズ推定量 $\hat{\theta}$ の事後共分散行列である．

(3) 赤池ベイズ情報量規準（ABIC）の適用

(2) 節では，観測データ分布と事前情報分布の事前平均 θ^* と，共分散行列の相対的な重み付けを決定するパラメータ λ が与えられれば，これに対応したベイズ推定量を求めることができるように，定式化されている．これらのパラメータを，超パラメータと言う．この節では，適切な θ^* や λ を決めるための規準である ABIC 導出の手順を説明する (Akaike, 1980)．

今観測値 y の真の分布を示す確率密度関数を $g(y)$ で表す．$g(y)$ は，決して我々には知り得ない関数であるが，これを最終的に精度良く予測することが求められる．

一方ベイズ推定では，データ分布 $f(y|\theta)$ と，事前分布 $f(\theta|\theta^*,\lambda)$ がある．ここで，θ が確率変数であることを考慮すると，このモデルの y の分布を平均的に表す分布は，

$$f(y|\theta^*,\lambda)=\int f(y|\theta)f(\theta|\theta^*,\lambda)d\theta \qquad (13)$$

であり，これはベイズ尤度関数として知られている．この $f(y|\theta^*,\lambda)$ と $g(y)$ が，なるべく一致するように θ^* や λ を決定する必要がある．

2つの確率密度関数の近さを測る規準の1つに，Kullback-Leibler の情報規準がある．今，真の分布を $g(y)$ で，モデル分布を $f(y|\theta^*,\lambda)$ で表すと，この規準は次のようになる．

$$B\big(g(y),f(y|\theta^*,\lambda)\big)=\int g(y)ln\frac{g(y)}{f(y|\theta^*,\lambda)}d\theta \qquad (14)$$

さらにこの規準は，次の性質を満足することを証明できる（坂元・石黒・北川, 1983）[8]．

$$B\big(g(y),f(y|\theta^*,\lambda)\big)\geq 0$$
$$B\big(g(y),f(y|\theta^*,\lambda)\big)=0\Leftrightarrow g(y)=f(y|\theta^*,\lambda) \qquad (15)$$

従って，K-L 規準をなるべく小さくするような超パラメータ θ^* や λ の選択をすることが望ましい．

ところで，真の分布 $g(y)$ は，決して求められないが，これに代わるものとして，観測データ $y_i(i,\ldots,n)$ がある．これを用いて，K-L 情報量を近似的に評価する方法を考える．今，一連の観測データ $y_i(i,\ldots,n)$ が独立に得られていると仮定すると，それぞれの生起の確率は 1/n である．これにより K-L 情報量は，

$$\begin{aligned}&B\big(g(y),f(y|\theta^*,\lambda)\big)\\&=\int g(y)\ln g(y)dy-\int g(y)\ln f(y|\theta^*,\lambda)dy\\&\approx \text{Constant}-\sum_{i=1}^{n}\frac{1}{n}\ln f(y_i|\theta^*,\lambda)\end{aligned} \qquad (16)$$

この近似では，真の分布 $g(y)$ は，観測データ $y_i(i,\ldots,n)$ によって，置き換えられている．

以上より，モデル $f(y|\theta^*,\lambda)$ をできる限り真の分布に近づけるには，式 (16) の第2項をできるだけ大きくすればよい．すなわち，次の量を最大にしてやればよい．

$$\begin{aligned}&\sum_{i=1}^{n}\frac{1}{n}\ln f(y_i|\theta^*,\lambda)\\&=\sum_{i=1}^{n}\frac{1}{n}\ln\int f(y_i|\theta)f(\theta|\theta^*,\lambda)d\theta\end{aligned} \qquad (17)$$

ABIC は，この量の最大化を考えているが，歴史的な経緯によりこれに−2を乗じた値の最小化を考える．さらに不偏推定量は，この値から超パラメータ数 $\dim(\theta^*,\lambda)$ を引いたものであるため，最終的に ABIC は次式のように与えられる．

$$\begin{aligned}&ABIC=-2(\text{最大対数尤度})+2(\text{超パラメータ数})\\&=-2\sum_{i=1}^{n}\frac{1}{n}\ln\int f(y_i|\theta)f(\theta|\theta^*,\lambda)d\theta+2\dim(\theta^*,\lambda)\end{aligned} \qquad (18)$$

式 (18) の観測方程式が線形の場合の積分は，次の形で与えられる．

$$\begin{aligned}&L(\sigma_\epsilon{}^2,\theta^*,\lambda^2|y)\\&=(2\pi)^{-n/2}(\sigma_\epsilon{}^2)^{-n/2}|\lambda^2V_\theta{}^{-1}|^{1/2}\\&\quad|V_\epsilon{}^{-1}|^{1/2}|X^TV_\epsilon{}^{-1}X+\lambda^2V_\theta{}^{-1}|^{-1/2}\\&\quad\exp\Big[-\frac{1}{2\sigma_\epsilon{}^2}\Big\{(y-X\theta)^TV_\epsilon{}^{-1}(y-X\theta)\\&\quad-\lambda^2(\theta-\theta^*)^TV_\theta{}^{-1}(\theta-\theta^*)\Big\}\Big]\end{aligned} \qquad (19)$$

