

水文確率分布の適合度と信頼性の評価指標に関する研究

A Study on the Index of Adaptability and Reliability of Hydrological Probability Distribution Model

許士 達広¹・横井 潤²
Tatuhiko KYOSHI and Jyun YOKOI

¹ 北海学園大学工学部 教授

² 北海学園大学工学研究科

要 旨

河川計画に用いられる水文確率値を求める上で、多くの確率分布モデルや母数推定法が利用されている。これらのうち何が最適であるか判定するために、主として SLSC や相関係数などの、データと分布関数の「適合度」を表す評価指標が用いられる。また最近は標本数の違いにより確率値が大きく変動しないようなものが望ましいという考え方があり、ジャックナイフ法などのリサンプリング法による「安定性」の指標も検討されている。

現在、どの確率分布モデルが望ましいか、またどのような評価指標により決定すべきかという問題について明確な答えがないのが実態である。このため、本研究では新たに分布の指標として、直線の信頼限界や母数の誤差分散を用いた「信頼性」の指標を提案する。さらに代表的な 4 つの確率分布について「適合度」、「安定性」、「信頼性」の各指標を計算して、主成分分析などにより各確率分布や評価指標の特性や類似性について考察を行なうものである。

《キーワード：水文確率分布；評価指標；信頼限界》

1. はじめに

現在の河川流域における様々な水工計画は、所要の安全度のもとに降雨量や河川流量データの確率統計解析を行い、その統計量（水文量確率値）を定めたうえで策定されている。この水文量確率値を求める上で、多くの確率分布モデルや母数推定法などが用いられている。これらの検討した水文量確率値のうちどれが最適であるかが水工計画上の大きな課題であり、それには最適化のための手法および評価指標が重要となる。

評価指標としては、一般的に確率分布関数の原データへの分布の「適合度」を重視するものとして、SLSC（標準最小二乗規準）や相関係数（X-COR, P-COR）、AIC などが用いられている。また、実際の計画においては標本値および標本数 N の違いによる確率値の変動が小さいこと（変動性最小）が望まれる。これをここでは分布の「安定性」と呼ぶこととし、これには、Jackknife 法やブートストラップ法といったリサンプリング法の推定誤差分散の指標が用いられている。

どの確率分布モデルが最適かということに対して様々な研究がなされ、客観的・定量的な評価について近年いくつかの提案がされているが明確な答えはなく、実用的には分布の「適合度」に関する指標 SLSC を一定の条件におきつつ、観測データ毎に各種のモデル計算を行い、いくつかの手法の結果を比較して水文量確率値の決定を行っているのが実情である。

本研究では、これら分布の適合度と安定性の両方に関係する第 3 の指標として、「水文量確率値の信頼区間」に着目し、分布の「信頼性」に関する新たな指標を提案する。さらに、確率分布の「適合度」および「安定性」、「信頼性」に関わる評価指標を算出し、各指標の水文量確率分布に対する適用性および各確率分布の指標に対する特性について比較して考察を行う。

2. 研究方法

本研究では評価指標の比較のため、既往の日雨量毎年最大値データに対し各確率分布モデルを用いて、各確率分布および評価指標の特性・類似性・関連性を検討する。評価指標は以下のものを対象とする。

- ①分布の適合度として、標準変量の理論値と実測値の相対誤差を表す「SLSC（標準最小二乗規準）」、相関係数「X-COR」および「P-COR」を用いる。
- ②水文量確率値の信頼区間について、水文量確率値と標準変量の直線回帰の考え方を用いて、「水文量確率値の信頼区間の指標 CI 」を提案し、算出する。
- ③水文量確率値の信頼区間は確率分布の母数の推定誤差の分散・共分散に依存することから「母数の誤差分散を利用した信頼区間の推定誤差 S_T 」を算出する。
- ④分布の安定性として、「Jackknife 法によるリサンプリング計算」を行い確率水文量の推定誤差分散を算出する。これは「水文量確率値の信頼区間」を表す指標でもあったと考えられる。

確率分布については、河川計画に用いられる降雨量などの水文量確率値を求める上で、現在用いられている水文量確率分布モデルの大半は、対数正規分布、極値分布、ガンマ分布の三種類に大別される。本研究では、それぞれの代表的な方法として、3 母数対数正規分布、グンベル分布、GEV 分布、および対数ピアソンⅢ型分布を検討対象とする。また、これらの母数の算定には 3 母数対数正規分布および対数ピアソンⅢ型分布は積率解を、グンベル分布および GEV 分布には L 積率解を用いる。

検討に使用するデータについては、北海道内 8 ヶ所（札幌市、旭川市、函館市、帯広市、釧路市、網走市、根室市、寿都町）の気象台観測の日雨量毎年最大値データを用いることとする。

3. 検討対象とする確率分布モデル

(1) 確率分布^{1) 6) 10)}

a) 3 母数対数正規分布

水文諸量のヒストグラムは、多くが単峰な非対称分布を示す。このような場合に、対数変換を用いると近似的に正規分布にみなすことができ、正規分布に関する多くの確率分布特性が利用できる。このとき、対数正規分布の確率密度関数は以下のように表される。

$$f(x) = \frac{1}{(x-a)\sqrt{2\pi}\sigma_y} \exp\left[-\frac{1}{2}\left(\frac{\ln(x-a)-\mu_y}{\sigma_y}\right)^2\right] \quad (x > a) \quad (3-1)$$

$$\text{ただし、水文学の対数 } y = \ln(x-a) \quad (3-2)$$

$$\text{また、この確率値 } y_p \text{ は標準変量 } Z_p \text{ の一次式、 } y_p = \mu_y + \sigma_y \cdot Z_p \quad (3-3)$$

x : 水文学量, a : 分布下限値, Z_p : 標準変量, μ_y : y の平均, σ_y : y の標準偏差で表すことができる。

ここで、標準変量 Z_p を求めるには以下の式を用いる。

$$Z_p = -\sqrt{\frac{t^2 [(4t+100)t+205]}{[(2t+56)t+192]t+131}} \quad (3-4)$$

非超過確率 $p \leq 0.5$ のとき

$$\text{係数 } t = -\ln(2p) \quad (3-5)$$

非超過確率 $p > 0.5$ のときは、式(3-5)が $t = -\ln[2(1-p)]$ となり、その時の標準変量 Z_p は、式(3-4)を $Z_p = -Z_{1-p}$ へ置き換えることにより求めることができる。

b) グンベル分布

最大値分布の第 I 型形式として、最大洪水流量の分布関数に用いることの有用性が知られている、グンベル分布の確率密度関数は以下のように表される。

$$f(x) = \frac{1}{a} \exp\left[-\frac{x-a}{a} - \exp\left(\frac{x-c}{a}\right)\right] \quad (-\infty < x < \infty) \quad (3-6)$$

上式中の各項目として、 x : 水文学量, a : 尺度母数, c : 位置母数

また、水文学量 x の確率値 x_p は標準変量 G_p の一次式で表すことができる。

$$x_p = c + a \cdot G_p \quad (3-7)$$

ここで、標準変量 G_p は非超過確率 p を用いて、次式で表される。

$$G_p = -\ln[-\ln(p)] \quad (3-8)$$

c) 対数ピアソンⅢ型分布

水文量の対数変換により水文量 x の対数 y がピアソンⅢ型分布に従うとき、分布は対数ピアソンⅢ型分布となり、そのときの確率密度関数は以下のように表される。

$$f(x) = \frac{1}{|a|\Gamma(b)x} \left(\frac{\ln x - c}{a} \right)^{b-1} \exp \left(-\frac{\ln x - c}{a} \right) \quad (3-9)$$

上式中の各項目として、 x ：水文量、 a ：尺度母数、 b ：形状母数、 c ：位置母数、 $\Gamma(\cdot)$ ：ガンマ関数

$$\text{ただし、水文量の対数変換は } y = \ln x \text{ である。} \quad (3-10)$$

この確率値 y_p は、標準変量 K_p の一次式で表すことができる。

$$y_p = \mu_y + \sigma_y \cdot K_p \quad (3-11)$$

このとき μ_y 、 σ_y が y の平均および標準偏差である。標準変量 K_p は次式で求められる。

$$K_p = \frac{2}{\gamma_y} \left(1 + \frac{\gamma_y Z_p}{6} - \frac{\gamma_y^2}{36} \right)^3 - \frac{2}{\gamma_y} \quad (3-12)$$

ここで、 γ_y ：ひずみ係数、 Z_p ： $N(0,1)$ に従う標準正規変量、 p ：非超過確率

d) GEV 分布（一般化極値分布）

3 種類の極値分布（グンベル分布、対数極値分布 A 型および B 型）を 1 つの式形に統一したものが GEV 分布（一般化極値分布）と呼ばれており、累積分布関数は以下のように表される。

$$F(x) = \exp \left\{ - \left[1 - \frac{k(x-c)}{a} \right]^{1/k} \right\} \quad (k \neq 0) \quad (3-13)$$

上式中の各項目として、 x ：水文量、 a ：尺度母数、 c ：位置母数、 k ：形状母数

また、水文量 x の確率値 x_p は、標準変量 GV_p の一次式で表すことができる。

$$x_p = c + \left(\frac{a}{k} \right) \cdot GV_p \quad (3-14)$$

この標準変量 GV_p は、非超過確率 p を用いて次式より求められる。

$$GV_p = 1 - \{ -\ln(p) \}^k \quad (3-15)$$

このように、これらの検討対象確率分布における水文量 x またはその対数 y の確率値は、一般的に標準変量 Z_p 、 G_p または K_p 、 GV_p の一次式で表され、直線回帰の問題として考えることができる。

(2) 水文量確率値

ここでは、前述した 4 つの確率分布モデルによる水文確率値について、比較したものを以下に示す。水文量確率は 1/100, 1/400 とし、北海道内 8 ヶ所（札幌市、旭川市、函館市、帯広市、釧路市、網走市、根室市、寿都町）の気象台観測の日雨量毎年最大値データから、最近の 20 年、40 年、60 年、100 年の期間のデータを用いて、各確率分布別および確率年別にその算定を行った。

表 3-1 確率分布モデルより算定した北海道内 8 ヶ所の水文量確率値

場所	データ年数	20ヶ年		40ヶ年		60ヶ年		100ヶ年	
	確率年数	1/100	1/400	1/100	1/400	1/100	1/400	1/100	1/400
札幌市	3母数対数正規分布	143.9	175.7	86.9	93.5	87.8	95.3	155.4	184.0
	Gumbel分布	141.7	168.7	92.5	102.9	94.6	106.8	158.3	188.4
	対数ピアソンIII型分布	155.4	199.0	89.0	97.3	90.7	100.3	166.3	205.9
	GEV分布	155.6	199.1	90.3	98.6	91.1	100.0	169.4	212.3
旭川市	3母数対数正規分布	88.7	103.1	83.0	90.7	137.7	169.7	173.9	222.8
	Gumbel分布	91.6	107.2	94.0	109.1	131.3	156.6	157.2	189.3
	対数ピアソンIII型分布	92.1	109.3	85.7	94.7	146.9	190.6	185.8	252.7
	GEV分布	91.4	106.8	83.2	89.3	149.0	196.5	188.5	262.7
函館市	3母数対数正規分布	121.2	130.9	121.6	132.5	119.5	131.1	119.1	130.5
	Gumbel分布	140.9	163.5	141.6	165.7	135.6	158.0	133.8	155.3
	対数ピアソンIII型分布	126.3	138.3	123.3	133.7	121.5	133.3	121.3	133.5
	GEV分布	120.8	127.5	121.4	129.3	120.2	129.7	120.7	131.0
帯広市	3母数対数正規分布	124.1	140.9	122.8	134.8	120.1	132.1	118.3	129.4
	Gumbel分布	127.9	146.5	136.5	157.8	133.9	155.4	132.6	153.5
	対数ピアソンIII型分布	126.3	144.9	125.7	139.2	123.7	137.5	120.6	132.5
	GEV分布	128.9	148.6	125.6	137.2	121.6	132.3	119.2	128.8
釧路市	3母数対数正規分布	119.4	132.2	120.3	129.9	117.9	127.0	115.9	124.5
	Gumbel分布	132.7	154.3	140.7	163.5	136.1	157.2	133.2	153.3
	対数ピアソンIII型分布	120.8	134.1	122.4	131.8	119.7	128.9	118.3	127.5
	GEV分布	120.2	131.0	119.4	125.8	118.1	124.9	115.5	121.7
網走市	3母数対数正規分布	71.4	76.0	104.3	117.3	113.2	132.4	107.1	124.4
	Gumbel分布	78.4	88.0	113.4	133.0	114.9	134.9	109.5	128.4
	対数ピアソンIII型分布	69.5	72.8	104.9	117.5	119.3	144.6	110.5	131.2
	GEV分布	72.4	76.9	109.6	125.5	123.2	152.8	113.1	135.9
根室市	3母数対数正規分布	106.5	116.5	105.4	112.6	142.7	160.7	151.8	176.9
	Gumbel分布	116.9	133.9	119.6	136.2	159.5	189.1	157.0	185.5
	対数ピアソンIII型分布	112.6	127.0	109.0	118.1	151.8	175.1	160.8	194.4
	GEV分布	112.1	124.5	105.3	110.5	147.9	166.7	160.9	193.5
寿都町	3母数対数正規分布	93.9	101.3	87.4	95.0	104.0	114.1	113.2	128.6
	Gumbel分布	108.9	126.0	98.5	113.5	117.5	136.7	120.6	141.1
	対数ピアソンIII型分布	97.2	105.9	89.4	97.9	107.7	119.6	119.7	140.6
	GEV分布	91.1	95.1	89.0	95.9	105.6	114.6	119.2	138.1
平均	3母数対数正規分布	108.6	122.1	104.0	113.3	117.9	132.8	131.8	152.6
	Gumbel分布	117.4	136.0	117.1	135.2	127.9	149.3	137.8	161.9
	対数ピアソンIII型分布	112.5	128.9	106.2	116.3	122.7	141.2	137.9	164.8
	GEV分布	111.6	126.2	105.5	114.0	122.1	139.7	138.3	165.5

北海道内 8 ヶ所の水文確率値の平均値を、確率年 1/100, 1/400 について 4 つの確率分布別に棒グラフ（図 3-1, 図 3-2）で示す。棒グラフの上の数字は、データ年数毎の 4 つの確率分布間における水文量確率値の大きさの順位を表し、その順位は値の小さいものから 1, 2, 3, 4 となっている。

表 3-1, 図 3-1, 図 3-2 から、データ年数毎の水文量確率値の大きさの傾向をみると、確率年に関係なく 20, 40, 60 ヶ年では小さい順に 3 母数対数正規分布, GEV 分布, 対数ピアソンⅢ型分布, グンベル分布の順であり、データ年数 100 年のみグンベル分布と, GEV 分布が入れ替わっている。

