

東京大学工学部 正員・藤野陽三

新日本製鉄 正員 小川昌洋

1. はじめに 信頼性解析・設計は確率的な立場、すなわち確率変数の分布形、パラメータ値を既知として定式化されることが多い。しかし現実には、我々の入手しうる情報は各変数に関して有限個の統計データに限られているのが普通である。したがって、これらの実験・観測データから精度よく母集団の分布形を決め、パラメータ値、特性値を推定することが要求される。

統計データが与えられた時、いくつかの理論分布形のデータへの適合度を比較し、最も適合度のよい分布形を選び、特性値を決めていくのが標準的なプロセスである。このための手法としては数多くある：統計学の分野で確立している尤度法、Kolmogorov-Smirnov法(K-S法)、工学、とくに自然環境外力の統計的解析に好んで使われる、経験的累積確率を用いるプロット法などが代表的なものである。なお、プロット法とは、データを小さい順に並べ、一つ一つに経験的累積確率を与え、これらを確率紙にプロットし、各種理論分布への適合度を比較し、最小二乗法等により統計パラメータを決める方法のことである。

本研究は、これら各種統計的手法の優劣を数値実験により明らかにしようとするものである。具体的には、分布形、平均、分散がわかっている母集団からモンテカルロ法によりデータを人工的に作成し、このデータに上記統計的手法を適用し、分布形判定の正答率、特性値の誤差の見地から比較した。

2. 分布形・統計パラメータ値の決定法 最尤法、プロット法、K-S法の3つの統計的手法を対象とする。

i) 最尤法(ML法)：候補理論分布 $F_j(x)$ に対し、尤度を最大にするようにパラメータ値を決める。尤度を最も大きくする理論分布が最適分布として選ばれ、この分布形のもとで特性値が評価される。

ii) プロット法：経験的累積確率の与え方として次の3通りを考える。

$$P_{(i)} = i/(n+1) \quad [\text{PL1法}] \quad P_{(i)} = (i-0.375)/(n+0.25) \quad [\text{PL2法}] \quad P_{(i)} = (i-0.5)/n \quad [\text{PL3法}]$$

なお、ここで n はデータ数、 i はデータの小さい方からの順位である。PL1法が最も普遍的であり、PL2法は正規分布に、PL3法は極値I型分布に適合すると考えられるデータに使われることが多い。統計データを $P_{(i)}$ を用いて確率紙にプロットし、最小二乗法によりパラメータ値を決める。最適分布はBowellが定義した $E_j = \sqrt{\frac{1}{n} \sum (x_i - x_j^*)^2} / \frac{1}{n} \sum x_i$ を用いれば、 E_j を最小にする理論分布 $F_j(x)$ が選ばれる。ここで n はデータ数、 x_i はデータ、 x_j^* は分布 $F_j(x)$ による理論値である。

iii) K-S法：標本平均と標本分散から $F_j(x)$ のパラメータ値を決定し、この $F_j(x)$ と、データから決まる累積確率分布 $S_n(x)$ (詳しくは参考文献1)を参照されたい)の差の最大値 $\max |S_n(x) - F_j(x)|$ を適合度の尺度とし、これが最も小さい分布を最適分布とする。

3. 対象とする分布形 構造強度、荷重の確率モデルとして用いられることの多い、次の6つの理論分布形を考える。1) 正規分布(N)、2) 対数正規分布(LN)、3) 極値I型最大値分布(EX1L)、4) 極値I型最小値分布(EX1S)、5) 極値II型分布(EX2)、6) 極値III型ワイブル分布(EX3) なお、平均は10、標準偏差を2とする。図1に各分布の累積確率分布を示す。

