

構造設計における確率場のモデル推定とそのばらつき

Modeling of Stochastic Fields for Structural Design

丸山 収*・鈴木 誠**

Osamu MARUYAMA and Makoto SUZUKI

* 工博 武蔵工業大学工学部土木工学科 (〒158 東京都世田谷区玉堤 1-28-1)

** 工博 清水建設(株) 和泉研究室 (〒100 東京都千代田区内幸町 2-2-2)

This paper investigates the following problem: given a random sample $\mathbf{X} = \{x_1, x_2, x_3, \dots, x_n\}^T$ from an unknown probability distribution $g(\mathbf{X})$, estimate the sampling distribution of prespecified probability distribution functions and evaluate the accuracy of each distribution models by AIC, on the bases of the observed data $\mathbf{X}$. General methods called "Bootstrap" and "EIC" (Extended Information Criterion) are shown to work satisfactory on a variety of model estimation problem.

Keywords: AIC, EIC, Bootstrap, random field, Monte Carlo simulation

1. はじめに

地盤物性値などのように、空間的に相関特性を有する場は、確率場として扱うことができる。その際、確率場をモデル化するための十分なサンプルを得ることは希であり、限られたサンプルからモデルの選定やそのパラメータを推定することが必要となる。

従来は、いくつかの分布特性モデルを仮定して最尤法によりモデルのパラメータを推定し、最大対数尤度を用いた AIC (Akaike Information Criterion)^{2,3)} により最良なモデルの選択を行っていた。しかしながら、サンプルには母集団の限られた情報しか含まれておらず、統計的推定の不確定性を避けることは困難である。すなわち、サンプル場の情報が母集団である確率場の性質を十分に表現するのは、統計学でいう大数の法則が成り立つという仮定のもとであり、実際の構造設計においては、数少ないデータから、母集団の性質を外挿しなければならないという点に問題がある。統計的推定の不確定性として主なものには、確率分布特性モデル自体が妥当なものであるか、またはモデル化が妥当であっても、そのパラメータの推定精度に関する検討が十分に行われているかの点が挙げられる。

また確率論を用いた構造設計において、入力データを確率変数や確率場として表現し、応答結果を確率論的に評価しようとするとき、前述した母集団のモデル化の影響は避けて通れないものである。すなわち母集団のモデル化誤差は、破壊確率の算定、物性値の推定結果に偏りを生じさせる要因となる。

本研究では、限られたサンプルから確率場をモデル化するときに含まれる不確定性を定量的に評価することを目的としている。そのため、まず従来からの最尤法と AIC の組合せによるモデル化についてまとめ、サンプル数がモデルである分布形の決定に与える影響を検討する。次に、ブートストラップによりパラメータ推定の不確定性を定量的に評価できることを示し、サンプル数による不確定性の差異を検討する。最後に、正規確率場を対象とし、EIC (Extended Information Criterion)²⁾ を用いたモデル選択基準について議論する。

2. モデルの選択基準

現実のデータを生成する真のモデルを $g(\mathbf{X})$ とし、それを近似したモデル $f(\mathbf{X}/\theta)$ を推定することを考える。一般的には、真のモデル $g(\mathbf{X})$ は未知であり、代わりにデータ $\mathbf{X} = \{x_1, x_2, x_3, \dots, x_n\}^T$ が与えられているのが現実である。したがって与えられた有限個のデータからモデル $f(\mathbf{X}/\theta)$ を推定することになるが、真のモデルに一致するモデルを見出すことは困難である。そこで、推定されるモデル $f(\mathbf{X}/\theta)$ について、その良否を検討するための評価基準が必要となる。

サンプルから推定されるモデルの良否を客観的に評価する基準のひとつに、式 (1) に示す KL 情報量がある。

$$I(g; f) = E_{\mathbf{X}} \{ \ln g(\mathbf{X}) \} - E_{\mathbf{X}} \{ \ln f(\mathbf{X}/\theta) \} \quad (1)$$

ここで、 $f(\mathbf{X}/\theta)$ は推定モデルの確率密度関数を示し、 $g(\mathbf{X})$ は真のモデルの確率密度関数を示している。式 (1) にお

第1項は、真の確率密度関数の平均対数尤度を示し、第2項は、モデルの真の確率密度関数に対する平均対数尤度を示している。また、式(1)は、式(2)で与えられる性質を有している。

$$I(g;f) \geq 0, \quad g(\mathbf{X}) = f(\mathbf{X}/\theta) \rightarrow I(g;f) = 0 \quad (2)$$

KL情報量は、真のモデルの平均対数尤度が、推定モデルの平均対数尤度よりも大きな値となることから常に正の値となる。すなわち、推定モデルが真のモデルと一致するときには最小値0となる。式(1)、(2)より、良い推定モデルを決定するためには、KL情報量をできるだけ小さくすれば良いことがわかる。言い換えれば、式(1)の第1項は、推定モデル $f(\mathbf{X}/\theta)$ には依存せず一定なので、第2項を最大とする $f(\mathbf{X}/\theta)$ を選択すれば最良の推定モデルを求めることができる。ただし、真のモデルの確率密度関数に対する平均対数尤度 $E_x\{\ln f(\mathbf{X}/\theta)\}$ を求めるためには、現実には未知の $g(\mathbf{X})$ を必要とするので、具体的に計算することはできない。