なお，この式ではデータ分布を多変量正規分布として記述しており，式 (18) で仮定している，個々の観測値が独立である場合は，この特殊なときである．

今，ABIC を用いて $\sigma_\epsilon{}^2$ を推定すると，結局式 (19) を最大化する $\sigma_\epsilon{}^2$ が，求めようとする $\sigma_\epsilon{}^2$ で，これは式 (19) の自然対数をとり，これを σ_ϵ^2 で微分して0と置くことにより容易に計算でき，次式のようになる．

$$\begin{aligned}\hat{\sigma_\epsilon{}^2}=\frac{1}{n}\big\{&(y-X\hat\theta)^TV_\epsilon{}^{-1}(y-X\hat\theta)\\&+\lambda^2(\hat\theta-\theta^*)^TV_\theta{}^{-1}(\hat\theta-\theta^*)\big\}\end{aligned} \qquad (20)$$

式 (20) を式 (19) に代入し，整理すると，

$$\begin{aligned}&\ln L(\theta^*,\lambda^2|y)=\\&-\frac{n}{2}\ln(2\pi)-\frac{n}{2}\ln(\hat{\sigma_\epsilon{}^2})+\frac{1}{2}\ln|\lambda^2V_\theta{}^{-1}|\\&+\frac{n}{2}|V_\epsilon{}^{-1}|-\frac{n}{2}\ln|X^TV_\epsilon{}^{-1}X+\lambda^2V_\theta{}^{-1}|\end{aligned} \qquad (21)$$

これを，式 (18) に代入すると，最終的に λ^2 と θ^* の決定に用いる ABIC を求めることができる．

$$\begin{aligned}&ABIC=n\ln(2\pi)+n\ln(\hat{\sigma_\epsilon{}^2})\\&\quad-\ln|\lambda^2V_\theta{}^{-1}|-\ln|V_\epsilon{}^{-1}|\\&\quad+\ln|X^TV_\epsilon{}^{-1}X+\lambda^2V_\theta{}^{-1}|+2\dim(\theta^*,\lambda)\end{aligned} \qquad (22)$$

(4) ABIC の情報エントロピーによる解釈

赤池ベイズ情報量規準（ABIC）は，最良の事前分布を決定する，非常に有用な道具である．この節では，情報エントロピーからその挙動をより明確に説明することを試みた．情報エントロピーは，当該状況のあいまいさを表現する尺度であり，これが大きいほど，状況はあいまいであることを表す．一般に n 次元正規分布の情報エントロピーは，次式で与えられることが知られている（有本, 1980）[9]．

$$H = \frac{n}{2}(1 + \ln 2\pi) + \frac{1}{2}\ln|V| \tag{23}$$

ここでは V はこの分布の共分散行列である．

以上の準備のもとで，式 (22) に示した ABIC を，次のように書き換えることができる．

$$\begin{aligned}
&\frac{1}{2}\mathrm{ABIC} \\
&= \left\{\frac{n}{2}\ln(2\pi) + \frac{1}{2}\ln \hat{\sigma_\varepsilon}^{2n}|V_\epsilon|\right\} \\
&+\left\{\frac{m}{2}\ln(2\pi) + \frac{1}{2}\ln \hat{\sigma_\varepsilon}^{2m}\lambda^{-2}|V_\theta|\right\} \\
&-\left\{\frac{m}{2}\ln(2\pi) + \frac{1}{2}\ln \hat{\sigma_\varepsilon}^{2m}\left|(X^T {V_\varepsilon}^{-1} X + \lambda^2 {V_\theta}^{-1})^{-1}\right|\right\} \\
&+\dim(\theta^*, \lambda)
\end{aligned} \tag{24}$$

上式を見ると $\frac{1}{2}$ABIC は，$\dim(\theta^*, \lambda)$, $\frac{n}{2}$, $\frac{m}{2}$ といった定常値を無視すると，次のようになっている．

$$\begin{aligned}
ABIC &= (\text{データ分布の情報エントロピー}) \\
&\quad + (\text{事前分布の情報エントロピー}) \\
&\quad - (\text{事後分布の情報エントロピー}) + (\text{定数}) \\
&= (\text{減少エントロピー})
\end{aligned} \tag{25}$$

以上より ABIC は一応，次のように解釈される．

> *ABIC* は，データ分布の持つあいまいさ（情報エントロピー）と，事前分布の持つあいまいさ（情報エントロピー）の総和と，事後分布の持つあいまいさ（情報エントロピー）の差が，最も大きくなるように，すなわち事前と事後におけるあいまいさ（情報エントロピー）の減少量が最大となるように，事前分布を選択する．

