

水文学の PLOTTING POSITION について

角 屋 睦

ON THE PLOTTING POSITIONS OF HYDROLOGIC DATA

by *Mutsumi KADOYA*

Synopsis

As a method of estimating the theoretical distribution of hydrologic data, an expedient using the concept of the plotting positions has been proposed by several authors.

In this paper, the difference is discussed between the plotting positions which mean the empirical probability of observed data and the plotting values appropriate for estimating the parameters of the theoretical distribution of hydrologic data. Moreover, the plotting values of the Extremes, the Log-Extremes and the Log-Normals in the hydrologic statistics are examined. And in addition, the theory of confidence limits of the plotting values and the plotting positions is discussed.

1. 序 言

一般にある分布にしたがうとみられる n ケの標本より、その分布に含まれる母数（常数）を推定するには、分布形さえ既知とすれば、これら母数と母平均 m , 母分散 σ^2 ないし母ヒズミ係数 c_s などの理論的な関係を求め、後これら特性値を適当に標本値 \bar{x} , s_s^2 , c_s' などとおきかえればよい。

$$\left. \begin{aligned} m &= f_1(n)E(\bar{x}), \\ \sigma^2 &= f_2(n)E(s_s^2), \\ c_s &= f_3(n)E(c_s'). \end{aligned} \right\} \dots\dots\dots (1)$$

そして一般に $f_1(n)=1$, $f_2(n)=n/(n-1)$ であることはよく知られているが、標本ヒズミ係数の分布を考えることは容易ではない。

水文統計では他の分野でのそれとは異なり非対称分布を考え、一般に小標本でありながら超過確率の非常に小さい値を問題にするなどの特性を有し、いろいろの問題があるわけで、ヒズミ係数の取扱いなど重要な課題となつている。

さてこのような問題を解決する一つの便宜的な方法として plotting position の概念を利用しようとする試みがある。元来 plotting position とは、データの確率紙上の打点位置を意味するものであつて、これより経験的分布関数を推定したり、データの適合性を簡単に検定するためのものであるが、これまた水文統計の基本的な問題の一つでありながら、必ずしも明解な結論が得られているわけではない。

筆者はこの plotting position の問題は元来順序統計学的な問題であつて、経験的分布関数を作るときの plotting position と、この便宜的な考え方を採用して理論的分布を推定する場合のそれとは一応区別して見るべきだと考え、両者の混同を防ぐため前者の plotting position なる名称に対し後者のそれを plotting value と呼ぶことにする。そして本文では、まず plotting value, plotting position のあり方を論じ、水文統計で主用される極値分布、対数極値分布、対数正規分布の場合の plotting value, およびこれらの信頼限界などについて述べることにする。

2. 従来 の 方 法

これまで plotting position 法として Hazen 法, Gumbel 法, Thomas 法などが採用されてきた。これらは前述のように使用目的によつて plotting value と plotting position とに区別すべきものと思われるが、従来はすべて plotting position と呼ばれ、その使用目的の如何によらず無批判に利用されてきているので、ここでは一括論議しておく。

2.1 Hazen 法

これは Hazen 法, (修正) カリフォルニヤ法, または略算法とも呼ばれ水文統計では広く利用されている方法である。この方法の理論的根拠はかなり薄いようであるが、一応つぎのような説明ができる。

観測によつて得られた n ケのデータを大きさの順序に, $x_1 < x_2 < \dots < x_n$ とすると, x_i が 2 ケ以上同じ値をとらない限り見かけ上

$$Pr\{x_{i-1} < x < x_{i+1}\} = 1/n$$

である。そこでこの分布の histogram を作るに, x_i を

$$Pr\{x_i - dx_i \leq x \leq x_i + dx_i\} = 1/n$$

となるような微小区間の中央値と考えると, Fig. 1 に示すように x_i の非超過確率 F_i は見かけ上

$$F_i = \frac{i-1}{n} + \frac{1}{2} \frac{1}{n} = \frac{i-(1/2)}{n} \dots \dots \dots (2)$$

とみることができる。これが Hazen Plot¹⁾ と呼ばれるものである。

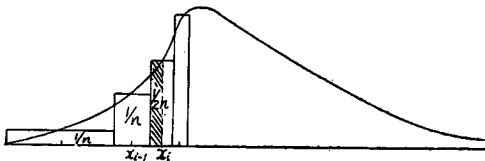


Fig. 1 Explanation of "Hazen Plot"

さて実際上の問題としては、どのようなデータでも x 軸上で x_i のすべてが中点として挟まれるように微小区間を連続的にとつていけるとは限らないから、この考え方はあくまでも概念的な意味しか持たない。ただこの考え方は、 n が大きいとき、観測値を適当に級別し、各中央値をもつて

それぞれの級の代表値として histogram を作る考え方に類同とみられ、簡易ながら一理ある考え方といふことができる。また考え方は全然別であるが、正規分布にしたがう大標本よりその平均値を推定する場合の sampling method として、比較的効率の高いことが F. Mosteller²⁾ によつて論じられているのも注目される。

2.2 Gumbel 法

E. J. Gumbel³⁾ は調整確率法 (adjustment frequency method) と称する plotting method をつぎのように提案した。すなわち観測によつて得られたデータを大きさの順序に列べ、その任意順位 x_i の非超過確率を考えると、これは sample size n と分布関数との関数として示されるはずであるが、 x_i もまた統計量である以上その非超過確率 F_i はわからない。そこで、この F_i の代表値として選ぶならば、最尤法の原理 (principle of maximum likelihood) を応用して mode の値を採用すべきだとした。しかしこれを任意の i について求めることはできないから、実用的には最大値、最小値の mode を求め中間順位のものには線形補間を行なう方法を示した。すなわち、

$$\left. \begin{aligned} n-1 &= -\frac{f'(\bar{x}_n)F(\bar{x}_n)}{f^2(\bar{x}_n)}, \\ n-1 &= \frac{f'(\bar{x}_1)[1-F(\bar{x}_1)]}{f^2(\bar{x}_1)} \end{aligned} \right\} \dots \dots \dots (3)$$

