

水文統計に有用な拡張一般化ガンマ分布の 提案とそのパラメータ推定

THE EXTENDED GENERALIZED GAMMA DISTRIBUTION WHICH IS USEFUL FOR HYDROLOGICAL STATISTICS AND ITS PARAMETER ESTIMATION METHOD

廣瀬 英雄¹

Hideo HIROSE

¹正会員 工博 九州工業大学情報工学部教授 (〒820-8502 福岡県飯塚市大字川津680-4)

In hydrologic frequency analysis, a variety of probability distribution functions, such as the gamma, log-normal, extreme-value, and log-gamma, are often examined whether they are appropriate to the flood data or not by using some hypothesis testing methods, because the properties of the flood data are not well understood even now. In order to find a property of the flood data, we can use another method: using the more general probability distribution function which includes many types of probability distributions used for hydrological statistics. The 4-parameter extended generalized gamma distribution function is one of such distribution functions. Two types of the extended distributions are proposed. The parameter estimation method is also introduced.

Key Words: *Extended generalized gamma distribution, Gamma distribution, Log-normal distribution, Extreme-value distribution, Gumbel distribution, Maximum likelihood estimation, Homotopy method*

1. まえがき

水文統計学に用いられる確率分布としては、対数正規、グンベル、ガンマ、対数ガンマ、一般化極値、パレート分布などがあり、実在する河川の観測データに対してそれぞれの確率分布関数をあてはめ、何らかの方法（例えば、尤度、最小2乗法、Kolmogorov-Smirnov検定（K-S検定）、SLSC（standard least-squares criterion）¹⁾など）でその河川データに適合と思われる分布を特定して、確率水文量の推定を行っている。毎年の最大流量のデータが過去にわたって蓄積されていることが多いことから、この場合河川データは極値分布のどれかに適合させるのが好ましいとは考えられるが、データの生成機構が不明な部分も多く、通常は上にあげたいくつかの確率分布に適合させて分布の特定を行うことが多い。

データの確率分布に対する型を見るには、上にあげたようにいくつかの確率分布に対してそれぞれ適

合度を調べる方法²⁾に対して、さまざまな特徴を合わせ持つ確率分布にまずデータをフィットさせ、推定されたパラメータからデータの性格を調べる方法が考えられる。例えば、一般化極値分布は、ワイブル分布、フレシェ分布、および（極限形ではあるが）グンベル分布の全ての特徴を合わせ持っているため、初めに一般化極値分布のパラメータの推定を行うことによって、3つの分布の中でのそのデータの性格をつかむことができる。

本稿では、このような後者の立場から、最初にあげた分布の多くを含むような確率分布を提示し、またその分布のパラメータ推定に関して考察する。提案される確率分布の形としては複数考えられるが、それらの利用法についても言及する。推定法としては最尤法を用いている。

2. 拡張一般化ガンマ分布

Stacy³⁾によって提案された一般化ガンマ分布の密

度関数,

$$f(x; a, b, p) = \frac{p}{a} \frac{1}{\Gamma(b)} \left(\frac{x}{a}\right)^{bp-1} \exp\left\{-\left(\frac{x}{a}\right)^p\right\},$$

$$(x \geq 0; a, b, p > 0) \quad (1)$$

は、指数、ワイブル、ガンマ分布等を含む魅力ある分布である。レイリー分布、カイ2乗分布もガンマ分布の一種と考えればこれらも含まれていることになる。更に、(1)の対数変換は、 $b \rightarrow \infty$ のときに正規分布に収束することから、(1)は $b \rightarrow \infty$ のときに対数正規分布に収束する。

Harter⁴⁾は(1)に位置パラメータを付加することによって、密度関数

$$f(x; a, b, c, p)$$

$$= \frac{p}{a} \frac{1}{\Gamma(b)} \left(\frac{x-c}{a}\right)^{bp-1} \exp\left\{-\left(\frac{x-c}{a}\right)^p\right\},$$

$$(x \geq c; a, b, p > 0) \quad (2)$$

の4パラメータのモデルを提案している。

これらのモデルでは形状パラメータ、 b, p に正の条件が付加されているが、StacyとMihram⁵⁾は(1)で形状パラメータに負の値を許すことでモデルの拡張を行った。 $1/\lambda^2 = b$ と変換することで、 $b \rightarrow \infty$ の取り扱いが容易になる⁶⁾。