3. 例題による考察

本章では，簡単な逆問題を解析し，これによって，先に述べた ABIC の挙動と，情報エントロピーによる解釈を考察，説明する．

3.1 例題の説明

ここに取り上げた例題は，**図-1** に示すような，2 つの連結したばねのモデルであり，それぞれ節点 1 と 2 に

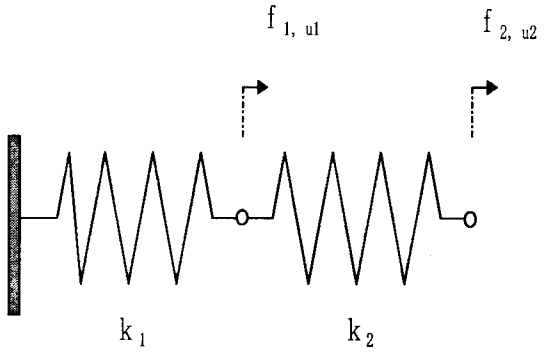


図-1 例題：バネモデル

力 f_1 と f_2 を作用させたときの変位を，u_1 と u_2 とする．この問題では 3 組の，力 f_1 と f_2 の組み合わせを作用させたときの，u_2 の値を観測し，この情報よりばね定数 k_1 と k_2 を推定する（佐藤, 1996）[10]．この問題を，逆問題の基本式，式 (1) に則して記述すると，各項は，次のようになる：

$$y = \begin{pmatrix} u_{21} \\ u_{22} \\ u_{23} \end{pmatrix} \quad \theta = \begin{pmatrix} 1/k_1 \\ 1/k_2 \end{pmatrix} \quad X = \begin{pmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \\ f_{31} & f_{32} \end{pmatrix}$$

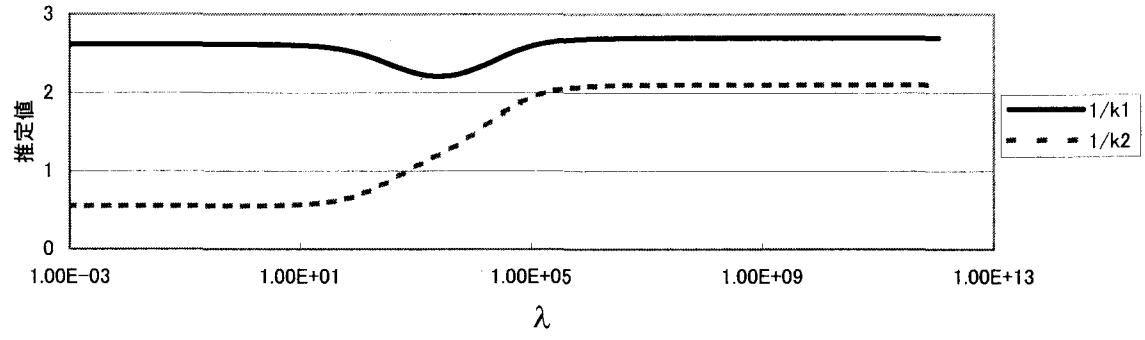
ここに示すケース・スタディーでは，3 組の事前情報と，4 種類の大きさの異なるノイズを持つデータを準備した．事前情報に関しては，PG,PM,PW（P : Prior, G : Good, M : Medium, W : Worse）の順に，事前平均値は真値から離れる．すなわち，PG では (1.8,1.4), PM では (2.7,2.1), PW では (3.6,2.8) を与えている．なお真値は，$1/k_1 = 1.8$, $1/k_2 = 1.4$ である．

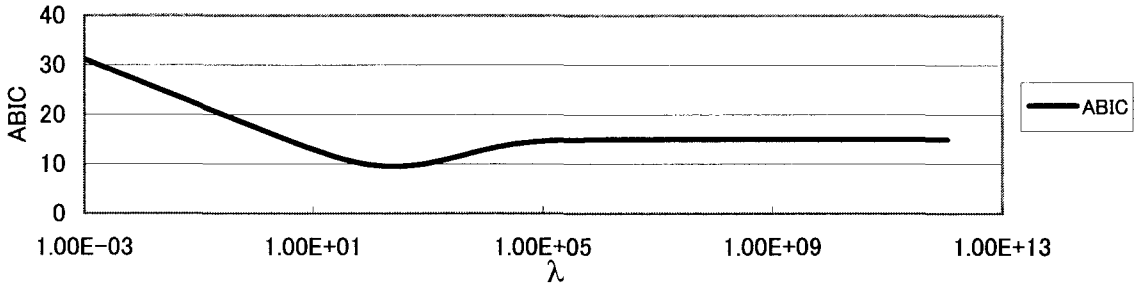
一方，ノイズについては，標準正規乱数 (N(0,1)) を基準にして，N1/2,N1,N3,N5 の順に観測値に付加されるノイズが大きくなる (1/2, 1 等は，計算値に，標準正規乱数を何倍して加えたかを示している)．またノイズの N0 は，ノイズの無い場合である．