そこで、一般に用いられているアプローチは、データ $\mathbf{X} = \{x_1, x_2, x_3, \dots, x_n\}^T$ は真のモデルの確率密度関数 $g(\mathbf{X})$ に従って発生していることを考える。そこで、データが $N \rightarrow \infty$ のときに大数の法則が成り立つとして、式(3)のように平均対数尤度を近似する。

$$\frac{1}{N} \sum_{i=1}^N \ln f(x_i) \cong E_x\{\ln f(\mathbf{X}/\theta)\} \quad (3)$$

詳細は文献2), 3)に委ねるが、赤池は式(4)に示すAICを最小とするモデルを推定することにより、近似的に最適なモデルが決定できることを提案している。モデルのパラメータ推定とモデルを評価するための式(3)による平均対数尤度の推定に、同じデータを2度用いたことによる平均的な偏りCが存在する。AICは、最尤法により推定されたモデルのパラメータ $\hat{\theta}$ を有する最大対数尤度 $\ell(\hat{\theta})$ を基にして、これらの偏りを補正した評価基準である。

$$\begin{aligned} \text{AIC} &= -2\ell(\hat{\theta}) + 2C \\ &= -2\ell(\hat{\theta}) + 2k \end{aligned} \quad (4)$$

ここで、 k はモデルのパラメータ数である。すなわち、AICは、モデルの平均対数尤度を、精度良く推定することを目指したものであり、基本的にはKL情報量の最小化と等価であると考えられる。

本研究では、AICを求めるにあたり、正規分布を母集団集合のモデルとした場合には、本来対数正規分布に従うサンプルであっても、正規分布に従うとして扱ってパラメータおよびAICを求めている。また、対数正規分布を母集団モデルと仮定した場合には、一度データを対数変換し、対数正規分布のパラメータを正規分布の関数空間で求めている。後者の場合、 $z = h(x) = \ln x$ と変換したデータを用いているが、原データを正規分布にあてはめた AIC_x と対数変換されたデータを正規分布にあてはめた AIC_z の比較を行うためには、次式のBox-Cox変換を行う

必要がある。

$$\text{AIC}_z' = \text{AIC}_z - 2 \ln \left| \frac{dh}{dx} \right| \quad (5)$$

ここで、 AIC_x と AIC_z' を比較し、小さい値を示すモデルが良いということになる。

3. 最尤法による統計モデルのパラメータ推定

本研究では、与えられたサンプルから $f(\mathbf{X}/\theta)$ のモデルパラメータを推定することが必要となる。ここでは、多次元対数正規分布および正規分布の場合について、最尤法によるモデルの分布形状を支配するパラメータを推定する手法を示す。

3.1 対数正規分布のパラメータ推定

確率変数 x が対数正規確率分布に従うとき、 N 次元対数正規確率密度関数は、次式で与えられる。

$$f(\mathbf{X}/\theta) = \frac{1}{\sqrt{(2\pi)^N} |\mathbf{Q}| \prod_{i=1}^N x_i} \exp \left\{ -\frac{1}{2} (\ln \mathbf{X} - \underline{\lambda})^T \mathbf{Q}^{-1} (\ln \mathbf{X} - \underline{\lambda}) \right\} \quad (6)$$

ここで、 $\mathbf{X} = \{x_1, x_2, x_3, \dots, x_n\}^T$ は測定値ベクトル、 $\mathbf{Q}$ は共分散マトリクスである。

対象とする対数正規確率場の平均値を μ_x 、分散値を σ_x^2 および相関係数を ρ とすると、 $\mathbf{X}$ の各成分を対数変換した測定値 $z = \ln x$ の平均値 $\lambda = E[\ln x]$ 、分散 $\zeta^2 = \sigma_{\ln x}^2$ および相関係数 ρ' は次式で与えられる⁹⁾。

$$\zeta^2 = \ln(1 + \delta^2) \quad (a)$$

$$\lambda = \ln \mu_x - \frac{1}{2} \zeta^2 \quad (b) \quad (7)$$

$$\rho' = \frac{\ln(1 + \rho \delta^2)}{\ln(1 + \delta^2)} \quad (c)$$

ここで、 $\delta = \sigma_x / \mu_x$ で与えられる変動係数である。

対数正規確率場 $\mathbf{X}$ における相関係数 ρ を式(8)とすると、式(6)における共分散行列 $\mathbf{Q}$ の要素 Q_{ij} は次式で与えられる。

$$\rho = \exp \left(-\frac{|x_i - x_j|}{a} \right) \quad (a)$$

$$\begin{aligned} Q_{ij} &= \zeta^2 \rho' \\ &= \ln \left\{ 1 + \left[\exp(-|x_i - x_j|/a) \right] \left[\exp(\zeta^2) - 1 \right] \right\} \end{aligned} \quad (b) \quad (8)$$