を解いて $F(\bar{x}_n), F(\bar{x}_1)$ を求め、任意の中間値の plotting position F_i

$$F_i = F(\bar{x}_1) + \frac{i-1}{n-1} [F(\bar{x}_n) - F(\bar{x}_1)] \dots\dots\dots (4)$$

とする方法である。

ただし、 $F(\bar{x})$ は mode \bar{x} の非超過確率で、 $f(\bar{x})$ 及び $f'(\bar{x})$ はそれぞれ

$$f(\bar{x}) = \left. \frac{d}{dx} F(x) \right|_{x=\bar{x}}, \quad f'(\bar{x}) = \left. \frac{d}{dx} f(x) \right|_{x=\bar{x}}.$$

この方法では plotting position は $F(x)$ の形によつて異なり、極値分布基本形 (Gumbel 分布) では、

$$F(\xi) = \exp\{-\exp(-\xi)\}, \dots\dots\dots (5)$$

$$\left. \begin{aligned} F_n &= \exp(1/n), \text{ あるいは } \xi_n = \lg n, \\ n &= \frac{-\lg F_1 - (1-F_1)}{F_1(-\lg F_1)}, \end{aligned} \right\} \dots\dots\dots (6)$$

正規分布の場合には、

$$F(\xi) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\xi} e^{-\xi^2} d\xi, \dots\dots\dots (7)$$

$$n-1 = 2 \xi_n \frac{F_n}{f_n} = -2 \xi_1 \frac{1-F_1}{f_1}, \dots\dots\dots (8)$$

さてこの考え方は筆者のいう plotting value の範疇に属するものであり、分布関数の理論解の推定に plotting position の考え方を導入する緒をつけたものとして注目されるのであるが、厳密な意味での最尤解を求めたものではない。そして最近では Gumbel 自身もあまり用いていないようである。

2 3 Thomas 法

これは筆者のいう plotting position そのものであつて、いま観測によつて得られた n ケのデータを大きさの順序に列べて、これを sample size n なる 1 組の順序統計量とみると、小さい方より i 番目順序統計量の分布は一般に、

$$dP_i = \frac{\Gamma(n+1)}{\Gamma(i)\Gamma(n-i+1)} F_i^{i-1} (1-F_i)^{n-i} dF_i \dots\dots\dots (9)$$

で示される。これを変量 x を含めず (non-parametric) に plotting position F_i の分布とみて F_i の期待値を求めると、

$$E(F_i) = \int_0^1 F_i dP_i = \frac{\Gamma(n+1)}{\Gamma(i)\Gamma(n-i+1)} \int_0^1 F_i^i (1-F_i)^{n-i} dF_i = \frac{i}{n+1} \dots\dots\dots (10)$$

これをもつて i 番目順序統計量の plotting position とする方法で、水文統計では Thomas plot⁴⁾ と呼んでいる。この考え方は理論的にもつともすつきりした方法であり、後述の最小分散の条件を満足している。なお (10) 式は以上の誘導とは別個に小川^{5), 6)} などによつて確率論的にも求められている形でもある。

3. 水文量の plotting value

まず便宜的に分布関数の理論解推定に利用するための plotting value について考える。いま観測によつて得られた n ケのデータを、その大きさの順序に $x_1 \leq x_2 \leq x_3 \leq \dots \leq x_n$ とし、これを分布形 (母ヒズミ

係数)が既知の母分布より得られた sample size n の 1 組の順序統計量とみなすと, i 番目順序統計量の分布は周知のように次式で示される。

$$dp_i = \frac{\Gamma(n+1)}{\Gamma(i)\Gamma(n-i+1)} F(x_i)^{i-1} [1-F(x_i)]^{n-i} f(x_i) dx_i \dots\dots\dots(11)$$

この式には母分布の常数 (parameter) がすべて含まれるわけであるが, 水文学の非対称分布を想定して, 一般に 3 ヶの常数が含まれているものとする, 分布形 (母ヒズミ係数) を既知としているからさし当つて 2 ヶの常数が未知である。そこでいま x のかわりにその線形変換変数 y

$$y = a(x-b), \text{ ただし, } a, b; \text{ 常数} \dots\dots\dots(12)$$

を考えると, 2 ヶの常数 a および b が未知であつても (11) 式のかわりに y の分布

$$dp_i = \frac{\Gamma(n+1)}{\Gamma(i)\Gamma(n-i+1)} F(y_i)^{i-1} [1-F(y_i)]^{n-i} f(y_i) dy_i \dots\dots\dots(13)$$

を考えれば充分である。これが plotting value であつて, i および n の関数で示される。さて問題はこのような y_i の分布を代表させる \hat{y}_i ではどうあるべきかということになる。

さて Fig. 2 において x, y を linear にとられたものとする (12) 式は直線式を表わすが, いま x_i のかわりにその大きさの順位のみに着目すると, 所与の i 及び n について y_i の分布は (13) 式で示されるから, 結局問題は*

$$\sum_i (y_i - \hat{y}_i)^2 = \min \dots\dots\dots(14)$$

となるような \hat{y}_i を求めればよい。このような解は, 明らかに,

$$\hat{y}_i = E(y_i) \dots\dots\dots(15)$$

であり, これを plotting value とすればよい。爾後これを expected plotting value と呼ぶことにする。

理論的には $E(y_i)$ を結ぶ線は必ずしも直線にはならないはずであるが, y_i に最小自乗法的に挿入した直線と, $E(y_i)$ 間に最小自乗法的に挿入した直線とは平均的な意味ではば一致するはずであり, その間に若干の差はあつても実用上無視できると考えてよいであろう。そしてまた単独順序統計量の分布を考えている以上, これ以上の最良解は得られないはずである。

