

単 変 量 解 析

神 田 徹

ま え が き

水文学において、その現象の確率統計法則の解明に重点を置く研究分野が米国で stochastic hydrology と名付けられてから約10年を経た。この間におけるこの分野の研究の発展は目ざましく、今日、水文システムにおける多くの問題の解決を可能にしているのは周知の通りである。

これらの情勢を反映して昭和49年度より、土木学会水理委員会において確率統計水文学研究小委員会（委員長；東工大、日野幹雄教授）が組織され、確率統計水文学における最近の進歩と今後の発展に関する研究が行なわれている。この研究グループは分担課題によって大きく2つの班で構成されている。すなわち

- ① 水文統計解析（主査：大阪大学、室田明教授）
- ② 水文確率過程解析（主査：東工大、日野幹雄教授）

筆者の属する第1の分担課題の内容について、全般的な位置づけを一望しておく。分担課題「水文統計解析」はさらにその理論体系から次のように分類される。

- (i) 1変数分布理論 (ii) 2変数分布理論 (iii) 多変数分布理論 (iv) 多変数 $\leftrightarrow$ 1変数分布理論

一方、水文量に固有の、かつ重要な確率統計特性を挙げれば次のような性質である。

- (i) 分布の非正規性、(ii) 非定常性 (iii) 時間的・空間的従属性。さらに、実用的見地から設定される時間単位の大きさによって、その統計的性質や解析法が大きく影響されることも水文現象の特質である。

これらの項目によって分類される各分野について、その研究段階・動向を巨視的に表示すれば表-1のごとくである。とくに、□印は現在重点的に研究が進められている分野であり、×印は研究方針の確立が急がれる分野である。実用化のレベルは理論研究、理論の適用性（理論的な確率モデルが水文量に適合するか否か）、水工問題への応用（治水計画、水資源計画等での利用、たとえば、確率降雨の決定、流量シミュレーション）の3段階に大別されているが、1変数問題では分布の正規性、非正規性を問わず理論的基盤が確立されていることおよび水文統計の発端が洪水のピーク流量（1変数）の推定であったことから実用化について充分の実績がある。これに対して、2変数以上の取扱いでは研究の歴史は浅く、実際問題へ応用される例も少ない。明らかに多変数問題として扱わねばならない問題、たとえば多地点降雨量の相関回帰解析等では分布の正規性あるいは生起過程の独立性の仮定のもとに適用はなされているが、その理論的根拠にはまだ目をつぶっているという段階であろう。そこで本来、極値分布やDAD解析のように多変数問題でありながら、その極限形式への漸近的な性質が明らかにされている場合には、近似的に1変数問題として解析し実用に供していることが多い。こうした意味から、2変数以上の確率分布に関する研究が活発化しているのであるが、その方向も1変数から2変数……という方向と一般的多変数解析理論から任意の数の変数分布に向うものがあり、前者に属するものとして現在開発されている2変数分布理論がある。

一方、水文量の従属性や非定常性のうち、前者については相関・回帰理論や確率過程論の手法と関連して研究が進みつつあるが、この場合にも分布の非正規性が大きな障害である。後者に関しては現段階

表 - 1 各分野の研究レベル

研究レベル 変数の数	実 用 化 の レ ベ ル				従 属 性		非定常性	各種の△t に対する総 合化
	確率統計理論の 研究段階		理論の適用 性検討の段 階	水工問題へ の応用の段 階	時間 的 従 属 性	空間 的 従 属 性		
	正規分布	非正規 分布						
1 変 数	○	○	○	○	□		□	□
1↔多変数	○	□	□	□	□		×	□
2 変 数	○	□	□	×	□		×	□
多 変 数	○	×	□	□	□		×	×

- : ほとんど研究完了の分野
 □ : 活発に研究中の分野
 × : 研究の緒についたばかりの分野

では有力な手法が確立されておらず、今後の重要な研究課題である。

以上が「水文統計解析」についてのぐくあらましの研究動向であり、本研修会で論述される各分野の全般的な把握のための参考とされたい。

さて、本文で取扱う単変数統計解析に含まれる問題を大別すれば次の通りである。

I. 水文現象の確率構造

- A. 確率モデルの選択
- B. 確率モデルにおけるパラメータの推定および統計量の特性
- C. 確率モデルによる資料の発生

II. 水文計画論

- A. 水工計画上の目的の指標化・定式化
- B. 評価手法の開発・確立

Iは水文過程に内在する確率統計法則を解析することである。Aには確率分布モデルと確率過程モデルの選択が含まれ、前者に対してはBの項では確率密度関数のパラメータ、後者に対しては時系列の持続性を表すパラメータの推定が対象となる。IIは洪水、渇水等の水文現象について水工計画上の指標を定式化し、Iによる成果を用いて計画論へ適用するものである。

つぎに、多変数解析に対する単変数解析の位置づけとしては、前述の意味から後者は今日つぎの2つの性格をもつものと考えてよいだろう。

1. 理論および手法にさらに厳密な修正、改良が加えられるべきもの。
2. 現在開発されつつある確率統計（OR，制御理論を含む）手法の基礎的性格をもつもの。

すなわち、従来単変数解析で処理していた問題についてさらに厳密な解析が必要な場合や水工計画の指標として新たな統計量が対象となる場合に、単変数統計解析法自体を改良する方向と、さらに複雑な確率モデルによる手法を開発する方向とがある。前者の方向では従来通り理論的發展と計算機性能の高度化が有力な手段となる。後者に関しては、たとえば降雨量や流量のごとき最も基本的な水文量の確率

分布に対して一般的な法則を与えておく努力がなされねばならないであろう。

本文は以上の観点から単変数統計解析に関する基本的な確率統計特性と近年の解析手法を述べるものであるが、その内容は上記の分類において、主としてⅠ-A、B（ただし、時系列モデルは1次の自己回帰モデルのみが扱われる。）およびⅡに関連する極値分布とリターン・ペリオドの問題である。

I 1 変数確率分布

1. 確率分布の選択

1-1. 概 説

水文量の確率（頻度）分布に関する研究は数多く報告されており、降水量や流量に対して経験的にいくつかの確率分布が採用されていることは周知の通りである。しかしながら、その頻度分布特性についての一般的な事実としては、たとえば水文量の頻度分布曲線は非対称性を示すことが多いというような定性的な事実だけであって、特定の分布が水文量の潜在的な分布として存在するという仮定を正当化する先験的情報は何もない。⁽¹⁾ 確率分布モデルを標本にあてはめる場合、大きさ20～60程度の小標本に対しては、従来の適合度検定法では特定の分布を指定することは不可能であるし、また分布形状の多様性が水文現象の複雑な機構に由来する限りその統一的表現自体が本質的に可能であるかどうかとも疑問であろう。しかしながら、現象を構成する特定の基本的要素、たとえば時間スケールに関して確率法則のより一般的な性質を明らかにしようとする最近の研究^{(2), (3)}に見られるように、水文量の確率統計的性質に関する理論的な接近に期待が持たれるとともに、一方実用上の観点からは、水文量への適合性に関して一般性の高い分布関数を設定し資料集積の統一化をはかることが、水文現象の広域的な把握、これは換言すれば局所性の解明のためにも要請されるであろう。

この方面に関する研究としては、米国のWater Resources Councilは1966年に洪水の頻度解析法に関する研究のWork Groupを組織し、その成果を報告している。^{(4), (5)} この報告では、洪水の頻度分布に次の6つの確率分布の適用がなされ、検討されている。

- (1) 2-パラメータ・ガンマ分布 (2) Gumbel 分布 (3) log-Gumbel 分布 (4) 対数正規分布 (5) log-ピアソン-Ⅲ型分布 (6) Hazen法

これらの分布関数の適用結果から得られた結論として、この報告は最適な確率分布形や解析手順は将来にわたって固定すべきではないとしながらも、log-ピアソン-Ⅲ型分布を洪水の頻度解析の基準法として推奨している。このような勧告もあって、米国では極値、（低水）時系列の両者に対してピアソン-Ⅲ型分布、またはlog-ピアソン-Ⅲ型分布は対数正規分布とともに多く用いられているが、近年、従来から用いられている極値分布、指数分布、 n -乗根分布の他にベータ分布やtwo-distribution分布等の適用性も検討されている。

わが国では水文量の頻度解析法に関する系統的な調査、研究は、いくつかの気象要素に対するもの⁽⁶⁾を除いてはほとんど見られないが、上に挙げた確率分布型は、種々の単位時間の降水量や流量に個々に適用されている。

本稿ではまず、確率分布の選択に関連する方法を述べたのち、種々の水文量への適用範囲の広い対数正規分布とピアソン-Ⅲ型分布について、その解析法と最近の改良法を述べる。

1-2. ピアソン系分布の判別基準

観測値の頻度分布に最も適合する確率分布型の選択法として、確率分布がピアソン系分布に属する場合に用いられる以下のような方法がある。

ピアソン系分布はKarl Pearson によって提示された一組の確率分布であり、この分布型は現象の頻度分布に関する理論的基盤には立脚していないが、数学的表現が種々の現象の分布に適合する自由度を有することが最大の特徴である。

ピアソン系分布は微分方程式

$$\frac{1}{y} \frac{dy}{dx} = (x+a) / (px^2 + qx + r) \quad \dots\dots\dots (1)$$

ここに、 a, p, q, r は定数。

の解、 $y=f(x)$ で定義され、定数 a, p, q, r の値の取り方によって12の型に分類できる。⁽⁷⁾水文学の頻度分布へのあてはめには、次の型がよく用いられる。

a. I型分布(ベータ分布)

$$y = y_0 \left(1 + \frac{x}{a_1}\right)^{m_1} \left(1 - \frac{x}{a_2}\right)^{m_2} \quad \dots\dots\dots (2)$$

ここに $m_1/a_1 = m_2/a_2$

b. III型分布(ガンマ分布)

$$y = y_0 \left(1 + \frac{x}{a}\right)^{ra} \exp(-rx) \quad \dots\dots\dots (3)$$

あるいは変数変換をすれば

$$y = \frac{1}{\Gamma(\lambda)} x^{\lambda-1} e^{-x} \quad \dots\dots\dots (4)$$

c. VII型分布(正規分布)

$$y = y_0 \exp \{ -(x-a)^2 / 2\sigma^2 \} \quad \dots\dots\dots (5)$$

(2)~(5)式において、確率変数 x を除く記号は定数である。

12の型のうち標本の分布にどの型が適合するかの判別基準として Pearson は次のパラメーター、 κ を用いている。

$$\kappa = \frac{\beta_1^2 (\beta_2 + 3)^2}{4(2\beta_2 - 3\beta_1^2 - 6)(4\beta_2 - 3\beta_1^2)} \quad \dots\dots\dots (6)$$

ここに $\beta_1 = \frac{\mu_3}{\mu_2^{3/2}}$; 歪み係数

$\beta_2 = \frac{\mu_4}{\mu_2^2}$; 尖り係数

μ_r ; 平均値のまわりの r 次の積率

$\kappa, \beta_1^2, \beta_2$ とピアソン系分布の主要な型との対応関係は表-2の通りである。

型が決まれば、観測値についての積率(標本積率) m_r をその型の積率 μ_r に等しいと置いて分布のパラメータを求める。前に挙げた型に対するパラメータは次のように与えられる(Pearsonの方法)。

表-2 ピアソン系分布関数の判別基準

型	確率密度関数	分布の上・下限	モード	κ	$\beta_1 \quad \beta_2$
I	$y = y_0 \left(1 + \frac{x}{a_1}\right)^{m_1} \left(1 - \frac{x}{a_2}\right)^{m_2}$	両端あり (非対称)	原点	< 0	
II	$y = y_0 \left(1 - \frac{x^2}{a^2}\right)^m$	両端あり (対称)	原点	0	$0 \quad \approx 3$
III	$y = y_0 \left(1 + \frac{x}{a}\right)^{ra} e^{-rx}$	一端のみあり (非対称)	原点	∞	$(2\beta_2 - 3\beta_1^2 - 6 = 0)$
IV	$y = y_0 \left(1 + \frac{x^2}{a^2}\right)^{-m} \cdot \exp(-\nu \tan^{-1} \frac{x}{a})$	両端なし (非対称)	$-\frac{\nu a}{2m}$	$1 < \kappa < \infty$	
V	$y = y_0 x^{-p} e^{-r/x}$	一端のみあり (非対称)	$\frac{r}{p}$	1	
VI	$y = y_0 (x - a)^{q_2} x^{-q_1}$	一端のみあり (非対称)	$\frac{ap_1}{p_1 - p_2}$	$1 < \kappa < \infty$	
VII	$y = y_0 \exp\left(-\frac{x^2}{2\sigma^2}\right)$	両端なし (対称)	原点	0	$0 \quad 3$

a) I 型

$$\begin{aligned}
 y_0 &= \frac{1}{b} \frac{m_1^{m_1} m_2^{m_2}}{(m_1 + m_2)^{m_1 + m_2}} \cdot \frac{\Gamma(m_1 + m_2 + 2)}{\Gamma(m_1 + 1) \Gamma(m_2 + 1)} \\
 m_1 &= \frac{1}{2} \left[r - 2 - \frac{1}{2} \left\{ r^2 - \frac{16r^2(r+1)}{\beta_1^2(r+2)^2 + 16(r+1)} \right\}^{\frac{1}{2}} \right] \\
 m_2 &= \frac{1}{2} \left[r - 2 + \frac{1}{2} \left\{ r^2 - \frac{16r^2(r+1)}{\beta_1^2(r+2)^2 + 16(r+1)} \right\}^{\frac{1}{2}} \right] \\
 r &= \frac{6(\beta_2 - \beta_1^2 - 1)}{3\beta_1^2 - 2\beta_2 + 6} \\
 \frac{a_1}{m_1} &= \frac{a_2}{m_2} = \frac{b}{m_1 + m_2} \\
 b &= \frac{1}{2} \mu_2^{\frac{1}{2}} \left\{ \beta_1^2(r+2)^2 + 16(r+1) \right\}^{\frac{1}{2}}
 \end{aligned}
 \quad \left. \begin{array}{l} \\ \\ \\ \\ \\ \end{array} \right\} \dots\dots\dots (7)$$

b) III 型

$$\begin{aligned}
 y_0 &= \frac{p^{p+1}}{a e^p \Gamma(p+1)} \\
 r &= 2\mu_2/\mu_3, \quad ra = 4(\mu_2^3/\mu_3^2) - 1
 \end{aligned}
 \quad \left. \begin{array}{l} \\ \end{array} \right\} \dots\dots\dots (8)$$

c) VII 型

$$\left. \begin{aligned} y_0 &= 1 / (\sqrt{2\pi} \cdot \sigma) \\ \sigma &= \mu_2^{1/2} \end{aligned} \right\} \dots\dots\dots (9)$$