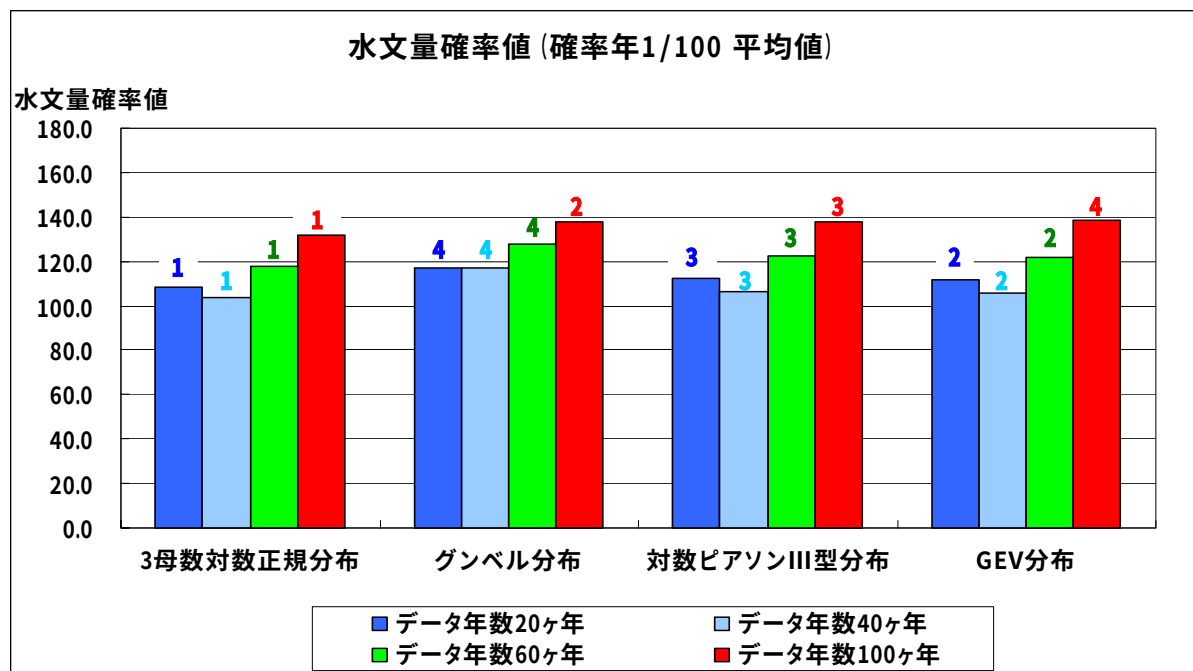


図 3-1 各確率分布におけるデータ年数別の水文量確率値（平均値）（確率年 1/100）

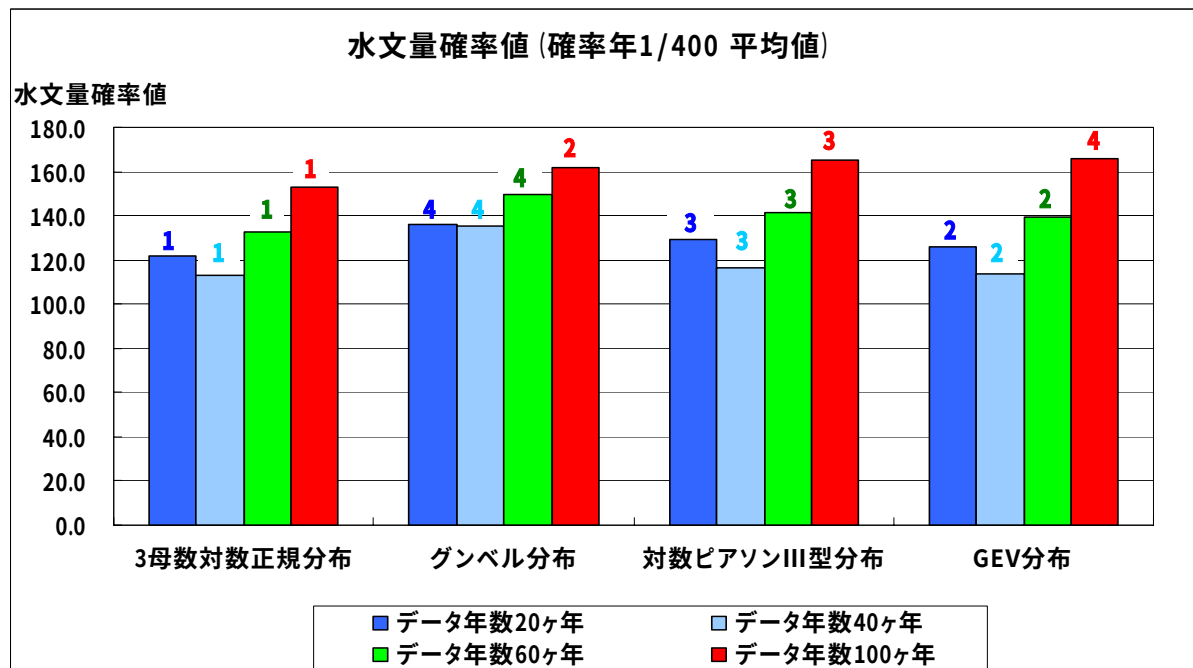


図 3-2 各確率分布におけるデータ年数別の水文量確率値（平均値）（確率年 1/400）

4. 適合度の評価指標

(1) 確率分布の適合度を表す指標^{1) 2)}

a) SLSC (標準最小二乗規準 Standard Least-Squares Criterion)

SLSC は、確率分布モデルのデータへの適合度を表す評価基準として、次式によって定義される。

$$SLSC = \frac{\sqrt{\zeta^2}}{|S_{0.99} - S_{0.01}|} \quad (4-1)$$

$$\zeta^2 = \frac{1}{N} \sum_{i=1}^N (S_i - r_i)^2 \quad (4-2)$$

$S_{0.99}$, $S_{0.01}$: 非超過確率を 0.99 及び 0.01 とした時の当該確率分布の標準変量

S_i : 順序統計量データを推定母数で変換した標準変量

r_i : プロットングポジションに対応した理論確率水文量を推定母数で変換した標準変量

SLSC はその算出される値が小さいほどよく適合していると考えられ、従来 SLSC 0.02 以下であれば分布の適合度は良好であり、SLSC > 0.03 であれば他の分布形ないし母数推定法を考慮すべきとされている。しかし、近年の研究において宝らは、河川流量において満足すべき適合度の基準として SLSC 0.04 以下とすることが適当であるとしている⁷⁾。本研究においても、北海道内 8 ヶ所の日雨量毎年最大値データをもとに SLSC を算出したが、次ページ表 4-1 に示すように 8 箇所を平均した値では SLSC はすべての分布で 0.02 以上である。

b) 重相関係数(X-COR および P-COR)

重相関係数は、同じプロットングポジションの原データの順序統計量 x_i と理論統計量 x_i^* 、および同じプロットングポジションの原データの非超過確率 p_i と理論統計量の非超過確率 p_i^* の相関関係を表すものである。

$$COR = \frac{\sum_{i=1}^N (X_i - \bar{X})(Y_i - \bar{Y})}{\left[\left\{ \sum_{i=1}^N (X_i - \bar{X})^2 \right\} \left\{ \sum_{i=1}^N (Y_i - \bar{Y})^2 \right\} \right]^{1/2}} \quad (4-3)$$

ここに、 N は資料の個数、 (X_i, Y_i) は i 番目における原データと、理論統計量の組の値

$\bar{X}$, $\bar{Y}$ はそれぞれの変数の平均値

式(4-3)において、それぞれの因子 (X_i, Y_i) の取り方により X-COR, P-COR が使われる。上述の(4-3)のような一般的な相関係数の式において、X-COR は X に順序統計量(原データ) x_i 、 Y にその理論値(推定値) $\hat{x}_i$ を置いた場合の相関を表していることになる。また P-COR は、 X を原データ x_i のプロットングポジションに対する非超過確率 p_i 、 Y を x_i から理論式で逆算した推定非超過確率 $\hat{p}_i$ とした場合の相関を表している。

X-COR, P-COR の両指標とも、値の絶対値が大きく 1 に近いものが良好と判断される。重相関係数が 1 の時は図上では $y = x$ すなわち 45° の直線上となる。この X-COR および P-COR は、実際の水文量確率値推定において、検討分布の適合度の判断のための指標として SLSC 同様に用いられる。

(2) 確率分布の適合度を表す指標の比較(SLSC, X-COR および P-COR)

ここでは、北海道内 8 ヶ所の最近 20 年データおよび 100 年データを用いて、分布の適合度の評価指標として用いられる SLSC, X-COR および P-COR の 3 指標の各確率分布およびプロットングポジション方法ごとの値を算出し、それらの平均値から 3 指標間の比較を行う。

表 4-1 確率分布の適合度を表す 3 指標の比較(SLSC, X-COR, P-COR)

データ年数	確率分布	SLSC			X-COR			P-COR		
		Cunnane	Hazen	Weibull	Cunnane	Hazen	Weibull	Cunnane	Hazen	Weibull
20	3母数対数正規分布	0.040	0.041	0.043	0.983	0.982	0.984	0.986	0.986	0.986
	グンベル分布	0.042	0.045	0.043	0.973	0.971	0.977	0.985	0.985	0.985
	対数ピアソンIII型分布	0.038	0.038	0.041	0.982	0.980	0.984	0.986	0.986	0.986
	GEV分布	0.037	0.039	0.038	0.984	0.982	0.985	0.987	0.987	0.987
100	3母数対数正規分布	0.025	0.026	0.023	0.993	0.992	0.994	0.996	0.996	0.996
	グンベル分布	0.037	0.039	0.034	0.983	0.981	0.986	0.995	0.995	0.995
	対数ピアソンIII型分布	0.022	0.023	0.021	0.991	0.990	0.994	0.997	0.997	0.997
	GEV分布	0.028	0.030	0.023	0.992	0.990	0.994	0.997	0.997	0.997

以下のグラフは、データ年数 100 年の Cunnane 式について 3 指標の比較を表した例である。棒グラフの上の数字は、3 指標 (SLSC, X-COR, P-COR) ごとにその大きさについて分布の順位を表している。X-COR および P-COR については、値が大きい方が良好であるため、順位は値の大きいものから 1, 2, 3, 4 となっている。SLSC はその逆で、値の小さいものが良好と判断できるため、順位は値の小さいものから 1, 2, 3, 4 となっている。

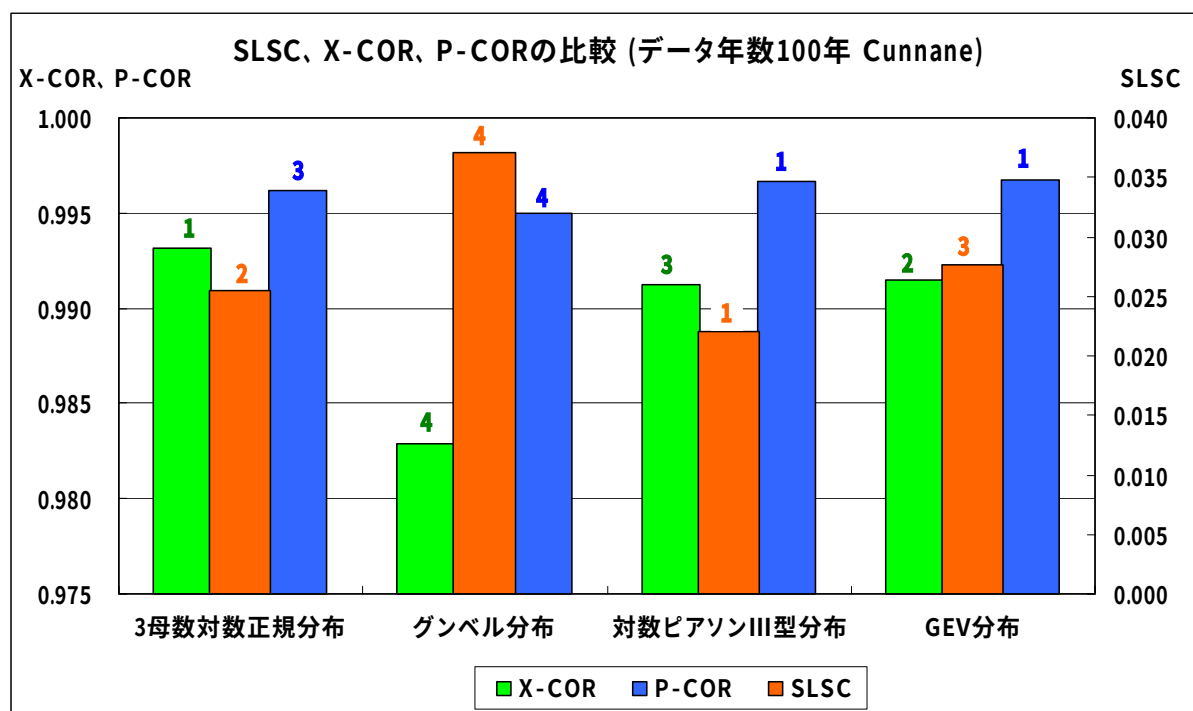


図 4-1 分布の適合度を表す指標 SLSC, X-COR, P-COR の比較(データ年数 100 ヶ年 Cunnane)

表 4-1, 図 4-1 より、確率分布全体の値の大きさの順位から傾向を見ると、X-COR ではデータ年数 100 年で 3 母数対数正規分布, GEV 分布の順, データ年数 20 年では GEV 分布, 3 母数対数正規分布の順に適合度が良好であり、グンベル分布が最小で適合度が劣っている。P-COR については同様に、GEV

分布と 3 母数対数正規分布が良好であり、グンベル分布が最小となっているが値の差は小さい。SLSC については、データ年数 100 年では、対数ピアソンⅢ型分布が良好で、グンベル分布が最下位となっている。データ年数 20 年では GEV 分布と対数ピアソンⅢ型分布が同程度で良好である。全体的にみて、グンベル分布の適合度が悪いほかは 3 分布の適合度に大きな差は無い。またプロットングポジションの違いによる順位の違いはあまり見られない。

5. 水文量確率値の信頼区間

(1) 水文量確率値における推定誤差分散と信頼区間の関係^{1) 10)}

水文量確率値の正確さの評価・推定精度を表す指標として、標本から算定される水文量確率値に対する信頼区間および推定誤差分散 $Var(\hat{x}_p)$ を用いることができる⁹⁾。この推定誤差分散 $Var(\hat{x}_p)$ の平方根は Standard Error (推定誤差) $\sqrt{Var(\hat{x}_p)}$ と言われ、標本、確率分布形、母数推定法により異なる値をとる。これは現在ほとんど実用されていないが、適合度とは違った「信頼性」の観点からの評価指標として、本研究で検討する。

平均値 μ_x の母集団から無作為に抽出した変数 x を n 個抽出したときに、標本平均 $\bar{x}$ 、標本から求まる標準偏差が S_x であるならば、母平均 μ_x の $100(1-\alpha)\%$ の信頼区間は、次のように表される。

$$\bar{x} - \frac{S_x}{\sqrt{n}} t_{1-\alpha/2, n-1} \leq \mu_x \leq \bar{x} + \frac{S_x}{\sqrt{n}} t_{1-\alpha/2, n-1} \quad (5-1)$$

ここで、 $t_{1-\alpha/2, n-1}$ は自由度 $n-1$ に対するスチューデントの t 分布の $100(1-\alpha)\%$ の臨界値である。また、 $S_x/\sqrt{n}$ は、平均値 μ_x の推定値 $\bar{x}$ の分散の平方根を表しており、標本平均の $\sqrt{Var(\hat{x}_p)}$ を推定している。一般的に、水文量確率値 $\hat{x}_p$ が推定誤差分散 $Var(\hat{x}_p)$ をもつ t 分布に従うのであれば、そのとき、確率値の $100(1-\alpha)\%$ 信頼区間は、次のように表わすことができる。

$$\hat{x}_p - t_{1-\alpha/2} \times \sqrt{Var(\hat{x}_p)} \leq x_p \leq \hat{x}_p + t_{1-\alpha/2} \times \sqrt{Var(\hat{x}_p)} \quad (5-2)$$