一般にこの種の問題を厳密に考えるには単独順序統計量の分布を考えるより, 各順位統計量の同時分布の形で考えるべきものと思われるが, 実用上そのような解を求めることは至難なことであろう。ここではその差は無視できるとして, Gumbel, Thomas 流に単独順序統計量の分布を考えたわけである。plotting position については後述する。

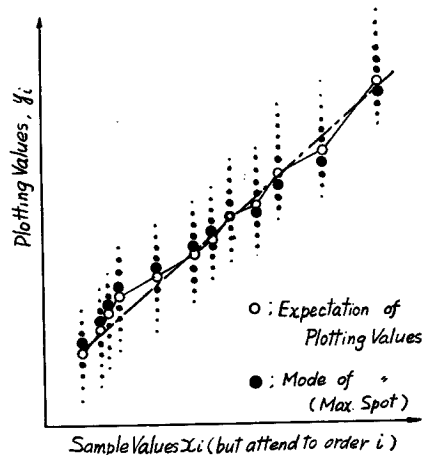


Fig. 2 Explanation of plotting value

* parameter 推定上の問題であるから,

$$E(y_i) = a[E(x_i) - b]$$

の条件の下で,

$$\sum_i [y_i - a(x_i - b)]^2 = \min$$

になるよう a, b を求めればよいが, この y_i がわからない。この y_i が plotting value であり, これを与えようとするわけである。

以上の論点に立つて水文統計で主用される極値分布, 対数正規分布の plotting value について以下若干考察してみる。

4. 極値分布の plotting value

4.1 極値分布基本形の plotting value

極値分布基本形, すなわち Gumbel 分布

$$\left. \begin{aligned} F(\xi) &= \exp(-e^{-\xi}), \\ \xi &= \alpha(x-x_0), \quad \infty > x > -\infty \end{aligned} \right\} \dots\dots\dots(16)$$

ただし, α, x_0 ; 常数

にしたがう i 番目順序統計量の基準変換変量 ξ_i は (13) 式, すなわち,

$$dp_i = \frac{\Gamma(n+1)}{\Gamma(i)\Gamma(n-i+1)} \exp[-(i-1)e^{-\xi}] \cdot \{1 - \exp(-e^{-\xi})\}^{n-i} \cdot \exp(-\xi - e^{-\xi}) d\xi$$

にしたがって分布するわけであるが, 前節の所論にしたがって, expected plotting value $E(\xi_i)$ を求めるには次式を解けばよい。

$$\begin{aligned} E(\xi_i) &= \frac{\Gamma(n+1)}{\Gamma(i)\Gamma(n-i+1)} \int_{-\infty}^{\infty} \xi \cdot \exp[-\xi - ie^{-\xi}] \cdot \{1 - \exp(-e^{-\xi})\}^{n-i} d\xi \\ &= \frac{\Gamma(n+1)}{\Gamma(i)\Gamma(n-i+1)} \sum_{r=0}^{n-i} (-1)^r r_{n-i} C_r \int_{-\infty}^{\infty} \xi \cdot \exp[-\xi - (i+r)e^{-\xi}] d\xi \dots\dots\dots(17) \end{aligned}$$

上式右辺の積分は $e^{-\xi} = x$ とおけば,

$$\begin{aligned} \int_{-\infty}^{\infty} \xi \cdot \exp[-\xi - (i+r)e^{-\xi}] d\xi &= \int_0^{\infty} -(1gx) \cdot e^{-(i+r)x} \cdot dx \\ &= (-1) \frac{d}{dt} \int_0^{\infty} x^{i-1} e^{-(i+r)x} dx \Big|_{t=1} = -\frac{d}{dt} [\Gamma(i) \cdot (i+r)^{-i}] = \frac{1}{i+r} [\gamma + 1g(i+r)], \end{aligned}$$

ただし, $\gamma \approx 0.5772 \dots\dots$ で Euler const. である。

したがって, (17) 式は,

$$E(\xi_i) = \frac{\Gamma(n+1)}{\Gamma(i)\Gamma(n-i+1)} \sum_{r=0}^{n-i} (-1)^r r_{n-i} C_r \frac{1}{i+r} [\gamma + 1g(i+r)]. \dots\dots\dots(18)$$

なお分散を知るに必要な $E(\xi_i^2)$ を同様にして求めておくとつぎのようになる。

$$E(\xi_i^2) = \frac{\Gamma(n+1)}{\Gamma(i)\Gamma(n-i+1)} \sum_{r=0}^{n-i} (-1)^r r_{n-i} C_r \frac{1}{i+r} \left\{ \frac{\pi^2}{6} + [\gamma + 1g(i+r)]^2 \right\}. \dots\dots\dots(19)$$

4.2 対数極値分布 A 型の plotting value

Gumbel 分布よりさらにヒズミ度の大きい分布に適用されるものとして, 筆者⁷⁾ が先に提案した対数極値分布 A 型の場合,

$$\left. \begin{aligned} F(y) &= \exp(-y^{-k}) = \exp(-e^{-\xi}), \\ y &= (x+b)/(x_0+b), \quad \infty > x > -b \\ \xi &= k \lg y = \alpha \log y, \end{aligned} \right\} \dots\dots\dots(20)$$