Pearson による上記の積率法では、小標本に対して推定の有効性が満たされないことに注意すべきである。

1-3. 正規化

水文量に限らず統計量の確率分布は正規分布に従わない場合が多い。しかるに確率・統計理論は一般に変量の正規性を前提としているから、水文量の頻度解析や時系列モデルによる統計解析においては、それらの理論の応用には限界がある。この制約を直接避けるために用いられる有力な方法が変数の正規化である。正規化の方法には、大別して確率変数の関数変換による方法と経験的変換法とがある。

a) 関数変換による正規化

これは正規分布に従わない確率変数 x を

$$y = f(x)$$

によって変換したとき、変換変数 y が正規分布に従うように関数 $f(x)$ を決定する方法である。このとき、 x の確率密度関数 $p(x)$ は次式で与えられる。

$$p(x) = p(y) \frac{dy}{dx} = \frac{1}{\sqrt{2\pi} \sigma_y} \exp \left\{ -\frac{(f(x) - \mu_y)^2}{2\sigma_y^2} \right\} d(f(x))/dx \quad (10)$$

ここに μ_y, σ_y^2 はそれぞれ変換値 $y = f(x)$ についての平均値、分散である。

一般的な関数変換の型式は次の通りである。

$$y = f(x) = r + \delta \cdot g \left(\frac{x - \mu}{\lambda} \right) \quad \dots\dots\dots (11)$$

あるいは

$$\frac{y - r}{\delta} = g \left(\frac{x - \mu}{\lambda} \right) \quad \dots\dots\dots (12)$$

ここに、 r, μ は位置パラメータ、 λ, δ は尺度パラメータ、 g は関数である。これらのパラメータと関数 g を決定するための実用的な方法は次の通りである。

まず、 $g(x)$ の型を選ぶために、各種の確率紙に資料をプロットし、プロットされた点の直線性から関数形を推定する。

つぎに(10)式から x の積率はパラメータ、 $\mu_y, \sigma_y, r, \mu, \lambda, \delta$ の関数として表わされる。すなわち

$$\mu'_{x,i} \text{ or } \mu_{x,i} = f_i(\mu_y, \sigma_y, r, \mu, \lambda, \delta) \quad \dots\dots\dots (13)$$

$$(i = 1, 2, \dots\dots\dots n)$$

ここに、 $\mu'_{x,i}, \mu_{x,i}$ はそれぞれ x の原点のまわりの i 次の積率、平均値のまわりの i 次の積率であり、 n は未知パラメータ (μ_y, σ_y など) の数である。(13)式の左辺の積率を標本積率と置いてこの式を解けば、パラメータ、 $\mu_y, \sigma_y, r, \mu, \lambda, \delta$ が決定される。

以上の手順によって変数変換式(11) (または(12)) が定まれば、原標本 $\{x\}$ の個々の値をこの式を用いて変換した資料 $\{y\}$ は平均値 μ_y 、分散 σ_y^2 の正規分布に従うことが保証される。

さて、正規化のために最もよく用いられる変換関数は次のような対数変換式、およびべき乗変換式で

あろう。

$$y = a + b \log(x + c) \quad \dots\dots\dots (14)$$

$$y = a(x + b)^c \quad \dots\dots\dots (15)$$

変換式を求める場合に注意すべきは、含まれるパラメータ(上式では a, b, c)の数を増すことが必ずしも意味のあることではない点である。上述の積率法からもわかる通り、パラメータの増加とともに高次の積率の値が必要となるが、通常の標本の大きさでは、高次になるにしたがい、その推定の誤差が急激に増加するからである。一方、最も簡単な場合として変換式にパラメータを含まない場合(たとえば上式で $y = \log x, y = x^{1/3}$)には、 x の確率密度関数は2つのパラメータ、すなわち変換変数の平均値と分散のみによって規定され、これらを標本から推定するには既存の標準的理論が使えるという利点がある。

水量にしばしば適用されるピアソンⅢ分布(ガンマ分布)に対しては、次式のWilson-Hilferty変換がある。(誘導は後述する。)

$$y = \frac{6}{\beta_1} \left\{ \left(\frac{\beta_1}{2} x + 1 \right)^{1/3} + \frac{\beta_1^2}{36} - 1 \right\} \quad \dots\dots\dots (16)$$

b) 経験的変換法

正規型とみられない変数 x について、 N 個の観測値を大きさの順にならべて、

$$x_{(1)} \leq x_{(2)} \leq x_{(3)} \quad \dots\dots\dots \leq x_{(N-1)} \leq x_{(N)}$$

とするとき、これらを標準化正規変量 $t(N(0, 1))$ の値 $t_1, t_2, \dots\dots\dots, t_{N-1}, t_N$ に1対1変換を行なうためには、各々の $x_{(i)}$ ($i = 1, 2, \dots\dots\dots N$)を次式の関係

$$\frac{i}{N+1} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t_i} \exp\left(-\frac{u^2}{2}\right) du \quad \dots\dots\dots (17)$$

$$(i = 1, 2, \dots\dots\dots, N)$$

を用いて、 t_i ($i = 1, 2, \dots\dots\dots, N$)にそれぞれ変換すればよい。上式の左辺はデータのplotting position法として周知の通り、任意の確率分布に従う変数 x の実現値が N 個与えられたとき、ある x が i 番目の順序標本値 $x_{(i)}$ 以下となる確率を順序統計量に関する理論から得たものである。

2. 対数正規分布

2-1. 一般的特性

対数正規分布は代表的な非対称確率分布の1つであり、その形式の分類、特性および実用上の問題等は文献⁽⁹⁾に詳しい。水量については種々の形式の対数正規分布のうち、次の形式が最も有用である。すなわち確率変数を x とすると、次式による変換量 y 、

$$y = \log(x - a) \quad \dots\dots\dots (18)$$

が正規分布となる場合である。この形式は3-パラメータ対数正規分布と呼ばれ、次式

$$y = \log x \quad \dots\dots\dots (19)$$

による形式(これを2-パラメータ対数正規分布と呼ぶ)は、(18式)において $a = 0$ とする特別の場合である。

(a) 2-パラメータ型

まず対数正規分布のうち最も簡単な2-パラメータ型(式(19))についてその特性を要約すれば次の通

りである。

式(19)の変数 x の確率密度関数は、

$$f(x) = \frac{1}{x \sigma_y \sqrt{2\pi}} \exp \left[\frac{(\log x - \mu_y)^2}{2 \sigma_y^2} \right] \quad \dots\dots\dots (20)$$

で表わされる。ここに、 μ_y , σ_y はそれぞれ $\log x$ についての平均値、標準偏差である。

変数 x についての平均値 μ_x , 標準偏差 σ_x と μ_y , σ_y との関係は次の通りである。

x についての原点のまわりの r 次の積率は

$$\mu_r' = \exp(\tau \mu_y + r^2 \sigma_y^2 / 2) \quad \dots\dots\dots (21)$$

であるから、平均値 μ_x , 分散 σ_x^2 はそれぞれ次のようになる。

$$\mu_x = \exp(\mu_y + \sigma_y^2 / 2) \quad \dots\dots\dots (22)$$

$$\left. \begin{aligned} \sigma_x^2 &= \mu_x^2 \{ \exp(\sigma_y^2) - 1 \} \\ &= \exp(2\mu_y + \sigma_y^2) \{ \exp(\sigma_y^2) - 1 \} \end{aligned} \right\} \quad \dots\dots\dots (23)$$

変動係数は

$$\eta_x = \sigma_x / \mu_x = (\exp(\sigma_y^2) - 1)^{1/2} \quad \dots\dots\dots (24)$$

また、歪み係数 β_1 , 尖り係数 β_2 はそれぞれ次の通りである。

$$\begin{aligned} \beta_1 &= \mu_3 / \sigma_x^3 = (\mu_3' - 3\mu_1' \mu_2' + 2\mu_1'^3) / \sigma_x^3 \\ &= \eta_x^3 + 3\eta_x = \frac{\exp(3\sigma_y^2) - 3\exp(\sigma_y^2) + 2}{\{\exp(\sigma_y^2) - 1\}^{3/2}} \end{aligned} \quad \dots\dots\dots (25)$$

$$\left. \begin{aligned} \beta_2 &= \mu_4 / \sigma_x^4 = (\mu_4' - 4\mu_3' \mu_1' + 6\mu_2' \mu_1'^2 - 3\mu_1'^4) / \sigma_x^4 \\ &= \eta_x^8 + 6\eta_x^6 + 15\eta_x^4 + 16\eta_x^2 + 3 \end{aligned} \right\} \quad \dots\dots\dots (26)$$

また y は正規分布にしたがうから、 y の median は $\zeta_y = \mu_y$ であり、 x の median, ζ_x とは $\zeta_y = \log \zeta_x$ の関係があるから

$$\zeta_x = \exp(\mu_y) \quad \dots\dots\dots (27)$$

したがって

$$\mu_x / \zeta_x = \exp(\sigma_y^2 / 2) \quad \dots\dots\dots (28)$$

上式のごとく η_x , β_1 , β_2 , μ_x / ζ_x は σ_y のみの関数であり、 σ_y が x の分布の変動を規定するパラメータとなることから σ_y はしばしば "変動の指標" (index of variability) とされる。

mode, median, 平均値 (mean) はそれぞれ次のように与えられる。

$$\left. \begin{aligned} \text{mode} &; \log_{10} x = \xi - 2.3026 \sigma^2 \\ \text{median} &; \log_{10} x = \xi \\ \text{mean} &; \log_{10} x = \xi + 1.1513 \sigma^2 \end{aligned} \right\} \quad \dots\dots\dots (29)$$

ここに ξ , σ^2 はそれぞれ $\log_{10} x$ についての平均値、分散である。

上記の2つのパラメータ μ_y , σ_y の推定法には、最尤推定法、積率法、quantiles法、図式法等がある。

(b) 3-パラメータ型

2-パラメータの場合と同様にして、パラメータの間には次の関係が与えられる。

$$\mu_x = a + \exp \left[\mu_y + \sigma_y^2 / 2 \right] \quad \dots\dots\dots (30)$$

$$\sigma_x^2 = \exp (2 \mu_y + \sigma_y^2) \{ \exp (\sigma_y^2) - 1 \} \quad \dots\dots\dots (31)$$

$$\beta_1 = \frac{\exp [3 \sigma_y^2] - 3 \exp [\sigma_y^2] + 2}{\{ \exp [\sigma_y^2] - 1 \}^{3/2}} \quad \dots\dots\dots (32)$$

$$\left. \begin{array}{l} \text{mode} \quad ; \quad \log_{10} (x - a) = \xi - 2.3026 \sigma^2 \\ \text{median} ; \quad \log_{10} (x - a) = \xi \\ \text{mean} \quad ; \quad \log_{10} (x - a) = \xi + 1.1513 \sigma^2 \end{array} \right\} \quad \dots\dots\dots (33)$$

ここに、 ξ 、 σ^2 はそれぞれ $\log_{10} (x - a)$ についての平均値、分散である。

また、 x の median を ζ_x で表わせば

$$\zeta_x - a = \exp (\mu_y) \quad \dots\dots\dots (34)$$

2-パラメータの場合の、歪み係数と変動係数との関係式(25)に対応する関係は、変数 x に関して次の通りである。

$$(i) \quad \beta_1 > \eta_x^3 + 3 \eta_x \quad a > 0$$

$$(ii) \quad \beta_1 = \eta_x^3 + 3 \eta_x \quad a = 0$$

$$(iii) \quad \beta_1 < \eta_x^3 + 3 \eta_x \quad a < 0$$

すなわち、正の歪みをもつ対数正規分布において、対数確率紙上で、上に凹、下に凹のプロットが得られる場合にはそれぞれ(i)、(iii)の関係があり、直線上にプロットされる場合は(ii)の関係が成立する。したがってまず、原資料を対数確率紙にプロットすることによって、あらかじめ a の符号を知ることができる。

