

## 極値Ⅱ型分布の特性と推定値の信頼区間

CHARACTERISTICS OF THE FISHER-TIPPETT TYPE II DISTRIBUTION  
AND THEIR CONFIDENCE INTERVALS

合田良実\*・小野澤昌己\*\*

By Yoshimi GODA and Masami ONOZAWA

The Fisher-Tippett type II distribution is presented in a three-parameter formula, which converges to the FT type I as the shape parameter approaches infinity. The best plotting position formula with least bias of return values is selected on the basis of Monte Carlo simulations with 10 000 samples for each sample size and shape parameter. The confidence intervals of the estimates of the scale and location parameters based on the least square method are tabulated for the shape parameters of 2.5, 3.33, 5.0, and 10.0. An empirical formula is derived for the standard error of the return values for the above values of the shape parameter. The range of variation of the correlation coefficient between the extreme data and its reduced variate is evaluated for selection of the best fitting distribution.

*Keywords: extreme statistics, FT-II distribution, plotting position, confidence interval*

## 1. まえがき

高波や大雨, 強風などのデータを統計解析して100年再現確率値などを求める際には, 各種の極値分布関数をあてはめ, データに最もよく適合する分布関数を用いることが多い. こうした分布関数の1つに極値Ⅱ型分布があり, 強風<sup>1)~3)</sup>や高波<sup>4)</sup>の極値資料の解析に利用されている. 極値Ⅱ型分布はフレッシュ分布ともいわれ, 標本中の最大値あるいは最小値に対して成立する3種類の極値分布の1つとして Fisher と Tippett が1928年に導いたものである<sup>5a)</sup>. また, 角屋は先にこの分布を対数極値分布A型と名付けて積率法による解法を公表し<sup>6)</sup>, さらに異常値の棄却基準についても検討した<sup>7)</sup>.

3種類の極値分布の中でもⅠ型分布(グンベル分布あるいは二重指数分布)が頻りに利用されているのに対し, 極値Ⅱ型分布は利用度が低い. これは分布関数の特性が利用者に十分に理解されていなかったことや, 解析方法が煩雑であったことが一因と思われる. しかし, 著者の一人が別途発表<sup>8)</sup>しているように, 日本における年最大

日雨量の極値資料は母集団が極値Ⅱ型分布(形状母数 $k=10$ )である可能性が指摘され, 今後は極値Ⅱ型分布を用いた検討が必要になると考えられる. 今回, この分布関数の特性および各種推定値の信頼区間について数値的検討を行ったので, ここに報告する次第である.

## 2. 極値Ⅱ型分布関数の標準形

今までに取り扱われた極値Ⅱ型分布は2母数型のものが多いけれども, 本来の形は3母数型である. Simiu と Filliben<sup>9)</sup>は3母数型として形状母数が無限大のときに極値Ⅰ型分布に漸近することを論じているが, 関数形が不適切なため証明がすっきりしていない. ここでは Gumbel<sup>5b)</sup>を参考にして次のような関数形を提案する.

$$F(x) = \exp\{-[1+(x-B)/kA]^{-k}\} \dots \dots \dots (1)$$

$$f(x) = [1+(x-B)/kA]^{-1+k} F(x)/A \dots \dots \dots (2)$$

ここに,  $F(x)$  は分布関数,  $f(x)$  は確率密度関数,  $A$ ,  $B$ ,  $k$  はそれぞれ尺度, 位置, および形状母数である. 式(1)の形であれば,  $k \rightarrow \infty$ のときに極値Ⅰ型分布となることが自明である. なお式(1)は Jenkinson<sup>9)</sup>による一般化極値分布(GEV)の形状母数  $x$  を  $x = -1/k$  に置き換えたものと同一である.