一方、(2)で、 $1/k = p$ と変換することにより、パラメータを求めるための数値計算が安定になることが考えられる⁷⁾。これはちょうど、ワイブル分布の形でパラメータを推定するよりも一般化極値分布の形でパラメータを推定した方が安定することと同じことが期待されるからである。

これらの変換を同時に行ったモデルで、変数変換を

$$a = \frac{\sigma}{|\lambda k|}, c = \mu - \frac{\sigma}{|\lambda k|}, p = \frac{1}{|k|}, b = \frac{1}{\lambda^2} \quad (3)$$

のように行えば、

$$f(x; \lambda, k, \sigma, \mu) = \frac{|\lambda|}{\sigma} \frac{1}{\Gamma(1/\lambda^2)}$$

$$\times \left\{1 + \lambda k \left(\frac{x-\mu}{\sigma}\right)\right\}^{1/(\lambda^2 k) - 1}$$

$$\times \exp\left[-\left\{1 + \lambda k \left(\frac{x-\mu}{\sigma}\right)\right\}^{1/k}\right],$$

$$1 + \lambda k \left(\frac{x-\mu}{\sigma}\right) \geq 0; \lambda \neq 0, k \neq 0, \sigma > 0 \quad (4)$$

の密度関数が得られる⁸⁾。このモデルを拡張一般化ガンマ分布A (Extended generalized gamma distribution A, 以下EGG4.Aと略する) と呼ぼう。

このモデルは、 $\lambda=1$ で一般化極値分布 (以下GEV) になり、その中で $k \rightarrow \infty$ とすればグンベル分布が得られる。

一方、変数変換を

$$a = \frac{|\lambda| \sigma}{|k|}, c = \mu - \frac{\sigma}{|\lambda k|}, p = \frac{1}{|k|}, b = \frac{1}{\lambda^2} \quad (5)$$

のように行えば、

$$f(x; \lambda, k, \sigma, \mu) = \frac{1}{\sigma |\lambda| \Gamma(1/\lambda^2)}$$

$$\times \left\{\frac{1}{\lambda^2} + \frac{k}{\lambda} \left(\frac{x-\mu}{\sigma}\right)\right\}^{1/(\lambda^2 k) - 1}$$

$$\times \exp\left[-\left\{\frac{1}{\lambda^2} + \frac{k}{\lambda} \left(\frac{x-\mu}{\sigma}\right)\right\}^{1/k}\right]$$

$$1 + \lambda k \left(\frac{x-\mu}{\sigma}\right) \geq 0; \lambda \neq 0, k \neq 0, \sigma > 0 \quad (6)$$

の拡張一般化ガンマ分布B (Extended generalized gamma distribution B, 以下EGG4.Bと略する) が得られる。このモデルは、 $k=1$ で3パラメータ拡張一般化ガンマ分布 (以下EGG3) になり、 $\lambda=1$ で一般化極値分布になる。従って、モデルA よりも適用性が高いように直感的には考えられるが、後で示すようにパラメータ推定時の数値的安定性の観点から不利であることがある。

3. パラメータ推定法

最尤法を用いてパラメータ推定を行う場合、未知パラメータの数が多くなれば数値計算は容易ではなくなる。推定法として、最尤法だけでなく、モーメント法やPWM法⁹⁾が用いられているのはこのためでもある。また、位置パラメータを含んだ確率分布には、Smithが詳しく述べているように非正則性の問題¹⁰⁾やHirose, Laiの指摘したパラメータ発散の問題¹¹⁾が存在するため、推定は注意深く行われなければならない。

Hiroseは最尤推定値が安定的に求められるアルゴリズムとしてホモトピー法を利用することを提案している¹²⁾。パラメータ数が多く、拡張されたEGG4.A, EGG4.Bのような場合にはこの方法が有効であるので、本稿でもホモトピー法による推定法を簡単に説明する。

ホモトピー法は、 $f(\theta)=0$ の解を求めたいときに、 $g(\theta)=0$ の解があらかじめ自明である関数 $g(\theta)$ と $f(\theta)$ とを連続的に、

$$h(t, \theta) = tf(\theta) + (1-t)g(\theta) = 0 \quad (7)$$

のように接続して、 $h(t, \theta) = 0$ の解を

$t = 0$ ($g(\theta) = 0$) から $t = 1$ ($f(\theta) = 0$) まで連続的に追跡する方法である。

$$g(\theta) = \nabla \log L(\theta) - \nabla \log L(\theta_0),$$

$$(\theta_0 \text{ は任意, } L \text{ は尤度関数}) \quad (8)$$

として、 $h = 0$ を t で微分すれば、

$$\theta(t) = -[h_{\theta}(t, \theta(t))]^{-1} h_t(t, \theta(t)) \quad (9)$$

が得られ、この微分方程式にオイラー法を適用すれば、

$$\theta^{(j+1)} = \theta^{(j)} - \delta(J^{(j)})^{-1} \nabla \log L(\theta_0), \quad j = 0, 1, \dots, \quad (10)$$