ここで、 a は、対数正規確率場における相関距離である。

また、 $\mathbf{X}$ の各要素を対数変換して得られるベクトル $\mathbf{Z}$ は、式(9)の N 次元正規確率密度関数に従う。

$$f(\mathbf{Z}/\theta) = \frac{1}{\sqrt{(2\pi)^N} |\mathbf{Q}|} \exp \left\{ -\frac{1}{2} (\mathbf{Z} - \underline{\lambda})^T \mathbf{Q}^{-1} (\mathbf{Z} - \underline{\lambda}) \right\} \quad (9)$$

式(9)は、正規確率密度関数で与えられるので、式(7)の関係を考慮して対数正規分布のパラメータベクトル $\theta = \{\lambda$

$\zeta^2 a)^T$ の推定値を、最尤法により求めることができる。実際の計算では、ガウスニュートン法を適用し、次式で与えられる負の対数尤度を最小化するようにパラメータを推定する。

$$L(\mathbf{Z}/\boldsymbol{\theta}) = -\ln f(\mathbf{Z}/\boldsymbol{\theta}) \\ = \frac{N}{2} \ln(2\pi) + \frac{1}{2} \ln |\mathbf{Q}| + \frac{1}{2} (\mathbf{Z} - \underline{\lambda})^T \mathbf{Q}^{-1} (\mathbf{Z} - \underline{\lambda}) \quad (10)$$

3.2 正規確率場のパラメータ推定

モデルが正規確率場で与えられるときは、以下のようにしてパラメータの推定を行う。確率変数 $\mathbf{X}$ が正規確率分布に従うとき、 N 次元正規確率密度関数は、次式で与えられる。

$$f(\mathbf{X}/\boldsymbol{\theta}) = \frac{1}{\sqrt{(2\pi)^N |\mathbf{Q}|}} \exp \left\{ -\frac{1}{2} (\mathbf{X} - \underline{\mu}')^T \mathbf{Q}^{-1} (\mathbf{X} - \underline{\mu}') \right\} \quad (11)$$

ここで、正規確率場の平均値 $\underline{\mu}$ 、分散 σ^2 および相関係数 ρ として、推定パラメータベクトル $\boldsymbol{\theta} = \{\underline{\mu} \ \sigma^2 \ \mathbf{a}'\}^T$ を求める場合には、式(10)に対応して、 $L(\mathbf{X}/\boldsymbol{\theta})$ の尤度関数を推定パラメータに関して最小化すればよい。なお、正規確率場における相関係数を次式で与えるものとする。

$$\rho = \exp \left(-\frac{|x_i - x_j|}{a'} \right) \quad (12)$$

4. ブートストラップ法

データから確率場をモデル化する際には、母集団から得られている1組のサンプルから、そのデータに基づいて分布を仮定し、パラメータを推定することになる。しかしながら、得られた推定結果は、サンプル数およびそのサンプルの観測点配置により、真のパラメータに対する偏りを有している。ここでは、得られたデータから推定結果のばらつきを調べるために、Efron⁴⁾の提案するブートストラップ法を用いる。ブートストラップ法は、数値実験法であるモンテカルロ法に基づいた手法であり、次のようなステップで解析を行う。

- ① 真のモデルからサンプル $\mathbf{X} = \{x_1 \ x_2 \ x_3 \ \cdots \ x_n\}^T$ が、得られているものとする。このサンプルから最尤法によりパラメータを推定し、サンプルの確率密度関数 $f(\mathbf{X}/\boldsymbol{\theta})$ を求める。
- ② $f(\mathbf{X}/\boldsymbol{\theta})$ から、ランダムサンプル（ブートストラップサンプル） $\mathbf{X}^* = \{x_1^* \ x_2^* \ x_3^* \ \cdots \ x_n^*\}^T$ を求める。次に求められたサンプルから、最尤法により $f(\mathbf{X}/\boldsymbol{\theta})$ のパラメータを求め、 $\boldsymbol{\theta}_1$ とする。
- ③ ②のブートストラップを n 回行い、 n セットのパラメータの推定値を求める。
- ④ これを母集団からのサンプルによる推定値の不確定性を見なすこととして、推定結果の分布を調べる。

この方法は、得られた情報をもとにして、モデルを決定し、そのモデルから得られるサンプルから、推定パラメータのばらつきを調べる方法である。したがって、サンプル数の推定結果に与える影響を評価することができる。そのため、本研究では詳細な検討を行うまでには至っていないが、観測点配置による推定結果の不確定性を求めることが可能である。

また、ブートストラップ法により、最大対数尤度の母集団に対する平均的な偏り C を、精度良く推定できる手法EICが提案されている。AICでは、最大対数尤度の平均的な偏り $C = k$ （パラメータ数）と近似しているが、EICでは、式(13)に示すように、 C^* を次式で評価する。

$$\text{EIC} = -2\ell(\hat{\boldsymbol{\theta}}) + 2C^* \\ C^* = E_{\mathbf{X}'} \{ \ln f(\mathbf{X}'/\mathbf{X}') - \ln f(\mathbf{X}/\mathbf{X}') \} \quad (13)$$