極値の確率等に適用される場合には、パラメータが推定されれば、任意の超過（または非超過）確率に対応する値は次式から計算できる。

$$\log (x - a) = \langle \log (x - a) \rangle + t \langle S_{\log (x - a)} \rangle \quad \dots\dots\dots (35)$$

ここに、 $\langle \log (x - a) \rangle$ 、 $\langle S_{\log (x - a)} \rangle$ はそれぞれ、 $\log (x - a)$ についての平均値、標準偏差であり、 t は frequency factor ($N(0, 1)$) である。

この分布では平均値、標準偏差に加えて式(18)に含まれるパラメータ a の推定が必要であるが、これらのパラメータ推定法は標本の平均値、標準偏差、歪み係数を用いる積率法が従来一般的であり、石原、高瀬による方法⁽¹⁰⁾、Matalasによる方法⁽¹¹⁾等がある。また、 a の推定に両端部の経験的確率を利用する岩井法⁽¹²⁾や角屋による改良法⁽¹³⁾がある。これらの推定法における最大の関心は歪み係数の推定であって、とくに水文量のように観測資料の数が少ない場合にはその標本誤差が大きくなることである。標本の大きさが100以下になれば歪み係数の推定は悲観的であるという報告がある。

一方、洪水流量のような極値ではなく低水流況を対象とする場合には、水文過程の持続性を再現する stochastic model を構成する必要がある。このような時系列モデルには、上述のパラメータの他に自己相関係数が含まれるから、その適切な値をモデルに導入しなければならない。

以下に、これらの問題に関する2、3の方法の概要を述べる。

2-2. パラメータ推定法

A. 積率変換法

変数 x が3-パラメータ対数正規分布に従い、その時系列が1次のマルコフ過程で表わされる場合について、Matalas⁽¹⁴⁾による資料の模擬発生法は次の通りである。

(18)式による変換変数 y が次式の 1 次マルコフモデルで表わされるものとする。

$$(y_{i+1} - \mu_y) = \rho_y(1) (y_i - \mu_y) + [1 - \rho_y^2(1)]^{1/2} \sigma_y \varepsilon_{i+1} \quad \dots\dots\dots (36)$$

ここに、 $\rho_y(1)$ は y についての lag 1 の自己相関係数、 ε_{i+1} は独立な正規標準化変量 ($N(0, 1)$) である。

x に関しては模擬発生モデルは次式で表わされる。

$$x_{i+1} = a + \{ \exp[\mu_y(1 - \rho_y(1))] \} (x_i - a)^{\rho_y(1)} \delta_{i+1} \quad \dots\dots\dots (37)$$

ここに

$$\delta_{i+1} = \exp \{ [1 - \rho_y^2(1)]^{1/2} \sigma_y \varepsilon_{i+1} \} \quad \dots\dots\dots (38)$$

x_i と δ_{i+1} が独立であるという仮定の下では

$$\rho_x(1) = \{ \exp[\sigma_y^2 \rho_y(1)] - 1 \} / \{ \exp[\sigma_y^2] - 1 \} \quad \dots\dots\dots (39)$$

となる。

x に関する平均値 μ_x ，標準偏差 σ_x ，歪み係数 β_1 および自己相関係数 $\rho_x(1)$ の推定値を原資料から得れば，(30)，(31)，(32)，(39)式から a ， μ_y ， σ_y ， $\rho_y(1)$ が求められる。この μ_y ， σ_y ， $\rho_y(1)$ を用いて(36)式から y 一時系列が発生でき，次式の逆変換

$$x = a + \exp(y)$$

から， x 一時系列が発生できる。この x の系列は原資料の μ_x ， σ_x ， β_1 ， $\rho_x(1)$ を再現しうる。

従来は，原資料 (x 一時系列) の対数値をとり，その対数変換資料から a ， μ_y ， σ_y ， $\rho_y(1)$ を求める近似法が用いられてきたが，この近似法では，発生された x 一時系列の μ_x ， σ_x ， β_x ， $\rho_x(1)$ は観測資料の値を再現しないことに注意しなければならない。Matalas⁽¹⁵⁾によって示された上記の方法と近似法とによる発生資料の比較がBurgess⁽¹⁶⁾やCodner & MacMahon⁽¹⁶⁾によって行なわれており，両者による相異は貯水池計画等に相当な影響を及ぼすことが示されている。

B. パラメータ a の推定について

前述の通りパラメータ推定において歪み係数の推定には大きな標本誤差が伴う。この点から，Sangal & Biswas⁽¹⁷⁾は歪み係数のかわりにmedianを用いる次の方法を提案している。

(18)式における x についての平均値 μ_x ，分散 σ_x^2 ，と変換変数 y の平均値 μ_y ，分散 σ_y^2 ，パラメータ a との間には前述の通り(30)，(31)式の関係がある。また x の median は(34)式で表わされる。

(30)，(31)，(34)式から次式が得られる。

$$\begin{aligned} & 2a^3(\mu_x - \zeta_x) + a^2(\sigma_x^2 + \zeta_x^2 - 5\mu_x^2 + 4\mu_x\zeta_x) + 2a(2\mu_x^3 - \zeta_x\sigma_x^2 \\ & - \zeta_x\mu_x^2 - \mu_x\zeta_x^2) + \zeta_x^2\sigma_x^2 - \mu_x^4 + \mu_x^2\zeta_x^2 = 0 \quad \dots\dots\dots (40) \end{aligned}$$

上式がパラメータ a を μ_x ， ζ_x ， σ_x に関係づける基本式である。この式は a に関する 3 次式であるが，1 つの実根と 2 つの虚根が存在することが証明できる。

σ_y が微小であると仮定すれば (ほとんどの水文資料でこの仮定が許される)， a に関して近似的であるが簡単な表現が得られる。すなわち，(30)，(31)式において σ_y の高次の項を無視すれば，

$$\mu_x = a + (1 + \sigma_y^2/2) \exp \mu_y \quad \dots\dots\dots (41)$$

$$\sigma_x^2 = \exp(2\mu_y) \cdot \sigma_y^2 \quad \dots\dots\dots (42)$$

(34)式を(41), (42)式に代入して σ_y^2 を消去すれば

$$a = \zeta_x - \frac{\sigma_x^2}{2(\mu_x - \zeta_x)} \quad \dots\dots\dots (43)$$

この式によってパラメータ a を決定することができる。上式によって a を推定する場合、 μ_x, σ_x^2 は標本積率を用いればよいか、median, ζ_x については小標本の場合は超過確率50%の値は信頼性が低いので、資料のmiddle fifthの平均値、すなわち超過確率40%から60%の資料の平均値として推定することができる。

(43)式によれば、medianが平均値に近い値の場合に a の値は敏感に変化するので、このような場合には次のような方法を試みることができる。(1) 資料を正規確率紙にプロットすれば、多くの場合良い近似で直線上に乗る。(2) medianの値を平均値よりも小さいが、ごく近い値、たとえば平均値の99%の値に仮定して a を求める。

Sangal & Biswas は、以上のように与えられる3-パラメータ対数正規分布およびその他の確率分布を、年流量、月流量に適用して、各分布の適合性を比較するとともに、パラメータ a の推定に関する上記の方法が良好な結果を与えることを確かめている。

2-3. 時系列モデルの自己相関特性

3-パラメータ対数正規分布において(39)式がlag 1の自己相関係数について得られるのに対して、Mejia & Rodrigues-Iturbe⁽¹⁸⁾は全てのlagについて成立する次式を誘導している。

$$\rho_x(\tau) = \frac{\exp(\sigma_y^2 \rho_y^{|\tau|}(1)) - 1}{\exp(\sigma_y^2) - 1} \quad \dots\dots\dots (44)$$

したがって前述のMatalasの方法で発生した x -時系列の自己相関構造は(44)式のようになるはずである。

(44)式を級数展開すると

$$\rho_x(\tau) = \frac{1}{[\exp(\sigma_y^2) - 1]} \sum_{i=1}^{\infty} \frac{[\sigma_y^2 \rho_y^{|\tau|}(1)]^i}{i!}$$

となる。 τ の大きな値に対して上式は

$$\rho_x(\tau) = \frac{\sigma_y^2 \rho_y^{|\tau|}(1)}{\exp(\sigma_y^2) - 1} \quad \dots\dots\dots (45)$$

となり、これはARMA(1, 1)-model (Autoregressive Moving Average)の相関構造に近似する。

原時系列がマルコフ過程であるという仮説のもとでは、(44)式で与えられる相関構造は1より大きなlagに対して歪んでおり、その数値例は表-3のごとくである。この表は原時系列がマルコフ過程である場合の相関係数の理論値と、 y を正規マルコフ過程としてMatalasの方法で x -時系列に変換したときに得られる相関係数の値とを比較したものである。ただし、 $\rho_x(1) = 0.2$, $\sigma_y = 0.3$ である。表から明らかに、 $\rho_x(1)$ のみの再現しか保証しない通常の方法では、lagの増加にともなう相関係数の減衰がかなり緩やかになる。

そこで原時系列の自己相関構造を再現するためには、次の相関係数 $\rho_y(\tau)$ を用いて正規過程 y を発生すべきである。

$$\rho_y(\tau) = \log \{ 1 + \rho_x(\tau) [\exp(\sigma_y^2) - 1] \} / \sigma_y^2 \quad \dots\dots\dots (46)$$

表-3 Values for the Correlation of a
Log Normally Distributed Process for
Different Generating Procedures

Lag	Theoretical	$\rho_s(\tau)^*$	$\rho_x(\tau)^{**}$
1	2.00×10^{-1}	2.02×10^{-1}	2.02×10^{-1}
2	4.00×10^{-2}	4.12×10^{-2}	4.01×10^{-2}
3	8.00×10^{-3}	8.52×10^{-3}	8.00×10^{-3}
4	1.60×10^{-3}	1.77×10^{-3}	1.60×10^{-3}
5	3.20×10^{-4}	3.66×10^{-3}	3.20×10^{-4}
10	1.02×10^{-7}	1.40×10^{-3}	1.02×10^{-7}
15	3.28×10^{-11}	5.31×10^{-3}	3.28×10^{-11}
50	1.12×10^{-35}	6.50×10^{-3}	1.12×10^{-35}
100	1.27×10^{-70}	4.64×10^{-3}	1.27×10^{-70}

* Values of the correlation function when the original process is generated from the transformation $x = e^y + a$ of a Markovian process preserving the first serial correlation coefficient of the original data.

** Values of the correlation function when the original process is generated from the transformation $x = e^y + a$ of an ARMA(1, 1) process with parameters $\rho_x(1) = 0.2$ and $\sigma_y = 0.3$.

ここで、 $\rho_x(\tau)$ は原時系列の自己相関係数である。このような一般的な自己相関構造を再現するための方法が Mejia & Rodrigues-Iturbe¹⁹⁾によって提案されている。

原時系列がマルコフ型である場合には、(46)式は

$$\rho_y(\tau) = \log \{ 1 + \rho_x^{\tau+1}(1) [\exp(\sigma_y^2) - 1] / \sigma_y^2 \} \dots\dots\dots (47)$$

となり、この式は漸近的に次式で表わされる。

$$\rho_y(\tau) = \rho_x^{\tau+1}(1) [\exp(\sigma_y^2) - 1] / \sigma_y^2 \dots\dots\dots (48)$$

この式を用いるのが y -時系列を発生するための近似的方法である。

$\rho_x(1)$, σ_y が与えられた時系列について、ARMA(1, 1)-modelを用いれば、(48)式の $\rho_y(\tau)$ を再現することができる。したがって $x = \exp(y) + a$ の自己相関構造は

$$\rho_x(\tau) = \frac{\exp[\sigma_y^2 \cdot \rho_y(\tau)] - 1}{\exp(\sigma_y^2) - 1} \dots\dots\dots (49)$$

となる。ここに $\rho_y(\tau)$ は(48)式で与えられる。

表-3の最右欄はこの方法による $\rho_x(\tau)$ の値であり、自己相関係数の減衰特性の保持はきわめて良好である。なお、大きな τ の値に対する $\rho_x(\tau)$ の値は非常に小さいが、 $\rho_x(\tau)$ の減衰率は range の特性とくに Hurst's H に重要な影響を及ぼすことに注意すべきである。

lag-1 の自己相関パラメータの推定法に関しては Frost & Clarke²⁰⁾ は最尤推定法を用いた改良手法を提示している。

3. ピアソン-III型分布

3-1. パラメータ推定法

ピアソン-III型分布はガンマ分布とも呼ばれ、1-パラメータ(4式)、2-パラメータ(3式)および3-パラメータの分布型がある。より一般的な3-パラメータ分布の確率密度関数は次式で表わされる。

$$f(x) = \frac{1}{\Gamma(b+1)} \left(\frac{x-m}{a}\right)^b \exp\left[-\left(\frac{x-m}{a}\right)\right] \quad \dots\dots\dots (50)$$