が得られる。これは非線形方程式の解法であるニュートン法とアルゴリズムはほとんど同じであるため、ニュートン法のプログラムが開発されていればそれをそのまま使うことができる。ここに、 J は $\nabla \log L(\theta)$ のヤコビアンである。

しかしながら、補助パラメータ t が単調に増加できないようなケース（図-1参照）では上のような単純な方法ではなく、更に単調増加が約束される補助パラメータ s （解曲線の出発点からの距離）を用いることが多い。曲線を $C(s)$ で表すとき、 $h = 0$ の s による微分は、

$$h'(C(s)) \cdot \dot{C}(s) = 0 \quad (11)$$

となるので、この微分方程式を解けばよい。この方法を予測子修正子ホモトピー法と呼ぶ。この方法では、図-1のようなケースの回避ばかりでなく、より早く、安定的に解を求めることのできる利点がある。

4. 解析例

(1) EGG4.AとEGG4.Bとの比較

いくつかの公開されている信頼性分野のデータに、EGG4.A, EGG4.Bをフィットさせ、パラメータの推定値を求めてみた結果の一部を表-1に示す。参考のため、表にはGEV, EGG3のパラメータの推定値も併記している。これらの例は、パラメータの推定値が定義域の内部に存在するケースである。

表によれば、EGG4.Aの型でパラメタライズした尺度、位置パラメータの推定値はGEVやEGG3のそれらの推定値に近い値を示しているが、EGG4.Bの推定値とは離れている傾向が見られる。従って、これらの例からは、分布の傾向を推定されたパラメータから把握するときには、EGG4.BよりもEGG4.Aの確率分布モデルの方が好ましいと考えられる。た

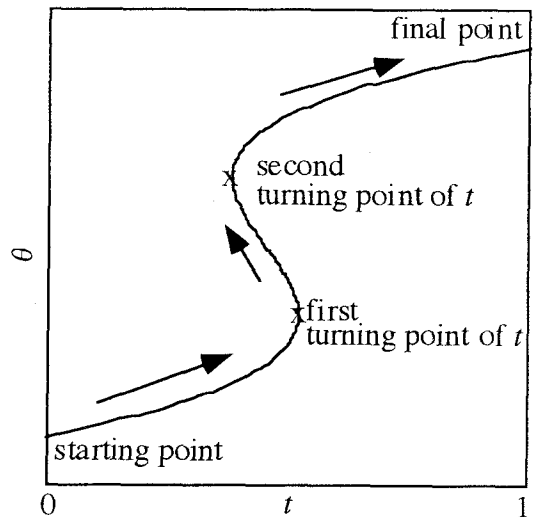


図-1 ホモトピー法による解の追跡

だ、必ずこのような傾向になる訳ではなく、多くの場合そのような傾向があるということである。

一方、数値計算の安定性からいけば、EGG4.Bの場合には、 $\lambda = 1$ から出発するケースと、 $k = 1$ から出発するケースの両方が使えるにもかかわらず、計算に失敗する場合がEGG4.Aのモデルを用いたときよりも多い傾向があることを筆者は経験した。従って、この面からもEGG4.Aの確率分布モデルの方が好ましいと考えられる。

2つの形状パラメータについては両モデルとも同じ値なので同等である。また、対数尤度についても両モデルとも同じ値である。EGG4.Aでの推定値が求められれば、EGG4.Bでの推定値も単純な変換によって求めることができるし、逆も同様である。

(2) EGG4の分布形状とEGG4の有用性

次にEGG4の分布の形状の豊かさを見るために実際の公開されている信頼性データに分布をあてはめ、その形状を見てみた。それを図-2に示す。パラメータの推定値を定義域の境界にまで許せば、非常に豊かな分布形状が得られている。従って、従来の極値分布で表すことが困難な分布の表現に用いられる可能性がある。