ここで、 C^* はブートストラップ法により求められる平均的な偏り、 $\ell(\hat{\boldsymbol{\theta}})$ は最尤法で推定されたパラメータにより求められる最大対数尤度である。すなわち、EICでは、評価を行う際に、真のモデルの確率密度関数に置き換えて、数値シミュレーションにより推定することを行っている。

5. 数値計算例

5.1 分布特性の推定

ここでは、予め設定した対数正規分布をから得られるサンプルを用いて、モデルとして対数正規分布および正規分布の分布特性モデルを仮定して、それぞれのモデルの良否について検討を行う。まず、CASE1では、対数正規確率密度関数において、 $\mu_x = 10.0$ 、 $\sigma_x^2 = 1.005$ 、 $(\lambda = 2.3025$ 、 $\zeta = 0.1)$ 、 $a = 5.0$ に従うサンプルを1次元空間座標上で1m間隔で、200個作成した（図-1）。図-1に示すサンプル値のヒストグラムを図-2.aに示す。同様に2mの等間隔で100個のサンプリングが行われた場合（図-2.b）、5m間隔で40個（図-2.c）および10m間隔で20個（図-2.d）のサンプリングが行われた場合についてヒストグラムを示している。CASE2では、 $\mu_x = 10.0$ 、 $\sigma_x^2 = 28.403$ 、 $(\lambda = 2.178$ 、 $\zeta = 0.5)$ 、 $a = 5.0$ として、同様な計算を行った。図-3にサンプル値を示し、図-4に各サンプル数ごとのヒストグラムを示す。

図-2.aより、CASE1の場合は、分布形状は正規分布に近い形状をしていることがわかる。これに対し、CASE2の場合には、CASE1と異なり、歪んだ形状となっている。ここで両者とも、サンプル数が少なくなると、目標とした確率場が表現できていないことがわかる。

次に、100個、40個および20個のサンプル数ごとに、正規分布および対数正規分布として、分布のパラメータを推定し、AICによりモデルの判定を行った。その際、各分布形ごとに前述した最尤法によるパラメータの推定法を用いている。

表-1および表-2に、各ケースにおけるAICおよびパラメータの推定結果を示す。ここで、サンプル数が大きくなると目標とした確率場を表現していることがわかる。言

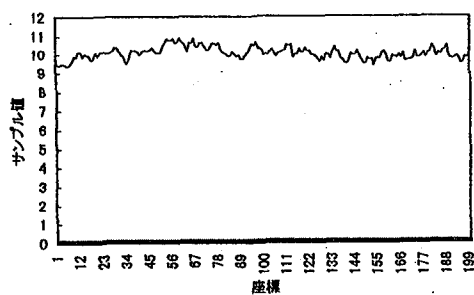
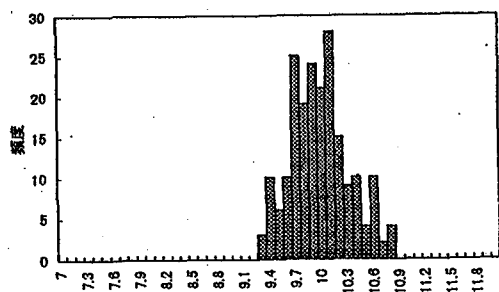
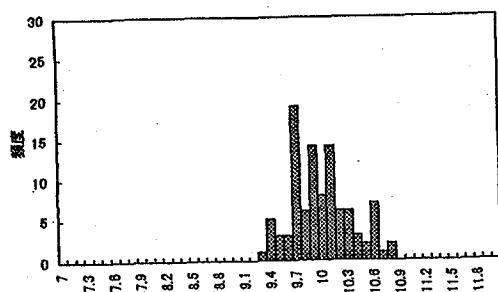


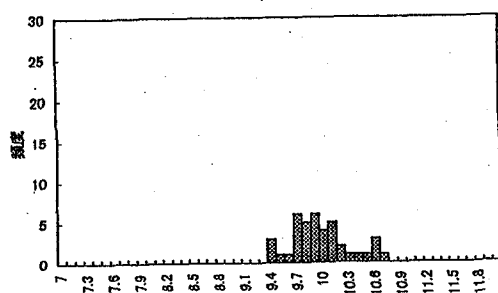
図-1 サンプル場 (CASE 1)



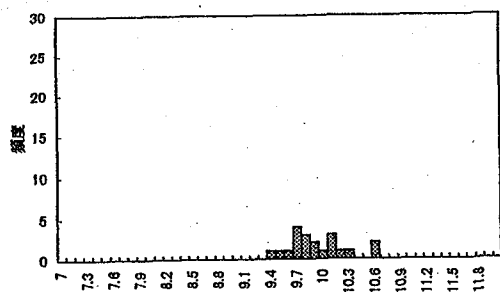
(a) . 200個のサンプル



(b) . 100個のサンプル



(c) . 40個のサンプル



(d) . 20個のサンプル

図-2 サンプルのヒストグラム (CASE 1)