ただし、 $k = \alpha \log e$, x_0, b ; 常数,

の plotting value は x の線形変換変量 y について考えるわけで、前節の所論によつてその expected plotting value $E(y_i)$ を求めると、つぎのようになる。

$$\begin{aligned} E(y_i) &= \frac{\Gamma(n+1)}{\Gamma(i)\Gamma(n-i+1)} k \int_0^\infty y^{-k} \cdot \exp(-iy^{-k}) \cdot \{1 - \exp(-y^{-k})\}^{n-i} dy \\ &= \frac{\Gamma(n+1)}{\Gamma(i)\Gamma(n-i+1)} k \sum_{r=0}^{n-i} (-1)^r {}_{n-i}C_r \int_0^\infty y^{-k} \exp[-(i+r)y^{-k}] dy. \dots\dots\dots(21) \end{aligned}$$

右辺の積分は $(i+r)y^{-k} = z$ とおいて、

$$\begin{aligned} k \int_0^\infty y^{-k} \exp[-(i+r)y^{-k}] dy &= (i+r)^{-1+(1/k)} \int_0^\infty z^{-1/k} e^{-z} dz \\ &= (i+r)^{-1+(1/k)} \Gamma(1-(1/k)). \end{aligned}$$

したがつて、(21) 式は、

$$E(y_i) = \frac{\Gamma(n+1)}{\Gamma(i)\Gamma(n-i+1)} \Gamma(1-(1/k)) \sum_{r=0}^{n-i} (-1)^r {}_{n-i}C_r (i+r)^{-1+(1/k)}. \dots\dots\dots(22)$$

同様にして、 $E(y_i^2)$ をも求めておくと、

$$E(y_i^2) = \frac{\Gamma(n+1)}{\Gamma(i)\Gamma(n-i+1)} \Gamma(1-(2/k)) \sum_{r=0}^{n-i} (-1)^r {}_{n-i}C_r (i+r)^{-1+(2/k)}. \dots\dots\dots(23)$$

4.3 最大値 $i=n$ の plotting value

Gumbel 分布の場合、(18) 式で $i=n$ とおけば最大値の expected plotting value が得られる。

$$E(\xi_n) = \gamma + \lg n, \dots\dots\dots(24)$$

$$\sigma^2 \xi_n = E(\xi_n^2) - [E(\xi_n)]^2 = \frac{\pi^2}{6}. \dots\dots\dots(25)$$

対数極値分布A型の場合も同様で、(22) 及び (23) 式より、

$$E(y_n) = n^{1/k} \cdot \Gamma(1-(1/k)), \dots\dots\dots(26)$$

$$\sigma^2 y_n = n^{2/k} [\Gamma(1-(2/k)) - \Gamma^2(1-(1/k))]. \dots\dots\dots(27)$$

なお、ここで従来の plotting position と比較するために (26) 式を基準形に直すと、

$$\begin{aligned} A\xi_n &= k \lg [E(y_n)] = \lg n + k \lg \Gamma(1-(1/k)) \\ &= \lg n + \gamma + \sum_{r=2}^{\infty} r^{-1} \cdot S(r) \cdot (1/k)^{r-1}, \dots\dots\dots(28) \end{aligned}$$

ただし、 $S(r) = \lim_{m \rightarrow \infty} (1 + 2^{-r} + 3^{-r} + \dots + m^{-r})$, $r \geq 2$

すなわち、対数極値分布A型の基準化 expected plotting value は、Gumbel 分布のそれより $\sum_{r=2}^{\infty} r^{-1} \cdot S(r) \cdot (1/k)^{r-1}$ だけ大きく、また $\lim_{1/k \rightarrow 0} A\xi_n = E(\xi_n)$ となることは Gumbel 分布が対数極値分布で $1/k \rightarrow 0$ の special case であることよりも当然のことである。

さて Gumbel の調整確率法によれば、Gumbel 分布の場合、(6) 式より、

$$\sigma \hat{\xi}_n = \lg n. \dots\dots\dots(29)$$

対数極値分布A型の場合にも同様にして求めると,

$$\sigma_A \hat{\xi}_n = \lg n - \lg \{1 + (1/k)\} = \lg n - \sum_{r=1}^{\infty} (-1)^{r+1} \frac{1}{r} \left(\frac{1}{k}\right)^r \dots\dots\dots(30)$$

また Hazen plot $H \hat{\xi}_n$, Thomas plot $T \hat{\xi}_n$ はそれぞれつぎのようになる。

$$\left. \begin{aligned} H F_n &= \{n - (1/2)\} / n, \\ H \hat{\xi}_n &= -\lg \left[-\lg \left(1 - \frac{1}{2n}\right) \right] = \lg n + \lg 2 - \frac{1}{4n} - \frac{5}{96n^2} \dots\dots \end{aligned} \right\} \dots\dots\dots(31)$$

$$\left. \begin{aligned} T F_n &= n / (n+1), \\ T \hat{\xi}_n &= -\lg \left[\lg \left(1 + \frac{1}{n}\right) \right] = \lg n + \frac{1}{2n} - \frac{5}{24n^2} + \dots\dots \end{aligned} \right\} \dots\dots\dots(32)$$

以上よりみると、従来の plotting position 法では Hazen plot が expected plotting value, あるいは対数極値分布A型の基準化 expected plotting value とよい近似を示しているのが注目される。すなわち対数極値分布A型の場合では $1/k$ の値により若干異なるが、 $1/k \approx 0.1$ ではほぼ一致した値を示しており、この分布では $1/k < 1/3$ であることを考慮すると、Hazen plot の根拠がかなり薄いにもかかわらず近似度が良好なことは、式形の簡単なことと併せて一応注目されてよいであろう。これに対し Gumbel 分布の場合の Gumbel plot は Thomas plot とほとんど一致し、ともに expected plotting value と $r \approx 0.5772$ ないしこれに近い偏りがあるが、対数極値分布A型の場合には Gumbel plot の偏りは $1/k$ の大なるほど益々大きいといえる。Fig. 3 はこれらの結果を示したものである。