$\sqrt{Var(\hat{x}_p)}$: 推定誤差

式(5-2)は、分布の水文量確率値およびパラメータから、近似的に信頼区間を推定する式である。なお、標本数が大きい場合、式(5-2)におけるスチューデントの t 分布は正規分布に近い値となり、 $t_{1-\alpha/2, n-1}$ は $Z_{1-\alpha/2}$ に置き換えることができる。この誤差分散は水文確率値の「信頼性」を表す指標と考えられるが、推定誤差 $\sqrt{Var(\hat{x}_p)}$ の算出にはいくつかの方法が考えられ、分布の「安定性」に関連する指標もある。次章以降それらについて検討する。

6. 確率分布の安定性を表す指標

(1) Jackknife 法による誤差分散^{1) 2)}

水文資料の蓄積が進んでいくなかで、河川計画の面では、新たな標本が加わることによって水文量確率値が大きく変動しないような確率分布モデルと母数推定法を選定することが望まれている。誤差分散は新たなデータの付け加えによる確率値の変動性を表し、信頼性ととも安定性に関する指標としても考えられる。特に安定性を示す誤差分散算定の方法として、リサンプリング法がある。リサンプリング法は、1 組のデータセット (標本) から部分的にデータの抽出をしたり、繰返しを許して元の標本と同

じデータ個数だけ抽出するという操作を反復して多数のデータセットを作り、元の標本に対して存在する統計量の偏倚を補正し、統計量の推定誤差を求める手法である。代表的なリサンプリング法として Jackknife 法があげられる。Jackknife 法による誤差分散算定の流れをフローチャートで表わすと以下の図のようになる。

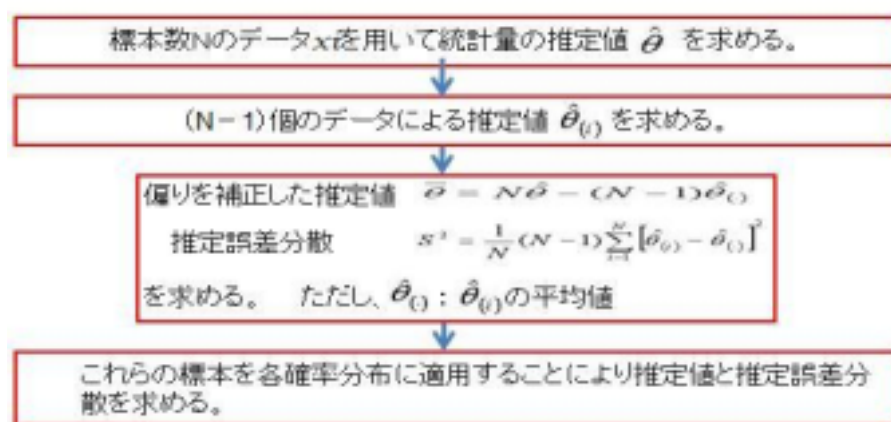


図 6-1 Jackknife 法による計算の流れ

Jackknife 法として、データを 1 つ減らした数で確率値を求めたものの分散を誤差分散とし、この値が小さい方が信頼性は高いと考えられる。また、データが入れ替わった場合の確率値の変動を示すので、分布の安定性を示す指標として用いられる。

(2) 各確率分布における Jackknife 誤差分散の比較

ここでは、最近のデータ 20 年、100 年および確率年 1/100、1/400 の条件で算出した 3 母数対数正規分布、グンベル分布、対数ピアソンⅢ型分布、GEV 分布の Jackknife 誤差分散の値を比較する。結果は表 6-1 であり、このうち確率年 1/100 の場合を図 6-2 に示す。

表 6-1 各確率分布における Jackknife 誤差分散

確率年	確率分布	Jackknife 誤差分散			
		3母数対数正規分布	グンベル分布	対数ピアソンⅢ型分布	GEV分布
1/100	データ年数				
	データ20年	10.20	9.91	12.89	13.98
	データ100年	5.55	5.74	7.51	7.98
1/400	データ20年	15.28	12.51	21.25	24.21
	データ100年	7.83	7.29	12.20	13.90

図 6-2 の棒グラフの上の数字は、Jackknife 誤差分散の値の大きさの確率分布間における順位を表している。Jackknife 誤差分散については、値が小さい方が分布の安定性が良好と判断できるため、順位は値の小さいものから 1, 2, 3, 4 としている。グンベル分布は 20 年の 1/100 と、20 年と 100 年の 1/400 で最小となり、良好である。3 母数対数正規分布もグンベル分布に次いで、値が小さく良好である。対数ピアソンⅢ型分布はデータ 20 年、100 年とも、各確率分布間で順位が 3 番目であり、GEV 分布がいずれも最大となる。グンベル分布と GEV 分布は同じ極値分布であるが、Jackknife 誤差分散は正反対の性質を示している。データ年数による違いの比較では、データ 20 年の値がいずれのケースもデータ 100 年よりも大きい。

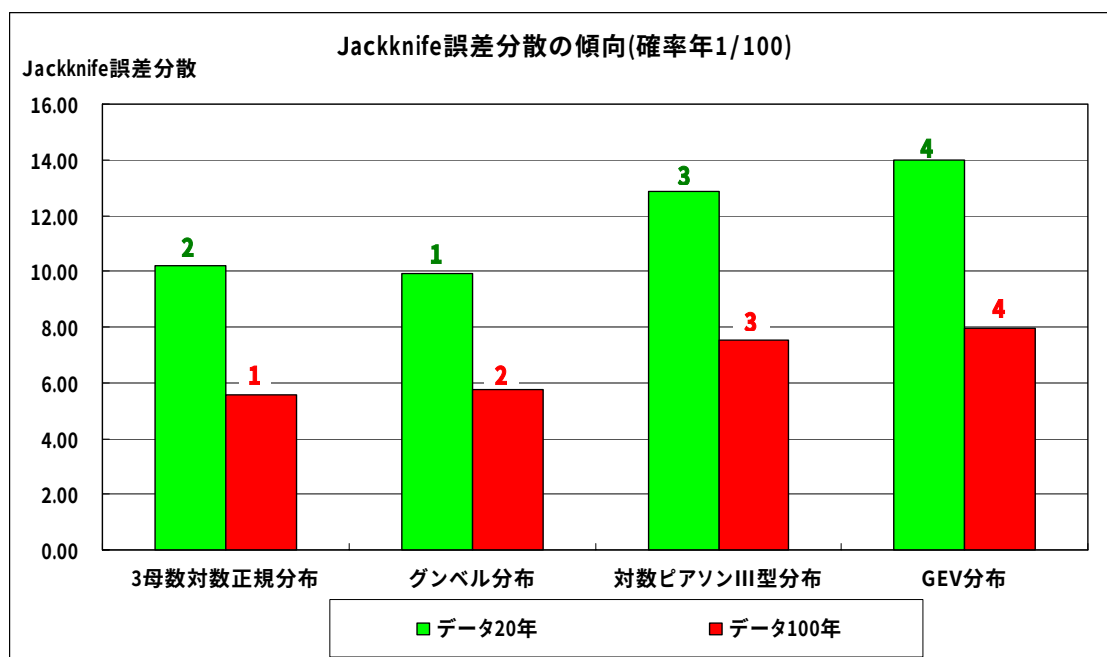


図 6-2 各確率分布における Jackknife 誤差分散の比較(確率年 1/100)

7. 母数推定誤差分散からの確率値の推定誤差

(1) 推定誤差 S_T の基本式⁹⁾

水文量確率値 x_p の $100(1-\alpha)\%$ 信頼区間は、確率値の誤差分散 $Var(\hat{x}_p)$ を用いて、

$$\hat{x}_p \pm t_{1-\alpha/2} \times \sqrt{Var(\hat{x}_p)} \quad (7-1)$$

で表される。

誤差分散 $Var(\hat{x}_p)$ は水文量確率値を推定する分布の母数の分散・共分散に依存し、それらから近似的に推定することができる。これによって求める誤差分散 $Var(\hat{x}_p)$ を、ここでは S_T^2 と表す。一般的に、再現期間 T 年の確率値 x_T の誤差分散は積率の推定値を用いて以下のように導かれる。

もし確率値 x_T が、分布における 1 次、2 次、3 次の積率と再現期間 T 年を用いて、以下のように表されるとする。

$$x_T = f(m_1', m_2, m_3, T)$$

ただし、 m_1' は原点周りの積率、 m_r は平均値周りの積率である。

このとき、確率値 x_T の誤差分散 S_T^2 は、分散 Var 、共分散 cov を用いて、

$$\begin{aligned} S_T^2 = & \left(\frac{\partial x_T}{\partial m_1'} \right)^2 \text{var } m_1' + \left(\frac{\partial x_T}{\partial m_2} \right)^2 \text{var } m_2 + \left(\frac{\partial x_T}{\partial m_3} \right)^2 \text{var } m_3 + 2 \frac{\partial x_T}{\partial m_1'} \cdot \frac{\partial x_T}{\partial m_2} \text{cov}(m_1', m_2) \\ & + 2 \frac{\partial x_T}{\partial m_1'} \cdot \frac{\partial x_T}{\partial m_3} \text{cov}(m_1', m_3) + 2 \frac{\partial x_T}{\partial m_2} \cdot \frac{\partial x_T}{\partial m_3} \text{cov}(m_2, m_3) \end{aligned} \quad (7-2)$$

となる。実際には確率値 x_T は積率を用いて、

$$x_T = m_1' + K \sqrt{m_2} \quad (7-3)$$

のように表される．ここで， K はひずみ係数 g の関数で， $g = m_3 / m_2^{3/2}$ である．
したがって，

$$\frac{\partial x_T}{\partial m_1'} = 1 \quad \frac{\partial x_T}{\partial m_2} = \frac{K}{2\sqrt{m_2}} + \sqrt{m_2} \frac{\partial K}{\partial m_2} = \frac{K}{2\sqrt{m_2}} - \frac{3g}{2\sqrt{m_2}} \cdot \frac{\partial K}{\partial g}$$

また，

$$\frac{\partial x_T}{\partial m_3} = \frac{\partial x_T}{\partial K} \cdot \frac{\partial K}{\partial g} \cdot \frac{\partial g}{\partial m_3} = \frac{1}{m_2} \cdot \frac{\partial K}{\partial g}$$

よって，誤差分散 S_T^2 は，

$$\begin{aligned} S_T^2 = \text{var } m_1' &+ \left[\frac{1}{2\sqrt{m_2}} \left(K - 3g \frac{\partial K}{\partial g} \right) \right]^2 \text{var } m_2 + \frac{1}{m_2^2} \left(\frac{\partial K}{\partial g} \right)^2 \text{var } m_3 + \frac{1}{\sqrt{m_2}} \left(K - 3g \frac{\partial K}{\partial g} \right) \text{cov}(m_1', m_2) \\ &+ \frac{2}{m_2} \cdot \frac{\partial K}{\partial g} \text{cov}(m_1', m_3) + \frac{1}{m_2^{3/2}} \cdot \frac{\partial K}{\partial g} \left(K - 3g \cdot \frac{\partial K}{\partial g} \right) \text{cov}(m_2, m_3) \end{aligned} \quad (7-4)$$

ここで，既往の研究により，母集団のモーメント

$$\mu_r' = \int_{-\infty}^{\infty} x^r p(x) dx \quad \mu_r = \int_{-\infty}^{\infty} (x - \mu_r')^r p(x) dx$$

を用いて，以下のように置き換えることができる．

$$\begin{aligned} \text{var } m_1' &= \mu_2 / n & \text{var } m_2 &= \frac{1}{n} \{ \mu_4 - \mu_2^2 \} \\ \text{var } m_3 &= \frac{1}{n} \{ \mu_6 - \mu_3^2 - 6\mu_4\mu_2 + 9\mu_2^3 \} & \text{cov}(m_1', m_2) &= \mu_3 / n \\ \text{cov}(m_1', m_3) &= \frac{1}{n} \{ \mu_4 - 3\mu_2^2 \} & \text{cov}(m_2, m_3) &= \frac{1}{n} \{ \mu_5 - 4\mu_3\mu_2 \} \end{aligned}$$

さらに，以下のようにひずみ度，尖り度などを用いて，単純化する．

$$\begin{aligned} \gamma_1 &= \mu_3 / \mu_2^{3/2} & \gamma_2 &= \mu_4 / \mu_2^2 \\ \gamma_3 &= \mu_5 / \mu_2^{5/2} & \gamma_4 &= \mu_6 / \mu_2^3 \end{aligned}$$

結果として，再現期間 T 年の確率値 x_T の誤差分散 S_T^2 は，以下のように表される．

$$\begin{aligned} S_T^2 &= \frac{\mu_2}{n} \left\{ 1 + K\gamma_1 + \frac{K^2}{4} [\gamma_2 - 1] + \frac{\partial K}{\partial \gamma_1} [2\gamma_2 - 3\gamma_1^2 - 6 + K(\gamma_3 - 6\gamma_1\gamma_2 / 4 - 10\gamma_1 / 4)] \right. \\ &\quad \left. + \left(\frac{\partial K}{\partial \gamma_1} \right)^2 [\gamma_4 - 3\gamma_3\gamma_1 - 6\gamma_2 - 9\gamma_1^2\gamma_2 / 4 + 35\gamma_1^2 / 4 + 9] \right\} \end{aligned} \quad (7-5)$$

これはどのような再現期間 T 年に対しても、それぞれの確率分布のデータ n 、2 次モーメント μ^2 、 $\gamma_1 \sim \gamma_4$ 、係数 K 、 $\partial K / \partial \gamma_1$ を用いて推定誤差 S_T が求まることを意味している。
 もしも、係数 K が、ひずみ係数 γ_1 に依存しない確率分布では

$$S_T = \delta \sqrt{\frac{\mu_2}{n}} \quad (7-6)$$

となる。

$$\text{ここで、係数 } \delta \text{ は、} \delta = \left\{ 1 + \gamma_1 K + (\gamma_2 - 1) \frac{K^2}{4} \right\}^{1/2} \quad (7-7)$$

で表される。例えば、正規分布においては、

$$\begin{aligned} \mu_2 &= \sigma^2, & \mu_3 &= 0, & \mu_4 &= 3\sigma^2 \text{ (ここで、} \sigma \text{: 母標準偏差) より} \\ \gamma_1 &= \mu_3 / \mu_2^{3/2} = 0 & \gamma_2 &= \mu_4 / \mu_2^2 = 3\sigma^2 / \sigma^2 = 3 \text{ となることから、係数 } \delta \text{ は,} \end{aligned}$$

$$\delta = \left\{ 1 + \frac{K^2}{2} \right\}^{1/2} \quad (7-8)$$

となる。ここで、係数 K は、正規分布の標準変量 Z_p であり、また、 μ_2 を標本分散 S_x^2 とすると、正規分布における確率値 x_p の誤差分散は、

$$S_T^2 = \frac{S_x^2}{n} \left(1 + \frac{Z_p^2}{2} \right) \quad (7-9)$$

となる。

(2) 各分布の推定誤差の算定方法^{9) 10) 11)}

a) 3 母数対数正規分布

無作為に抽出された変数 x が正規分布に従うとき、水文量確率値 $\hat{x}_p$ の算定は以下ようになる。

$$\hat{x}_p = \bar{x} + Z_p S_x \quad (7-10)$$

また、漸近的にこの水文量確率値 $\hat{x}_p$ の推定誤差分散 S_T^2 は以下のように表すことができる。

$$S_T^2 = \frac{S_x^2}{n} \left(1 + \frac{1}{2} Z_p^2 \right)$$

ゆえに、この水文量確率値 $\hat{x}_p$ の $100(1-\alpha)\%$ 信頼区間は近似的に以下ようになる。

$$(\bar{x} + Z_p S_x) \pm Z_{1-\alpha/2} \sqrt{\frac{S_x^2}{n} \left(1 + \frac{1}{2} Z_p^2 \right)} \quad (7-11)$$