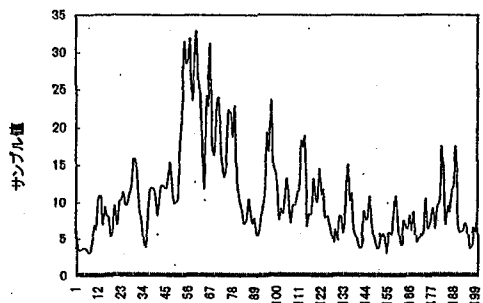
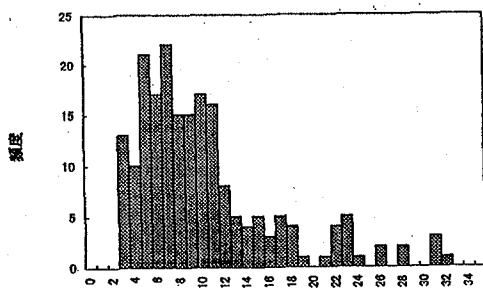
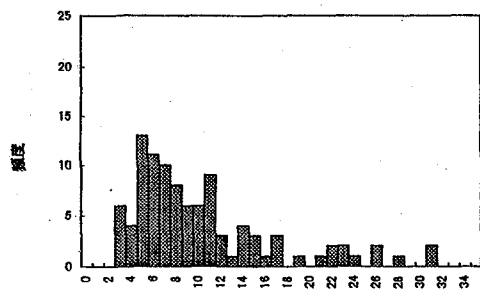


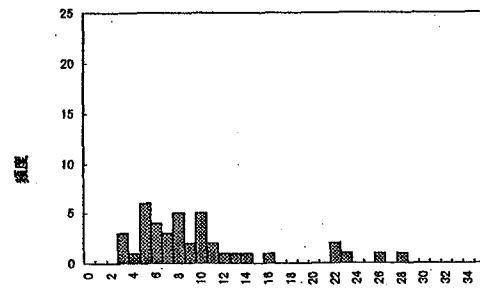
図-3 サンプル場 (CASE 2)



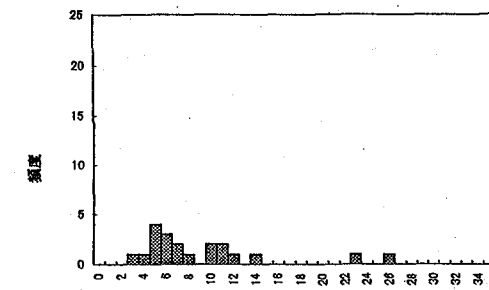
(a) . 200個のサンプル



(b) . 100個のサンプル



(c) . 40個のサンプル



(d) . 20個のサンプル

図-4 サンプルのヒストグラム (CASE 2)

い換えると少ないサンプル数では、1組のサンプルから母集団の分布特性を推定するときに、大きな不確実性を有していることになる。この傾向は、特に相関距離aの推定結果に顕著である。ここで、AICにより正規分布および対数正規分布のモデルの良否を比較する際には、同一のサンプル数の場合にのみ比較が可能である。また、ここでは省略したが、同一サンプル数の場合であっても観測点配置による影響を調べることが必要である。

CASE1の場合の対数正規分布と正規分布のAICを比較すると、対数正規分布の方が多少小さくなっているが、その差は小さいことから、これらに分布形の差はないと言える。

CASE2においては、対数正規分布のAICが正規分布のAICに比べて小さく、対数正規分布の方が良いモデルであ

表-1 AIC およびパラメータの推定結果 (CASE1)
(目標分布：平均：10.0、分散：1.005、相関距離：5.0)

サンプル数	正規分布	対数正規分布
20	AIC=58.05	AIC=56.77
	平均：9.845	平均：9.847
	分散：1.027	分散：0.956
	相関距離：13.99	相関距離：13.63
40	AIC=113.827	AIC=113.280
	平均：9.987	平均：9.995
	分散：1.159	分散：1.105
	相関距離：7.174	相関距離：6.553
100	AIC=245.787	AIC=243.911
	平均：10.078	平均：10.0232
	分散：1.197	分散：1.153
	相関距離：5.138	相関距離：4.908

表-2 AIC およびパラメータの推定結果 (CASE2)
(目標分布：平均：10.0、分散：28.403、相関距離：5.0)

サンプル数	正規分布	対数正規分布
20	AIC=127.81	AIC=114.85
	平均：9.413	平均：9.333
	分散：34.094	分散：24.832
	相関距離：14.494	相関距離：13.410
40	AIC=246.34	AIC=232.821
	平均：10.144	平均：10.087
	分散：39.107	分散：32.426
	相関距離：9.928	相関距離：6.338
100	AIC=585.53	AIC=543.60
	平均：10.605	平均：10.540
	分散：40.907	分散：36.172
	相関距離：6.074	相関距離：4.84

ることを示している。これは、図-2および図-4に示す発生させたサンプルの分布からも理解できる結果であり、モデルのあてはめ具合が正しく行われていることがわかる。また両ケースとも推定された相関距離に着目すると、正規分布と仮定した方が大きめの値を推定していることがわかる。この傾向は、分布形の差異が明らかなCASE2の方に大きい。