4.4 最小値 $i=1$ の plotting value

最小値や中間値の expected plotting value は、それぞれ (18) 及び (22) 式で $i=1$ などの値を適当に与えることによつて求めることができるわけであるが、実際には両式とも $n \rightarrow$ 大、 $i \rightarrow 1$ になるにつれて煩雑でかつ膨大な計算を必要とし、その結果は $i=n$ のように簡単に

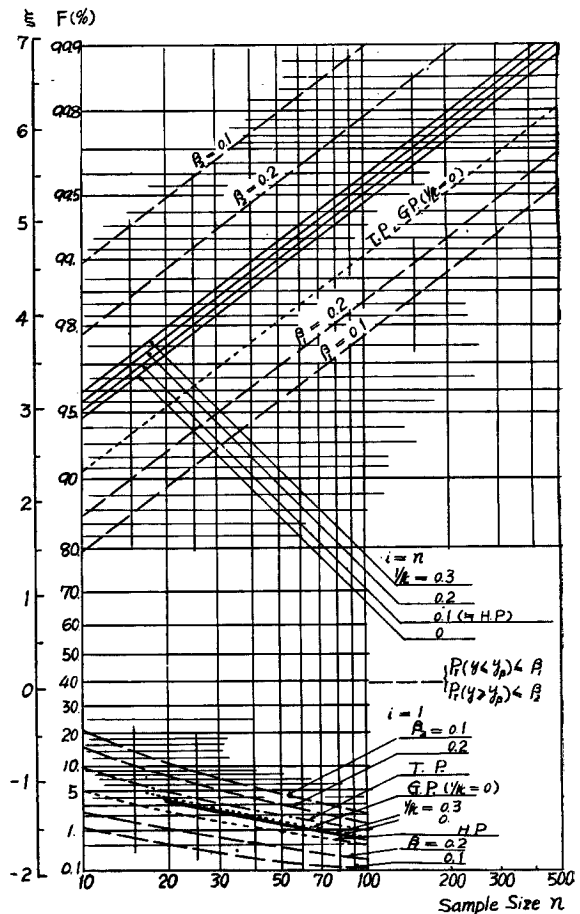


Fig. 3 Plotting values of the Extremes and its confidence limits (provided $i=n$ and 1)

は表わせない。そこで $i=1$ の場合について数値計算した結果を示すと Fig. 3 のようである。なお数値は $1/k$; 0 (Gumbel 分布) (0.02) 0.30, n ; 20 (2) 40, 40 (5) 100 まで小数位 5 位までのものを数表化してあるが、ここでは省略する。また Fig. 3 には Gumbel plot (ただし $1/k \rightarrow 0$ のみ), Thomas plot, Hazen plot の値も比較のため示してある。

さて, Fig. 3 よりすれば最小値 $i=1$ の場合は $1/k > 0$ (対数極値分布 A 型) のときでも, $1/k$ の差異による (expected) plotting value の差異は僅少でほとんど等値を示している。従来 plotting position 法では Gumbel plot ($1/k \rightarrow 0$) が非常に近い値を示しているが, 全般にいずれの方法によるも大差ないようである。

5. 対数正規分布の plotting value $\exp(-\xi^2)d\xi$

水文学の非対称分布によく適用される対数正規分布にはいろいろの形式が考えられるが, ここではこれまで最も主用されている形式, 下限常数 ($-b$) を含む Slade 型分布の plotting value について考える。

$$F(\xi) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\xi} \exp(-\xi^2) d\xi, \dots\dots\dots (23)$$

$$\xi = k \lg y = a \log y,$$

$$y = (x+b)/(x_0+b),$$

ただし, $k = a \log e$, x_0, b ; 常数。

この場合も観測によつて得られたデータを大きさの順序に列べたものを sample size n の 1 組の順序統計量とみなすと, その i 番目順序統計量の plotting value y_i の分布は (13) 式で示され, その expected plotting value は,

$$E(y_i) = \frac{\Gamma(n+1)}{\Gamma(i)\Gamma(n-i+1)} \int_0^{\infty} y_i \cdot F(y_i)^{i-1} \cdot [1-F(y_i)]^{n-i} \cdot f(y_i) dy_i \dots\dots\dots (34)$$

を計算すればよいはずであるが, その式中の $F(y_i)$ 自体が被積分形で示されるので, これを explicit に求めることは容易ではない。そこでとくに $i=n$ 及び 1 の場合について (34) 式を直接数値積分した結果を Fig. 4 に示した。

Fig. 4 は (a) $i=n$, (b) $i=1$ の場合の expected plotting value, 及び従来 plotting position 法によるものを基準変量 (正規変量) ξ に直した結果が示してあり, 同図で $1/k' = 1/\sqrt{2}k$ である。

これよりみれば, $i=n$ の場合, 従来の方法では Hazen plot が expected plotting value にかなりよい近似を示し, $1/k' \approx 0.6$ ではほとんど一致しているに対し, Gumbel plot ($1/k \rightarrow 0$) は Thomas plot によく近似しているが, expected plotting value よりの偏りが大きい。