ここに、 m は位置パラメータ(x の下限境界)、 a は尺度パラメータ、 b は形状パラメータであり、 $m=0$ の場合が2-パラメータ分布である。分布が正に歪む場合には $a>0$ 、 $x \geq m$ であり、負の歪みの場合にはその逆である。

m 、 a 、 b と x の積率とは次式の関係がある。

$$\mu_x = m + a(b+1) \quad \dots\dots\dots (51)$$

$$\sigma_x = a(b+1)^{1/2} \quad \dots\dots\dots (52)$$

$$\beta_1 = 2/(b+1)^{1/2} \quad \dots\dots\dots (53)$$

$$\beta_2 = 6/(b+1) + 3 \quad \dots\dots\dots (54)$$

これらのパラメータの推定法には積率法と最尤法がある。

A. 積率法

標本から計算した積率推定値、 $\bar{\mu}_x$ 、 $\bar{\sigma}_x$ 、 $\bar{\beta}_1$ および(51)、(52)、(53)式からパラメータ m 、 a 、 b の推定値を得る。 m 、 a 、 b の推定値は $\bar{\mu}_x$ 、 $\bar{\sigma}_x$ および $\bar{\beta}_2$ から求めることもできるが、 $\bar{\beta}_2$ の推定誤差が大きいため $\bar{\beta}_2$ は用いられない。

B. 最尤推定法

尤度関数 $L(x)$ は次式で定義される。

$$L(x) = \sum_{i=1}^N \ln f(x_i)$$

ここに、 $f(x_i)$ は $x=x_i$ に対する $f(x)$ の値を表わし、 N は標本の大きさである。

m 、 a 、 b の最尤推定値は次の3つの非線形方程式を連立に解けば決定できる。

$$\left. \begin{aligned} \frac{\partial L}{\partial m} &= -b \sum_{i=1}^N \left\{ \frac{1}{(x_i - m)} \right\} + \frac{N}{a} = 0 \\ \frac{\partial L}{\partial a} &= -\frac{N}{a} (b+1) + \frac{1}{a^2} \sum_{i=1}^N (x_i - m) = 0 \\ \frac{\partial L}{\partial b} &= -N \frac{d \ln \Gamma(b+1)}{db} + \sum_{i=1}^N \ln \left(\frac{x_i - m}{a} \right) = 0 \end{aligned} \right\} \quad \dots\dots\dots (55)$$

得られた推定値 m 、 a 、 b と(51)、(52)、(53)式から μ_x 、 σ_x 、 β_1 の推定値が求められる。

積率法と最尤法の優劣に関しては、前者を主唱するPearsonと後者を主唱するFisherの議論以来、約半世紀が過ぎた。少なくとも有効性に関して最尤法が積率法より優れることが認められているけれども、水文学の分野での最尤法の利用はきわめて少ない。この主たる理由は最尤法における計算の困難さにあったが、今や高性能の計算機の出現によってその解を比較的容易に得ることができるようになった。

Matalas & Wallis⁽²⁾は、ピアソン-III型分布について、計算機シミュレーションによって両者の方法の適合性を検討し、以下のような結果を得ている。

もし、 μ_x 、 σ_x 、 β_1 の推定のみに興味をもつとすれば、最尤法よりもむしろ積率法の方が便利である

かもしれない。なぜならば、 μ_x, σ_x, β_1 の積率推定には確率分布関数の仮定を要しないからである。しかしながら、主要な関心は、 μ_x, σ_x, β_1 ではなく種々の確率レベル、 P に対する x の値 $x(P)$ 、あるいは lower bound、 m の値の推定にある。このような推定値に対しては、積率法、最尤法のいずれでも分布関数の仮定が必要である。

さて、最尤法は積率法よりもパラメータ、 m, a, b および $x(P)$ の推定値の偏り(bias)と変動(variation)において優れており、とくに小標本の場合、 P が $N/(N+1)$ よりも大きくなるとともに、その傾向は顕著になる。

積率法では $x(P)$ の推定値の偏りは標本の大きさ N の増加とともに減少し、また同一の N に対しては Return Period、 $T(=1/P)$ の増加とともに増加する。したがって T —年確率の水文量の推定値は事実上 $T' < T$ のときの水文量の推定値である。

C. 最小二乗法

Snyder⁽²²⁾ はパラメータの推定に非線形最小二乗法を用いる方法を報告している。この方法によれば通常適用される分布形に対しては共通の手法でパラメータを適合させることができるとして、ピアソン—III型分布、Gumbel分布、正規分布について積率法との比較を行なっている。またこの方法では、plotting position や異常標本(sample outlier)に関する問題も消滅すると述べている。

3-2. Wilson—Hilferty変換とその修正法

ピアソン—III型分布は対数正規分布とともに種々の水文量の頻度分布に適合する代表的な非対称分布であるが、観測値のプロットが確率紙上で直線に乗るように変量のスケールを変換することは容易でないために、その実用性が対数正規分布に比べて若干劣ることは確かである。この分布に従う確率変量の模擬発生法には、正規マルコフ過程の m 項の二乗和から発生させるYevdjovich⁽²³⁾の方法があるが歪み係数の選択に制限がある。Thomas & Fiering⁽²⁴⁾ は1次の自己回帰型モデル、

$$x_{i+1} = \rho_x(1)x_i + \xi_{i+1}\sigma_x(1 - \rho_x(1))^{1/2} \quad \dots\dots\dots (56)$$

において、ピアソン—III型分布にしたがう変量 x を発生するために、 ξ_{i+1} に次式で定義される量を用いた。

$$\xi_{i+1} = \frac{2}{\beta_\xi} \left\{ \left(1 + \frac{\beta_\xi t_{i+1}}{6} - \frac{\beta_\xi^2}{36} \right)^3 \right\} \quad \dots\dots\dots (57)$$

ここに、 β_ξ は ξ の歪み係数で、 x の歪み係数 β_x と次式で関係づけられる。

$$\beta_\xi = \frac{1 - \rho_x^3(1)}{\{1 - \rho_x^2(1)\}^{3/2}} \beta_x \quad \dots\dots\dots (58)$$

56式において、 t_{i+1} が $N(0, 1)$ の正規分布に従うときには ξ_{i+1} は近似的に、平均値が0、分散が1、歪み係数が β_ξ のピアソン—III型分布に従うことがわかっている。56式の偶発成分にこの ξ_{i+1} を用いることによって、1次のマルコフ過程は3次の定常性をもつから、 $\mu_x, \sigma_x, \beta_x, \rho_x(1)$ をよく再現する x —時系列を模擬発生できる。

57式はWilson—Hilferty変換⁽²⁵⁾と呼ばれ次のように誘導される。

1—パラメータ・ピアソン—III型分布

$$f(x) = \frac{1}{\Gamma(\lambda)} x^{\lambda-1} e^{-x} \quad \dots\dots\dots (59)$$

にしたがう変数 x について、Wilson & Hilferty は χ^2 —分布の性質を用いて、次式の変換、

$$y = (x/\lambda)^{1/3} \quad \dots\dots\dots (60)$$

による y が次の平均値, 分散をもつ正規分布に近似的にしたがうことを導いた。

$$\mu_y = 1 - 1/(9\lambda) \quad \dots\dots\dots (61)$$

$$\sigma_y^2 = 1/(9\lambda) \quad \dots\dots\dots (62)$$

したがって

$$t = \frac{(\frac{x}{\lambda})^{1/3} - (1 - \frac{1}{9\lambda})}{(\frac{1}{9\lambda})^{1/2}} = 3\lambda^{1/2} \{ (\frac{x}{\lambda})^{1/3} + \frac{1}{9\lambda} - 1 \} \quad \dots\dots\dots (63)$$

による t は充分の近似度で $N(0, 1)$ にしたがう。59式の分布について, 平均値, 標準偏差, 歪み係数はそれぞれ

$$\mu_x = \lambda, \quad \sigma_x = \lambda^{1/2}, \quad \beta_1 = 2/\lambda^{1/2} \quad \dots\dots\dots (64)$$

で与えられ, また x の規準化変量を

$$\xi = (x - \mu_x)/\sigma_x \quad \dots\dots\dots (65)$$

とすれば, 64, 65式を63式に代入すれば次式を得る。

$$t = \frac{6}{\beta_1} \left\{ \left(\frac{\beta_1}{2} \xi + 1 \right)^{1/3} + \frac{\beta_1^2}{36} - 1 \right\} \quad \dots\dots\dots (66)$$

上式は57式と同一である。

66式は前述(1-2)のごとく, ピアソン-III型分布(ガンマ分布)にしたがう確率変数の正規化変換に用いることができる。

さて, この変換式57を用いた59式のモデルは, ピアソン-III型分布にしたがう水文量に対して広く用いられてきたが,⁶⁶⁾ 歪み係数の値が大きい場合には観測値の積率を精度よく再現しないことが見出された。すなわち, 57式において β_ξ が増せば, ξ_{i+1} がガンマ変量の理論的下限值よりも小さな値をとることが多くなるためであり, この影響は β_ξ が 3.0 ~ 4.0 より大きな値で顕著になる。

この欠点を補正するために, Kirby⁶⁷⁾ は57式における定数を β_ξ に関係するパラメータに置きかえて次の変換式を提示した。

$$\xi^M = A \left\{ \max [H, 1 - (G/6)^2 + (G/6)\zeta]^3 - B \right\} \quad \dots\dots\dots (67)$$

ここに, A, B, G, H は ξ^M が規準化され, 所要の β_ξ および正しい下限値 $(-2/\beta_\xi)$ をもつように選ばれたパラメータで, その数値は表-4および

$$H = \left\{ B - \left[(2/\beta_\xi)/A \right] \right\}^{1/3} \quad \dots\dots\dots (68)$$

で与えられる。表-4で線形の内挿は精度が落ちるので次の近似式を用いる。

$$A = \max (2/\beta_\xi, 0.40) \quad \dots\dots\dots (69)$$

$$B = 1 + 0.0144 \max (0, \beta_\xi - 2.25)^2 \quad \dots\dots\dots (70)$$

$$G = \beta_\xi - 0.063 \max (0, \beta_\xi - 1)^{1.85} \quad \dots\dots\dots (71)$$

この修正Wilson-Hilferty法によれば, 少なくとも $\beta_\xi \leq 9.0$ の範囲でピアソン-III型分布への近似は良好であることをKirbyは確かめている。ただし, この方法では計算がかなり面倒になるので適切な計算機サブルーチンを用意しておく必要があろう。

表-4 Parameters of the Modified Wilson-Hilferty Transformation

Skew	G	A	B
0. 25000	0. 24986	8. 00461	1. 00000
0. 50000	0. 49886	4. 00916	1. 00000
0. 75000	0. 74624	2. 68022	1. 00000
1. 00000	0. 99133	2. 01775	0. 99998
1. 25000	1. 23360	1. 62176	0. 99993
1. 50000	1. 47245	1. 35917	0. 99981
1. 75000	1. 70695	1. 17326	0. 99967
2. 00000	1. 93608	1. 03571	0. 99968
2. 25000	2. 15899	0. 93062	1. 00012
2. 50000	2. 37513	0. 84832	1. 00133
2. 75000	2. 58418	0. 78258	1. 00371
3. 00000	2. 78602	0. 72920	1. 00764
3. 25000	2. 98071	0. 68526	1. 01342
3. 50000	3. 16776	0. 64876	1. 02133
3. 75000	3. 34819	0. 61802	1. 03166
4. 00000	3. 52191	0. 59193	1. 04453
4. 25000	3. 68922	0. 56962	1. 06008
4. 50000	3. 85037	0. 55041	1. 07836
4. 75000	4. 00587	0. 53375	1. 09935
5. 00000	4. 15577	0. 51924	1. 12311
5. 25000	4. 30047	0. 50655	1. 14956
5. 50000	4. 43948	0. 49549	1. 17893
5. 75000	4. 57429	0. 48567	1. 21071
6. 00000	4. 70462	0. 47699	1. 24499
6. 25000	4. 83073	0. 46929	1. 28169
6. 50000	4. 95285	0. 46244	1. 32069
6. 75000	5. 07113	0. 45635	1. 36192
7. 00000	5. 18586	0. 45091	1. 40523
7. 25000	5. 29718	0. 44605	1. 45053
7. 50000	5. 40527	0. 44170	1. 49773
7. 75000	5. 51028	0. 43781	1. 54671
8. 00000	5. 61238	0. 43434	1. 59738
8. 25000	5. 71171	0. 43123	1. 64964
8. 50000	5. 80843	0. 42844	1. 70335
8. 75000	5. 90259	0. 42596	1. 75852
9. 00000	5. 99436	0. 42375	1. 81500
9. 25000	6. 08388	0. 42178	1. 87268
9. 50000	6. 17109	0. 42004	1. 93162
9. 75000	6. 25601	0. 41853	1. 99199