さらにこれらの結果は、2 母数対数正規分布の水文量確率値に対する信頼区間を算出するにあたり、同様の方法を用いることが出来る。もし、水文量 x が対数正規に従うならば、水文量の対数 y も正規分布に従う。その時の水文量確率値 $\hat{x}_p$ は、以下のように表わされる。

$$\hat{x}_p = \exp(\mu_y + Z_p \sigma_y)$$

また、最尤法による水文量確率値 $\hat{x}_p$ に対する推定誤差分散 S_T^2 の算定方法は、以下のようになる。

$$S_T^2 \approx \hat{x}_p^2 \left[\frac{\sigma_y^2}{n} \left(1 + \frac{1}{2} Z_p^2 \right) \right] \quad (7-12)$$

このとき、対数正規分布で表される水文量確率値 $\hat{x}_p$ の $100(1-\alpha)\%$ 信頼区間は、近似的に次のように表される。

$$\exp \left[(\bar{y} + Z_p S_y) \pm Z_{1-\alpha/2} \sqrt{\frac{S_y^2}{n} \left(1 + \frac{1}{2} Z_p^2 \right)} \right] \quad (7-13)$$

b) グンベル分布の推定誤差の算定方法

グンベル分布において、非超過確率 p に対する水文量確率値の推定誤差分散 S_T^2 は、積率法により決定した尺度母数と標準変量の 2 つのパラメータをもとに、漸近的に以下の式から概算することができる^{9) 10)}。

$$S_T^2 = \frac{\alpha^2 (1.11 + 0.52 y + 0.61 y^2)}{n} \quad (7-14)$$

ここで、 n ：データ年数、 α ：前章 3-4-2 の積率法によるグンベル分布の尺度母数 α

y ：グンベル分布の標準変量 $G_p = -\ln[-\ln(p)]$

また、L 積率法を使用した場合の概算は、以下のようになる。

$$S_T^2 = \frac{\alpha^2 \left[(1.1128 - 0.9066/n) - (0.4574 - 1.1722/n)y + (0.8046 - 0.1855/n)y^2 \right]}{n-1} \quad (7-15)$$

さらに上式は、PWM（確率重みつき積率法）を用いた場合でも論理的な概算を提供できるものとなっている。

c) 対数ピアソンⅢ型分布の推定誤差の算定方法

正規分布に従う水文量確率値の信頼区間の考え方は、以下に示すひずみ係数を加味したファクター η を使用することで、近似的にピアソンⅢ型分布の信頼区間に対しても拡張することができる。

このファクター η は、近似的にピアソンⅢ型分布と正規分布の水文量確率の分散の比率により、以下のように概算される^{9) 10)}。

$$\eta = \left[\frac{\text{Var}(\hat{y}_p)}{\text{Var}(\hat{x}_p)} \right]^{1/2} = \sqrt{\frac{1 + \gamma K_p + \frac{1}{2}(1 + \frac{3}{4}\gamma^2)K_p^2}{1 + \frac{1}{2}Z_p^2}} \quad (7-16)$$

ここで、 K_p はピアソンⅢ型分布の水文量に対する標準変量であり、3章の式(3-12)より求めることができる。また、 Z_p は正規分布の水文量に対する標準変量を表し、式(3-4)より求められる。近似的に、ピアソンⅢ型分布に対する水文量確率値の $100(1-\alpha)\%$ 信頼区間は、水文量の対数 y を用いて以下のように概算される。

$$\hat{y}_p + \eta(\zeta_{1-\alpha,p} - Z_p)S_y < y_p < \hat{y}_p + \eta(\zeta_{\alpha,p} - Z_p)S_y \quad (7-17)$$

ここで、 $\hat{y}_p = \bar{y} + K_p S_y$

使用される係数 $\zeta_{\alpha,p}$ および $\zeta_{1-\alpha,p}$ は次式により定義される。

$$\zeta_{\alpha,p} = \frac{Z_p + Z_\alpha \sqrt{\frac{1}{n} + \frac{Z_p^2}{2(n-1)} - \frac{Z_\alpha^2}{2n(n-1)}}}{1 - \frac{Z_\alpha^2}{2(n-1)}} \quad (7-18)$$

また、 $\zeta_{1-\alpha,p}$ は $100(1-\alpha)\%$ の信頼区間に対する係数を表しており、式(7-18)において、信頼区間を算出するために用いる正規分布に基づく値 Z_α が $Z_{\alpha-1}$ となるため $-Z_\alpha$ と置きかえればこの係数は算出することができる。いま、水文量確率値の対数 $y_p = \ln x_p$ の関係により、式(7-17)は

$$\hat{y}_p + \eta(\zeta_{1-\alpha,p} - Z_p)S_y < \ln x_p < \hat{y}_p + \eta(\zeta_{\alpha,p} - Z_p)S_y$$

と表すことができる。さらに、水文量確率値 x_p に対して左辺と右辺は

$$\exp\{\hat{y}_p + \eta(\zeta_{1-\alpha,p} - Z_p)S_y\} < x_p < \exp\{\hat{y}_p + \eta(\zeta_{\alpha,p} - Z_p)S_y\}$$

となる。

この水文量確率値 x_p についての信頼区間を $\hat{x}_p - S_{T,x} < x_p < \hat{x}_p + S_{T,x}$ とすると、

$$S_{T,x} = \frac{1}{2} [\exp\{\hat{y}_p + \eta(\zeta_{\alpha,p} - Z_p)S_y\} - \exp\{\hat{y}_p + \eta(\zeta_{1-\alpha,p} - Z_p)S_y\}]$$

また、 $\hat{x}_p = \exp \hat{y}_p$ より、以下のようなになる。

$$S_{T,x} = \frac{1}{2} \hat{x}_p [\exp\{\eta(\zeta_{\alpha,p} - Z_p)S_y\} - \exp\{\eta(\zeta_{1-\alpha,p} - Z_p)S_y\}] \quad (7-19)$$

上式により、対数ピアソンⅢ型分布の水文量確率値 $\hat{x}_p$ の信頼区間の指標として、この推定誤差 $S_{T,x}$ を求める。

d) GEV 分布の推定誤差の算定方法

GEV 分布では、L 積率法により決定される位置と尺度に関する 2 つのパラメータより、局部的に固有な値すなわち形状母数 k が形成される。この固有な形状母数 k から、非超過確率 p に対する水文量確率値の推定誤差分散 S_T^2 を近似的に概算する方法は、以下の式による⁹⁾¹⁰⁾。

$$S_T^2 = \frac{\alpha^2 (c_1 + c_2 y + c_3 y^2)}{n} \quad (7-20)$$

$$y: \text{GEV 分布の標準変量 } GV_p = 1 - \{-\ln(p)\}^k \quad (k \neq 0)$$

ここで、 n ：データ年数、 α ：前節 3-4-4 の積率法による GEV 分布の尺度母数 a

また、係数 c_1 、 c_2 、 c_3 は形状母数 k によって決定され、漸近的に算出できる。

(3) 各確率分布の推定誤差 S_T の比較

ここでは、3 母数対数正規分布、グンベル分布、対数ピアソンⅢ型分布、GEV 分布のそれぞれの推定誤差 S_T を算出し、確率分布間における推定誤差の値の傾向について考察を行う。最近のデータ 20 年、100 年および確率年 1/100、1/400 の条件のもとに算出した推定誤差 S_T を表に示す。

表 7-1 各確率分布における推定誤差 S_T

確率年	データ年数	確率分布			
		3母数対数正規分布	グンベル分布	対数ピアソンⅢ型分布	GEV分布
1/100	20ヶ年	6.966	12.322	12.123	3.326
	100ヶ年	4.191	6.977	2.824	1.819
1/400	20ヶ年	8.042	16.040	19.386	3.348
	100ヶ年	4.838	9.092	4.510	1.941

グラフは確率 1/100 のもので、棒グラフの上の数字は、確率分布間における推定誤差 S_T の値の大きさの順位を表している。推定誤差 S_T は値が小さいほど信頼性が高いことから、順位は小さいものから 1, 2, 3, 4 としている。

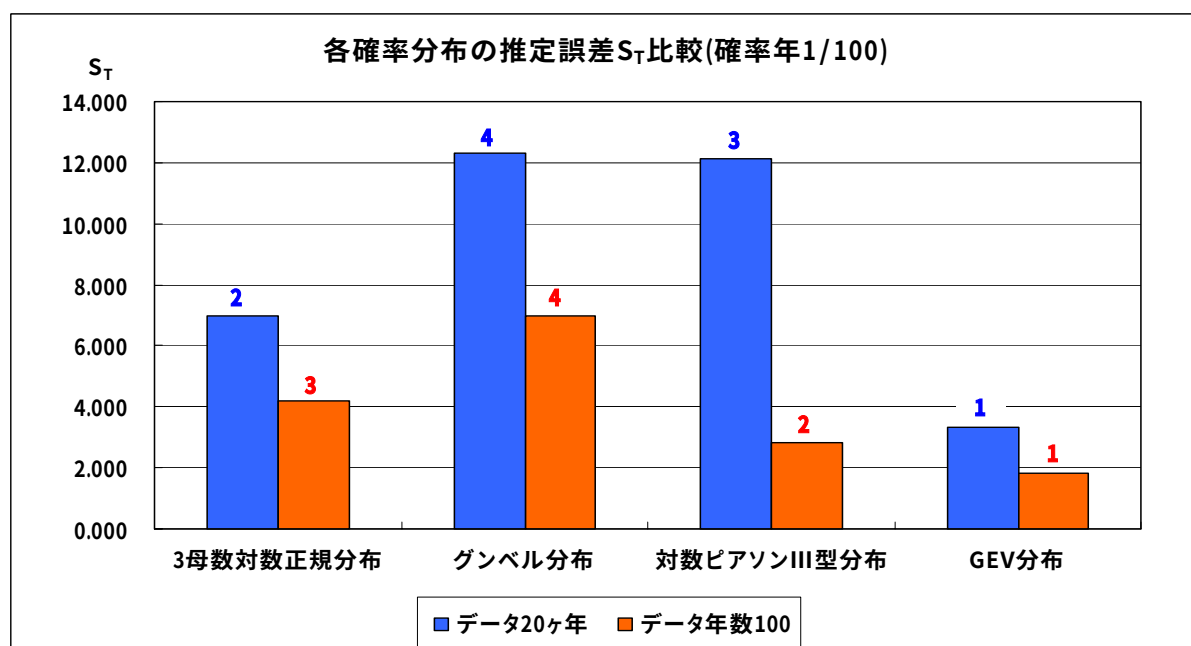


図 7-1 各確率分布における推定誤差 S_T の比較(確率年 1/100)

推定誤差 S_T の値の大きさの順位の比較を見ると、確率分布間全体の順位の中で、GEV 分布がデータ 20 年、100 年とも最も小さく信頼性が良好である。グンベル分布は、データ 100 年において各確率分布間で S_T が最大であり、データ 20 年においても確率年 1/100 で最大と、信頼性が低い。対数ピアソンⅢ型分布は、データ 100 年において信頼性は中程度であるのに対し、データ 20 年の確率値 1/400 では最大となっている。どのケースもデータ 20 年の推定誤差 S_T が、データ 100 年の推定誤差 S_T よりも大きくなっている。

8. 回帰直線による確率値の信頼区間の考え方と水文量確率値への適用

3 章で示した通り、主要な確率分布は標準変量の一次式で表わされる。本研究では、これを利用して回帰直線の信頼区間から確率水文量の信頼性を推定・予測する方法を新たに提案する。

(1) 回帰直線および予測値の信頼区間^{3) 4)}

一般に直線の回帰式は平均値を通る直線として以下のように表される。

$$y = ax + b = \bar{y} + a(x - \bar{x}) \quad (8-1)$$

(a, b : 定数 $\bar{x}, \bar{y}$: 平均値)

また、この時の直線の信頼区間は、推定された統計量（直線上の推定値）からその誤差の標準偏差を定数倍した値を加えたり減じたりすることによって得られる。定数はその統計量の理論分布によって与えられ、 t 分布を用いることにより以下のように表される。

$$\hat{y}_0 = \bar{y} \pm t(n-2, \alpha) Se \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}} \quad (8-2)$$

$t(n-2, \alpha)$: t 分布の自由度および片側超過確率 $\alpha/2$ に対する臨界値

x_0 : 予測値における変数

Se は誤差の標準偏差、 S_{xx} は偏差平方和を表し、それぞれ以下の式により求められる。

$$Se = \sqrt{\frac{\sum_{i=1}^N (\hat{y}_i - y_i)^2}{n-2}} \quad (8-3)$$

$$S_{xx} = \sum_{i=1}^N (x_i - \bar{x})^2 \quad (8-4)$$

上式(8-3)の誤差の標準偏差 Se は、予測値 $\hat{y}_i$ と標本データ y_i との残差に関係し、直線の適合度に関する指標を表している。

前項の考え方をもとに「新たな予測値」の信頼区間³⁾として以下の式が用いられる。

$$\hat{y}_0 = \bar{y} \pm t(n-2, \alpha) Se \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}} \quad (8-5)$$

$t(n-2, \alpha)$: 分布の自由度および片側超過確率 $\alpha/2$ に対する臨界値

x_0 : 予測値における変数

(2) グラフ表示でみる信頼区間⁴⁾

前項の直線回帰式の信頼区間 式(8-2) および新たな予測値の信頼区間 式(8-5) は、一般に以下の図のように表される。

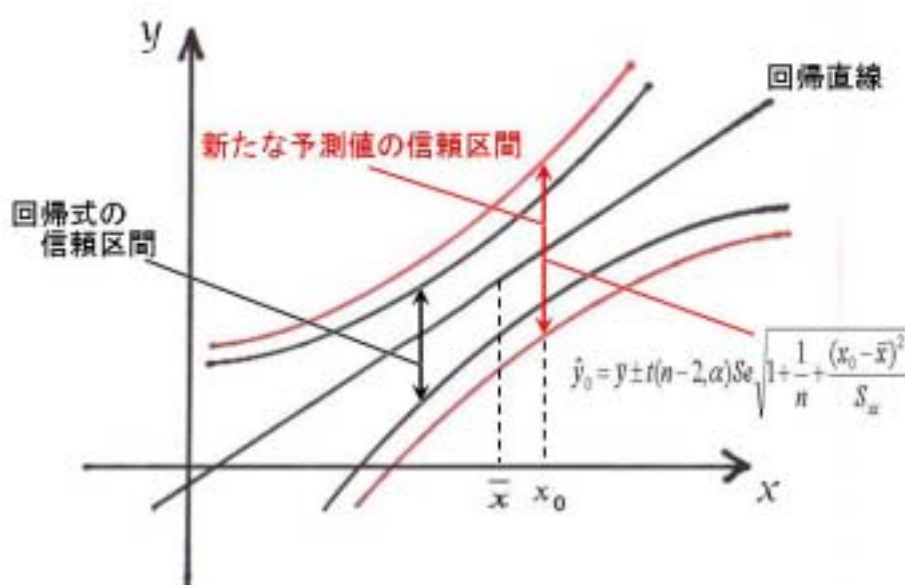


図 8-1 グラフ表示による信頼区間の概念

直線回帰式の信頼区間は上図では黒の曲線で表され、その外側の赤の曲線が予測値の信頼区間を表している。これらは、回帰直線の平均 $\bar{x}$ から予測値における変数 x_0 が遠くなると信頼区間が広がっていくことを示している。また、新たな予測値の信頼区間は、新しい標本の分散が加わるため、変数 x が同一であっても回帰直線の信頼区間より範囲は広がる。