さて、EGG4.A, EGG4.Bの対数尤度に関しては、GEVやEGG3のそれよりわずかに大きい値となっていることが多い。通常パラメータ数を増やすことによって尤度は大きくなるので、最適なパラメータ数を決定する判断にはよく赤池の情報量基準(AIC)¹³⁾が用いられる。パラメータ数1の増加に対して尤度の増加が1に比べてかなり小さければパラメータ数を増やすことには消極的になった方が好ましい。従って、拡張一般化ガンマ分布を使用するこ

表-1 文献データでのパラメータ推定値

データ	モデル	推定値				
		λ	k	σ	μ	$\log L$
*1	EGG4.A	2.751	.03747	.2246	3.661	-6.80599
	EGG4.B	2.751	.03747	.02968	1.770	-6.80599
	GEV	1	-.1593	.2730	3.364	-7.49358
	EGG3	-.7177	1	.4172	3.160	-7.00027
*2	EGG4.A	.7457	.01849	.2755	1.399	-14.2394
	EGG4.B	.7457	.01849	.4955	17.36	-14.2394
	GEV	1	.08435	.2729	1.642	-14.2853
	EGG3	-.2781	1	.3145	1.507	-14.8607
*3	EGG4.A	1.432	.08907	.2255	1.394	-2.00359
	EGG4.B	1.432	.08907	.1100	.4882	-2.00359
	GEV	1	.04692	.2215	1.249	-2.08236
	EGG3	-.4158	1	.2687	1.130	-2.04942
*4	EGG4.A	1.666	.2758	63.31	151.0	-205.833
	EGG4.B	1.666	.2758	22.82	62.87	-205.833
	GEV	1	.4547	46.28	104.3	-206.258
	EGG3	.2961	1	43.81	92.68	-206.759
*5	EGG4.A	1.359	.3235	44.43	138.0	-196.390
	EGG4.B	1.359	.3235	24.05	91.68	-196.390
	GEV	1	.4292	35.84	114.3	-196.518
	EGG3	.2650	1	34.03	104.8	-196.889

*1 Data Case #2 in Kako (1986)¹⁴⁾

*2 Data Case #1 in Smith and Naylor (1987)¹⁵⁾

*3 Data Case #2 in Smith and Naylor (1987)¹⁵⁾

*4 Data Case #1 in Harter and Moore (1966)¹⁶⁾

*5 Data Case #2 in Harter and Moore (1966)¹⁶⁾

表-2 St Mary's Riverデータでのパラメータ推定値

確率分布モデル	最尤推定値				
	位置パラメータ	尺度パラメータ	形状パラメータ1	形状パラメータ2	$\log L$
対数正規	67.0553	319.877	.399337		-451.363
グンベル	349.583	108.344			-451.597
ガンマ	156.979	79.1781	3.23805		-451.223
対数ガンマ	326.863	.0579907	.407425		-451.532
ワイブル	182.649	258.111	1.66826		-451.383
GEV	440.760	154.719	.599427		-451.383
EGG3	413.362	142.478		.555723	-451.223
EGG4.A	296.494	74.6077	.851181	.675017	-451.213
EGG4.B	451.625	163.740	.851181	.675017	-451.213

とは、パラメータを増やすことによって劇的に分布への適合の程度が改善される効果を見るというよりも、データの適合する特徴ある分布を見るのに適していると思われる。

(3) St Mary's Riverデータの解析
次に、水文統計に使われているSt Mary's Riverデータ²⁾に関して、さまざまな確率分布に適合させた結果を比較してみた。それを表-2に示す。4パラ