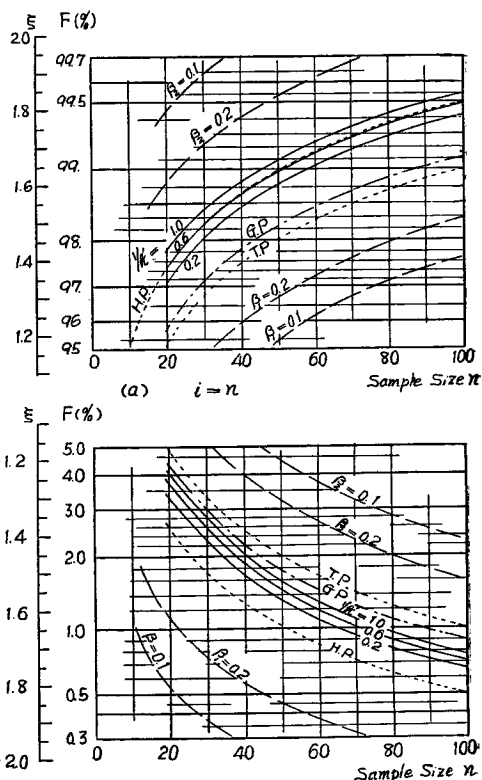


Fig. 4 Plotting values of the Log-Normal and its confidence limits (Provided $i=n$ and 1)

$i=1$ の場合にはこれとは異なり, Hazen plot は expected plotting value よりかなり小さく, 逆に Gumbel plot ($1/k \rightarrow 0$) の方がよい近似を示している。なお同図では示していないが, Gumbel plot で $1/k$ が大きくなると expected plotting value よりの偏りは大きくなる。以上の諸傾向はすべて極値分布の場合と類同である。

6. 水文量の plotting position — 経験的分布関数の推定

一般に観測によつて得られた n ケのデータよりその経験的分布関数 $nG(x)$ を推定するには, 通常⁹⁾ データの値を大きさの順序に列べ, $x_1 \leq x_2 \leq \dots \leq x_n$ とすると,

$$nG(x) = \begin{cases} 0 & ; x < x_1, \\ i/n & ; x_i \leq x \leq x_{i+1}, i=1, 2, \dots, n-1, \\ 1 & ; x_n < x \end{cases} \dots\dots\dots(35)$$

である。これよりすれば, $x_i \leq x \leq x_{i+1}$ となるような x の経験的分布関数が i/n であること, 逆にいえば $(i-1)/n \leq nG(x_i) \leq i/n$ であるということであり, それ以上のことはわからない。むろん真の確率 $F(x_i)$ が必ずしも $nG(x_i)$ というわけではない。とくに最大値の $nG(x_n)$ は $(n-1)/n \leq nG(x_n) < 1$ であつて, これを確率紙上に plot する場合, どこに plot すればよいかは不明である。そこで実際上は, $x_i \leq x \leq x_{i+1}$ なる x を適当に選んで(最大値 x_n は考えないで)これを i/n に対応させて plot し, これに適当に曲線(または直線)を挿入して経験的分布関数とすることがよく行なわれている。そしてこのようにして得られた $nG(x)$ が母分布 $F(x)$ にどれだけ近いかをいろいろ検定する方法がとられているわけである。

もとより上述の方法でよいわけではあるが, このように経験的分布関数を作るにも問題がある。観測によつて得られた x_i , したがつてその経験的分布関数 $nG(x_i)$ もまた一つの統計量であつて, あるいは(11)式

$$dp_i = \frac{\Gamma(n+1)}{\Gamma(i)\Gamma(n-i+1)} F(x_i)^{i-1} [1-F(x_i)]^{n-i} dF(x_i) \dots\dots\dots(11) \text{前出}$$

にしたがつて分布する $F(x_i)$ の実現値とみることができ。さてこのように考えると, $nG(x_i)$ をどのような $F(x_i)$ の実現値とみればよいかという問題になる。したがつて3節と同様の所論により, この場合は分布形に無関係に $F(x_i)$ そのものの分布を考えれば充分であつて,

$$\sum_i (F_i - \hat{F}_i) = \min., \dots\dots\dots(36)$$

したがつて

$$G = \hat{F}_i = E(F_i) = \frac{i}{n+1}, \dots\dots\dots(37)$$

すなわち, Thomas plot をもつて確率紙上の plotting position とすればよい。

この結果は3~5節で述べた母分布の常数を推定するための expected plotting value に対応する $F\{E(x_i)\}$ と異なり一見不合理に見えることであるが, 前者が分布形を既知とする parametric の問題であるに対し, 後者は分布形にとらわれずに経験的分布関数を作る non-parametric の問題であること, あるいは χ^2 -検定にみられるように $Pr(\chi^2 > \chi_0^2)$ がとくに小さいのはもちろん, 大きいのもおかしいとする奇異な問題と一見相通ずる問題のように思われ, 統計学上の本質的な問題の一つと考えるのは不都合であろうか。もとより χ^2 を最小, すなわち経験分布関数によく適合するような分布をもつて母分布とみなせようとする立場をとるならば, 直接経験的分布関数より, あるいは(37)式で plot された諸点を smooth curve で結べばこと足りるはずである。

7. plotting value および plotting position の信頼限界

plotting value および plotting position の問題は元来観測によつて得られた $x_i, (F(x_i))$ が母分布のどの点に対応するかということであつて、これに対し本文では i 番目順序統計量の分布を考え、推定誤りを最小にするには expected plotting value, position をもつてその解とすべきだと論じたわけであるが、もとより、 x_i が常に expected plotting value, position に対応するというものではない。そこで plotting value y_i および plotting position F_i の信頼域をつぎのように考えることができる。

$$\left. \begin{aligned} P_r(y_i \leq y_{\beta_1} \text{ or } F_i \leq F_{\beta_1}) &\leq \beta_1, \\ P_r(y_i \geq y_{\beta_2} \text{ or } F_i \geq F_{\beta_2}) &\leq \beta_2, \end{aligned} \right\} \dots\dots\dots(38)$$