(3) 信頼係数 h_0 および信頼区間との関係

新たな予測値による信頼区間,

$$\hat{y}_0 = \bar{y} \pm t(n-2, \alpha) Se \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}} \quad (8-6)$$

における、赤い枠で囲まれた $\sqrt{\quad}$ の部分の数式を、本研究では信頼区間の広がりを表す指標、信頼係数 h_0 として以下のように定義する。

$$\text{信頼係数 } h_0 = \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}} \quad (8-7)$$

また、上式の信頼係数 h_0 において、式中の

$$\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \quad (8-8)$$

は統計学の分野ではテコ比といわれ、観測値による回帰直線の変動性に関する因子を表す。したがっ

て、この信頼係数 h_0 は信頼性の指標として考えられ、以下の性質を持っている。

- ① 予測値における変数 x_0 が、 x の平均 $\bar{x}$ からどのくらい離れているかにより変化する。
- ② 誤差の標準偏差 Se とは別に、信頼区間の広がり示し、予測値の信頼区間に影響する。
- ③ 信頼係数 h_0 の値が小さいほど信頼区間が狭くなり、信頼性が高い。

(4) 水文確率値への適用、信頼区間の指標 CI について

回帰直線における信頼区間の理論を水文量確率値に適用する。水文量確率値 x_p は、一般的に標準変量 ε の一次式 $x_p = \bar{x} + \sigma \cdot \varepsilon$ で表わされる。いま、下図のように Y 軸を水文量 x 、 x 軸を標準変量 ε （各検討対象確率分布モデルの Z_p 、 G_p 、 K_p 、 GV_p ）とする。

このとき、水文量 x の水文量確率値における信頼区間は、「 t 分布の自由度 $n-2$ および片側超過確率 $\alpha/2$ に対する臨界値 $t(n-2, \alpha)$ 」と「観測データごとに定まる分布の適合度を表す誤差の標準偏差 Se 」及び「求める確率分布の標準変量に対する信頼係数 h_0 」の積によって表される。

$$\text{水文量確率値の信頼区間} = t(n-2, \alpha) \times Se \times h_0 \quad (8-9)$$

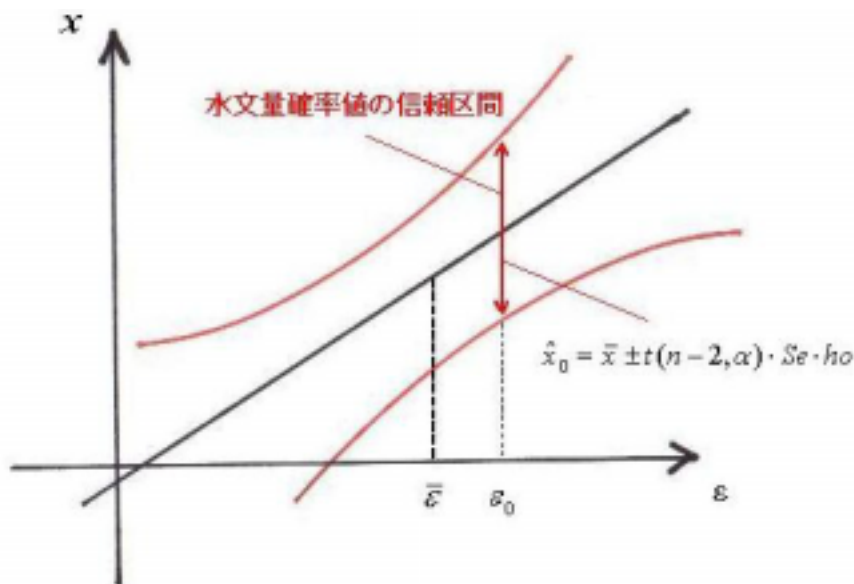


図 8-2 水文量確率値の信頼区間の概念図

誤差の標準偏差 Se 、信頼係数 h_0 およびそれに伴う残差平方和 $S_{\varepsilon\varepsilon}$ は次式より求める。

$$\text{誤差の標準偏差 } Se = \sqrt{\frac{\sum (\hat{x}_i - x_i)^2}{n-2}} \quad (8-10)$$

$$\text{信頼係数 } h_0 = \sqrt{1 + \frac{1}{n} + \frac{(\varepsilon^* - \bar{\varepsilon})^2}{S_{\varepsilon\varepsilon}}} \quad (8-11)$$

$$\text{残差平方和 } S_{\varepsilon\varepsilon} = \sum (\varepsilon_i - \bar{\varepsilon})^2 \quad (8-12)$$

ε_i : 標準変量, ε^* : 求める確率値 x_p における標準変量,

$\bar{\varepsilon}$: 標準変量の平均値 x_i : 水文量実測値 $\hat{x}_i$: 水文量推定値

ここで、 t 分布の臨界値 $t(n-2, \alpha)$ は、水文確率分布の種類や水文データの値に関係しない。本研究では臨界値を除く次式を信頼区間の指標 CI (Confidence Interval) と定義する。

$$CI = Se \times h_0 \quad (8-13)$$

この信頼区間の指標 CI の値が小さいものほど信頼性が高い。

(5) 確率値の信頼区間の算定に用いるプロットイングポジション¹⁾

水文量確率値の信頼区間の算定に用いる、水文量の測定値毎の非超過確率 p は、プロットイングポジションによって表すことができる。非超過確率 p は次式で示される。

$$p \equiv F(x_i) = \frac{i - \alpha}{n + 1 - 2\alpha} \quad (8-14)$$

n : 標本数, x_i : 標本を値の小さいものから順に並べた時の i 番目の順序標本値

$F(x_i)$: x_i のプロットイングポジション α : 0~1 で表わされる定数

Weibull 公式 : $\alpha = 0$ Hazen 公式 : $\alpha = 1/2$ Cunnane 公式 : $\alpha = 2/5$

(6) 信頼区間 h_0 の算定までの流れについて

ここでは計算例として、3 母数対数正規分布を用いた信頼区間 h_0 の算定方法について説明する。

a) プロットイングポジションによる非超過確率 p を 式(8-15)から、また、その時の p を用いて 式(8-16), (8-17)より、3 母数対数正規分布の標準変量 Z_p を算出する。

$$p \equiv F(x_i) = \frac{i - \alpha}{n + 1 - 2\alpha} \quad (8-15)$$

$$Z_p = -\sqrt{\frac{t^2[(4t + 100)t + 205]}{[(2t + 56)t + 192]t + 131}} \quad (8-16)$$

$$\text{非超過確率 } p \leq 0.5 \text{ の場合, 係数 } t \text{ は } t = -\ln(2p) \quad (8-17)$$

また、非超過確率 $p > 0.5$ の場合は、係数 $t = -\ln[(2(1-p))]$ となり、その時の標準変量 Z_p は、 $Z_p = -Z_{1-p}$ に置き換えることにより求められる。

b) 信頼係数 h_0 は、求めようとする確率年に対する標準変量 Z_p^* と標準変量 Z_p の平均 $\bar{Z}_p$ 、標準変量の偏差平方和 S_{zz} より、次式で算出する。

$$h_0 = \sqrt{1 + \frac{1}{n} + \frac{(Z_p^* - \bar{Z}_p)^2}{S_{zz}}} \quad (8-18)$$

c) 他の分布形の場合は、信頼係数 h_0 の算定に Z_p の代わりにその分布の標準変量 ε (G_p あるいは K_p , GV_p) を用いればよい。なお、各確率分布の信頼係数 h_0 に対して以下のことが言える。

①3 母数対数正規分布、グンベル分布において、標準変量 (Z_p , G_p) が標本数 n と求めようとする確率のみにより定まることから、信頼係数 h_0 は水文データに影響されない。

②対数ピアソンⅢ型分布に関して、式(8-19)の標準変量 K_p にひずみ係数 γ_y が含まれるため、水文データ毎に信頼係数 h_0 の値は異なる。

$$K_p = \frac{2}{\gamma_y} \left(1 + \frac{\gamma_y Z_p}{6} - \frac{\gamma_y^2}{36} \right)^3 - \frac{2}{\gamma_y} \quad (8-19)$$

③GEV 分布に関して、式(8-20)の標準変量 GV_p に形状母数 k が含まれているため、水文データ毎に信頼係数 h_0 の値は異なる。

$$GV_p = 1 - \{-\ln(p)\}^k \quad (8-20)$$

(7) 確率分布の信頼係数 h_0

式(8-18)において定義した信頼係数 h_0 について、各確率分布、各プロットングポジションにおける値の変化をデータ年数と確率年の条件を変えて算出し比較する。ここでは、データ年数は最近の 20 年および 100 年、確率年数は 1/100 と 1/400 とする。

表 8-1 確率分布、プロットング方法別の信頼係数 h_0

確率年	データ年数	確率分布	信頼係数 h_0		
			Cunnane	Hazen	Weibull
1/100	20ヶ年	3母数対数正規分布	1.1633	1.1569	1.1868
		グンベル分布	1.2819	1.2666	1.3363
		対数ピアソンⅢ型分布	1.1955	1.1871	1.2257
		GEV分布	1.2320	1.2204	1.2737
	100ヶ年	3母数対数正規分布	1.0323	1.0319	1.0338
		グンベル分布	1.0554	1.0542	1.0593
		対数ピアソンⅢ型分布	1.0379	1.0374	1.0398
		GEV分布	1.0537	1.0524	1.0581
1/400	20ヶ年	3母数対数正規分布	1.2213	1.2123	1.2538
		グンベル分布	1.4563	1.4324	1.5404
		対数ピアソンⅢ型分布	1.2799	1.2679	1.3231
		GEV分布	1.3565	1.3387	1.4195
	100ヶ年	3母数対数正規分布	1.0445	1.0439	1.0466
		グンベル分布	1.0944	1.0924	1.1012
		対数ピアソンⅢ型分布	1.0553	1.0545	1.0581
		GEV分布	1.0962	1.0937	1.1045

次ページ図 8-3 は確率 1/100、データ年数 100 年の例で、棒グラフの上の数字は、プロットング方法毎の信頼係数 h_0 の値の大きさの確率分布間における順位を表している。信頼係数 h_0 は値が小さいほど信頼性が高いことから、順位は小さいものから 1, 2, 3, 4 となっている。プロットング方法ごとに図中の棒グラフを色分けで示したが、確率分布の違いによる h_0 の値の順位は、プロットング方法に影響をほとんど受けなかった。

h_0 の値は、小さい順から 3 母数対数正規分布、対数ピアソンⅢ型分布、GEV 分布、グンベル分布の順であり、GEV 分布とグンベル分布は大きい。この傾向は、データ年数や確率年数が違っててもほとんど変わり無く、データ年数 100 年の確率値 1/400 におけるグンベル分布と GEV 分布の値の順位が他のケースと違うのみである。グンベル分布と GEV 分布は同じ極値分布であることから、 h_0 は極値分布で大き

いといえる。

各確率分布ごとにおけるプロット方法の違いによる h_0 の値の傾向は、小さい順に Hazen, Cunnane, Weibull である。これは確率分布、データ年数、確率年が異なっても同じであった。

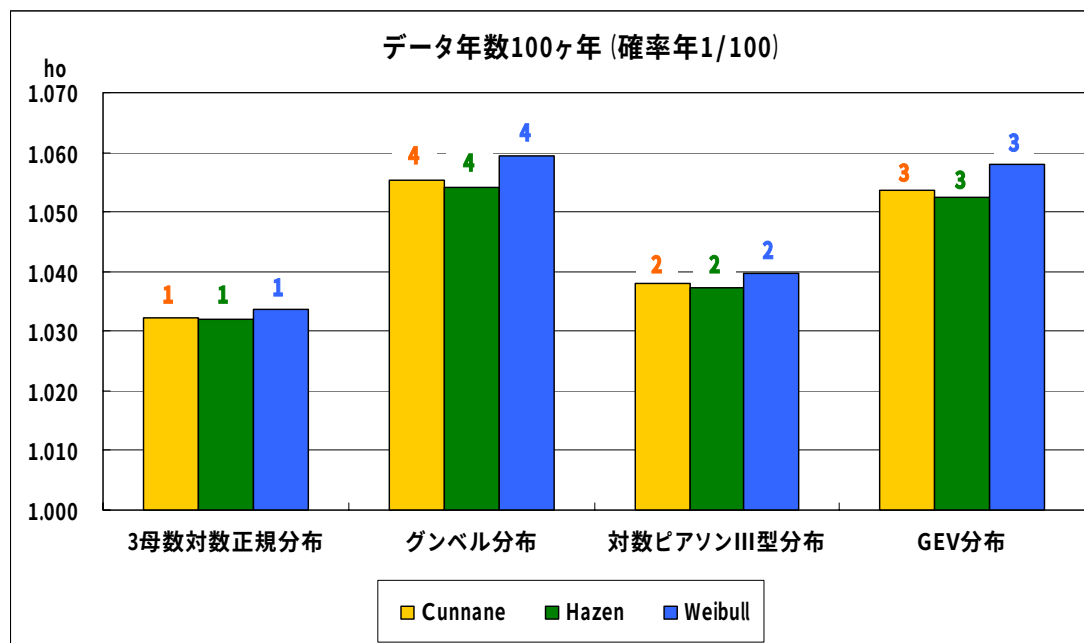


図 8-3 各確率分布，プロット方法別の信頼係数 h_0 （データ年数 100 ヶ年，確率年 1/100）

(8) 信頼区間の指標 CI について

式(8-13)において定義した信頼区間の指標 CI について、各確率分布、各プロット方法ポジションによる値の変化を、データ年数 n と確率年の条件を変えて算出し比較を行う。ここではデータ年数は最近の 20 年および 100 年を用い、確率年数は 1/100 と 1/400 について傾向を見る。

表 8-2 各確率分布，プロット方法別の信頼区間の指標 CI

確率年	データ年数	確率分布	信頼区間の指標 CI		
			Cunnane	Hazen	Weibull
1/100	20	3母数対数正規分布	3.4333	3.4232	4.1047
		グンベル分布	4.5975	4.8057	4.8744
		対数ピアソンIII型分布	4.2852	4.5309	4.4979
		GEV分布	3.4797	3.6412	3.9628
	100	3母数対数正規分布	2.8033	2.9041	2.7828
		グンベル分布	4.1653	4.3358	3.8592
		対数ピアソンIII型分布	4.1722	4.4911	3.5857
		GEV分布	3.1015	3.3789	2.6509
1/400	20	3母数対数正規分布	3.6044	3.5874	4.3364
		グンベル分布	5.2230	5.4347	5.6190
		対数ピアソンIII型分布	4.5858	4.8369	4.8550
		GEV分布	3.8196	3.9832	4.4507
	100	3母数対数正規分布	2.8364	2.9380	2.8174
		グンベル分布	4.3195	4.4930	4.0117
		対数ピアソンIII型分布	4.2434	4.5667	3.6504
		GEV分布	3.2541	3.5431	2.7939

次ページの図 8-4 において、棒グラフの上の数字は、確率分布間全体における信頼区間の指標 CI の値の大きさの順位を表している。信頼区間の指標 CI は、値が小さいほど信頼性が高いことから、順位は小さいものから 1, 2, 3, 4 となっている。

確率分布の違いによる CI の値の大きさの順位は、プロット方法ポジション及びデータ年数により

わずかな違いは生ずるが、ほとんどの傾向の違いはない。順位の傾向を見ると、3 対数正規分布、次いで GEV 分布が小さく信頼性は良い。グンベル分布、および対数ピアソンIII型は、各データ年数およびプロットング方法において値が大きく、信頼性が小さい。