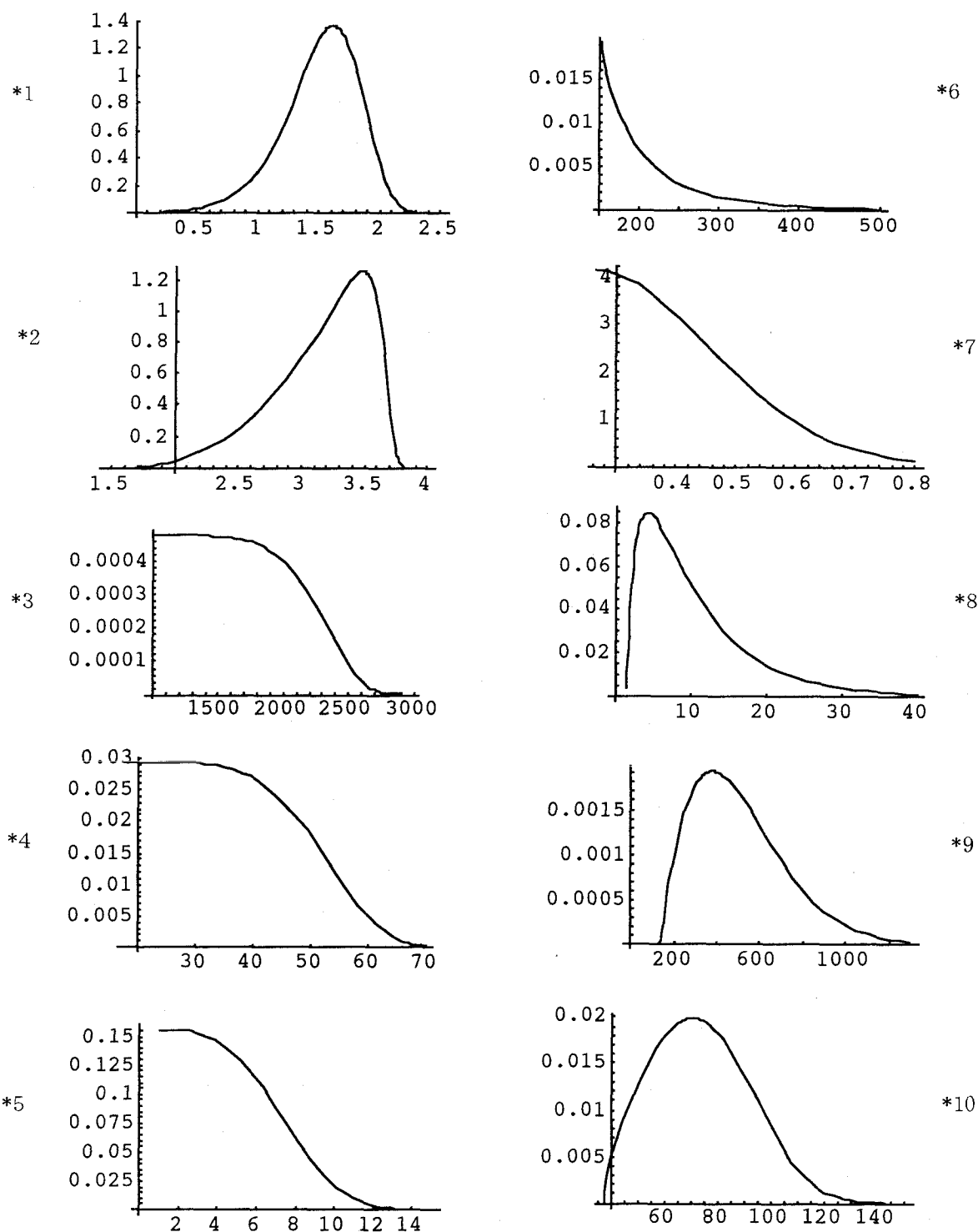


図-2 EGG4の豊かな分布例

- *1 Data Case #1 in Smith and Naylor (1987)¹⁵⁾
- *2 Data Case #2 in Kako (1986)¹⁴⁾
- *3 Data Case in Bilikan, Moore and Petrick(1979)¹⁷⁾
- *4 Data Case in Jiang and Murthy (1997)¹⁸⁾
- *5 Data Case in Hill (1963)¹⁹⁾

- *6 Data Case in Cohen, Whitten and Ding (1985)²⁰⁾
- *7 Data Case in Dumonceaux and Antle (1973)²¹⁾
- *8 Data Case in Wilson and Worcester (1945)²²⁾
- *9 Data Case in Lambert (1964)²³⁾
- *10 Data Case #1 in Harter and Moore (1966)¹⁶⁾

メータのEGG4であっても3パラメータの分布から尤度はそれほど増えていない。また、分布の適合度から見ても3パラメータの尤度にも分布間でそれほどはっきりした違いはなく、どの分布が最もフィットしているとはっきり断言できない（あえて最もふさわしい分布を抽出するとすればガンマ分布、EGG3である）。このような場合には、EGG4のパラメータにも曖昧な結果が出ている。このような場合、AICの観点から考えても従来のガンマ分布からの対応で十分であろう。

5. まとめ

さまざまな特徴を合わせ持つ確率分布にデータをフィットさせ、推定されたパラメータからデータの性格を調べる方法として、4パラメータの拡張一般化ガンマ分布の利用を提案した。その場合、提案できる複数の分布関数、EGG4.A、EGG4.Bについて調べた結果、EGG4.BよりもEGG4.Aの方が、ガンマ、一般化極値、対数正規などの分布のパラメータと類似した結果が出ることが多く、またパラメータ推定時の数値計算の安定性の上からも優れていることが分かった。EGG4は豊かな分布形状を持つが、尤度の改善の程度は大きくはない。