4. 計算機実験による標本統計量の特徴

われわれが取扱う観測資料は1組の標本にすぎないから、その資料からは母集団のパラメータの1つの推定値が得られるだけである。大きさ n の標本における、母集団平均値、標準偏差、歪み係数の推定値をそれぞれ $\bar{X}$, S , G で表わせば、 n の関数である $\bar{X}$, S , G の標本特性値は一般に、推定法、周辺分布関数、確率過程の発生機構に関係する。Wallis, Matalas, Slack⁽²⁸⁾は、推定法として積率法を用い、 n 個の変数は独立かつ均質であるという仮定の下に、モンテ・カルロ法によって、 $\bar{X}$, S , G の

の標本特性を数値実験的に求めた。確率分布関数は次の6つの分布である。

- (i) 正規分布 (ii) 3-パラメータ・対数正規分布 (iii) 3-パラメータ・ピアソン-III型分布
(iv) Gumbel分布 (v) Weibull分布 (vi) Pareto (ピアソン-IV型) 分布

大きさ n の標本 $Y_1, Y_2, \dots, Y_n$ について、平均値 $\bar{X}$ 、標準偏差 S 、歪み係数 G を次式で定義する。

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n Y_i \quad \dots\dots\dots (72)$$

$$S = \left[\sum_{i=1}^n Y_i^2 / n - \bar{Y}^2 \right]^{1/2} \quad \dots\dots\dots (73)$$

$$G = \left[\sum_{i=1}^n Y_i^3 / n - 3 \bar{Y} S^2 - \bar{Y}^3 \right] / S^3 \quad \dots\dots\dots (74)$$

100,000組の、大きさ n の標本から、 $\bar{X}$ 、 S 、 G の各々について分布関数および平均値、標準誤差、歪み係数、尖り係数を求め、その結果から、 $\bar{X}$ 、 S 、 G の分布が標本の大きさ n 、用いる確率分布関数 $f(Y)$ 、およびその歪み係数 $r(Y)$ とどのような関係にあるかが明らかにされた。

母数の推定に関しては、 $\bar{X}$ は母集団の $\mu(Y)$ の不偏推定値であるが、 S および G は母集団の $\sigma(Y)$ 、 $r(Y)$ に対する偏った推定値である。 S 、 G の偏り (bias ratio) をそれぞれ

$$\alpha(S) = \sigma(Y) / \tilde{\sigma}(Y) \quad \dots\dots\dots (75)$$

$$\alpha(G) = r(Y) / \tilde{r}(Y) \quad \tilde{r}(Y) > 0 \quad \dots\dots\dots (76)$$

で表わすとき、 $\alpha(S)$ 、 $\alpha(G)$ の値はそれぞれ表-5、6のごとくである。ここに、 $\tilde{\sigma}(Y)$ 、 $\tilde{r}(Y)$ はそれぞれ、100,000組の S 、 G の平均値である。 $r(Y)$ の推定値として、Water Resources Council²⁹⁾ および Interagency Committee on Water Resources³⁰⁾ はそれぞれ次の値を勧告している。

$$G' = n^{1/2} \left(\frac{(n-1)^{1/2}}{n-2} \right) G \quad \dots\dots\dots (77)$$

$$G^* = n^{1/2} \left(\frac{(n-1)^{1/2}}{n-2} \right) (1 + 8.5/n) G \quad \dots\dots\dots (78)$$

ここに、 G は(74)式で与えられる。したがって、(78)式によれば bias ratio は次式で与えられる。

$$\alpha^*(G) = n^{1/2} \left(\frac{(n-1)^{1/2}}{n-2} \right) (1 + 8.5/n) \quad \dots\dots\dots (79)$$

$\alpha^*(G)$ は表-6の最下行に示されているが、表中の数値から(79)式は $r(Y)$ の狭い範囲、たとえば対数正規分布に対しては、 $1/2 < r(Y) < 2$ において近似的な不偏推定値であることがわかる。しかし、 $r(Y) < 2^{1/2}$ においては $\alpha(G)$ はほとんど分布に関係しない値をとる。

$\bar{X}$ 、 S 、 G の分布の例 ($n=10$ 、 $r(Y)=3$) が図-1~4であるが、 G の分布は上限境界値の存在を示唆することが注目される。

この境界値に関して Kirby³¹⁾ は、標本歪み係数は母集団の確率分布に関係せず、標本の大きさのみに関係する上、下限境界をもつことを理論的に導いた。同様の結果が変動係数(全資料の値が正の場合)、最大値の平均値からの偏差および範囲について得られ、これらを代数的境界(algebraic bound)と呼んだ。各々の境界値は標本の大きさ n の関数として表-7のように与えられる。このような成果は極値の統計解析を行なう上できわめて重要であり、また異常標本(outlier)の処理においても有効な知

表-5 Bias Factors for Standard Deviation

Distribution	Skew, r	Sequence Length n								
		10	20	30	40	50	60	70	80	90
Normal	0.0	1.084	1.040	1.026	1.019	1.016	1.013	1.011	1.010	1.009
Gumbel	1.14	1.108	1.053	1.035	1.027	1.021	1.018	1.015	1.013	1.012
Log normal	0.25	1.085	1.040	1.026	1.020	1.016	1.013	1.011	1.010	1.009
	0.50	1.088	1.042	1.028	1.021	1.016	1.014	1.012	1.010	1.009
	0.71	1.093	1.044	1.029	1.022	1.017	1.014	1.012	1.011	1.010
	1.00	1.101	1.048	1.032	1.024	1.019	1.016	1.014	1.012	1.011
	1.14	1.104	1.050	1.033	1.025	1.020	1.017	1.015	1.013	1.011
	1.41	1.116	1.057	1.038	1.029	1.023	1.019	1.017	1.015	1.013
	2.00	1.141	1.072	1.049	1.038	1.031	1.026	1.023	1.020	1.018
	3.00	1.186	1.100	1.071	1.056	1.047	1.041	1.036	1.032	1.029
	4.00	1.233	1.131	1.095	1.076	1.065	1.057	1.050	1.045	1.041
	5.00	1.276	1.161	1.119	1.096	1.083	1.073	1.065	1.059	1.054
	10.00	1.452	1.284	1.221	1.184	1.163	1.147	1.133	1.123	1.114
	15.00	1.581	1.378	1.299	1.254	1.226	1.206	1.188	1.175	1.164
Pareto	3.00	1.191	1.104	1.073	1.057	1.047	1.040	1.034	1.030	1.027
	4.00	1.232	1.133	1.097	1.077	1.064	1.055	1.049	1.043	1.039
	5.00	1.265	1.157	1.116	1.094	1.079	1.069	1.061	1.055	1.050
	10.00	1.354	1.224	1.173	1.145	1.125	1.111	1.101	1.092	1.085
	15.00	1.392	1.253	1.199	1.168	1.147	1.131	1.119	1.110	1.102
Pearson	0.25	1.084	1.039	1.026	1.019	1.015	1.013	1.011	1.009	1.008
	0.50	1.088	1.042	1.028	1.021	1.016	1.013	1.011	1.010	1.009
	0.71	1.091	1.044	1.029	1.021	1.017	1.014	1.012	1.011	1.009
	1.00	1.099	1.048	1.032	1.024	1.019	1.016	1.014	1.012	1.010
	1.14	1.104	1.051	1.033	1.025	1.020	1.017	1.014	1.013	1.011
	1.41	1.112	1.055	1.037	1.027	1.022	1.018	1.016	1.014	1.012
	2.00	1.138	1.070	1.047	1.036	1.029	1.025	1.021	1.018	1.016
	3.00	1.201	1.104	1.071	1.054	1.044	1.036	1.031	1.027	1.025
	4.00	1.289	1.153	1.106	1.082	1.067	1.057	1.049	1.044	1.039
	5.00	1.390	1.212	1.146	1.114	1.094	1.080	1.069	1.061	1.055
Weibull	0.25	1.080	1.037	1.024	1.018	1.014	1.012	1.010	1.009	1.008
	0.50	1.083	1.038	1.024	1.018	1.015	1.012	1.011	1.009	1.008
	0.71	1.087	1.041	1.026	1.020	1.016	1.013	1.011	1.010	1.009
	1.00	1.095	1.045	1.029	1.022	1.017	1.014	1.012	1.011	1.010
	1.14	1.100	1.047	1.031	1.023	1.018	1.015	1.013	1.011	1.010
	1.41	1.111	1.053	1.035	1.026	1.021	1.017	1.015	1.013	1.012
	2.00	1.140	1.070	1.046	1.035	1.028	1.024	1.020	1.018	1.016
	3.00	1.199	1.104	1.071	1.055	1.044	1.038	1.033	1.029	1.026
	4.00	1.263	1.144	1.100	1.078	1.064	1.055	1.048	1.043	1.039
	5.00	1.327	1.184	1.131	1.104	1.086	1.075	1.066	1.058	1.053
	10.00	1.622	1.380	1.285	1.234	1.200	1.177	1.159	1.144	1.133
	15.00	1.861	1.551	1.422	1.352	1.304	1.273	1.247	1.225	1.209

Bias ratio is $\alpha(S)$ for standard deviation. The ratio is the population value over the mean of 100,000 samples.

表-6 Bias Factors for Skew

	Skew, r	Sequence Length n								
		10	20	30	40	50	60	70	80	90
Distribution										
Gumbel	1.14	2.172	1.541	1.355	1.269	1.217	1.183	1.156	1.137	1.123
Log normal	0.25	1.903	1.381	1.246	1.183	1.141	1.116	1.099	1.086	1.076
	0.50	1.960	1.413	1.267	1.198	1.156	1.129	1.109	1.096	1.085
	0.71	2.019	1.450	1.295	1.220	1.176	1.147	1.126	1.111	1.099
	1.00	2.100	1.499	1.331	1.249	1.201	1.168	1.144	1.126	1.113
	1.14	2.161	1.534	1.359	1.268	1.221	1.188	1.163	1.144	1.139
	1.41	2.251	1.595	1.404	1.309	1.252	1.218	1.189	1.168	1.151
	2.00	2.528	1.773	1.545	1.428	1.358	1.307	1.276	1.248	1.220
	3.00	3.066	2.120	1.827	1.665	1.573	1.506	1.455	1.414	1.381
	4.00	3.641	2.498	2.134	1.931	1.813	1.727	1.659	1.607	1.563
	5.00	4.234	2.888	2.453	2.209	2.064	1.959	1.876	1.811	1.757
	10.00	7.247	4.880	4.087	3.636	3.362	3.161	3.002	2.876	2.773
	15.00	10.239	6.857	5.710	5.055	4.654	4.359	4.126	3.940	3.788
Pareto	3.00	2.744	1.954	1.701	1.570	1.484	1.425	1.381	1.345	1.316
	4.00	3.464	2.429	2.089	1.910	1.791	1.708	1.646	1.595	1.553
	5.00	4.202	2.922	2.495	2.269	2.118	2.013	1.933	1.868	1.813
	10.00	7.975	5.463	4.611	4.154	3.846	3.629	3.464	3.328	3.215
	15.00	11.784	8.038	6.762	6.075	5.613	5.285	5.035	4.830	4.659
Pearson	0.25	1.868	1.359	1.232	1.169	1.129	1.103	1.088	1.078	1.066
	0.50	1.925	1.407	1.255	1.186	1.146	1.119	1.103	1.088	1.080
	0.71	1.969	1.416	1.270	1.201	1.160	1.132	1.112	1.098	1.087
	1.00	1.963	1.430	1.279	1.207	1.165	1.138	1.119	1.104	1.093
	1.14	1.972	1.441	1.291	1.216	1.174	1.145	1.125	1.109	1.096
	1.41	1.978	1.450	1.302	1.226	1.182	1.151	1.130	1.113	1.100
	2.00	2.054	1.519	1.354	1.273	1.225	1.192	1.166	1.147	1.132
	3.00	2.233	1.650	1.464	1.364	1.302	1.260	1.228	1.204	1.186
	4.00	2.466	1.811	1.590	1.473	1.398	1.345	1.307	1.276	1.252
	5.00	2.735	1.982	1.724	1.588	1.499	1.436	1.389	1.352	1.323
Weibull	0.25	1.863	1.359	1.224	1.160	1.125	1.104	1.090	1.077	1.068
	0.50	1.778	1.326	1.205	1.148	1.116	1.097	1.083	1.072	1.063
	0.71	1.777	1.333	1.213	1.155	1.124	1.104	1.090	1.078	1.070
	1.00	1.795	1.349	1.226	1.165	1.132	1.110	1.095	1.082	1.073
	1.14	1.819	1.366	1.239	1.176	1.141	1.118	1.102	1.088	1.079
	1.41	1.874	1.402	1.265	1.197	1.158	1.133	1.114	1.099	1.088
	2.00	2.057	1.521	1.357	1.274	1.224	1.191	1.166	1.147	1.132
	3.00	2.441	1.770	1.553	1.438	1.367	1.319	1.282	1.253	1.230
	4.00	2.873	2.053	1.778	1.631	1.538	1.473	1.424	1.384	1.353
	5.00	3.325	2.352	2.019	1.838	1.722	1.642	1.579	1.529	1.490
	10.00	5.661	3.904	3.279	2.934	2.708	2.547	2.422	2.320	2.240
	15.00	7.990	5.463	4.548	4.041	3.706	3.466	3.280	3.129	3.008
Bias ratio $\alpha^*(G)$ for skew	—	2.194	1.543	1.352	1.260	1.207	1.171	1.146	1.128	1.113

Bias ratio is $\alpha(G)$ for skew. The ratio is the population value over the mean of 100,000 samples.