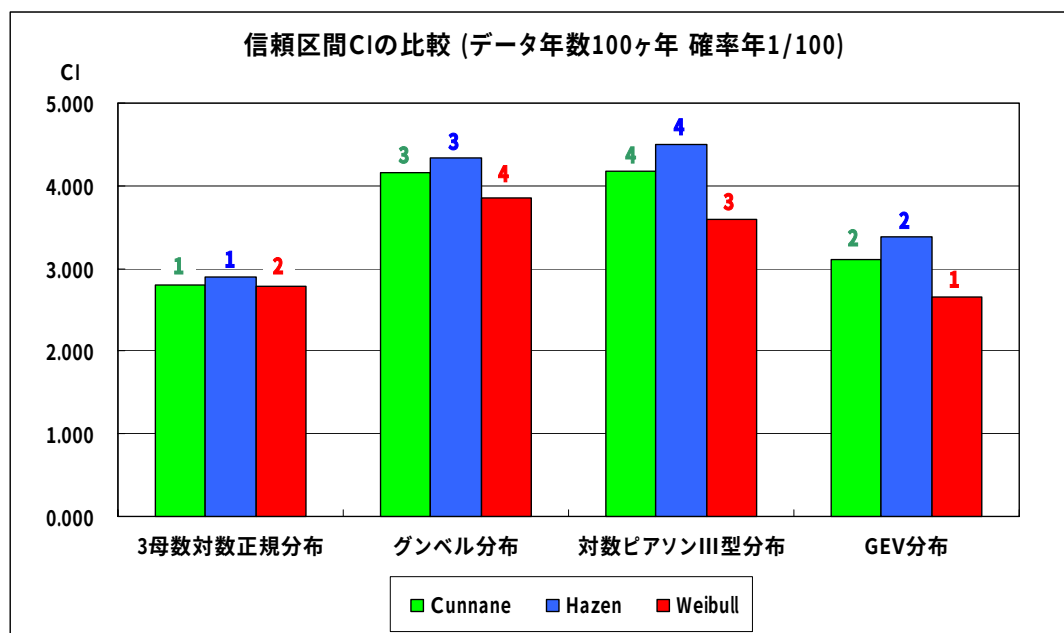


図 8-4 各確率分布，プロットング別の信頼区間の指標 CI（データ年数 100 ヶ年，確率年 1/100）

(9) 信頼係数 h_0 ，誤差の標準偏差 Se ，信頼区間の指標 CI の関係

前項において定義した信頼区間の指標 h_0 ， Se ， CI の関係について考察を行う。データ年数は最近の 20 年および 100 年データを用い、確率年数は 1/100 と 1/400 について傾向を見る。

表 8-3 信頼係数 h_0 ，誤差の標準偏差 Se ，信頼区間の指標 CI の関係

確率年	データ年数	確率分布	信頼係数 h_0			誤差の標準偏差 Se			信頼区間の指標 CI		
			Cunnane	Hazen	Weibull	Cunnane	Hazen	Weibull	Cunnane	Hazen	Weibull
1/100	20	3母数対数正規分布	1.1633	1.1569	1.1868	2.9513	2.9591	3.4586	3.4333	3.4232	4.1047
		グンベル分布	1.2819	1.2666	1.3363	3.5864	3.7940	3.6478	4.5975	4.8057	4.8744
		対数ピアソンIII型分布	1.1955	1.1871	1.2257	3.5868	3.8190	3.6702	4.2852	4.5309	4.4979
		GEV分布	1.2320	1.2204	1.2737	2.8367	2.9954	3.0899	3.4797	3.6412	3.9628
	100	3母数対数正規分布	1.0323	1.0319	1.0338	2.7155	2.8143	2.6919	2.8033	2.9041	2.7828
		グンベル分布	1.0554	1.0542	1.0593	3.9468	4.1128	3.6431	4.1653	4.3358	3.8592
		対数ピアソンIII型分布	1.0379	1.0374	1.0398	4.0183	4.3276	3.4470	4.1722	4.4911	3.5857
		GEV分布	1.0537	1.0524	1.0581	2.9291	3.1947	2.4912	3.1015	3.3789	2.6509
1/400	20	3母数対数正規分布	1.2213	1.2123	1.2538	2.9513	2.9591	3.4586	3.6044	3.5874	4.3364
		グンベル分布	1.4563	1.4324	1.5404	3.5864	3.7940	3.6478	5.2230	5.4347	5.6190
		対数ピアソンIII型分布	1.2799	1.2679	1.3231	3.5868	3.8190	3.6702	4.5858	4.8369	4.8550
		GEV分布	1.3565	1.3387	1.4195	2.8367	2.9954	3.0899	3.8196	3.9832	4.4507
	100	3母数対数正規分布	1.0445	1.0439	1.0466	2.7155	2.8143	2.6919	2.8364	2.9380	2.8174
		グンベル分布	1.0944	1.0924	1.1012	3.9468	4.1128	3.6431	4.3195	4.4930	4.0117
		対数ピアソンIII型分布	1.0553	1.0545	1.0581	4.0183	4.3276	3.4470	4.2434	4.5667	3.6504
		GEV分布	1.0962	1.0937	1.1045	2.9291	3.1947	2.4912	3.2541	3.5431	2.7939

以下に確率 1/100，データ年数 100 年，Cunnane 公式のグラフを示す。棒グラフの上の数字は、確率分布間における指標毎の値の大きさの順位を表している。ここで 3つの指標 CI ， h_0 ， Se は、値が小さいほど信頼性および適合性が高いことから、順位は小さいものから 1，2，3，4 となっている。

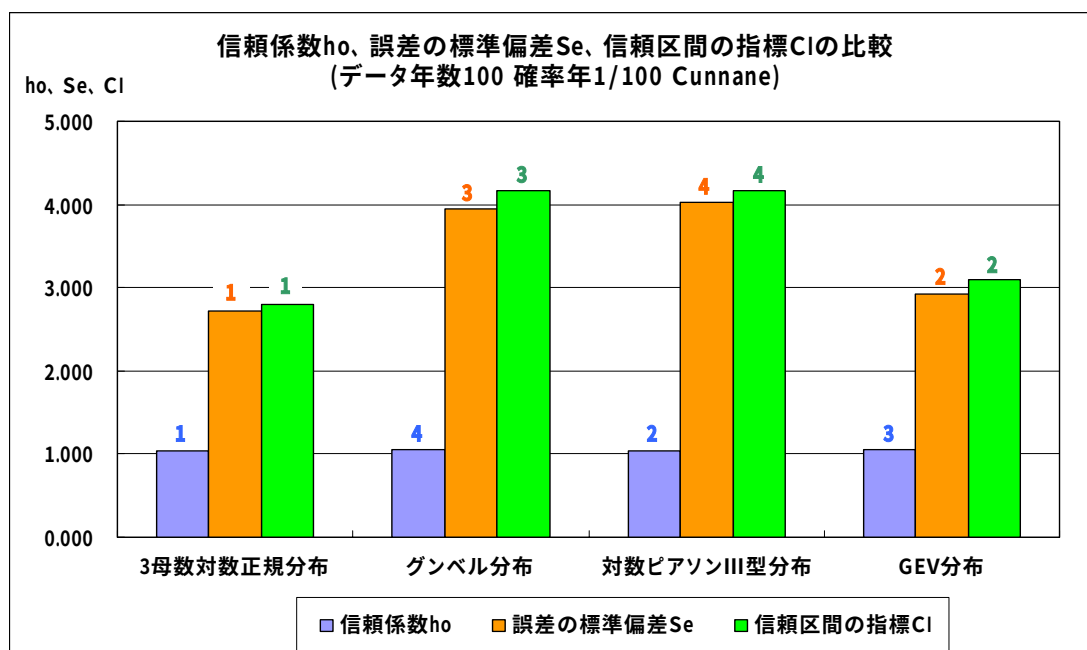


図 8-5 h_0 , Se , CI の比較

表 8-3, 図 8-5 より分かるように, 確率分布の違いによる信頼係数 h_0 の違いは, 誤差の標準偏差 Se の違いに比べて小さいため, 信頼区間の指標 CI は h_0 より Se の影響を大きく受けている. 確率分布間全体の値の順位の傾向から, Se の値の大きさの順位が小さい3母数対数正規分布とGEV分布が, CI においても値の順位が小さい. CI が1番良好であるのが3母数対数正規分布であり, GEV分布は, h_0 の値の順位は大きいほうであるが, Se の値の大きさの順位が小さいため, 3母数対数正規分布に次いで CI が小さく良好である. グンベル分布と対数ピアソンIII型分布は, Se および CI の値の順位が大きくなっている.

9. 各評価指標の類似性および分布関数との関係

(1) 評価指標間の相関¹²⁾

前章までに4つの確率分布に対して道内8ヶ所の観測所ごとにデータ年数20年, 40年, 60年, 100年の年最大日雨量を用いて8つの指標を求めた. これらは適合度関係3つ(X-COR, P-COR, SLSC), 安定性に関する誤差分散1つ(Jackknife誤差分散), 回帰から求めた信頼性の誤差分散 CI 関連3つ(CI , h_0 , Se), 母数から求めた誤差分散1つ(S_T)である. 8種類の評価指標のデータ年数20年の例を表 9-1 に示す.

評価指標間の類似性を知るために, ここでは単相関係数を用いることにより考察を行う. 例としてデータ年数100年, 確率年1/100の各評価指標間の単相関係数の結果を表 9-2 に示す. この表においてオレンジ色で示しているものは強い相関を表し, 水色は弱い相関を表している. X-CORとP-CORは他の評価指標との相関をとると負の相関を多く示しているが, これは, X-CORとP-CORは値が大きいものが指標として良好であり, 値が小さいものが良好である他の評価指標とは反対の相関を示すためである. また, X-CORとP-CORの両者の相関関係は正の相関を表す.

表 9-1 北海道内の年最大日雨量から算出した確率分布評価指標(データ年数 20 ヶ年, 確率年 1/100)

		CI	S_T	Jackknife	SLSC	X-COR	P-COR	ho	Se
3 母数対数正規分布	札幌市	3.968	10.34	22.4	0.034	0.993	0.992	1.163	3.411
	旭川市	2.272	5.90	12.5	0.032	0.992	0.991	1.163	1.953
	函館市	4.971	8.25	4.9	0.045	0.976	0.987	1.163	4.273
	帯広市	2.724	7.06	14.8	0.032	0.992	0.991	1.163	2.342
	釧路市	2.831	7.99	10.6	0.024	0.992	0.997	1.163	2.433
	網走市	2.102	3.67	3.7	0.044	0.979	0.975	1.163	1.807
	根室市	4.392	6.27	5.6	0.057	0.968	0.973	1.163	3.775
	寿都町	4.206	6.26	7.1	0.051	0.97	0.981	1.163	3.615
グンベル分布	札幌市	4.509	17.81	19.1	0.028	0.99	0.991	1.282	3.517
	旭川市	2.373	10.32	10.0	0.026	0.992	0.991	1.282	1.851
	函館市	7.967	14.91	8.2	0.059	0.952	0.984	1.282	6.215
	帯広市	2.839	12.34	12.1	0.026	0.992	0.991	1.282	2.215
	釧路市	4.373	14.32	11.3	0.034	0.984	0.996	1.282	3.411
	網走市	2.778	6.35	6.0	0.049	0.969	0.976	1.282	2.167
	根室市	5.282	11.20	7.4	0.052	0.963	0.974	1.282	4.120
	寿都町	6.659	11.32	5.2	0.065	0.941	0.978	1.282	5.195
GEV 分布	札幌市	4.316	1.33	31.0	0.026	0.992	0.992	1.367	3.158
	旭川市	2.370	1.73	15.9	0.026	0.992	0.991	1.280	1.852
	函館市	4.716	2.64	7.9	0.051	0.979	0.988	1.155	4.083
	帯広市	2.874	1.79	17.4	0.026	0.991	0.991	1.290	2.227
	釧路市	2.786	2.49	14.3	0.029	0.993	0.998	1.197	2.327
	網走市	2.181	1.54	6.9	0.036	0.978	0.976	1.191	1.832
	根室市	4.741	3.32	11.6	0.049	0.968	0.975	1.239	3.827
	寿都町	3.854	2.25	6.8	0.052	0.975	0.984	1.138	3.387
対数ピアソンIII型分布	札幌市	3.738	28.98	29.1	0.029	0.993	0.992	1.208	3.095
	旭川市	2.207	10.70	14.2	0.03	0.992	0.991	1.187	1.858
	函館市	7.795	12.40	9.2	0.041	0.974	0.987	1.194	6.527
	帯広市	2.693	10.95	15.8	0.029	0.992	0.991	1.196	2.252
	釧路市	4.030	11.71	13.9	0.021	0.992	0.998	1.185	3.401
	網走市	2.734	4.37	6.3	0.047	0.976	0.972	1.215	2.250
	根室市	4.539	8.54	7.8	0.055	0.967	0.974	1.181	3.843
	寿都町	6.546	9.32	6.8	0.048	0.968	0.982	1.197	5.468

表 9-2 各評価指標の相関関係(データ年数 100 ヶ年, 確率年 1/100)

n= 31

	CI	S_T	Jackknife	SLSC	X-COR	P-COR	ho	Se
CI	1	0.31356741	0.3826549	0.25638	-0.517	-0.0921	0.40995	0.99944
S_T	0.31357	1	0.173797	0.41884	-0.3965	-0.2355	0.19464	0.31084
Jackknife	0.38265	0.17379702	1	-0.2884	0.01895	0.16265	0.43208	0.37199
SLSC	0.25638	0.41883867	-0.288447	1	-0.847	-0.5246	0.25683	0.24794
X-COR	-0.517	-0.3965098	0.0189507	-0.847	1	0.44096	-0.4161	-0.5079
P-COR	-0.0921	-0.2354829	0.1626479	-0.5246	0.44096	1	-0.1175	-0.0885
ho	0.40995	0.19464405	0.4320762	0.25683	-0.4161	-0.1175	1	0.38058
Se	0.99944	0.31084423	0.3719923	0.24794	-0.5079	-0.0885	0.38058	1

(2) 確率分布による適合度, 安定性, 信頼区間の指標の大きさの比較

水文量確率値に対する分布の適合度, 安定性, 信頼区間の指標の類似性について, 最近のデータ 20, 40, 60, 100 年に対して確率値 1/100 で各指標を算出し, その平均値の大きさの比較から考察を行う。

ここで, 対数ピアソンIII型分布の S_T が臨界値 Z_α を含む形で算出されているため, Jackknife 誤差分散と CI は正規分布の臨界値 Z_α ($\alpha=0.05$) を乗じて求めている。図 9-1 はデータ年数 100 年の例で, 棒グラフの上の数字は, 確率分布間における指標毎の値の大きさの順位を表している。用いる各指標は全てが値の小さいものほど良好であることから, 順位は小さいものから 1, 2, 3, 4 となっている。

表 9-3 データ別の水文量確率値に対する分布の適合度，安定性，信頼区間の指標(確率年 1/100)

確率年1/100 データ年数	確率分布 評価指標	3母数対数正規分布	対数ピアソンIII型分布	グンベル分布	GEV分布
20	Jackknife誤差分散 $\times Z_\alpha$	20.043	25.324	19.478	27.461
	$S_T \times Z_\alpha$	13.688	12.123	24.213	6.536
	$CI \times Z_\alpha$	6.746	8.421	9.034	6.838
	SLSC	0.040	0.038	0.042	0.037
40	Jackknife誤差分散 $\times Z_\alpha$	8.720	11.962	11.864	11.078
	$S_T \times Z_\alpha$	9.393	4.526	17.170	5.069
	$CI \times Z_\alpha$	5.680	10.100	9.156	5.610
	SLSC	0.036	0.035	0.050	0.039
60	Jackknife誤差分散 $\times Z_\alpha$	11.348	15.917	13.559	16.310
	$S_T \times Z_\alpha$	10.845	4.141	15.792	4.195
	$CI \times Z_\alpha$	5.608	9.627	8.583	5.851
	SLSC	0.031	0.028	0.042	0.033
100	Jackknife誤差分散 $\times Z_\alpha$	12.464	17.882	13.559	18.913
	$S_T \times Z_\alpha$	8.235	2.824	13.711	3.574
	$CI \times Z_\alpha$	5.508	8.198	8.185	6.094
	SLSC	0.025	0.022	0.037	0.028