Fig.1は, 形状母数  $k=2.5, 3.33, 5, 10$  のときの確率密度関数の形状を示したもので, 参考のために  $k=\infty$

\* 正会員 工博 横浜国立大学教授 工学部建設学科  
(〒240 横浜市保土ヶ谷区常盤台156)

\*\* 正会員 日本テトラポッド(株)応用水理研究所  
(〒300 土浦市東中買町2-7)

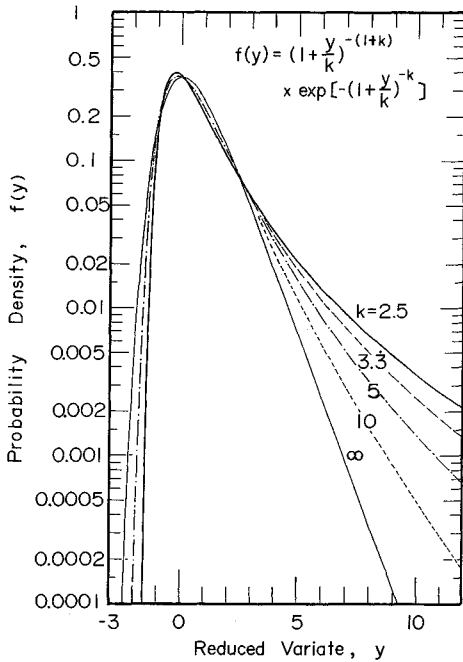


Fig.1 Probability density of the Fisher-Tippett type II distribution.

に相当する極値 I 型分布の形状も示している。この形状母数の値は  $1/k=0, 0.1, 0.2, 0.3$  および  $0.4$  となるように選んだもので、本ノートでは一貫して使用している。図中の変数  $y$  は基準化変数であり、 $y=(x-B)/A$  である。ここに示されるように、極値 II 型分布は  $1/k$  の値が大きくなるにつれて分布の裾が非常に広がるのが特徴である。たとえば、非超過確率が 99% となるのが極値 I 型分布では  $y=4.60$  であるのに対し、極値 II 型分布では  $1/k=0.1, 0.2, 0.3, 0.4$  に対してそれぞれ  $y=5.84, 7.54, 9.92, 13.24$  となる。

なお、母集団平均および標準偏差はよく知られているようにガンマ関数  $\Gamma(x)$  を使って次のように表わされる。

$$E[x]=kA[\Gamma(1-1/k)-1]+B \dots\dots\dots (3)$$

$$\sigma(x)=kA[\Gamma(1-2/k)-\Gamma^2(1-1/k)]^{1/2} \dots\dots\dots (4)$$

また、最多値は次式で与えられる。

$$x_{mode}=kA\{[k/(1+k)]^{1/k}-1\}+B \dots\dots\dots (5)$$

3. プロットイング・ポジション公式

標本を極値分布関数にあてはめる方法として最小 2 乗法を用いるときは、プロットイング・ポジションとして再現確率統計量や母数の推定値に偏りを生じないものを選ぶなければならない。ここでは、モンテカルロ法によっていろいろな大きさの標本を各 10 000 組ずつ抽出し、試行錯誤によってプロットイング・ポジション公式の定

数を次のように設定した。

$$\hat{F}_m=1-\frac{m-0.44-0.52/k}{N_T+0.12-0.11/k} \dots\dots\dots (6)$$

$$: m=1, 2, \dots, N$$

ただし、 $\hat{F}_m$  は降順で  $m$  番目のデータに対する非超過確率の推定値 ( $m=1$  が最大値)、 $N_T$  は対象期間中に発生したと想定される極値の総数、 $N$  は標本中のデータの個数すなわち標本の大きさである。式 (6) は、 $k \rightarrow \infty$  において極値 I 型分布に対して不偏性が確認<sup>10)</sup>されているグリーンゴルテン公式に一致するようにしてある。

式 (6) を使ったときの再現確率統計量の推定値の平均値の誤差は、 $A=1, B=5$  とし、再現期間の長さを標本の大きさの 10 倍に相当する値にとったとき、 $N=10 \sim 200$  の範囲の 10 000 組の全平均で  $-2.7\% \sim +0.3\%$  であった。ただし、誤差の大きいのは  $k=2.5$ 、次いで  $k=3.33$  の場合であり、 $k=5.0$  と  $10.0$  については  $-0.6\% \sim +0.3\%$  であった。