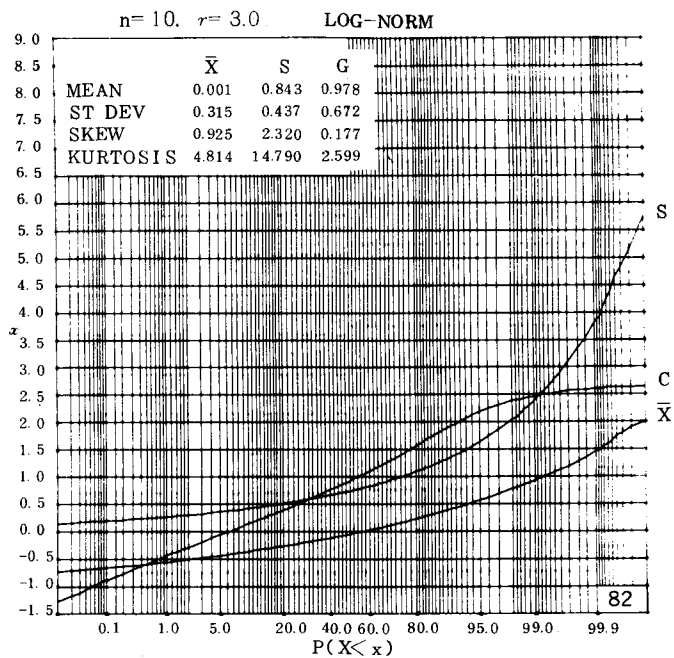


图-1 Distribution functions of sample statistics for the log normal distribution

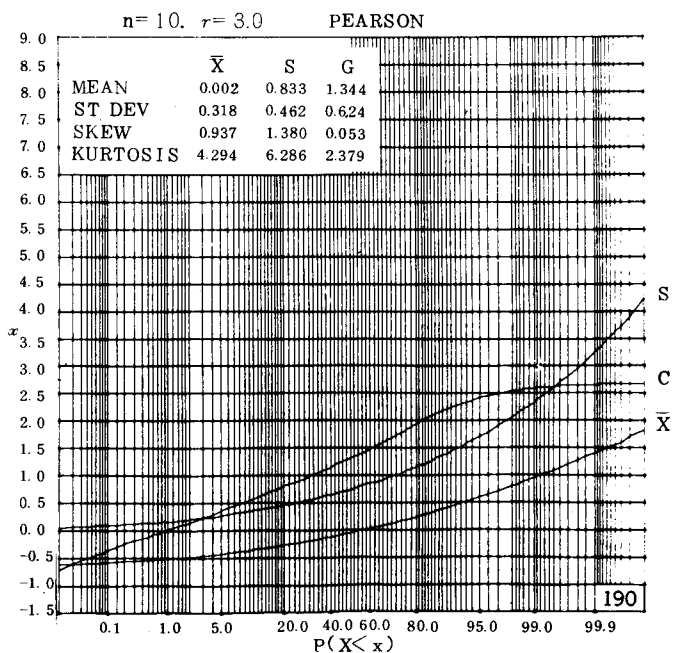


图-2 Distribution functions of sample statistics for the Pearson distribution

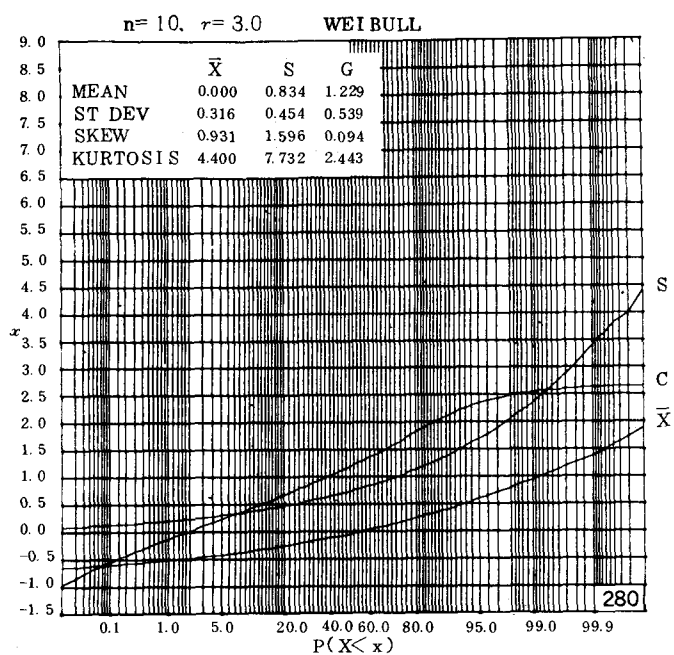


图-3 分布函数 of 样本统计量 for the Weibull 分布

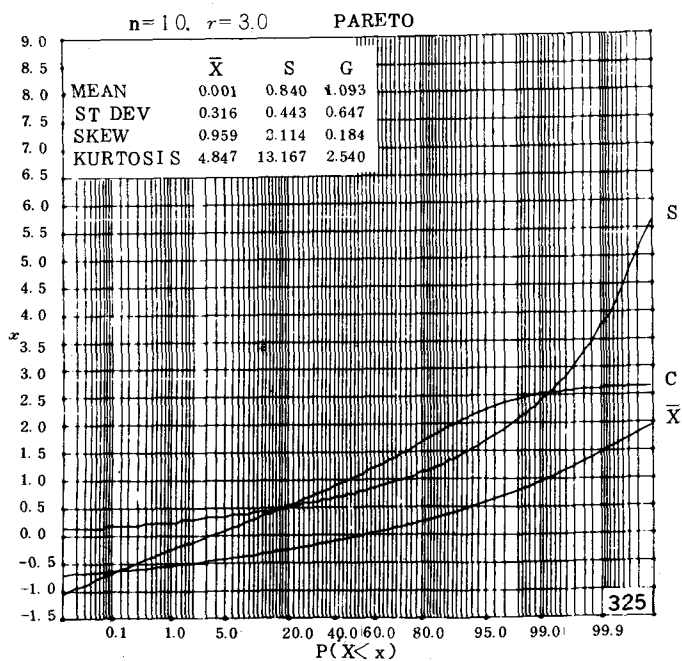


图-4 分布函数 of 样本统计量 for the Pareto 分布

表-7 Algebraic Bounds of Sample Statistics for Sample Size n

Statistic	Upper Bound	Lower Bound
Coefficient of variation $s_n/\bar{x} (x_i > 0)$	$(n-1)^{1/2}$	0
Sample skew $\Sigma(x_i - \bar{x})^3 / ns_n^3$	$(n-2)/(n-1)^{1/2}$	$-(n-2)/(n-1)^{1/2}$
Standardiznd maximum deviate $(x_{max} - \bar{x})/s_n$	$(n-1)^{1/2}$	$(n-1)^{1/2}$
Standardized range $(x_{max} - x_{min})/s_n$	$(2n)^{1/2}$	$2(n \text{ is even})$

見となる。前述の計算機シミュレーションの結果を総合するとき、一層の成果を期待することができよう。

II 極値分布とリターン・ピリオド

水工計画の中で、治水計画における代表的な水文量の指標はピーク流量であり、実際には河川の許容流量を超えるピーク流量の生起頻度、あるいは再現期間の確率分布特性が問題となる。したがって、洪水問題に対しては次の2つの観点から研究が行なわれている。

- (1) 主としてピーク値の分布特性に注目したもの。
 - (2) 主として洪水の時間的発生過程に注目したもの。
- (1)は従来、極値分布の問題として取扱われているものであり、(2)はリターン・ピリオドにも関連するものである。

1. 極値分布

水文資料はその摘出方法によって次のように分類される。

- A. 全水文資料(complete duration series)
- B. 部分的な水文資料(partial - duration series)
- C. 極値水文資料(extreme - value series)
 - 最大値資料(maximum-value series)
 - 最小値資料(minimum-value series)

通常、極値分布とは極値水文資料に対して適用される分布形式であり、大きさ n の標本における最大値 x_n 、あるいは最小値 x_1 の分布の $n \rightarrow \infty$ のときの極限形式を意味する。この問題に関しては極値極限分布に3つの形式の存在することを示した Fréchet (Type 2)⁶²⁾、Fisher & Tippett (Type 1, 3)⁶³⁾ の研究をはじめとして、Weibull、Gumbel、角屋らによって実用的な形式が得られている。

1-1. 最大値分布

順序統計量の理論によれば、確率分布関数 $F(x)$ をもつ全水文資料から得られた大きさ n の標本の最大値 x の分布関数 $P(x)$ は次式で与えられる。

$$\left. \begin{aligned} dP(x) &= n [F(x)]^{n-1} f(x) dx \\ P(x) &= [F(x)]^n \end{aligned} \right\} \dots\dots\dots (80)$$

ここに、 $F(x) = \int_{-\infty}^x f(x) dx$

いま, n の値が十分大きく, かつ標本最大値を問題にしているから $F(x)$ は 1 に近いとすると,

$$\ell_n F(x) = [F(x) - 1] + O[F(x) - 1] \quad \dots\dots\dots (81)$$

であるから, 式(80)は漸近的に次式で与えられる。

$$\lim_{n \rightarrow \infty} P(x) = \exp \{ -n [1 - F(x)] \} \quad \dots\dots\dots (82)$$

このような分布関数 $P(x)$ が存在する場合, これは一般に次の 3 形式で表わされる。

$$\text{Type 1: } P(z) = \exp(-e^{-z}) \quad -\infty < z < \infty \quad \dots\dots\dots (83)$$

$$\text{Type 2: } P(z) = \exp(-z^{-k}) \quad k > 0, \quad 0 < z < \infty \quad \dots\dots\dots (84)$$

$$\text{Type 3: } P(z) = \exp\{-(z)^k\} \quad k > 0, \quad -\infty < z < 0 \quad \dots\dots\dots (85)$$

ここに, $z = A(x - B)$ A, B, k は定数

いま, Type 2, 3において, x を変数変換すれば Type 1 と同じ形式になるので, 角屋^(84, 85) はこれらを実用的に次の 3 形式にまとめている。

$$P(x) = \exp(-e^{-x}) \quad \dots\dots\dots (86)$$

$$\text{Gumbel 分布 (Type 1): } y = a(x - x_0), \quad -\infty < x < \infty \quad \dots\dots\dots (87)$$

$$\text{対数極値分布 A 型 (Type 2): } y = k \ell_n \frac{x+b}{x_0+b}, \quad -b < x < \infty, \quad k > 0 \quad \dots\dots (88)$$

$$\text{対数極値分布 B 型 (Type 2): } y = k \ell_n \frac{u-x_0}{u-x}, \quad -\infty < x < u, \quad k > 0 \quad \dots\dots (89)$$

ここに, y は標準極値変量と呼ばれ, a, k, x_0, b, u は定数である。

上記の極値分布は次のような母集団の分布関数 $F(x)$ を式(82)に代入して得られるから, 両者はそれぞれつぎのように対応する。

$$\text{Type 1} \quad \longleftrightarrow \quad \text{指数タイプ} \quad : \quad F(x) = 1 - \exp\{-a(x - x_0)\} \quad \dots\dots\dots (90)$$

$$\text{Type 2} \quad \longleftrightarrow \quad \text{Cauchy タイプ} \quad : \quad F(x) = 1 - \left(\frac{x+b}{x_0+b}\right)^{-k} \quad \dots\dots\dots (91)$$

$$\text{Type 3} \quad \longleftrightarrow \quad \text{有限タイプ} \quad : \quad F(x) = 1 - \left(\frac{u-x_0}{u-x}\right)^{-k} \quad \dots\dots\dots (92)$$

極限形式の極値分布には, 観測値の独立性や均質性の他に次のような制約があることに注意しなければならない。すなわち, 極値が得られた観測値の数が充分大きいこと, および母集団分布は上述の 3 つのタイプ (指数, Cauchy, 有限) のいずれかに属することである。実際にはこれらの全ての制約条件を満足できない場合が多い。

このような問題もあるので, 極限形式以外の分布形も極値分布に適用されており, その代表的なものが, I に述べた対数正規分布, ピアソン-III 型, log-ピアソン-III 型分布等である。また, ピアソン-III 型, 正規分布, 極値分布 Type-2, 3 および対数 Gumbel 分布等をその特別の場合として包括できる一般化されたガンマ分布が Mendel⁽⁸⁶⁾ によって次式のごとく提示されている。

$$f(x) = c \frac{\frac{a+1}{b^c}}{\Gamma(\frac{a+1}{c})} x^a \exp(-bx^c), \quad x, a \geq 0, b > 0, c > 0 \quad (93)$$