ところでこのような問題は y_i あるいは F_i の分布 (13) 式を考えればよく、(38) 式の解は、

$$\left. \begin{aligned} P_r(y_i \leq y_{\beta_1} \text{ or } F_i \leq F_{\beta_1}) &= \frac{\Gamma(n+1)}{\Gamma(i)\Gamma(n-t+1)} \int_0^{F_{\beta_1}} F^{i-1} (1-F)^{n-t} dF, \\ P_r(y_i \geq y_{\beta_2} \text{ or } F_i \geq F_{\beta_2}) &= \frac{\Gamma(n+1)}{\Gamma(i)\Gamma(n-t+1)} \int_{F_{\beta_2}}^1 F^{i-1} (1-F)^{n-t} dF \end{aligned} \right\} \dots\dots\dots(39)$$

を満足する F_{β_1}, F_{β_2} を求めることによつて得られる。上式は不完全 B-関数表を用いて容易に求めることができる。その主な結果を示すと Fig. 3~5 のようである。なお、 β の値としてはどのような値を指定

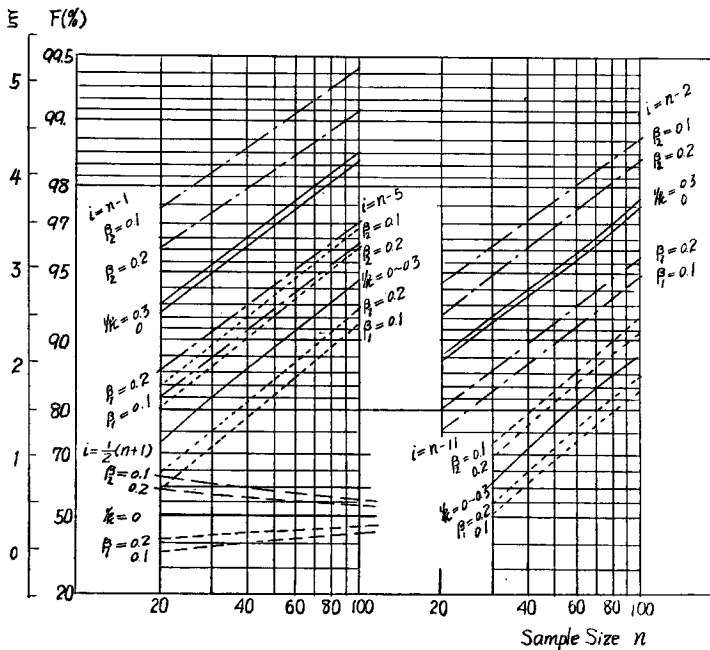


Fig. 5 Confidence limits of plotting values and plotting positions (used the probability paper of the Log-Extremes)

してもよいわけであるが、元来 (13) 式などは $F; 0 \sim 1$ の全領域にわたつて分布することを示しており、 β をあまり小さくするのは必ずしも実用的ではなく、 $\beta=10 \sim 20\%$ ぐらいで充分かと思われる。

なお従来の信頼帯の構成法としては経験的の分布関数を (35) 式より作り、

$$\left. \begin{aligned} \theta_1 &\leq F(x_i) \leq \theta_2, \\ \theta_1 &= {}_nG(x_i) - d_n > 0, \\ \theta_2 &= {}_nG(x_i) + d_n < 1 \end{aligned} \right\} \dots\dots\dots(40)$$

とし、 d_n を信頼係数 $1-\varepsilon$ に対応するものとして、

$$Pr\{\max |{}_nG(x_i) - F(x_i)| > d_n\} = \varepsilon \dots\dots\dots(41)$$

とする方法がある。 d_n の値は A. Kolmogoroff⁹⁾ によれば、次式で与えられている。

$$\left. \begin{aligned} d_n &= \lambda / \sqrt{n} \\ \varepsilon &= \sum_{r=-\infty}^{\infty} (-1)^r \exp(-2r^2\lambda^2). \end{aligned} \right\} \dots\dots\dots(42)$$

ex. $\varepsilon=0.5 ; \lambda=0.828, \quad \varepsilon=0.2 ; \lambda=1.073$

この方法を適用すると、超過確率が d_n より小さな範囲では、信頼係数 $1-\varepsilon$ を要求する限り、その超過確率については論ずることができなくなるので^{10), 11)} 水文統計では利用しがたい。著者の方法はこれらの難点を含まないのが特長である。

8. 本研究の利用について(結語にかえて)

本研究の適用については稿をあらためて論述する予定であるが、本文ではその基本方針について触れておく。

8.1 Gumbel 分布への適用

Gumbel 分布(16)式に含まれる常数 α, x_0 の推定は、

$$\left. \begin{aligned} \alpha &= S_\xi / \sigma_{x_0}, \\ x_0 &= E(x) - \frac{1}{\alpha} E(\xi) \end{aligned} \right\} \dots\dots\dots(43)$$

の形で求めることができるが⁷⁾、前述の結果を利用して、

$$\left. \begin{aligned} \alpha &= S_\xi / S_x, \\ x_0 &= \bar{x} - \frac{1}{\alpha} \bar{\xi}, \end{aligned} \right\} \dots\dots\dots(44)$$

ただし、 \bar{x}, S_x ; x の標本平均、標準偏差である。 $S_\xi, \bar{\xi}$ の表は sample size n の関数として求めることができ、この数表はすでに求めてある。Gumbel¹²⁾ は(37)式を利用する方法を示しているが、もとよりデータと最も適合性のよい分布を採用しようとする場合にはともかく、前述の所論よりすれば必ずしも十分なものとはいえない。