4. 標本の標準偏差の変動性

極値 II 型分布の母集団の標準偏差は式 (4) で与えられるけれども、これから抽出された標本の不偏標準偏差はこれよりもやや小さく、しかも標本ごとにばらつく。数値シミュレーションの結果から標本の平均値とその変動係数をまとめて Table 1 に示す。これは標本の大きさに各 10 000 組の作業を 2 回繰り返して平均し、それを  $N$  に対して滑らかに変化するように若干修正したものである。形状母数が  $k=2.5$  の場合の平均値がかなり小さいのは、この分布では Fig. 1 に示すように、分布の裾が非常に遠くまで伸びていることの影響と思われる。

ここに示したように、標本の標準偏差は母集団値と異なるので、母数推定に積率法を使用する場合には、標本の大きさの影響補正を行わないと尺度母数を過小に見積

Table 1 Mean and coefficient of variation of the unbiased standard deviation of sample data.

標本の大きさ N	極 値 II 型 分 布				極値 I 型分布
	k = 2.5	k = 3.33	k = 5.0	k = 10.0	
10	0.632 (1.25)	0.798 (0.79)	0.882 (0.55)	0.928 (0.41)	0.951 (0.32)
14	0.665 (1.12)	0.829 (0.71)	0.906 (0.49)	0.946 (0.36)	0.965 (0.27)
20	0.690 (0.99)	0.851 (0.63)	0.923 (0.42)	0.958 (0.31)	0.975 (0.23)
30	0.714 (0.87)	0.873 (0.54)	0.940 (0.36)	0.969 (0.25)	0.984 (0.19)
40	0.732 (0.80)	0.886 (0.49)	0.949 (0.32)	0.975 (0.22)	0.988 (0.16)
60	0.763 (0.74)	0.905 (0.44)	0.963 (0.28)	0.983 (0.18)	0.992 (0.13)
100	0.793 (0.66)	0.929 (0.38)	0.974 (0.23)	0.989 (0.15)	0.995 (0.10)
140	0.810 (0.61)	0.941 (0.35)	0.981 (0.20)	0.992 (0.13)	0.996 (0.09)
200	0.827 (0.56)	0.951 (0.32)	0.987 (0.18)	0.994 (0.11)	0.998 (0.07)
母集団値 $\sigma/A$	[3.8513]	[2.4341]	[1.8287]	[1.4921]	[1.2826]

注：1) 平均値は母集団値に対する比で表している。 2) 括弧内の数値は変動係数。

もる結果となる。

5. 母数推定値の信頼区間と再現確率統計量の標準偏差

前項と同じ資料に基づいて、尺度母数  $A$  および位置母数  $B$  を最小2乗法で推定した場合の信頼区間を Table 2 に示す。母数絶対値の影響を避けるため、前者については  $A/\hat{A}$ 、後者については  $(\hat{B}-B)/\hat{A}$  の補助統計量の形で示している。ただし、 $\hat{A}$ 、 $\hat{B}$  は推定値を表わす。

母数推定値の変動は、再現確率統計量の変動をもたらす。数値シミュレーション結果から再現期間  $R$  における再現確率統計量の推定値  $\hat{x}_R$  の標準偏差を次のような実験式の形に取りまとめた（なお、先に口頭で発表したもの<sup>11)</sup>は計算に誤りがあったので今回のものに差し替えさせていただく）。

$$\sigma(\hat{x}_R) = [1 + a(y_R - c + \alpha \ln \nu)^2]^{1/2} \sigma_x / N^{1/2} \dots\dots (7)$$

ここに、

$$a = a_1 \exp \{ a_2 [\ln(N\nu^{0.5}/N_0)]^2 - \kappa [\ln(\nu/\nu_0)]^2 \} \dots\dots (8)$$

ただし、 $\nu$  はデータ採択率  $= N/N_T$ 、 $\sigma_x$  は標本資料の不

偏標準偏差である。式中の各係数は  $N=10\sim 200$ 、 $\nu=0.25, 0.5, 1.0$  の数値実験結果に基づいて Table 3 のように設定した。なお、再現期間  $R$  に対する  $y_R$  は次式で算定される。

$$y_R = k \{-\ln[1 - 1/(\lambda R)]^{-1/k} - 1\} \dots\dots\dots (9)$$

ここに、 $\lambda$  は平均発生率であり、毎年最大値資料では  $\lambda=1$ 、 $\nu=1$  とおくことができる。

Fig. 2, 3 は、数値シミュレーション結果に対する実験式の適合状況の例であり、縦軸の  $\sigma_z$  は  $z = (\hat{x}_R - x_R) / \sigma_x$  の補助統計量の標準偏差である。Fig. 2 は  $k=10$  の極値

Table 3 Empirical coefficients for the standard deviation of the FT-II return value.