表-1 に示すように，3 組の事前情報と，5 種類のノイズの強さを組み合わせた，合計 15 組の場合について逆解析を実施した．

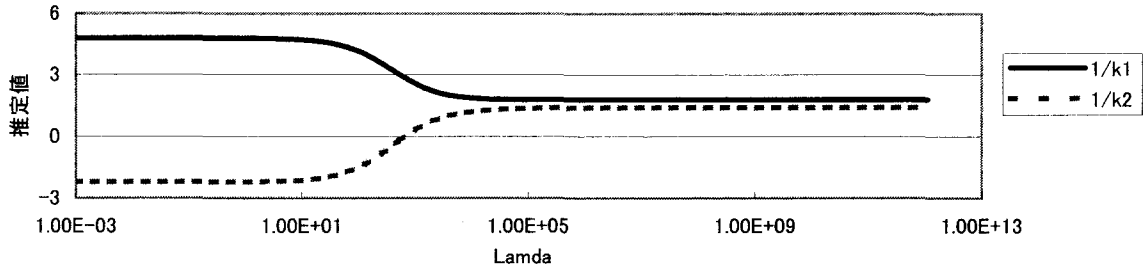
3.2 最良の事前分布の選択

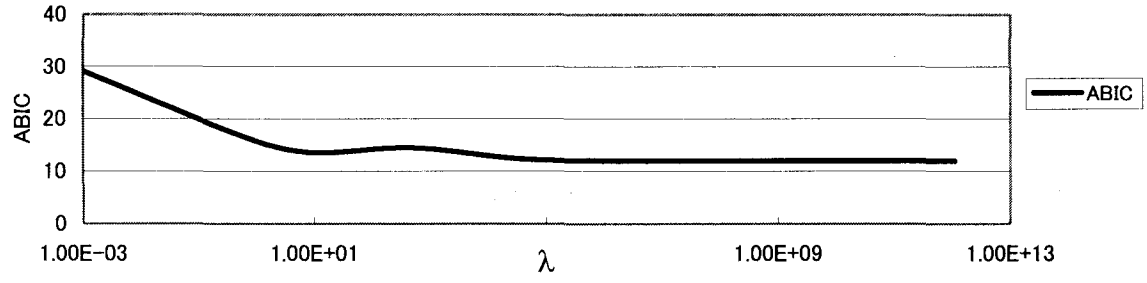
表-1 の観測データに含まれるノイズのレベルが同じケースを比較すると，それぞれ ABIC が最小値をとるとき，最良の事前分布を選択している．すなわち，ABIC が最良の事前分布の選択において，有効であることを示している．また，的確な事前分布が与えられる場合の方が，大きな λ^2 を取っている．これは事前情報に相対的に大きな重みを置く方が良いことを，ABIC が自動的に



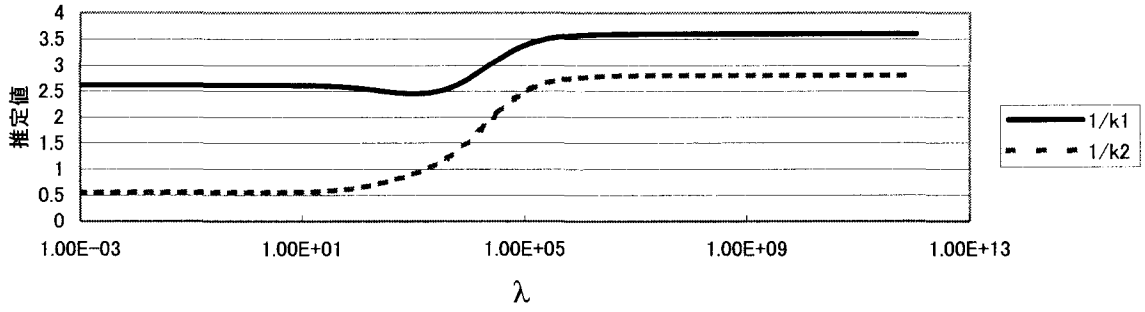


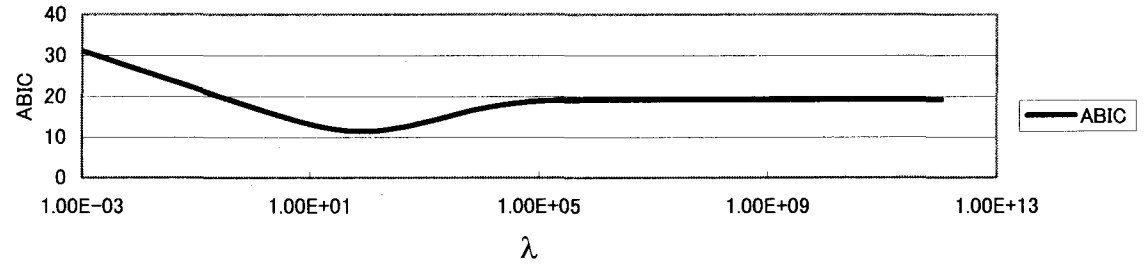
(a) 推定値、ABICとλ(ケース PM-N1)





(b) 推定値、ABICとλ(ケース PG-N3)





(c) 推定値、ABICとλ(ケース PW-N1)

図-2　推定値、ABIC とλ (ケース PW-N1)

表-1 各ケースでの ABIC, λ^2, 推定値及び標準偏差

ケース	ABIC	λ^2	$1/k_1$ (1.8)	標準偏差 $1/k_1$	$1/k_2$ (1.4)	標準偏差 $1/k_2$
PG-N0	5.027	1.64×10^2	1.84	0.374	1.48	0.46
PM-N0	9.718	5.12	1.88	1.405	1.56	1.63
PW-N0	12.088	1.28	1.74	1.841	1.72	2.131
PG-N1/2	3.729	6.55×10^2	1.82	0.205	1.42	0.26
PM-N1/2	9.066	2.56	2.09	1.653	1.25	1.914
PW-N1/2	11.627	1.28	2.14	1.843	1.19	2.131
PG-N1	5.059	1.37×10^9	1.8	1.499×10^{-4}	1.4	1.927×10^{-4}
PM-N1	9.514	2.56	2.41	1.653	0.82	1.914
PW-N1	11.563	6.4×10^{-1}	2.58	1.967	0.61	2.401
PG-N3	11.933	2.15×10^7	1.8	1.199×10^{-3}	1.4	1.541×10^{-3}
PM-N3	13.362	1.6×10^{-1}	4.68	2.079	-2.09	2.401
PW-N3	13.508	1.6×10^{-1}	4.69	2.079	-2.1	2.401
PG-N5	14.764	2	6.94	2.116	-4.97	2.443
PM-N5	14.643	2	6.95	2.116	-4.97	2.443
PW-N5	14.619	2	6.95	2.116	-4.97	2.443

調整していることを意味している.