参考文献

- 1) 高棹, 宝, 清水 : 琵琶湖流域水文データの基礎的分析, 京都大学防災研究所年報, 29, B-2, pp.157-171 (1986)
- 2) Kite, G.W. : *Frequency and Risk Analysis in Hydrology*, Water Resources Publications, Fort Collins, Colorado, USA (1988)
- 3) Stacy, E.W. : A generalization of the gamma distribution, *Annals of Mathematical Statistics*, 33, pp.1187-1192 (1962)
- 4) Harter, H.L. : Maximum-likelihood estimation of the parameters of a four-parameter generalized gamma population from complete and censored samples, *Technometrics*, 9, pp.159-165 (1967)
- 5) Stacy, E.W. and Mihram, G.A. : Parameter estimation for a generalized gamma distribution, *Technometrics*, 7, pp.349-358 (1965)
- 6) Prentice, R.L. : A log gamma model and its maximum likelihood estimation, *Biometrika*, 61, pp.539-544 (1974)
- 7) Hirose, H. : Parameter estimation in the extreme-value distributions using the continuation method, *Transactions of Information Processing Society of Japan*, 35, pp.1674-1681 (1994)
- 8) Hirose, H. : Maximum likelihood parameter estimation by model augmentation with applications to the extended four-parameter generalized gamma distribution, *Mathematics & Computers in Simulation*, to appear.

- 9) Hosking, J.R.M., Wallis, J.R., Wood, E.F. : Estimation of the generalized extreme-value distribution by the method of probability weighted moments, *Technometrics*, 27, pp.251-261 (1985)
- 10) Smith, R.L. : Maximum likelihood estimation in a class of nonregular cases, *Biometrika*, 72, pp.67-90 (1985)
- 11) Hirose, H. and Lai, T.L. : Inference from grouped data in three-parameter Weibull models with applications to breakdown-voltage experiments, *Technometrics*, 39, pp.199-210 (1997)
- 12) Hirose, H. : Parameter estimation for the 3-parameter gamma distribution using the continuation method, *IEEE Transactions on Reliability*, 47, pp.188-196 (1998)
- 13) Akaike, H. : Information theory and an extension of the maximum likelihood principle, 2nd *International Symposium on Information Theory*, Akademiai Kiado, Budapest, pp.267-281 (1973)
- 14) Kako, Y. : Estimation of Weibull parameters for BDV data of transformer oil by most likelihood methods, *Conference Internationale des Grands Reseaux Electriques, Working Group bf 15-01, Task Force 02* (1986)
- 15) Smith, R.L. and Naylor, J.C. : A comparison of maximum likelihood and Bayesian estimators for the three-parameter Weibull distribution, *Biometrika*, 73, pp.67-90 (1987)
- 16) Harter, H.L. and Moore, A.L. : Local-maximum-likelihood estimation of the parameters of three-parameter lognormal populations from complete and censored samples, *Journal of the American Statistical Association*, 61, pp.842-851 (1966)
- 17) Bilikan, J.E., Moore, A.H. and Petrick, G.L. : K sample ML ratio test for change of shape parameter, *IEEE Transactions on Reliability*, 28, pp.47-50 (1979)
- 18) Jiang and Murthy : Comment on: A general linear-regression analysis applied to the 3-parameter Weibull distribution, *IEEE Transactions on Reliability*, 46, pp.389-393 (1997)
- 19) Hill, M.B. : The Three-parameter lognormal distribution and Bayesian analysis of a point-source epidemic, *Journal of the American Statistical Association*, 58, pp.72-84 (1963)
- 20) Cohen, A. C., Whitten, B.J. and Ding, Y. : Modified moment estimation for the three-parameter lognormal distribution, *Journal of Quality Technology*, 17, pp.92-99 (1985)
- 21) Dumonceaux, R. and Antle, C.E. : Discrimination between the lognormal and the Weibull distributions, *Technometrics*, 15, pp.923-926 (1973)
- 22) Wilson, E.B. and Worcester, J. : The normal logarithmic transform, *Review of Economic Statistics*, 27, pp.17-22 (1945)
- 23) Lambert, J.A. : Estimation of parameters in the three-parameter lognormal distribution, *Australian Journal of Statistics*, 6, pp.29-32 (1964)

(2000.10. 2 受付)