形状母数	$a_1$	$a_2$	$N_0$	$\kappa$	$\nu_0$	$c$	$\alpha$
$k=2.5$	1.27	0.12	23	0.24	1.34	0.3	2.3
$k=3.33$	1.23	0.09	25	0.38	0.66	0.2	1.9
$k=5.0$	1.34	0.07	35	0.41	0.45	0.1	1.6
$k=10.0$	1.48	0.06	80	0.47	0.26	0	1.4

Table 2 Confidence intervals of the FT-II parameters estimated by the least square method.

形状母数	N	尺度母数 $A/\hat{A}$				$\sigma$	位置母数 $(\hat{B}-B)/\hat{A}$				$\sigma$
		2.5%	25%	75%	97.5%		2.5%	25%	75%	97.5%	
$k=2.5$	10	0.30	0.89	2.04	4.05	0.98	-1.27	-0.31	0.39	1.42	0.66
	14	0.31	0.89	1.90	3.45	0.82	-0.92	-0.25	0.37	1.21	0.53
	20	0.35	0.89	1.78	3.05	0.70	-0.66	-0.20	0.35	1.05	0.44
	30	0.37	0.89	1.65	2.65	0.59	-0.51	-0.15	0.33	0.92	0.37
	40	0.39	0.89	1.59	2.48	0.53	-0.47	-0.13	0.32	0.85	0.34
	60	0.42	0.89	1.51	2.25	0.47	-0.44	-0.10	0.30	0.75	0.30
	100	0.45	0.90	1.43	2.01	0.40	-0.43	-0.08	0.28	0.65	0.28
$k=3.33$	140	0.48	0.90	1.39	1.90	0.36	-0.43	-0.08	0.27	0.60	0.26
	200	0.49	0.90	1.35	1.80	0.33	-0.43	-0.07	0.27	0.56	0.25
	10	0.36	0.85	1.73	3.19	0.74	-1.09	-0.28	0.30	1.11	0.54
	14	0.38	0.86	1.62	2.73	0.61	-0.82	-0.23	0.27	0.91	0.43
	20	0.42	0.86	1.53	2.43	0.51	-0.62	-0.19	0.24	0.76	0.35
	30	0.46	0.87	1.43	2.14	0.43	-0.47	-0.15	0.21	0.65	0.28
	40	0.49	0.88	1.38	2.00	0.38	-0.40	-0.13	0.20	0.57	0.25
$k=5.0$	60	0.52	0.88	1.32	1.83	0.33	-0.34	-0.10	0.17	0.48	0.21
	100	0.55	0.89	1.26	1.65	0.28	-0.30	-0.09	0.15	0.39	0.18
	140	0.59	0.90	1.23	1.56	0.24	-0.28	-0.08	0.14	0.34	0.16
	200	0.62	0.91	1.20	1.49	0.22	-0.26	-0.07	0.12	0.30	0.14
	10	0.43	0.84	1.53	2.63	0.57	-0.96	-0.26	0.26	0.95	0.47
	14	0.46	0.85	1.44	2.29	0.47	-0.74	-0.21	0.22	0.76	0.37
	20	0.50	0.86	1.36	2.03	0.39	-0.57	-0.18	0.19	0.62	0.30
$k=10.0$	30	0.55	0.87	1.29	1.81	0.32	-0.44	-0.14	0.16	0.51	0.24
	40	0.59	0.88	1.25	1.70	0.28	-0.37	-0.12	0.15	0.44	0.21
	60	0.62	0.89	1.21	1.57	0.24	-0.30	-0.10	0.12	0.38	0.17
	100	0.67	0.91	1.16	1.44	0.19	-0.24	-0.08	0.09	0.28	0.13
	140	0.71	0.91	1.14	1.37	0.17	-0.21	-0.07	0.08	0.24	0.11
	200	0.74	0.92	1.12	1.32	0.15	-0.18	-0.06	0.07	0.20	0.10
	$k=10.0$	10	0.51	0.84	1.39	2.26	0.46	-0.85	-0.24	0.25	0.86
14		0.54	0.86	1.32	1.98	0.37	-0.68	-0.19	0.21	0.69	0.34
20		0.59	0.87	1.26	1.77	0.30	-0.53	-0.16	0.18	0.55	0.27
30		0.64	0.88	1.20	1.60	0.24	-0.42	-0.13	0.15	0.44	0.22
40		0.67	0.90	1.18	1.51	0.21	-0.36	-0.11	0.13	0.38	0.19
60		0.71	0.91	1.14	1.41	0.18	-0.28	-0.09	0.10	0.31	0.15
100		0.76	0.92	1.11	1.31	0.14	-0.22	-0.07	0.08	0.23	0.12
$k=10.0$	140	0.79	0.93	1.09	1.26	0.12	-0.19	-0.06	0.07	0.20	0.10
	200	0.82	0.94	1.08	1.22	0.10	-0.16	-0.05	0.06	0.16	0.08