推定値そのものも, ノイズの特に大きい N3,N5 は別として, 他のケースでは, 一応妥当な推定値が得られた.

3.3 ABIC の情報エントロピーによる考察

ABIC の挙動の説明のため, 先に設定したケースの内, PM-N1, PG-N3, PW-N1 の 3 つのケースに着目してみよう.

事前情報を $1/k_1 = 2.7$, $1/k_2 = 2.1$ として, ベイズ推定量を求めた PM-N1 のケースでは, **図-2(a)** に示すように, λ^2 が小さい間, k_1 と k_2 は式 (10) の第 1 項を最小とする, すなわち観測値と計算値の残差二乗和を最小とする値を与える. 一方, λ^2 が非常に大きくなると, 式 (10) の第 2 項が支配的となるため, k_1 と k_2 はこの事前平均値をとる. この 2 者の間で λ^2 がある大きさのとき, k_1 と k_2 は, 前者から後者へ緩やかに移動する. ABIC を最小とする k_1 と k_2 は, このような移動の中間で表れる.

次にケース PG-N3, すなわち観測データのノイズが大きく, しかし事前平均値が適切である場合を見る. **図-2(b)** を見ると, ABIC の最小値は, λ^2 の大きいとき, すなわち k_1 と k_2 が事前平均値に達してから表れる.

最後に PW-N1 のケース, すなわち事前平均値がまったく不適切で, ノイズが比較的小さいときは, PG-N1 のケースとほとんど同じ ABIC の挙動が見られる (**図-2(c)**). ただし, このときの λ^2 は, 6.4×10^{-1} であり, PW-N1 の λ^2 が 2.56 であったのと比べると, 観測データを事前平均より重視していることがわかる.

次に ABIC の挙動が, なぜ以上述べたようになるのか考える. このため, ABIC と λ^2 の関係を図示したものを**図－3, 図－4** に示す.

まず, 推定される誤差分布 ${\hat{\sigma}_\epsilon}^2$ は式 (20) より与えられる. この第 1 項は, 観測値と計算値の誤差二乗和をデータ数で除した値である. この値は, **図-3(a)** の左の図に示すように λ^2 が小さいとき $\hat{\theta}$ は, 最小二乗解となり最小値を示し, λ^2 が大きくなると $\hat{\theta} = \theta^*$ となり, 最大値を示し, λ^2 の増加にともなって前者から後者へ移行する.

一方, 第 2 項は $\hat{\theta}$ と事前平均値 θ^* の差の二乗和に λ^2 を乗じた値であり, θ^* が最小二乗解のときは λ^2 が小さいので小さく, λ^2 の増加とともに増加するが, $\hat{\theta}$ が θ^* に接近すると再び減少し, **図-3(a)** の真中の図に示すような, 上向き V 字形の挙動をする.

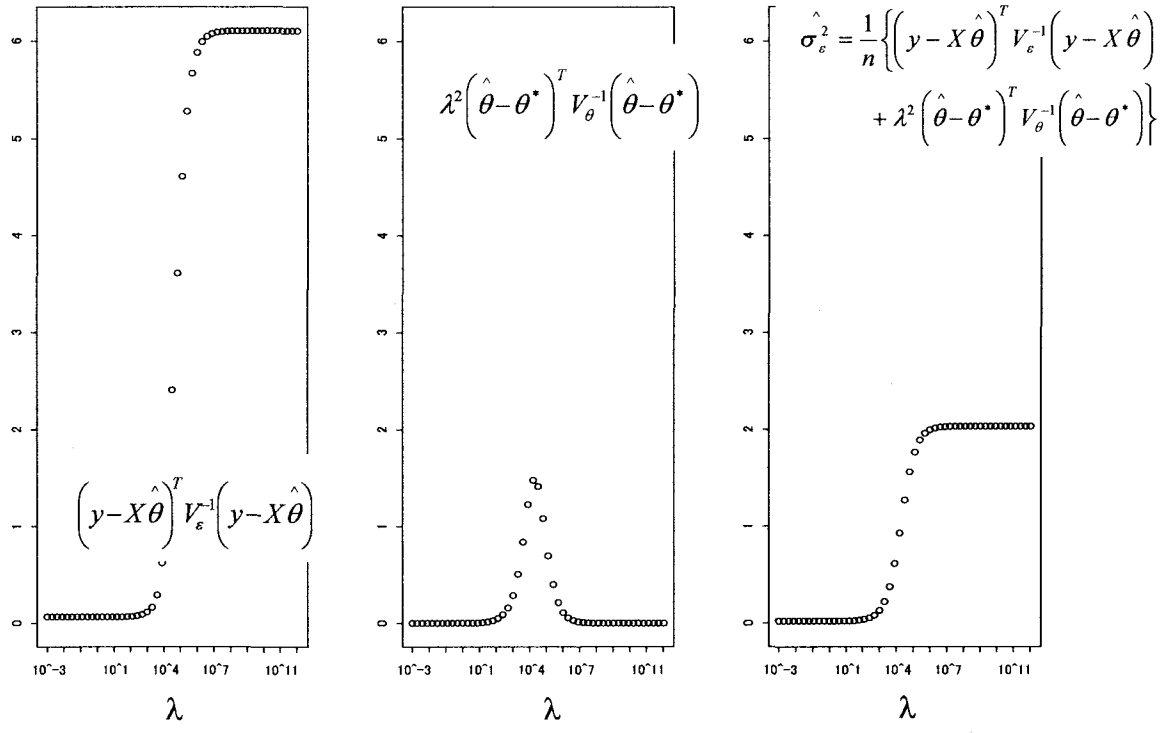
以上 2 つの項の和である ${\hat{\sigma}_\epsilon}^2$ は, **図-3(a)** の右の図に示すように, 全体的には第 1 項の挙動に支配された挙動をする.