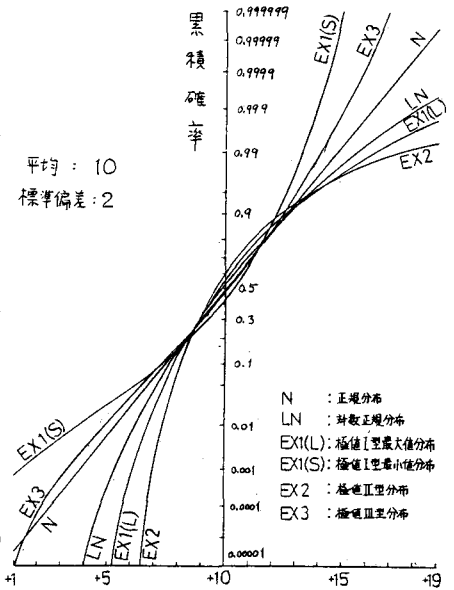


図1 対象とした6つの分布形状

4. 分布形判定に関する数値実験		母集団分布	N	LN	判定分布		EX1L	EX1S	EX2	EX3	母集団分布	N	LN	判定分布		EX1L	EX1S	EX2	EX3
(1-a)		N	27	30	6	9	3	25			N	32	23	10	14	4	17		
の分布に従い、平均10、		LN	13	36	30	1	10	10			LN	16	28	27	2	20	7		
標準偏差2をもつ母集団		EX1L	4	21	34	0	36	5			EX1L	7	15	20	1	54	3		
からのサンプルデータ		EX1S	11	3	1	59	0	26			EX1S	10	2	1	68	0	19		
をモンテカルロ法に		EX2	5	7	19	0	68	1			EX2	5	8	14	0	72	1		
より作成する、このデータ		EX3	21	12	4	16	1	46			EX3	25	11	2	35	2	25		

a) ML法による

b) PL1法による

表1: 100回のシミュレーションでの、最適分布の頻度

の手法を適用し、最も適合すると判定された分布形を調べた。データ数を20個とし、シミュレーションを100回行った結果を、例として表1に示す。表1(a)はML法による判定結果、b)はPL1法による結果である。表1(a)より母集団がN分布の場合、N分布と正しく判定されるのは100回中、27回であることがわかる。EX3分布の場合、ML法では100回中46回正しく判定するのに対し、PL1法では25回のみである。データ数がふえれば正答率は高まるが、その傾向は方法によって、

母集団分布	ML	PL1	PL2	PL3	K-S	高評価 ◎ とてもよい
N	△	○	○	○	○	○ まあまあ
LN	○	△	○	○	○	△ おおまかによい
EX1L	○	△	○	△	○	×
EX1S	○	○	○	○	○	×
EX2	○	○	△	△	×	×
EX3	◎	×	×	×	×	×

また分布形によってかなり異なる。5つの方法を定性的に比較したのが表2である。どの分布形に対しても共通によいのはML法である。

表2 分布形判定から見た5つの手法の比較

5. 特性値(fractile)の誤差

4.の数値実験と基本的には同じ考えで、特性値の誤差に関する数値実験を行った。乱数を用いて作られたデータに最も適合すると判断された分布形のもとで特性値を求め、真の値

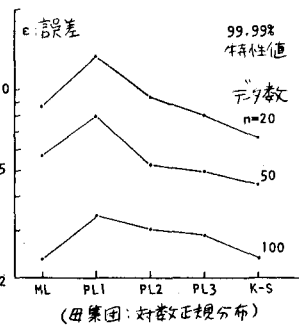
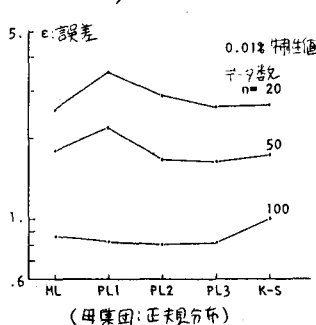


図2: 特性値の二乗平均誤差(シミュレーション100回)

との二乗平均誤差Eをもつ分布形に対し求めた。母集団を正規分布とし、0.01%特性値を求めたときの誤差Eの値を図2(a)に、対数正規分布で99.99%特性値を求めたときのE値を図2(b)に示す。データ数が増えれば誤差は減少する。方法によって誤差Eに差があることがわかる。種々の数値実験に基づき、各方法を、上側、下側特性値にわけて定性的に比較したのが表3(a)、b)である。特性値の誤差という点からも最尤法(ML法)が安定して精度が高いといえる。プロット法の中では最もよく使われるPL1法は劣っており、プロット法の中ではPL3法が優れている。

評価	母集団分布	ML	PL1	PL2	PL3	K-S
◎ とてもよい	N	○	△	○	○	△
○ まあまあ	LN	○	×	△	○	◎
△ おおまかによい	EX1L	○	×	○	◎	◎
×	EX1S	◎	×	○	△	×
×	EX2	◎	×	○	○	×
×	EX3	○	×	○	○	△

表3 a) 上側特性値の誤差からみた手法の比較

6. まとめ 数値実験を行い、母集団の分布形、特性値の決定に用いられる5つの統計的手法を比較した。統計学的に確立している最尤法が、比較した方法の中では最も優れている、プロット法の中ではHazenのプロットを用いるPL3法が優れていることが明らかにされた。

母集団分布	ML	PL1	PL2	PL3	K-S
N	◎	×	○	◎	○
LN	◎	×	○	○	×
EX1L	◎	×	○	○	×
EX1S	○	○	○	○	△
EX2	◎	×	△	△	×
EX3	○	×	○	○	△

表3 b) 下側特性値の場合

<参考文献> 1) Ang, Tang (伊藤, 亀田訳) 土木・建築のための確率・統計の基礎, 丸善, 1978, pp.274-277.