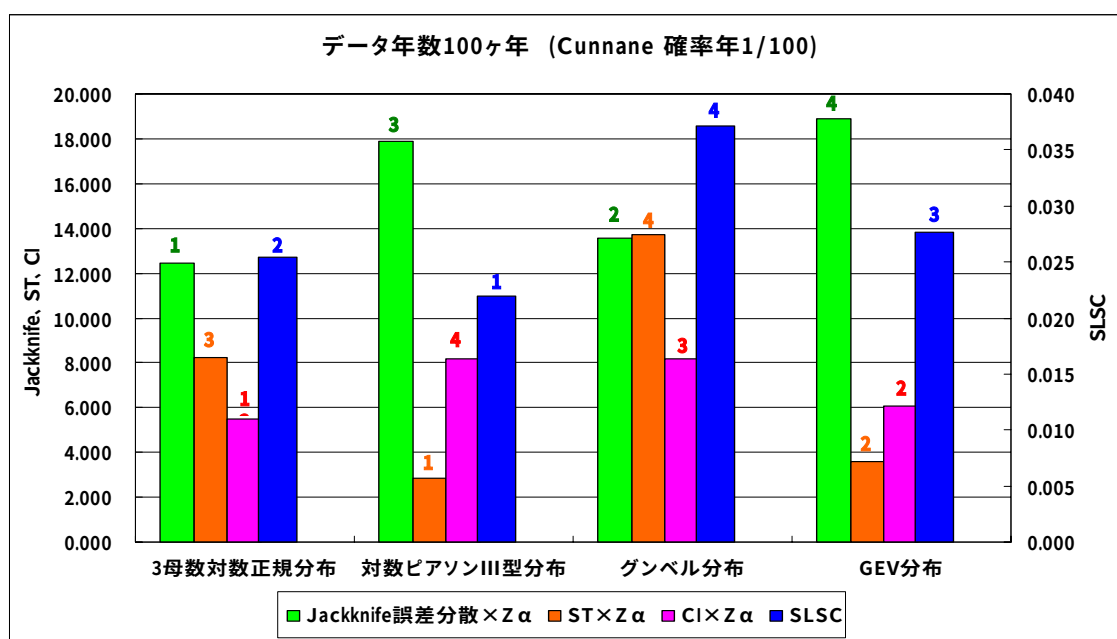


図 9-1 分布の適合度，安定性，信頼区間の指標の比較(データ 100 ヶ年，確率年 1/100)

表 9-3 および図 9-1 における適合度，安定性，信頼区間の指標の値の大きさの順位から指標間の類似性について傾向を見ると，3母数対数正規分布は安定性を示す Jackknife 誤差分散と CI が良好で，SLSC もやや良い．対数ピアソンIII型分布は S_T と SLSC が良い．GEV 分布は S_T と CI が比較的良好，グンベル分布は Jackknife 誤差分散以外はあまり良くない．

指標については SLSC と $S_T \times Z_\alpha$ の値の大きさの順位の傾向がやや似ている．また信頼度の指標 $CI \times Z_\alpha$ は Jackknife 誤差分散にやや近いが明確でなく，中間的な傾向を示している．傾向をより明確に示すために次節以降で主成分分析を行う．

(3) 主成分分析法^{12) 13)}

ここでは，各評価指標間の類似性と指標の特性，及び分布関数の特性と指標との関連性を知るために，主成分分析を用いる．主成分分析とは変数相互の関係から，新しい概念やファクター（主成分）を導き，変数やサンプル間の類似性や相対的位置関係（ポジショニング）を明らかにするものである．

この主成分分析の基本式は，次式で表わされる．

$$Z_1 = a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + a_{14}x_4 + \cdots + a_{1p}x_{1p}$$

$$Z_2 = a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + a_{24}x_4 + \cdots + a_{2p}x_{2p}$$

⋮

$$Z_p = a_{p1}x_1 + a_{p2}x_2 + \cdots + a_{pp}x_{pp} \quad (9-1)$$

Z_p : (各) 主成分, $x_1 \cdots x_p$; 各説明変数, p : 変数の数

$a_{p1} \cdots a_{pp}$: (各) 主成分に対する影響度を表す係数 (固有ベクトル)

上式は、説明変数が p 個ならば、主成分の数も p 個求まる事を表している。その内第 1 主成分は一般的に総合力を示すファクターである。他の主成分については、変数の数 p まで求めることができるが、第 1 主成分と第 2 主成分までの分析結果で類似性の判断などを行うことが多い。

本研究において各評価指標間の類似性を知るために、表 9-1 のような各評価指標のデータ群を直接用以て主成分分析を行うと、分析結果が膨大になる。また、指標毎に単位が異なり数値のオーダーが違うことから、分析結果に影響を与えないよう単位を合わせる必要がある。さらに、X-COR と P-COR は大きい方が良好、他の指標は小さい方が良好であるため、指標の良好性の方向をそろえなければならない。このため、指標毎に 4 つの分布の 8 観測所のデータについて平均 $\bar{x}$ および標準偏差 σ を算出し、 $Z = \frac{x - \bar{x}}{\sigma}$ として指標の標準化を行って数値のオーダーを合わせ、また、X-COR および P-COR 以外の指標はプラスマイナスを反対にして指標の良好性の方向を統一した。この標準化したデータ群を、分布毎に平均化し、データ年数別に示したのが表 9-4 である。

表 9-4 主成分分析に用いるデータ年数別の各評価指標のデータ群(確率年 1/100)

		標準化した各評価指標の平均値							
データ年数	確率分布	CI	S _T	Jackknife	SLSC	X-COR	P-COR	ho	Se
20	3母数対数正規分布	0.3290	0.2485	0.2373	-0.0590	0.1897	-0.0078	0.9588	0.2283
	グンベル分布	-0.4137	-0.6884	0.2815	-0.2644	-0.5504	-0.1014	-1.1137	-0.2735
	GEV分布	0.2993	1.0935	-0.3430	0.1874	0.2459	0.1170	-0.2421	0.3189
	対数ピアソンIII型分布	-0.2145	-0.6536	-0.1758	0.1360	0.1148	-0.0078	0.3970	-0.2738
40	3母数対数正規分布	0.5527	0.1392	0.6648	0.3724	0.3714	0.0000	0.9361	0.5229
	グンベル分布	-0.4293	-1.3270	-0.2913	-0.9265	-1.1370	-0.2408	-1.5120	-0.3625
	GEV分布	0.5726	0.9545	-0.0523	0.0626	0.4285	0.2007	0.4447	0.5693
	対数ピアソンIII型分布	-0.6960	0.2333	-0.3212	0.4916	0.3371	0.0401	0.1312	-0.7297
60	3母数対数正規分布	0.6585	0.0197	0.4016	0.3362	0.4039	-0.1686	0.9786	0.6241
	グンベル分布	-0.4241	-1.2888	0.0991	-1.0374	-1.0516	-0.3669	-1.0226	-0.3783
	GEV分布	0.5700	1.0284	-0.2773	0.0904	0.3399	0.3074	-0.3457	0.5978
	対数ピアソンIII型分布	-0.8043	0.2407	-0.2235	0.6109	0.3079	0.2281	0.3897	-0.8436
100	3母数対数正規分布	0.5788	-0.1058	0.3408	0.3286	0.5371	0.0074	0.9770	0.5554
	グンベル分布	-0.4622	-1.3446	0.2293	-1.0946	-1.0111	-0.6049	-0.8235	-0.4402
	GEV分布	0.3509	0.9487	-0.3163	0.0613	0.2894	0.3327	-0.6914	0.3828
	対数ピアソンIII型分布	-0.4675	0.5017	-0.2112	0.7457	0.2517	0.2657	0.5379	-0.4980

(4) 分析結果の寄与率

主成分分析の分析結果の精度を確認する方法として、累積寄与率(%)がある。ここで、寄与率(%)とは、分析対象のデータ（変数）が有していた情報が、求めた主成分にどのくらい集約されているかを判断する目安となるものである。この寄与率は、次式で表わされる。

$$p \text{ 番目の主成分の寄与率} = \frac{\lambda_p}{p} \times 100 (\%) \quad (9-2)$$

λ_p :固有値, p :変数の数

この寄与率を第1主成分から順に主成分 p まで足し合わせたものを累積寄与率という。累積寄与率が高いほど良い分析結果であり、分析が成功したかどうかについて明確な基準は無いが、一般的に第2主成分までの累積寄与率が50%以上であることが望ましいとされる。表 9-4 を用いて、データ年数別に主成分分析を行った結果、次表のような累積寄与率を得た。

表 9-5 主成分分析の分析結果の累積寄与率

データ年数	主成分No.	固有値	寄与率(%)	累積(%)
20	1	5.55	69.36	69.36
	2	1.38	17.20	86.55
	3	1.08	13.45	100.00
40	1	5.53	69.10	69.10
	2	1.75	21.89	91.00
	3	0.72	9.00	100.00
60	1	4.29	53.62	53.62
	2	2.38	29.71	83.32
	3	1.33	16.68	100.00
100	1	4.76	59.55	59.55
	2	2.01	25.10	84.65
	3	1.23	15.35	100.00

ここではデータ年数毎の累積寄与率を第3主成分まで示したが、各データ年数とも、第2主成分までの累積寄与率が80%以上であることから、これらの分析結果を用いても問題ない。

(5) 各指標の主成分への影響度合い(固有ベクトル)

式(9-1)における主成分 Z は目的変数（ここでは3母数対数正規分布などの分布関数）の評価をする際に用いることができる想像上の尺度である。主成分を説明する係数 $a_{p1} \cdots a_{pp}$ は固有ベクトルとも呼ばれ、変数（ここでは CI , S_T 等の各指標）の主成分への影響度合いを示す。

固有ベクトルの値が大きければ、その指標の主成分への影響が大きくなり、各指標の第1主成分、第2主成分などの固有値ベクトルの値を比較することにより、各指標間の類似性を知ることができる。

a) 第1主成分

第1主成分 Z_1 は通常、総合力を示す成分であり、ここでは分布関数の総合力を示すと考えられる。各指標のデータ年数ごとの標準化した第1主成分の固有ベクトルを表 9-6 に示す。

表 9-6 データ年数に対する主成分1の各指標の固有ベクトル

データ年数	主成分 1				平均
	20	40	60	100	
確率分布					
CI	0.3787	0.3178	0.2317	0.2879	0.3040
ST	0.3703	0.3889	0.4567	0.4168	0.4082
Jackknife	-0.2950	0.2703	-0.1257	-0.2070	-0.0893
SLSC	0.3551	0.3437	0.4342	0.4101	0.3858
X-COR	0.3973	0.3986	0.4762	0.4504	0.4306
P-COR	0.4054	0.3716	0.3840	0.4318	0.3982
ho	0.2376	0.4133	0.3369	0.2581	0.3115
Se	0.3573	0.2980	0.2138	0.2765	0.2864

また、縦軸に各評価指標名、横軸に第 1 主成分における各評価指標の標準化した固有ベクトルの値を示した例（データ年数 100 年）が図 9-2 である。

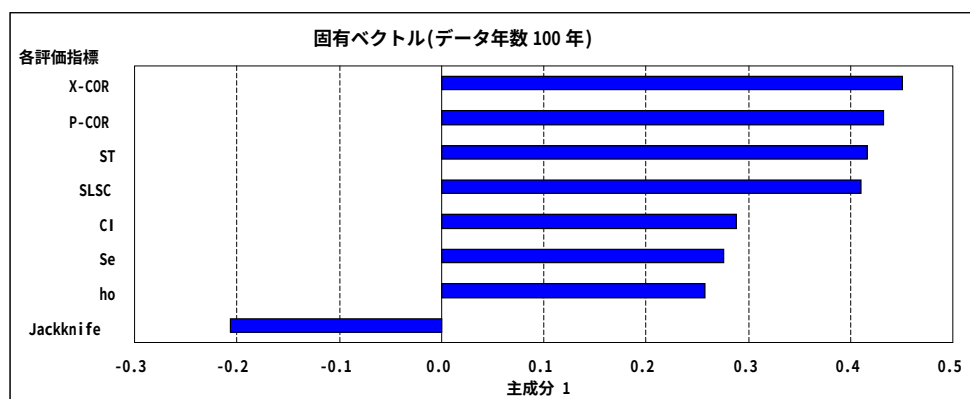


図 9-2 主成分 1 の各指標の固有ベクトル(データ年数 100 年)

表 9-6 により、データ年数別に第 1 主成分に対する各指標の固有ベクトルの傾向をみると、データ 20 年において、P-COR と X-COR の分布の適合度を表す指標が、主成分に与える影響の上位に位置しており、分布の安定性を表す指標である Jackknife 誤差分散がのみ他の評価指標とは逆の傾向を表している。データ 40 年では、全ての評価指標が同じ傾向を持ち、主成分に対する影響の上位は、信頼係数 h_0 、次いで P-COR、推定誤差 S_T の順となっている。データ 60 年では、影響の上位は X-COR、次いで S_T 、SLSC となっている。データ 100 年では図 9-2 に示すように、主成分に対する影響の上位は、X-COR、次いで P-COR、 S_T となっている。ここでも Jackknife 誤差分散は、他の指標とは別の傾向を表している。

次に、4 つデータ年数の平均の指標値から第 1 主成分における各指標の固有ベクトルの傾向を示す。

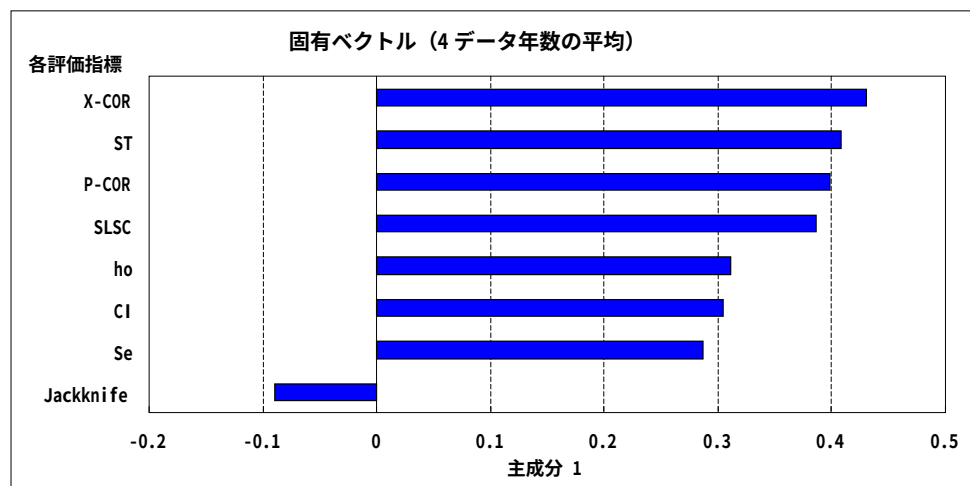


図 9-3 主成分 1 の各指標の固有ベクトル(4 データ年数の平均)