次に式 (24) の各情報エントロピーの挙動について述べる. まずデータ分布のエントロピーは, λ^2 の増加とともに ${\hat{\sigma}_\epsilon}^2$ とほぼ同様の挙動をする (**図-3(b)** の左の図). すなわち, λ^2 が小さいとき観測値と計算値の残差二乗和が小さいことからこのエントロピーは小さく, λ^2 が大きくなると増加するが, ある一定値に収束する.

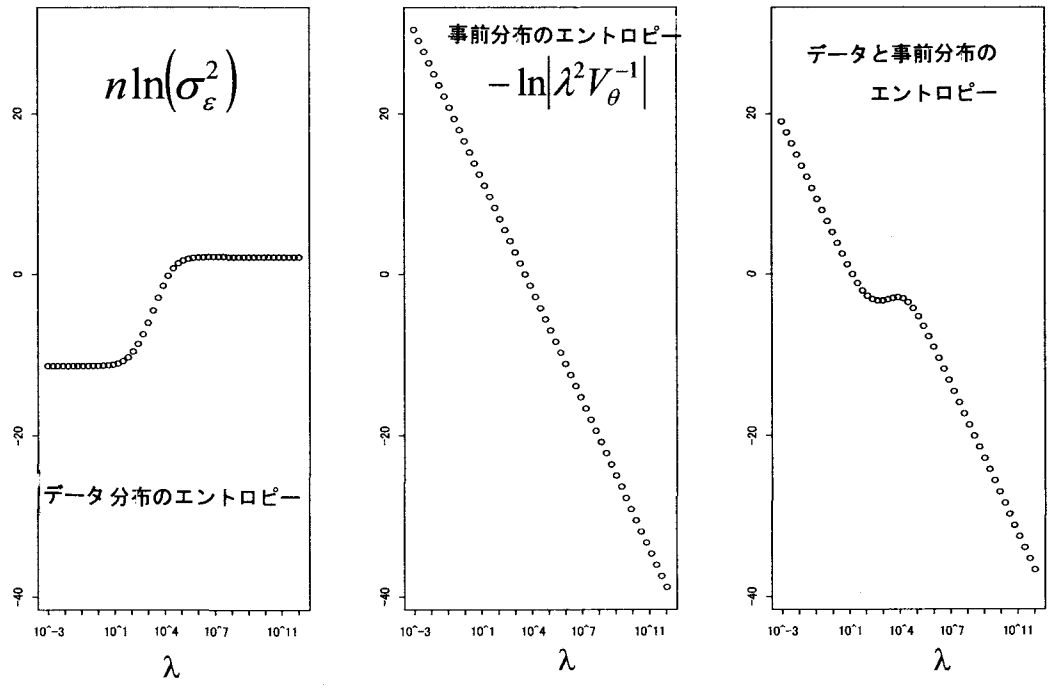
次に事前分布のエントロピーは λ^2 の増加にともない, 事前分散を低く設定していることになるため, その値は一様に減少していく (**図-3(b)** 中央の図).

データ分布によるエントロピーと事前分布によるそれの合計値は, 全体的には後者に支配され, λ^2 の増加とともに減少する (**図-3(b)** 右の図または**図-4(a)** 左の図). 一方, 事後分布のエントロピーに－ (マイナス) をつけたものは, λ^2 が小さいときは一定値, λ^2 が増加するにつれて増加する. (**図-4(a)** 中央の図)

結果的に減少エントロピーは, 上に示した 3 つの成分の和となり, ケース PM-N1 については, **図-4(a)** 右の図のようになる. すなわち, 全体的には λ^2 の増加にともない, データ分布と事前分布の情報エントロピーの和の減少と, 負の事後エントロピーの増加という, 2 つの成分のトレードオフ関係から, 最小値が現れる. しかし, 最小値が現れる付近の減少エントロピーの挙動は,