ここでは、対数正規分布と正規分布の範囲ではあるが、AICを用いてどちらのモデルのあてはめが良いかを検討した。その結果、CASE2のような対数正規分布の偏りの大きい場合には、対数正規分布のモデルを選択できることがわかった。しかしながら、CASE1のように、対数正規分布の特性が正規分布に近くなるような場合には、両者のAICの差はなく、どちらのモデルでも良いことが言える。また、このように分布形を比較する場合には、式(5)で示したBox-Cox変換を用いたAICが有効となる。

5.2 ブートストラップ法によるパラメータのばらつき

5.1では、母集団の分布特性について検討したが、もし確率密度関数が規定できたとしても、その分布形を支配するパラメータの推定精度により、表現される分布特性は異なるものとなる。特にサンプル数が少ないと、推定さ

表-3 ブートストラップ法による推定結果
(正規分布)

サンプル数	パラメータ	平均値	分散値
20	μ	-0.188	0.0973
	σ^2	0.790	0.1116
	a	10.930	38.193
40	μ	-0.03	0.0508
	σ^2	1.026	0.0843
	a	6.130	4.978
100	μ	0.052	0.0552
	σ^2	1.064	0.0573
	a	4.472	1.226

表-4 ブートストラップ法による推定結果
(対数正規分布)

サンプル数	パラメータ	平均値	分散値
20	μ	9.131	2.351
	σ^2	20.355	218.607
	a	11.102	29.163
40	μ	9.974	1.381
	σ^2	30.740	220.217
	a	5.921	3.729
100	μ	10.426	1.670
	σ^2	34.104	176.047
	a	4.416	0.987

れたパラメータの不確実性が大きい。

真のモデルからサンプルを得た場合にも、母集団の分布が未知であるので、サンプルから推定されたパラメータで母集団を近似する。すなわち、近似モデルを母集団とするサンプルから、パラメータを推定していることとなる。一般に分布の形状を規定するパラメータの推定値のばらつきを理論的に評価することは困難である。そこで、本解析の目指すところは、観測されたデータをもとにしたモデルからサンプルの抽出を行うことでばらつきの評価を行うことにある。ここでは、確率密度関数が与えられているという条件のもとで、サンプル数による推定されるパラメータのばらつきを、ブートストラップ法で検討する。仮定する分布形は、正規確率分布および対数正規分布であり、サンプル数がこれらのばらつきに与える影響を検討する。

表-3と表-4に正規分布と対数正規分布のパラメータのばらつきを、ブートストラップにより求めた結果を示す。正規分布として与えたパラメータは、 $\mu=0.0$ 、 $\sigma^2=1.0$ で $a=5.0$ としている。また、対数正規分布として与えたパラメータは $\mu=10.0$ 、 $\sigma^2=28.402(=2.177, =0.5)$ で $a=5.0$ として

いる。

表-3より、パラメータのうち μ と σ^2 の平均値はほぼ与えた値に等しく、その分散はサンプル数が多くなるに従って小さくなることがわかる。ここでは示していないが、これらのパラメータ μ は、ほぼ正規分布のようなばらつきを、 σ^2 は対数正規分布のようなばらつきを示している。また、 a はサンプル数が小さいときほど大きな平均値と分散となった。このような傾向は、サンプリングを、等間隔で行ったため、サンプル数20では最小間隔が10.0 mとなっていることが原因と考えられる。これらのヒストグラムを図-5に示す。この図よりもサンプル数が小さいとばらつきが大きいことがわかり、サンプル数による差は顕著である。ばらつきは多少偏りをもった対数正規分布している。

表-4より、正規分布と同様にパラメータのうち μ と σ^2 の平均値はほぼ与えた値に等しく、その分散はサンプル数が多くなるに従って小さくなることがわかる。また、これらのパラメータのばらつきの分布も正規分布と同様となる。図-6に対数正規分布の相関距離の推定結果のばらつきを示す。