8.2 対数極値分布A型への適用

対数極値分布A型(20)式に含まれる3常数は一応積率法によつて解かれているが⁷⁾、とくにヒズミ係数 C_s と $1/k$ の間には、

$$C_s = \frac{\Gamma_{(1-(3/k))} - 3\Gamma_{(1-(2/k))} \Gamma_{(1-(1/k))} + 2\Gamma_{(1-(1/k))}^3}{[\Gamma_{(1-(2/k))} - \Gamma_{(1-(1/k))}^2]^{3/2}} \dots\dots\dots(45)$$

の関係がある。ところで C_s をデータからどのように推定するか問題があつて、

$$C_s' = \frac{\frac{1}{n} \sum (x - \bar{x})^3}{\left[\frac{1}{n} \sum (x - \bar{x})^2 \right]^{3/2}} \dots \dots \dots (46)$$

をもつて C_s の推定値とするには偏りが大きいので、これには n の関数で表わされる補正係数が必要である。そこで、

$$x = (x_0 + b) \hat{y} - b$$

の関係を利用して、

$$C_s' = \frac{\frac{1}{n} \sum (\hat{y} - \bar{\hat{y}})^3}{\left[\frac{1}{n} \sum (\hat{y} - \bar{\hat{y}})^2 \right]^{3/2}} \dots \dots \dots (47)$$

を求めておけば、

$$\hat{C}_s = C_s' \frac{C_s}{C_s'} = C_s' (1 + \beta_s) \dots \dots \dots (48)$$

より求められる \hat{C}_s をもつて C_s の推定値とすることができる。(45) 式や β_s などの図表類は作製済である。

8.3 対数正規分布への適用

この場合の適用法も 8.2 と全く同様である。対数正規分布 (33) 式の積率解についてはすでに石原・高瀬¹³⁾ によって有用な数表が提示されている。ただヒズミ係数の修正係数 β_s について Hazen Plot を利用した便宜的な図表が併示されているが、Table 1 および Table 2 よりすればこの β_s はいくぶん小さな値を与えているようにも思われる。実用的には大差ないはずとも考えられるが、これについてはあらためて検討したいと考えている。

Table 1 Plotting values of the Log-Normals

i	n	$1/k'$	$\hat{\xi} = k \lg E(y)$	Hazen Plot		Gumbel Plot ($1/k=0$)		Thomas Plot	
				$H\hat{\xi}$	$H\hat{\xi} - \hat{\xi}$	$\sigma\hat{\xi}$	$\sigma\hat{\xi} - \hat{\xi}$	$T\hat{\xi}$	$T\hat{\xi} - \hat{\xi}$
n	100	0.2	1.786		0.035		-0.108		-0.142
		0.6	1.815	1.821	0.006	1.678	-0.137	1.644	-0.171
		1.0	1.845		-0.024		-0.167		-0.201
	40	0.2	1.545		0.040		-0.111		-0.152
		0.6	1.579	1.585	0.006	1.434	-0.145	1.393	-0.185
		1.0	1.617		-0.032		-0.183		-0.224
1	100	0.2	-1.753		-0.068		0.075		0.109
		0.6	-1.736	-1.821	-0.085	-1.678	0.058	-1.644	0.092
		1.0	-1.713		-0.108		0.035		0.069
	40	0.2	-1.509		-0.076		0.075		0.116
		0.6	-1.482	-1.585	-0.103	-1.434	0.048	-1.393	0.089
		1.0	-1.452		-0.133		0.018		0.059

Table 2 Range of plotting values of the Log-Normals

n	1/k'	$\hat{R} = \hat{\xi}_n - \hat{\xi}_1$	Hazen Plot		Gumbel Plot (1/k=0)		Thomas Plot	
			$H\hat{R}$	$H\hat{R} - \hat{R}$	$G\hat{R}$	$G\hat{R} - \hat{R}$	$T\hat{R}$	$T\hat{R} - \hat{R}$
100	0.2	3.539		0.103		-0.183		-0.251
	0.6	3.551	3.642	0.091	3.356	-0.195	3.288	-0.263
	1.0	3.558		0.084		-0.202		-0.270
40	0.2	3.054		0.116		-0.186		-0.268
	0.6	3.060	3.170	0.110	2.868	-0.192	2.786	-0.274
	1.0	3.069		0.101		-0.201		-0.283

参 考 文 献

- 1) A. Hazen : Flood Flow, Wiley (1930).
- 2) F. Mosteller : On Some Usefull Inefficient Statistics, A.M.S. Vol. 17 (1946).
- 3) E. J. Gumbel : Simplified Plotting of Statistical Observations, Trans. A.G.U., Vol. 27 (1945)
- 4) H. A. Thomas : Frequency of Minor Floods, J.B.S.C.E., Vol. 25 (1948).
- 5) 小川潤次郎 : ある Order Statistic の問題について, 統計数理研究所講究録, Vol. 3-13, 14 (1947), ただし 6) による.
- 6) 鈴木栄一 : 河川の流量資料の統計的取扱について, 電力気象連絡会彙報 Vol. 5 (1955).
- 7) M. Kadoya : The Extreme-Value Law and its Some Solution for Engineering Problems, Proc. 6th J.N.C.A.M., (1956).
- 8) 統計科学研究会 : 新編統計数値表, p. 98, 河出 (1952).
- 9) A. Kolmogoroff : Sulla Determinazione empirica di una Legge di Distribuzione. Giornale dell' Istituto degli Attuari, Annv 4 (1933).
- 10) T. Okubo, and H. Nishihara : Primary Power of the Hydro-Electric Power Plant, M.F.E., Kyoto Univ., Vol. 17-2 (1955).
- 11) 沢田敏男・角屋睦 : 水文統計学における 2, 3 の問題, コンクリートダム, Vol. 2-6 (1954).
- 12) E. J. Gumbel : Statistical Theory of Drought, Proc. A.S.C.E., Vol. 80 (1954).
- 13) 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 (1957).