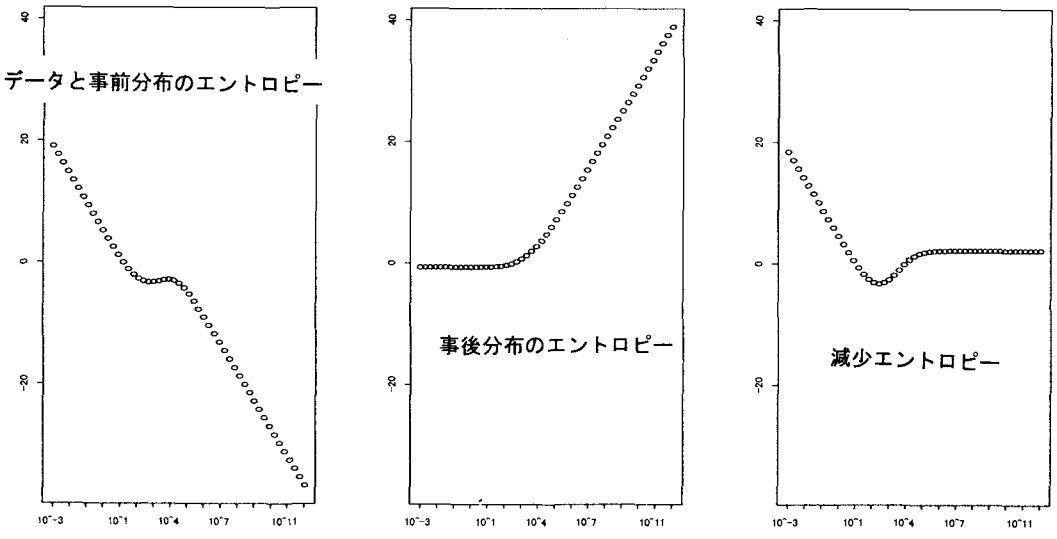


(a) データ分布のエントロピー(ケース PM-N1)

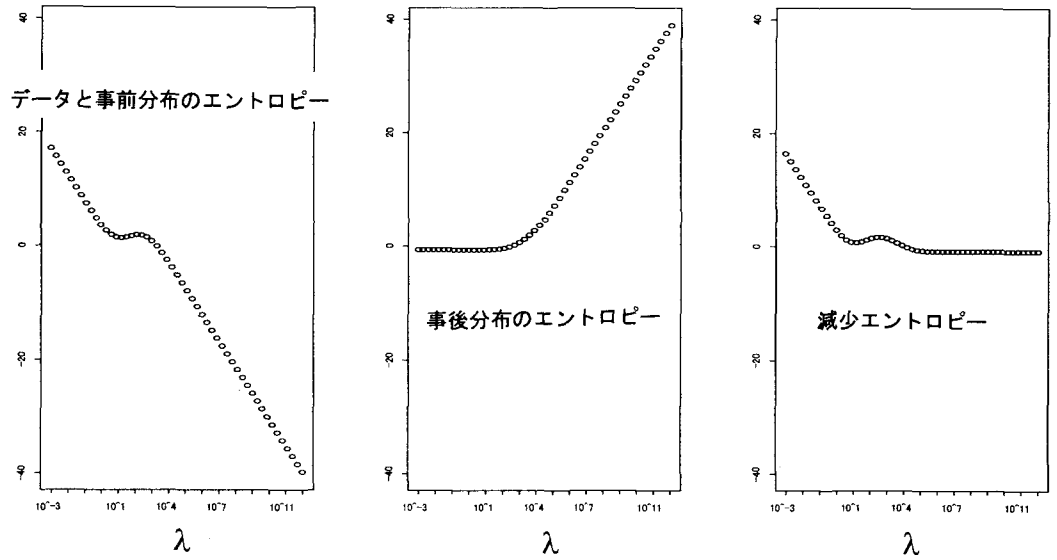


(b) データ分布と事前分布のエントロピー(ケース PM-N1)

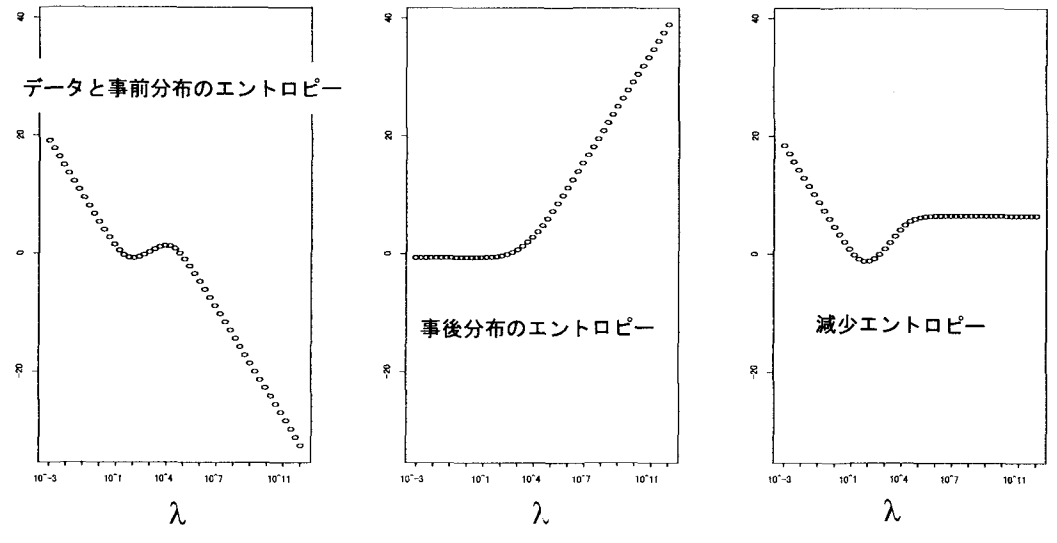
図3　データ分布と事前分布の情報エントロピー



(a) エントロピーの減少（ケース PM-N1）



(b) エントロピーの減少（ケース PG-N3）



(c) エントロピーの減少（ケース PW-N1）

図４　情報エントロピー減少の構造

各ケースで微妙に異なる.