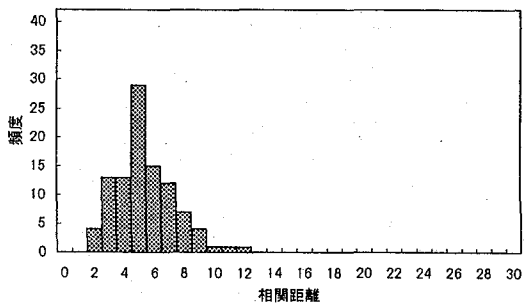
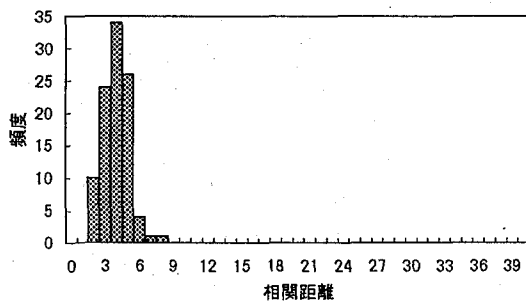
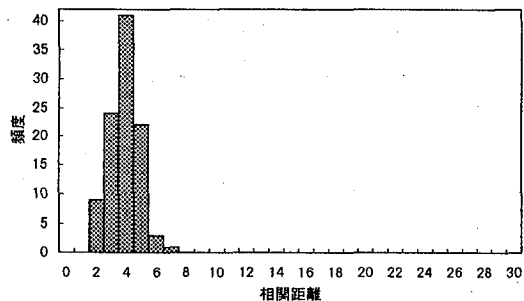
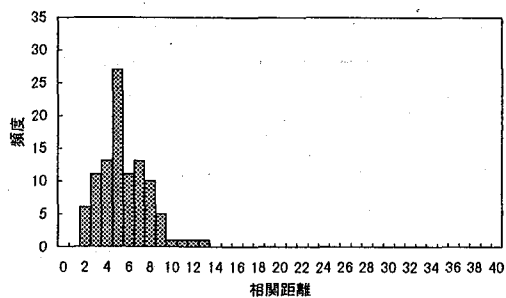
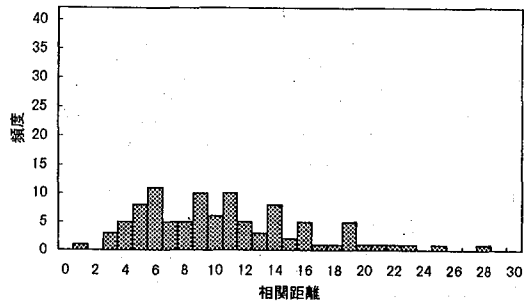
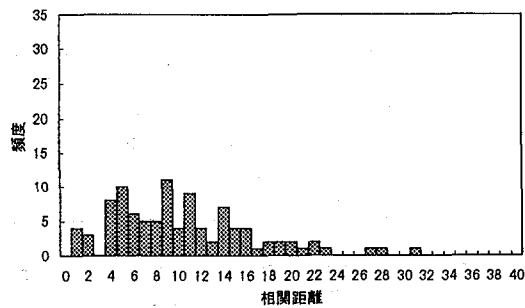


図-5 ブートストラップによる推定結果 (正規分布)

図-6 ブートストラップによる推定結果 (対数正規分布)

5.3 EICによるCの評価

EICは、式(14)に示すように最尤法により求められたパラメータによる最大対数尤度とモデルの真のモデルに対する平均対数尤度の偏りを、モンテカルロ法による数値シミュレーションにより評価する手法である。ここでは、比較的容易にモデル $f(\mathbf{X}/\theta)$ の真のモデルに対する平均対数尤度の理論解 $E_x[\ln f(\mathbf{X}/\theta)]$ が得られる正規確率場を対象として、Cの評価をEICにより行う。

正規確率場において、真の分布を規定する確率密度関数が式(14)、モデルの密度関数が式(15)で与えられる場合に、モデルの真の分布形に対する平均対数尤度は、 $f(\mathbf{X}/\theta)$ の $g(\mathbf{X})$ に対する期待値を求めることで、式(16)のように与えられる。

$$g(\mathbf{X}) = \frac{1}{\sqrt{(2\pi)^N |\mathbf{Q}|}} \exp \left\{ -\frac{1}{2} (\mathbf{X} - \underline{\mu})^T \mathbf{Q}^{-1} (\mathbf{X} - \underline{\mu}) \right\} \quad (14)$$

$$f(\mathbf{X}/\theta) = \frac{1}{\sqrt{(2\pi)^N |\mathbf{Q}|}} \exp \left\{ -\frac{1}{2} (\mathbf{X} - \underline{\mu}')^T \mathbf{Q}'^{-1} (\mathbf{X} - \underline{\mu}') \right\} \quad (15)$$

$$E_x[\ln f(\mathbf{X}/\theta)] = -\frac{N}{2} \ln(2\pi) - \frac{1}{2} \ln |\mathbf{Q}| + \frac{1}{2} \text{tr} \{ \mathbf{Q}'^{-1} (\mathbf{Q} + (\underline{\mu}' - \underline{\mu})(\underline{\mu}' - \underline{\mu})^T) \} \quad (16)$$

ここで、 tr は行列のトレースを表す。

式(16)の平均対数尤度の値を-2倍したものが、AICまたは、EICの求めようとしている値に相当する。AICおよびEICでは、最大対数尤度の推定値に、偏りの推定値CおよびC*の値の2倍を加えたものとなっている。本研究では、推定するパラメータ数は、平均値、分散および相関距離の3つのパラメータであるので、AICでは常に $k=3$ と評価することになる。

ここでは、EICを用いてC*の評価を行い、式(16)によ

表-5 平均対数尤度、AICおよびEICの比較

回数	平均対数尤度	AIC	EIC (C*)
1	347.92	370.44	367.26(1.41)
2	349.72	358.46	351.76(-0.36)
3	348.73	374.55	373.96(2.70)
4	348.06	351.70	350.52(2.41)
5	353.51	362.09	357.28(0.59)
6	357.69	306.28*	306.06(2.89)
7	347.26	341.02*	333.71(-0.66)
8	350.04	325.58	327.70(4.06)
9	348.60	336.96*	334.66(1.85)
10	346.91	347.72*	344.16(1.22)