ここに、 a, b, c は定数である。この分布関数は降雨特性に対して鈴木⁽⁸⁷⁾が示した超ガンマ型分布に他ならない。

1-2. 最小値分布

最大値分布と同様に、最小値分布の極限形式には次の3つの形式がある。

$$G(x) = \exp(-e^\xi) \quad \dots\dots\dots (94)$$

$$\text{Type 1 : } \xi = a(x - x_0) \quad -\infty < x < \infty \quad \dots\dots\dots (95)$$

$$\text{Type 2 : } \xi = k \ln \frac{u - x_0}{u - x} \quad -\infty < x < u, k > 0 \quad \dots\dots\dots (96)$$

$$\text{Type 3 : } \xi = k \ln \frac{x + b}{x_0 + b} \quad -b < x < \infty, k > 0 \quad \dots\dots\dots (97)$$

上式において、 $\xi = -y$ とおけば最大値分布と同じ形式になるから、結局最小値分布は最大値分布の対応形式を裏返しにした分布形を示す。また、水文量への適用性は Type 3 のみが Gumbel によって示されているが、他の形式はまず実用に供しがたいようである。

1-3. パラメータ推定法

積率法は、極値分布のように歪んだ分布に対しては過大あるいは過小な推定値を与えることが知られているが、^{(88), (89)} 実用的には最も簡便であるゆえにしばしば用いられてきた。最小二乗法も積率法より全体的に良好な適合を与えるものとして、Gumbel⁽⁴⁰⁾ 角屋⁽⁴¹⁾らによって用いられている。これらの方法では通常歪み係数の推定のために plotting position の概念が用いられる。

最尤法によれば、たとえば Gumbel 分布の定数推定の基本式は次式で与えられる。

$$L(x) = \sum_{i=1}^N \ln [a \exp(-y - e^{-y})] \quad \dots\dots\dots (98)$$

$$\frac{\partial L}{\partial a} = \frac{1}{a} \left(N - \sum_{i=1}^N y_i + \sum_{i=1}^N y_i \exp(-y_i) \right) = 0 \quad \dots\dots\dots (99)$$

$$\frac{\partial L}{\partial x_0} = \frac{1}{a} \left(-N + \sum_{i=1}^N \exp(-y_i) \right) = 0 \quad \dots\dots\dots (100)$$

ここに、 $y = a(x - x_0)$

式(99)、(100)を解くことによりパラメータ a, x_0 は原理的には決定しうる。この方法が前述の二方法と異なる点は方程式の中に標本の大きさ N の情報が明確に現われ、plotting position の概念を用いずとも良好な推定値が得られることである。

Kimball⁽⁴²⁾はこの方法を極値分布に初めて応用し、実用的計算法は Panchang & Aggarwal⁽⁴³⁾によって開発された。その後、Jenkins⁽⁴⁴⁾によって検討され、Lowery & Nash⁽⁴⁵⁾によって最も有効な方法であることが報告されている。

以上、極値分布の関数形を示すにとどめたが、その特性や解法に関しては文献(9)に詳述されているので参照されたい。

2. 極値水量の生起過程

極値洪水の生起過程に関しては、Hall & Howell⁽⁴⁶⁾, Shane & Lynn⁽⁴⁷⁾はその過程を独立なポアソン過程とみなし、後者ではランダム変数のランダムな個数の和の理論に基づいて部分的水文資料の解析法を示した。また、Kirby⁽⁴⁸⁾はピーク流量の生起をベルヌイ試行で表現し、洪水の生起時間間隔に関する確率分布を示している。1970年、Todorovic⁽⁴⁹⁾はある期間内での洪水の観測個数およびその観測結果がいずれもランダム変数である場合の極値の分布関数を得た。この研究は極値の確率特性に関する一般性の高い解析法を与えるものであり、ここに得られた数学的基礎はYevjevich & Todorovic,⁽⁵⁰⁾ Zelenhasic⁽⁵¹⁾の研究に應用されている。

ここでは極値分布に関するZelenhasicの研究を紹介する。

定常、独立な過程を考える。(Todorovicの論文ではまず非定常な過程の定式化から始まっているが、実用解を得るに至っていない。)水文量時系列は図-5のごとくモデル化される。

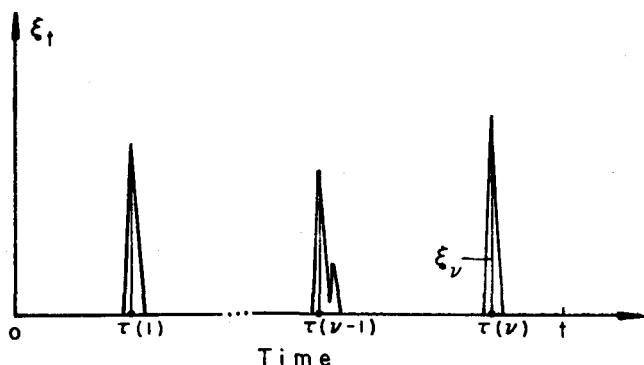


図-5 Intermittent process of flood hydrograph tops, with ξ_t the discharge above the base Q_b .

ある期間 $[0, t]$ に k 個の水文量 ξ_ν が発生したとすると、極値分布の確率分布関数 $F(x)$ は ξ_ν のすべてが x を超えない確率で与えられる。すなわち

$$P \left\{ \bigcap_{\nu=0}^k (\xi_\nu \leq x) \right\} \dots\dots\dots (101)$$

定常確率過程を考え、 $P(\xi_\nu \leq x)$ を ν に関して一定として

$$H(x) = P(\xi_\nu \leq x) \dots\dots\dots (102)$$

とすれば式(101)は

$$\{ H(x) \}^k \dots\dots\dots (103)$$

たとえば ξ_ν が指数分布に従うときは、

$$F_t(x) = (1 - e^{-\beta x})^k \dots\dots\dots (104)$$

$k \rightarrow \infty$ のとき式(104)は

$$F_t(x) \rightarrow \exp[-\lambda t e^{-\beta x}] \dots\dots\dots (105)$$

となり $t = \lambda = \beta = 1$ であれば Gumbel 分布に一致する。この近似 ($k \rightarrow \infty$) が十分精度よく成り立つためには、 λ の大小にもよるが一般に $k > 30$ 程度でなければならない。

一方、生起回数 k 自身も確率変量のはずである。Todorovic, Zelenhasic はこれを考慮した理論解を導いた。いま期間 t 内に k 回生起する事象を E_k^t で表わしておくと、生起回数の確率分布を考慮した極値分布関数は、

$$F_t(x) = \sum_{k=0}^{\infty} P \left\{ \left[\bigcap_{\nu=0}^k (\xi_{\nu} \leq x) \right] \cap E_k^t \right\} \quad \dots\dots\dots (106)$$

ただし、生起回数と、水文量強度 ξ_{ν} は独立としている。時系列的に独立とするときは、

$$F_t(x) = \sum_{k=0}^{\infty} \left\{ [H(x)]^k \cdot P(E_k^t) \right\} \quad \dots\dots\dots (107)$$

たとえば、 ξ_{ν} の確率密度関数がガンマ分布で表わされる場合を考える。今後 ξ_{ν} を x と書くことにし、密度関数を $h(x)$ と書くと、

$$\begin{cases} h(x) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} & \text{for } x \geq 0 \\ = 0 & \text{for } x < 0 \end{cases} \quad \dots\dots\dots (108)$$

α を正整数とするとときは、

$$H(x) = \int_0^x \frac{\beta^{\alpha}}{(\alpha-1)!} x^{\alpha-1} e^{-\beta x} dx \quad \dots\dots\dots (109)$$

$$= 1 - e^{-\beta x} \sum_{i=1}^{\alpha-1} \frac{(\beta x)^i}{i!} \quad \dots\dots\dots (110)$$

また E_k^t として Poisson 過程を仮定すれば、

$$P(E_k^t) = e^{-\lambda t} \frac{(\lambda t)^k}{k!} \quad \dots\dots\dots (111)$$

$$\therefore F_t(x) = e^{-\lambda t} \sum_{k=0}^{\infty} \frac{(\lambda t)^k}{k!} \cdot \left[1 - e^{-\beta x} \sum_{i=0}^{\alpha-1} \frac{(\beta x)^i}{i!} \right]^k \quad \dots\dots\dots (112)$$

式(112)はより簡単に表わせて、

$$F_t(x) = \exp \left[-\lambda t e^{-\beta x} \sum_{i=0}^{\alpha-1} \frac{(\beta x)^i}{i!} \right] \quad \dots\dots\dots (113)$$

たとえば $\alpha = 1$ に対して

$$F_t(x) = \exp \left[-\lambda t e^{-\beta x} \right] \quad \dots\dots\dots (114)$$

これは $t = \lambda = \beta = 1$ とすれば Gumbel 分布にはかならない。すなわち、洪水生起回数もまた確率的に変動するとすれば、生起頻度の大小に無関係に Double Exponential 分布が厳密に成立することが導けたわけである。

Zelenhasic は α もまた 1 つのパラメーターであるとして、上記の理論を観測水文資料に適用している。

3. リターン・ピリオド

3-1. Plotting Position

plotting position は確率分布関数による理論値の推定に代る方法として用いられるほかに、前述のように分布関数の定数推定のためにも用いられる。これには古くから多くの式が提案されているが、その主要な公式は次のようである。

$$A. \text{ California 公式 (1923) : } P_m = m/N \quad \dots\dots\dots (115)$$

$$B. \text{ Hazen 公式 (1930) : } P_m = (2m-1)/2N \quad \dots\dots\dots (116)$$

$$C. \text{ Weibull (1939) or Thomas 公式 (1948) : } P_m = m/(N+1) \quad \dots\dots\dots (117)$$

California式⁽⁵²⁾は plotting position として最初に採用された式であるが、この式は毎年最大値に対しては適当でなく、部分的水文資料に用い得ることが理論的に明らかにされている。^{(53), (54)}これ代って区間中央値を与える Hazen式⁽⁵⁵⁾が用いられるようになったが、その後、Chow,⁽⁵⁶⁾Benson⁽⁵⁷⁾らによって極値分布に対してはWeibull式⁽⁵⁸⁾(あるいはThomas式)がもっとも適切であることが示されている。なお、この式の名称については、欧米では最初の提案者の名前が用いられているのに対してわが国では通常Thomas式と呼ばれている。

角屋⁽⁵⁹⁾は、経験的分布関数を作るときの plotting position と理論的分布の定数推定のための plotting position とを区別すべきと考え、前者に対して後者の場合を plotting value と呼んだ。彼は資料 $\{x_i\}$ の線形変換量、 $y_i = a(x_i - b)$ について、その期待値を Gumbel 分布、対数極値分布 A 型に対してそれぞれ次のように得ている。

$$E(y_i)_G = \frac{\Gamma(n+1)}{\Gamma(i)\Gamma(n-i+1)} \sum_{r=0}^{n-i} (-1)^r {}_{n-i}C_r \frac{1}{i+r} [r + \ln(i+r)] \quad \dots\dots\dots (118)$$

$$E(y_i)_A = \frac{\Gamma(n+1)}{\Gamma(i)\Gamma(n-i+1)} \Gamma(1-(1/k)) \cdot \sum_{r=0}^{n-i} (-1)^r {}_{n-i}C_r (i+r)^{-1+1/k} \quad (119)$$

ただし、 $r = \text{Euler const}$ であり、 $1/k$ は分布の歪みの程度を示すパラメータである。任意の i 番目順序統計量の plotting value $E(y_i)$ は上式を n, i について解けばよいが、上式はいずれも n が大きくなるほど、 i が 1 に近づくほど計算は急激に煩雑となって、これを直接求めることは実用上不可能に近い。そこで実用的には $i = 1, n$ について $E(y_1), E(y_n)$ を求め、中間値については Gumbel がとった態度に準じて、次式に示すような確率線形補間方式を採用した。

$$P(y_i) = P\{E(y_i)\} + \frac{i-1}{n-1} [P\{E(y_n)\} - P\{E(y_1)\}] \quad \dots\dots\dots (120)$$

ここに、 $P(y_i)$ は y_i に対する非超過確率である。この方法により Gumbel、対数極値分布の定数推定のための数表を作成し、⁽⁶⁰⁾我国では広く利用されている。

3-2. リターン・ピリオドの推定法

リターン・ピリオド(再現期間)は次式で定義される。

$$T = 1 / \{ n [1 - P(x_u)] \} \quad \dots\dots\dots (121)$$

ここに、 n は水文量 x の年平均生起回数で、毎年最大値の場合には $n = 1$ であり、 $P(x_u)$ は $x \leq x_u$ となる非超過確率である。この定義から明らかに、リターン・ピリオドの推定問題は確率分布関数 $P(x)$ の選定とその定数推定の問題に帰着する。

極値水文資料の毎年最大値 (annual maximum series) によるリターン・ピリオド, T_M と部分的
水文資料の毎年超過値 (annual exceedance series) によるリターン・ピリオド T_E との間にはよく
知られた次の関係,

$$T_E = 1 / \{ \ln T_M - \ln (T_M - 1) \} \quad \dots\dots\dots (121)$$

があるが, Cunnane⁽⁶¹⁾ によれば10年以上のリターン・ピリオドの推定の場合, 毎年超過値よりも毎年最
大値を用いる方が有効で, とくに同じ推定精度を得るには前者の場合は後者の場合の1.65倍の年資料が
必要であることが示されている。