すなわち, **図-4(a),(b),(c)** に示した, ケース PM-N1, PG-N3, PW-N1 を比べると明らかであるが, データ分布と事前分布の情報エントロピーの和が, 極所的に凹む個所 (**図-4(a)(b)(c)** それぞれの左の図) の情報エントロピーが, 事後分布の λ^2 が小さいときの情報エントロピー一定のときの値より小さいとき, PM-N1, PW-N1 のようにデータ分布を重視した解となり, その逆のときは PG-N3 のように事前平均を重視した解となる.

さらに PM-N1 と PW-N1 を比較すると, それぞれの最小の ABIC を与えるときの λ^2 値は, 2.56 と 0.64 であり, 前者が相対的に事前平均値を重視する解を与えれおり, ここでも ABIC が適切なメカニズムを表していることがわかる.

最後に, 以上示した理論的説明と例題計算結果より分かるように, ここに示したベイズ推定の枠組みは, 観測データの性質がよいときには, 自動的に事前分布の寄与を消去するように調整する (すなわち, 非常に小さい λ^2 値が選択される) メカニズムを持っており, この場合の推定値は, 伝統的な統計的推定を用いるのと全く同じ結果を与える. その意味で, ベイズ推定と伝統的な統計的推定が同じ枠組みに統一され, 与えられるデータの性質に応じて, 自動的, 客観的に推定が行われることが理解できるであろう.

3.4 結論

本研究で得えられた結果を要約すると, 以下のようになる.

1. 赤池の情報量基準に基づいた拡張ベイズ法により, 観測値と事前分布を, 客観的かつ適切に調整することができる, その挙動の特徴を例題により明確し示すことができた.
2. 赤池の情報量基準は, データ分布のあいまいさ (情報エントロピー) と事前分布のあいまいさ (情報エントロピー) の総和と, 事後分布のあいまいさ (情報エントロピー) の差, つまりエントロピーの減少量を最大にするように, 事前分布を調整している, という情報エントロピー的な解釈を行うことができることを見出した.

この論文で示した例題は, 説明の都合上, きわめて単純なものであった. 本研究で示した考え方の, 確率場における自己相関関数の推定に応用した例を Honjo and Kazumba(2002)[11]に示しているので, 参照されたい.

謝辞

本研究は, 土木学会の逆問題研究委員会初代委員長を務められた故北原道弘東北大学教授を偲んで掲載される一連の論文の一つであることを明記する. 本研究の計算を担当した, 岐阜大学卒業生畔柳薫氏 (現 JR 東海建設株式会社) に深謝する.

参考文献

1) 本城勇介：地下水浸透流解析モデルのパラメータ推定：最適モデルの選択, 土質工学会講演集, No.24,pp.1647-1650, 1989.
2) 本城 勇介：逆解析における事前情報とモデルの選択 (その1) 及び (その2), 講座「地盤工学における逆解析」, 土と基礎, Vol.43, No.7, pp63-68 及び No.8, pp51-54, 1995.
3) 本城勇介・酒向一也・菊池喜昭：杭の水平地盤版力係数逆解析における各適切化手法の比較, 応用力学論文集 (土木学会), Vol.2, pp.73-81, 1999.
4) Akaike,H.: Likelihood and Bayes Procedure with discussion, Bayesian Statistic, J.M. Bernards, M.H. DeGroot, D.V. Lindly and A.F.m. Smith (eds), University Press, Valencia, pp.143-166, pp.185-203, 1982.
5) 赤池弘次：エントロピーとモデルの尤度, 日本物理学会誌 Vol.35, No.7, pp.608-614 1980.
6) 赤池弘次：事前分布の選択と応用, 「ベイズ統計学とその応用」(鈴木雪夫・国友直人編), 東京大学出版会, pp.81-98, 1989.
7) グロエッチェ,C (金子・山口・滝口訳)：数理科学における逆問題, 別冊数理科学, サイエンス社, pp.154, 1996.
8) 坂元慶行・石黒真木夫・北川源四郎, 情報量統計学, 共立出版株式会社, pp236, 1983.
9) 有本卓：確率・情報・エントロピー, 森北出版社, pp269, 1980.
10) 佐藤忠信：逆解析の手法, 講座「地盤工学における逆解析」, 土と基礎, Vol.43,No.4,pp.57-60, 1995.
11) Honjo, Y. and Kazumba, S.: Estimation of autocorrelation distance for modeling spatial variability of soil properties by random filed theory, 第 47 回地盤工学シンポジウム論文集 (地盤工学会), pp.279-286, 2002.

(2004 年 4 月 16 日 受付)