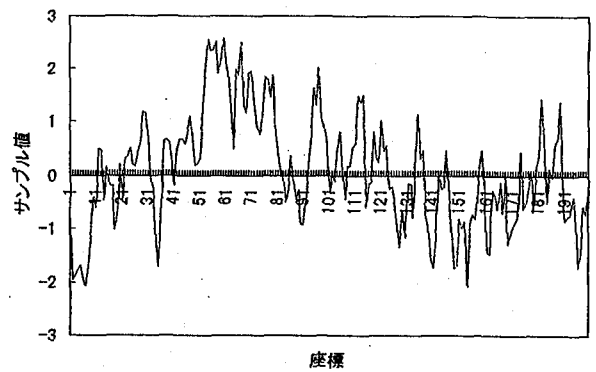


図-7 サンプル場

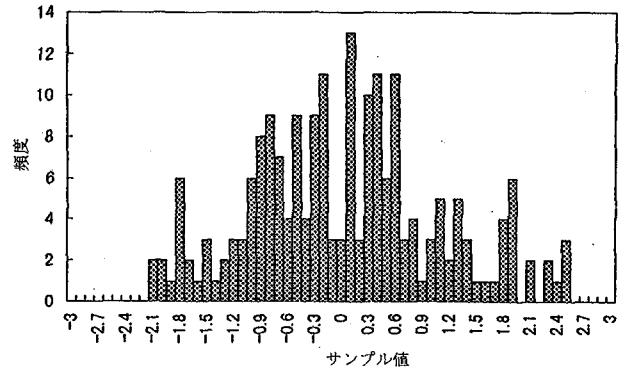


図-8 1つのサンプルデータのヒストグラム (200個)

るモデルの平均対数尤度と比較し、EICによる偏りC*の推定精度を検討することを目的としている。一般には、真の分布形は未知であるので、平均対数尤度の理論解を、求めることはできない。

まず正規確率密度関数に対して、 $\mu=0.0$ 、 $\sigma^2=1.0$ 、 $a=5.0$ として、200個のサンプル場を作成し、そのサンプル場から最尤法により正規分布のパラメータを推定し、モデルの密度関数を決定する。図-7に1サンプル、図-8にサンプルのヒストグラムを示す。次に、モデルの確率密度関数からブートストラップ法により200個のサンプルを100組作成して、各サンプルごとに分布パラメータを推定し、式(12)よりC*の評価を行う。この計算を10回を行い、各ケースごとにモデルの平均対数尤度の理論解、AICおよびEICの比較を行った。ここで、AICより 2×3 (パラメータ数) $=6$ を引いた値が、最大対数尤度である。

AICとEICの比較により、表-5における6ケースにおいてEICの方が、モデルの平均対数尤度に近い値を示していることがわかる。*はAICの優位性が認められる場合である。ここで、モデルの平均対数尤度は、シミュレートされるサンプルに依存するので、サンプルによってはC*の値が負となる場合もあった。この結果から判断すると、EICの優位性を明確に認めることはできない。しかし、AICはパラメータが最尤法で推定される場合にのみ適用可能であることにに対し、EICは最尤法以外の推定手法でモデ

ルが与えられた場合にも適用できるという利点がある。

6. まとめ

本研究では、限られたサンプルから確率場のモデルを決定するための手法について検討し、推定されるパラメータのばらつきについて定量的評価を行った。

その結果、確率場のモデルを選択するために Box-Cox 変換を併用した AIC の有用性を示し、サンプル数による推定パラメータのばらつきをブートストラップ法により評価できることを示した。本研究では、正規分布および対数正規分布を例題として取り上げたが、Box-Cox 変換を用いれば、他の分布特性に対しても AIC の評価が可能となる。これらの解析により、確率場のモデル化による不確定性を考慮するときの基礎データを整理することが可能となる。

また、正規確率場を対象とし、モデル選択基準である EIC を用いて、モデルの平均対数尤度を評価するための試算例を示した。ここでは、最尤法によるパラメータ推定を行っているために AIC と EIC の明確な良否を決定するまでには至らなかったが、EIC は最尤法以外の場合にも適用できるという利点がある。

今後の課題としては、最良なモデルを決定するためのサンプリング計画や、パラメータ推定の不確定性を考慮した構造系の確率応答評価手法を開発することを考えている。

最後に、本研究を行うにあたり、星谷勝教授（武蔵工業大学）との日頃の討議が有効であった。ここに記し、謝意を表します。

参考文献

- 1) 鈴木誠：空間的に分布する地盤物性値の統計量推定，JCOSSAR' 97 論文集，pp.707-710, 1997.
- 2) 北川源四郎・石黒真木夫・坂本慶行：情報量基準 AIC と EIC，信学技報，IT92-133, 1993.3.
- 3) 北川源四郎：時系列プログラミング，岩波書店，1994.4.
- 4) Efron, B：Bootstrap Methods：Another Look at the Jackknife, Annals of Statistics, Vol. 7, 1-26, 1976.
- 5) P.L. Liu and A.D. Kiureghian：Multivariate distribution Models with Prescribed Marginals and Covariances, Probabilistic Engrg. Mech., Vol.1, No.2, pp.105-113, 1986.

(1996 年 9 月 6 日受付)