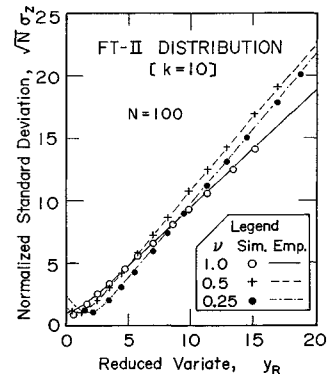


Fig. 2 Standard deviation of the auxiliary statistic  $z = (\hat{x}_R - x_R) / \sigma_x$  of the estimated return value.

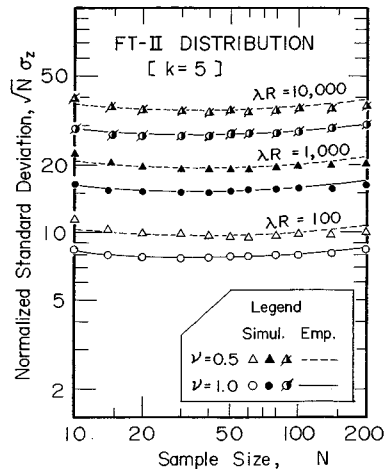


Fig. 3 Effect of sample size on the standard deviation of the auxiliary statistic  $z$ .

**Table 4** Empirical coefficients for the mean residue of correlation coefficient  $\Delta r_{\text{mean}}$ .

形状母数	係数 a	係数 b	係数 c
k = 2.5	-2.470+0.015 $\nu^{3/2}$	-0.1530-0.0052 $\nu^{5/2}$	0
k = 3.33	-2.462-0.009 $\nu^2$	-0.1933-0.0037 $\nu^{5/2}$	-0.007
k = 5.0	-2.463	-0.2110-0.0131 $\nu^{5/2}$	-0.019
k = 10.0	-2.437+0.028 $\nu^{5/2}$	-0.2280-0.0300 $\nu^{5/2}$	-0.033
k = $\infty$	-2.364+0.054 $\nu^{5/2}$	-0.2685-0.0457 $\nu^{5/2}$	-0.044

**Table 5** Empirical coefficients for the 95 % nonexceedance level of the residue of correlation coefficient  $\Delta r_{95\%}$ .

形状母数	係数 a	係数 b	係数 c
k = 2.5	-1.122-0.037 $\nu$	-0.3298+0.0105 $\nu^{1/4}$	0.016
k = 3.33	-1.306-0.105 $\nu^{3/2}$	-0.3001+0.0404 $\nu^{1/2}$	0
k = 5.0	-1.463-0.107 $\nu^{3/2}$	-0.2716+0.0517 $\nu^{1/4}$	-0.018
k = 10.0	-1.490-0.073 $\nu$	-0.2299-0.0099 $\nu^{5/2}$	-0.034
k = $\infty$	-1.444	-0.2733-0.0414 $\nu^{5/2}$	-0.045

II型分布で標本の大きさが  $N=100$  の場合についてデータ採択率による差を例示したものであり、Fig. 3 は  $k=5$  の場合について標本の大きさの影響をみたものである。