このような極値の確率分布から求められるリターン・ピリオドは, その定義から平均的に T 年に1度の
割合で期待される期間を意味するが, これに対して極値の生起時間間隔そのものの確率分布から再現期
間を考えようとする研究もあり, 前述2の極値の生起過程に関する研究もその1例である。Thom⁽⁶²⁾ は
事象の生起がPoisson分布にしたがうものとして実時間間隔 τ の分布を検討し, 平均再現期間 T と τ と
の関係を次式のように導いている。

$$\left. \begin{aligned} P(\tau) &= 1 - e^{-\tau/T} \\ T &= \tau \ln \left[\frac{1}{1 - P(\tau)} \right] \end{aligned} \right\} \quad \dots\dots\dots (122)$$

ここに, $P(\tau)$ は τ の非超過確率であり, τ は平均値 T をもつ指数分布に従うことがわかる。

また, 一般に年最大値の標本抽出期間は1水年であるが, この標本抽出期間が1水年より小さく, た
とえば, ある特定の季節, あるいは月間とすれば, リターン・ピリオドも当然変化する。⁽⁶³⁾ そこで, Stol⁽⁶⁴⁾
は年最大値 (Annual maxima) と月最大値 (Monthly maxima) の分析とそれらの合成事象の確率
評価を試みている。

一方, 菊地原⁽⁶⁵⁾ はリターン・ピリオドの概念を1地点から鉄道や道路のような長さをもった領域—線
形領域—to拡張し, その領域上の少なくとも1地点で再現期間 T_1 年以上の異常値が出現する時間間隔
の期待値 T_∞ を定義し, 次式の関係を導いてこれを東北本線の大雨に適用している。

$$T_\infty = \frac{T_1}{1 + m} \quad \dots\dots\dots (124)$$

ここに m は線形領域と地点再現期間 T_1 できまる定数で, 区間定数と呼ばれる。

結 語

ここに取扱われている問題は古くから多くの研究者によって成果が集積されており, 近年においても
飛躍的な進展は見られないようである。はじめに述べた通り, 渇水の時系列は勿論, 洪水ハイドログラ
フについても独立な1変数の問題として扱うことには限界があり, 時系列的連続性が導入されるべきであ
るし, また水文学の非定常特性の解明とその水工計画への導入も残された課題である。これらの方向か
らの接近を加えてより一般性の高い理論の展開が望まれるところである。

本文は土木学会確率統計水文学研究小委員会(委員長: 東工大・日野幹雄教授)の中間報告書をもとにし
て, 本研修会のためにまとめたものである。I は筆者, II は当委員会委員の端野道夫氏(徳島大学工学
部)および江藤剛治氏(近畿大学理工学部)の報告書を参考にさせていただいた。なお参考文献はその
一部だけを掲載した。

参 考 文 献

- (1) Matalas, N. C. and J. R. Wallis ; Eureka / It Fits a Pearson Type 3 Distribution, WRR (Water Resources Research), 9 (2), pp. 281-289, 1973.
- (2) 室田 明, 江藤剛治, 田中 剛 ; 水文量の和に関する統計的研究, 土木学会論文報告集, № 223, pp. 23 - 31, 1974.
- (3) 長尾正志 ; 豪雨の短時間配分に関する二変数ガンマ分布の応用, 土木学会第28回年次学術講演会講演概要集 (第2部), pp. 115 - 117, 昭和48年.
- (4) Water Resources Council ; A Uniform Technique for Determining Flood Flow Frequencies, Bull. 15, Hydrol. Comm., Water Resources Council, 1967.
- (5) Benson, M. A. ; Uniform Flood - Frequency Estimating Methods for Federal Agencies, WRR, 4 (5), pp. 891 - 908, 1968.
- (6) Suzuki, E. ; A Statistical and Climatological Study on the Rainfall in Japan, Papers in Met. and Geophysics, 18, 1967.
- (7) 統計工学ハンドブック, 技報堂, pp. 20 - 22, 1954.
- (8) 鈴木栄一 ; 気象統計学, pp. 25 - 28, 地人書館, 昭和43年.
- (9) 角屋 睦 ; 水文統計論, 水工学シリーズ, 土木学会水理委員会, pp. 27 - 39, 1964.
- (10) 石原藤次郎, 高瀬信忠 ; 対数正規分布とその積率による解法, 土木学会論文集, 47号, pp. 18 - 23, 1957.
- (11) Matalas, N. C. ; Probability Distribution of Low Flows, U. S. Geol. Surv. Prof. Pap. 434 - A, A 1 - A 27, 1963.
- (12) 岩井重久 ; 確率洪水推定法とその本邦河川への適用, 統計数理研究, 2 卷 3 号, pp. 21 - 36, 1949.
- (13) 角屋 睦 ; 対数正規分布の適用範囲, 定数について, 農業土木研究別冊 3 号, pp. 12 - 16, 1962.
- (14) Matalas, N. C. ; Mathematical Assessment of Synthetic Hydrology, WRR, 3 (4), pp. 937 - 945, 1967.
- (15) Burges, S. J. ; Some problems with Log - Normal Markov Runoff Models, Proc. ASCE, HY 9, pp. 1487 - 1496, 1972.
- (16) Codner, G. P. and T. A. McMahon ; Log - Normal Streamflow Generation Models Reexamined, Proc. ASCE, HY 9, pp. 1421 - 1431, 1973.
- (17) Sangal, B. P. and A. K. Biswas ; The 3 - Parameter Lognormal Distribution and Its Applications in Hydrology, WRR, 6 (2), pp. 505 - 515, 1970.
- (18) Mejia, J. A. and I. Rodriguez - Iturbe ; Correlation Links between Normal and Log Normal Processes, WRR, 10 (4), pp. 689 - 690, 1974.
- (19) Mejia, J. M. and I. Rodriguez - Iturbe ; On the Synthesis of Random Field Sampling from the Spectrum ; An Application to the Generation of Hydrologic Spatial Processes, WRR, 10 (4), pp. 705 - 711, 1974.
- (20) Frost, J. and R. T. Clarke ; Use of Cross Correlation between Hydrological Time Series to Improve Estimates of Lag One Autoregressive Parameters, WRR, 9 (4), pp. 906 - 917, 1974.

- (21) 前出(1)
- (22) Snyder, W. M. ; Fitting of Distribution Functions by Nonlinear Least Squares, WRR, 8(6), pp. 1423 - 1432, 1972.
- (23) Yevjevich, V. M. ; Stochastic Problems in the Design of Reservoirs, in Water Research, edited by A. V. Kneese and S. C. Smith, pp. 375 - 411, John Hopkins Press, 1966.
- (24) Thomas, H. A., Jr. and M. B. Fiering ; Statistical Analysis of Reservoir Storage - Yield Relations, Chap. 1, Operations Research in Water Quality Management, Final Report to Bureau of State Services, 1963.
- (25) Wilson, E. B. and M. M. Hilferty ; The Distribution of Chi - Square, Proc. Nat. Acad. Sci., U. S. A., Vol. 17, pp. 684 - 688, 1931.
- (26) 室田 明, 神田 徹 ; 利水を対象とした流量時系列の解析について, 第13回水理講演会講演集, pp. 13 - 18, 昭和44年.
- (27) Kirby, W. ; Computer - Oriented Wilson - Hilferty Transformation That Preserves the First Three Moments and the Lower Bound of the Pearson Type 3 Distribution, WRR, 8(5), pp. 1251 - 1254, 1972.
- (28) Wallis, J. R., N. C. Matalas and J. R. Slack ; Just a Moment /, WRR, 10(2), pp. 211 - 219, 1974.
- (29) 前出 4)
- (30) Interagency Committee on Water Resources, Methods of Flow Frequency Analysis, Bull. 13, Subcommittee on Hydrology, Government Printing Office, Washington, D. C., 1966.
- (31) Kirby, W. ; Algebraic Boundness of Sample Statistics, WRR, 10(2), pp. 220 - 222, 1974.
- (32) Fréchet, M. ; Sur la Loi de probabilité de l'écart Maximum (On the Probability Law of Maximum Error), Ann. Soc. Polonaise Math. (Cracow), Vol. 6, pp. 93 - 116, 1927.
- (33) Fisher, R. A. and L. H. C. Tippet ; Limiting Forms of the Frequency Distribution of the Smallest and Largest Member of a Sample, Proc. Cambridge Phil. Soc., Vol. 24, pp. 180 - 190, 1928.
- (34) 角屋 睦 ; 極値分布とその一解法, 農土研, Vol. 23, № 6, 昭和31年.
- (35) 角屋 睦 ; 極値 (最大値) 分布の一実用解, 農土研別冊, 第3号, 昭和37年.
- (36) Mendel, H. G. ; A New Proposal for Flood Frequency Computation, Proc. Second International Hydrology Symposium, Fort Collins, Vol. 1, pp. 325 - 333, 1972.
- (37) 前出 6)
- (38) Yevjevich, V. ; Probability and Statistics in Hydrology, Water Resources Publications, Fort Collins, Colorado, 1972.
- (39) Hershfield, D. H. ; An Empirical Comparison of the Predictive Value of

- Three Extreme - Value Procedures, J. Geophys. Res., Vol. 67, No. 5, pp. 1535 - 1542, 1962.
- (40) Gumbel, E. J. ; Statistical Theory of Droughts, Proc. ASCE, Vol. 80, No. 439, 1954.
- (41) 前出 35)
- (42) Kimball, B. F. ; Sufficient Statistical Estimation Functions for the parameters of the Distribution of Maximum Values, Ann. Math. Stats., Vol. 17, 1946.
- (43) Panchang, G. M. and V. P. Aggarwal ; Peak Flow Estimation by Method of Maximum Likelihood, Tech. Memo. HL 02, Government of India, Central Water and Power Research Station, Poona, India, 1962.
- (44) Jenkinson, A. F. ; Statistics of Extremes, in Estimation of Maximum Floods, World Meteorol. Organ. Rep. 98, Chap. 5, pp. 193 - 227, Geneva, 1969.
- (45) Lowery, M. D. and J. E. Nash ; A Comparison of Methods of Fitting the Double Exponential Distribution, J. Hydrology, Vol. 10, pp. 259 - 275, 1970.
- (46) Hall, W. A. and D. T. Howell ; Estimating Flood Probabilities within Specific Time Intervals, J. Hydrology Vol. 1, pp. 265 - 271, 1963.
- (47) Shane, R. M. and W. R. Lynn ; Mathematical Model for Flood Risk Evaluation, Proc. ASCE, HY 6, pp. 1 - 20, 1964.
- (48) Kirby, W. ; On the Random Occurrence of Major Floods, WRR, 5(4), 1969.
- (49) Todorovic, P. ; On Some Problems Involving Random Number of Random Variables, Annals of Mathematical Statistics, Vol. 41, No. 3, 1970.
- (50) Todorovic, P. and V. Yevjevich ; Stochastic Process of Precipitation, Colorado State University Hydrology Paper, No. 35, 1969.
- (51) Zelenhasic, E. ; Theoretical Probability Distributions for Flood Peaks, Colorado State University Hydrology Paper, No. 42, 1970.
- (52) Flow in California Streams, California State Dept. Pub. Works Bull. 5, Chap. 5, 1923.
- (53) Chow, V. T. ; Frequency Analysis of Hydrologic Data with Special Application to Rainfall Intensities, Univ. Illinois Eng. Exp. Sta. Bull, 414, 1953.
- (54) 石黒政儀 ; 非毎年最大値の確率降雨強度式算定法について, 水道協会雑誌, 第331号, 昭和37年.
- (55) Hazen, A. ; Flood Flow, A Study of Frequencies and Magnitudes, John Wiley & Sons, Inc., New York, 1930.
- (56) 前出 53)
- (57) Benson, M. A. ; Plotting Positions and Economics of Engineering Planning, Proc. ASCE, HY 6, 1962.
- (58) Weibull, W. ; A Statistical Theory of the Strength of Materials, Ing. Vetenskaps Akad. Handl. (Stockholm), Vol. 151, 1939.
- (59) 角屋 睦 ; 水文量の PLOTTING POSITION について, 京都防災研究所年報, 第3号, 昭和43年.

- (60) 前出 35)
- (61) Cummane, C. ; A Particular Comparison of Annual Maxima and Partial Duration Series Methods of Flood Frequency Prediction, J. Hydrology, Vol. 18, pp. 257 - 271, 1973.
- (62) Thom, H. C. ; A Time Interval Distribution for Excessive Rainfall, Proc. ASCE HY 7, 1959.
- (63) Murota, A. and M. Hashino ; Studies of a Stochastic Rainfall Model and Its Application to Sediment Transportation, Tech. Reports of the Osaka Univ., Vol. 19, № 873, Faculty of Eng., Osaka Univ., 1969.
- (64) Stol, PH. T. H. ; On the Decomposition of the Extreme Value Distribution of Daily Rainfall Depths and the Derivation of Probabilities of Compound Events, J. Hydrology, Vol. 14, pp. 181 - 196, 1971.
- (65) 菊地原英和 ; 線形領域の再現期間の理論とその計算, 第二報 — 計算の実施と検討 (東北本線の大
雨に対する適用), J. Meteo. Soc. of Japan, Vol. 47, № 3, pp. 133 - 144, 1969.