**6. 最小2乗法における相関関数の特性**

標本を各種の分布関数にあてはめるときは、最適およびあてはめ棄却の判断基準が必要である。このため、著者の一人は先に MIR 基準および REC 基準なるものを提案した<sup>12)</sup>。いずれも、最小2乗法を適用したときの  $x_r$  と  $y_r$  との間の相関係数と1との残差  $\Delta r = 1 - r$  を指標とするもので、前者は標本の値と母集団平均値  $\Delta r_{\text{mean}}$  との比が最小のものを最適と判断する基準であり、後者は標本の値が母集団における相関係数の残差の95%非超過確率値  $\Delta r_{95\%}$  を越える場合にその分布関数のあてはめを棄却する基準である。これらは極値I型分布、ワイブル分布、および対数正規分布について次のような形の実験式で取りまとめた。

$$\Delta r_{\text{mean}} = \exp [a + b \ln N + c (\ln N)^2] \dots \dots \dots (10)$$

95%非超過確率値についても同一形式である。今回の極値II型分布に対しても同様に取りまとめて係数値を求めた結果を Table 4, 5 に示す。

数値実験値に対するあてはめ誤差は、 $\Delta r_{\text{mean}}$  は  $\pm 2\%$  以内が97%、 $\pm 4\%$  以内が99%であり、 $\Delta r_{95\%}$  については  $\pm 2\%$  以内が63%、 $\pm 4\%$  以内が97%であっ

た。

なお、標本中の最大値の棄却基準については別論文<sup>8)</sup>に発表したので参照して頂きたい。

**7. あとがき**

今回の検討はすべて数値シミュレーションに基づくものなので、結果の数値はそれぞれ誤差を伴っており、理論計算値のような正確さは期待できない。特に、形状母数  $k$  が小さい場合は Table 1, 2 で明らかなように変動が大きいのので、数値の精度に留意しながら利用して頂ければ幸いである。本ノートを機縁として極値II型分布適用の検討事例が広まることを願う次第である。

**参考文献**

- 1) Thom, H. C. S. : Distributions of extreme winds over oceans, Proc. ASCE, Vol. 99, No. WW1, pp. 1~17, 1973.
- 2) Simiu, E. and Filliben, J. J. : Probability distribution of extreme wind speeds, Proc. ASCE, Vol. 102, No. ST 3, pp. 1861~1877, 1976.
- 3) 藤野陽三・伊藤 学・酒井利夫 : 年最大風速記録による設計基本風速の算定に関する研究, 土木学会論文報告集, 第305号, pp. 23~34, 1981.
- 4) Thom, H. C. S. : Extreme wave height distributions over oceans, Proc. ASCE, Vol. 99, No. WW 3, pp. 355~374, 1973.
- 5) Gumbel, E. J. : Statistics of Extremes, Columbia Univ. Pres., 1958 : a)-Sec. 5.1; b)-Sec. 7.1.3.
- 6) 角屋 睦 : 極値分布とその一解法, 農業土木研究, 第23巻, 第6号, pp. 28~35, 1956.
- 7) 角屋 睦 : 異常(確率)水文量とデータの棄却検定, 農業土木研究, 別冊第3号, pp. 23~27, 1962.
- 8) 合田良実 : 極値統計における異常値の棄却基準とその応用, 土木学会論文集, No. 417/II-13, 1990.
- 9) Jenkinson, A. F. : The frequency distribution of the annual maximum (or minimum) values of meteorological elements, Quart. J. Roy. Meteor. Soc., Vol. 81, pp. 158~171, 1955.
- 10) 合田良実 : 極値統計におけるプロットング公式ならびに推定値の信頼区間に関する数値的検討, 港湾技術研究所報告, 第27巻, 第3号, pp. 31~92, 1988.
- 11) 小野澤昌己・合田良実 : 極値II型分布の特性とその年最大風速資料に対する適用性について, 土木学会第44回年次学術講演会, 第I部門, 1989.
- 12) 合田良実・小舟浩治 : 波浪の極値統計における分布関数のあてはめ基準, 海岸工学論文集, 第36巻, pp. 135~139, 1989.

(1989. 8. 21・受付)