4 つのデータ年数の平均から第 1 主成分に対する各指標の固有ベクトルの傾向について見てみると、上位に位置するのは、X-COR、次いで推定誤差 S_T 、P-COR、SLSC であり、分布の適合度の指標と推定誤差 S_T が大きく影響していることがわかる。このことから、分布の適合度に関する指標 (X-COR、P-COR、SLSC) には類似性があり、 S_T についても同様な傾向があると判断できる。また、分布の安定性の指標である Jackknife 誤差分散は、他の指標とは逆の関係性を有している。適合度に関する指標が多いためこれが総合力を示す第 1 成分となっている可能性があり、第 1 主成分のプラス側を適合度の指標、マイナス側を安定性の指標という意味合いもあると考えられる。

b) 第2主成分

次に第2主成分の分析結果から、各評価指標間の類似性について傾向を見てみる。

表 9-7 データ年数に対する主成分2の各指標の固有ベクトル

データ年数 確率分布	主成分 2				平均
	20	40	60	100	
CI	0.3822	0.4788	0.4963	0.4668	0.4560
ST	0.1463	-0.2288	-0.1031	-0.2607	-0.1116
Jackknife	0.6118	0.4516	0.5621	0.5931	0.5547
SLSC	-0.4266	-0.3578	-0.0973	-0.1127	-0.2486
X-COR	0.0091	-0.2577	0.0220	0.1034	-0.0308
P-COR	-0.1941	-0.2376	-0.3636	-0.2327	-0.2570
ho	0.3159	-0.0564	0.2038	0.2715	0.1837
Se	0.3724	0.5102	0.4934	0.4594	0.4589

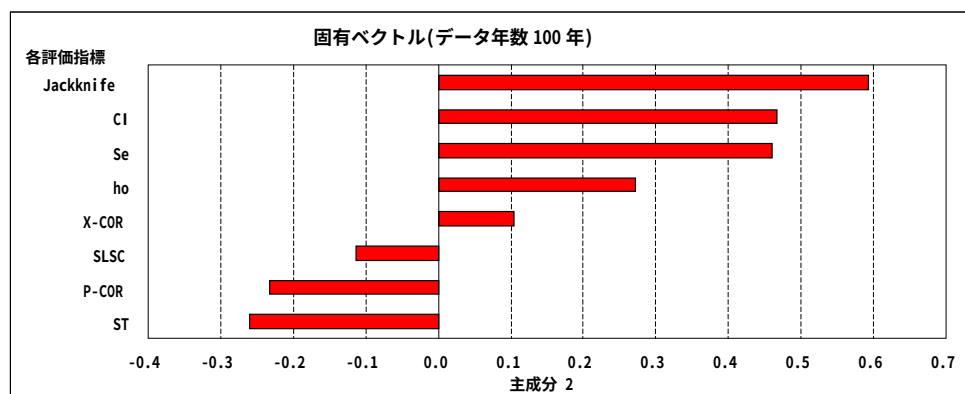


図 9-4 主成分2の各指標の固有ベクトル(データ年数 100 年)

これらから、データ年数別に第2主成分に対する各指標の固有ベクトルの傾向をみると、データ20年、60年、100年において、分布安定性の指標である Jackknife 誤差分散が第2主成分に与える影響の最上位に位置している。データ40年では、主成分に対する影響の上位は、誤差の標準偏差 Se 、次いで信頼区間の指標 CI 、Jackknife 誤差分散の順となっている。また、分布の適合度を表す指標である $SLSC$ 、 $P-COR$ などが、第2主成分に対するマイナスの傾向を表している。

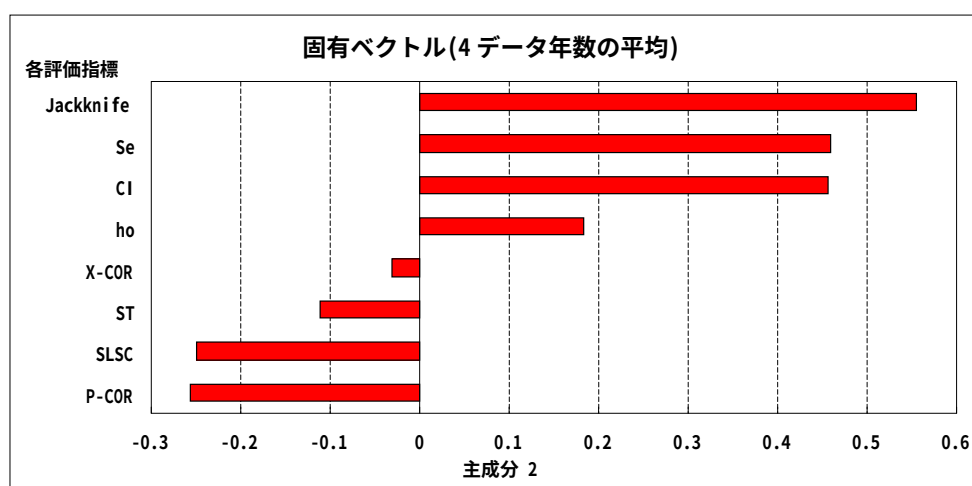


図 9-5 主成分2の各指標の固有ベクトル(データ全体の平均)

4つのデータ年数の平均より第2主成分に対する各指標の固有ベクトルについて見てみると、図 9-5 に示すように、上位に位置するのは Jackknife 誤差分散、次いで Se 、 CI 、 h_0 であり、分布の安定性を表す指標と本研究で提案した信頼区間の指標 CI 、それに関係する誤差の標準偏差 Se 、信頼係数 h_0 につ

いて類似性がある可能性がある」と判断できる。また、分布の適合度の指標を表す P-COR, SLSC, X-COR と推定誤差 S_T が他の指標とは逆の傾向を表している。これらのことから、第 2 主成分の分析結果は、主として分布の安定性と CI の信頼性が合わさった要素と、それ以外の要素との関連性を表しているものと思われる。

(6) 指標の類似性

この第 1 主成分を Y 軸、第 2 主成分を X 軸とし、各評価指標の相対的位置関係をデータ年数 100 年及び 40 年について点グラフで示し、各評価指標間の類似性について考察を行った。

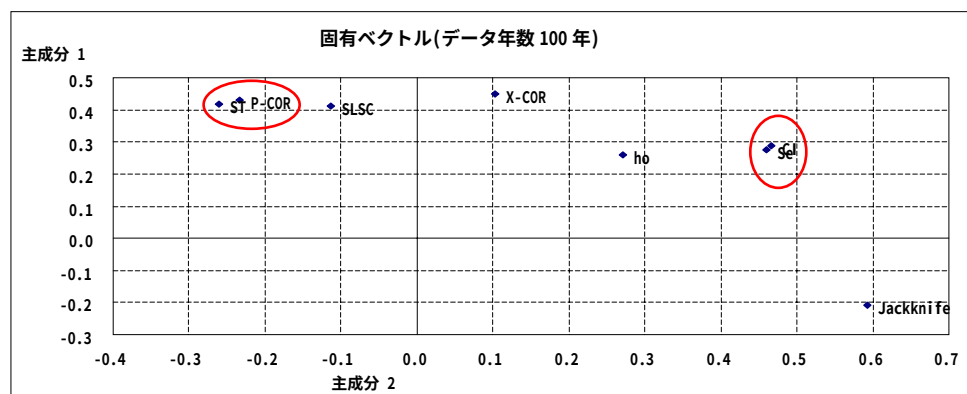


図 9-6 各評価指標の相対的位置関係(データ年数 100 年)

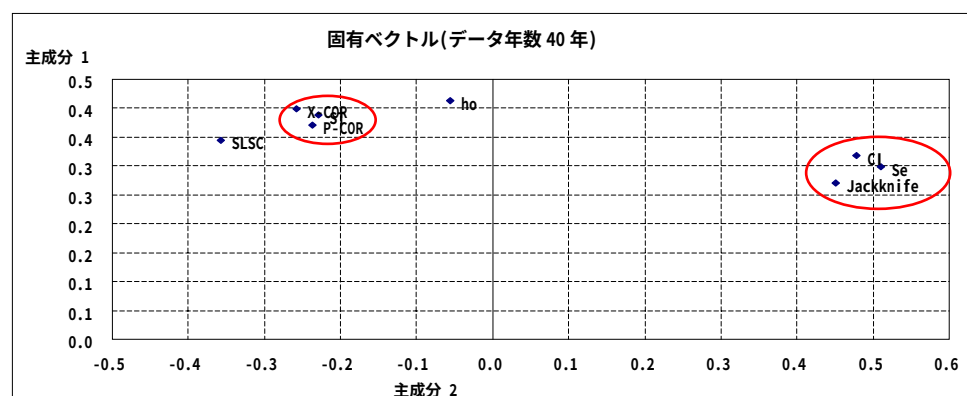


図 9-7 各評価指標の相対的位置関係(データ年数 40 年)

データ年数別に各評価指標の相対的位置関係について見てみると、例えばデータ 100 年では、CI と Se のほか、 S_T と P-COR が近い傾向を表し、Jackknife 誤差分散が、他の評価指標とは最も傾向が離れている。またデータ 40 年では、CI, Se, Jackknife 誤差分散が近い傾向を示し、分布の適合度である X-COR, P-COR と推定誤差 S_T が近い傾向で、データ年により違いがある事がわかる。

次ページの図 9-8 から 4 つの各評価指標のデータ年数の平均より相対的位置関係について見てみると、指標の類似性は次の 4 パターンに分けられる。一つは信頼区間の指標 CI・Se、二つ目は分布の適合度の指標である P-COR・SLSC・X-COR と推定誤差 S_T 、三つ目は信頼係数 h_0 、四つ目は分布の安定性の指標である Jackknife 誤差分散である。これから言えることは、分布の適合度の指標 P-COR と SLSC は極めて類似性があり、これらと X-COR, 推定誤差 S_T についても類似性がある。また、CI と Se に極めて類似性がみられ、信頼係数 h_0 は SLSC・P-COR・X-COR・ S_T のグループと CI・Se のグループの間に位置している。分布の安定性の指標である Jackknife 誤差分散は、他の指標とは別の傾向を表している。

Se と h_0 は $CI = Se \times h_0$ であるから、1つの指標 CI の一部と考えることができ、 CI はJackknife誤差分散とその他の適合度を示す指標の中間の性質を持つと考えることができる。

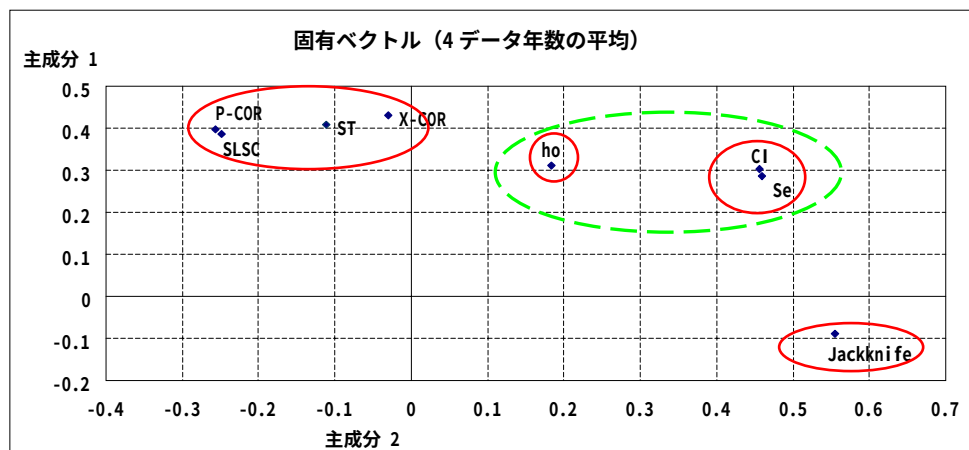


図 9-8 各評価指標の相対的位置関係(4 データ年数の平均)

(7) 主成分得点

最後に、主成分得点を用いて各評価指標間の類似性を確率分布から傾向を見ることにする。各確率分布に対する主成分得点 Z をデータ年数 20 年、40 年、60 年、100 年についてまとめたものが表 9-8 であり、データ年数 100 年で第 1 主成分を縦軸、第 2 主成分を横軸とする点グラフにした例が、図 9-9 である。また、4 つのデータ年数の平均を用いて点グラフを描いたものを図 9-10 に示す。

表 9-8 データ年数別の各確率分布に対する主成分得点

データ年数	確率分布	主成分得点			
		主成分 1	主成分 2	主成分 3	主成分 4
20	3母数対数正規分布	0.7945	1.6043	0.5342	0.0000
	グンベル分布	-3.1009	0.0545	-0.7442	0.0000
	GEV分布	2.5207	-0.5342	-0.9825	0.0000
	対数ピアソンIII型分布	-0.2143	-1.1246	1.1926	0.0000
40	3母数対数正規分布	1.7133	1.0506	0.8855	0.0000
	グンベル分布	-3.2558	0.7289	-0.1449	0.0000
	GEV分布	1.7355	0.1191	-1.1055	0.0000
	対数ピアソンIII型分布	-0.1930	-1.8987	0.3650	0.0000
60	3母数対数正規分布	0.8943	1.9725	0.7543	0.0000
	グンベル分布	-3.0671	0.0524	-0.2724	0.0000
	GEV分布	1.4738	-0.2423	-1.5143	0.0000
	対数ピアソンIII型分布	0.6990	-1.7826	1.0325	0.0000
100	3母数対数正規分布	1.1220	1.9212	0.4261	0.0000
	グンベル分布	-3.2585	0.0729	-0.1505	0.0000
	GEV分布	1.3252	-0.5845	-1.4497	0.0000
	対数ピアソンIII型分布	0.8113	-1.4095	1.1741	0.0000
平均	3母数対数正規分布	1.1310	1.6371	0.6500	0.0000
	グンベル分布	-3.1706	0.2272	-0.3280	0.0000
	GEV分布	1.7638	-0.3105	-1.2630	0.0000
	対数ピアソンIII型分布	0.2758	-1.5539	0.9410	0.0000

各確率分布に対する主成分得点の傾向をみると、3母数対数正規分布は分布の中で最も CI などの分布の信頼性の指標に関係性が強く、一定の適合度および安定性を持つ。対数ピアソンIII型分布は最も分布の適合度との関係性が大きい。また、GEV分布は分布の適合度に関係性が強く、分布の中で総合力が最も良い。グンベル分布は分布の安定性との関係性が強く、分布の中で総合力が最も悪い。

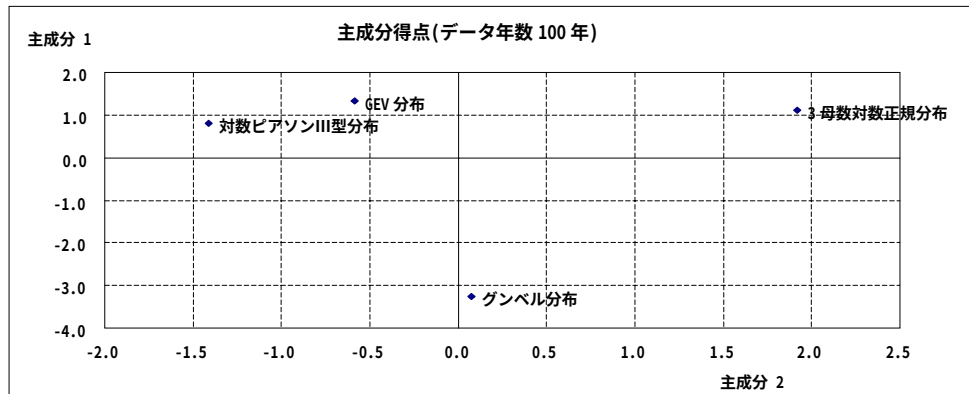


図 9-9 各確率分布に対する主成分得点(データ年数 100 年)

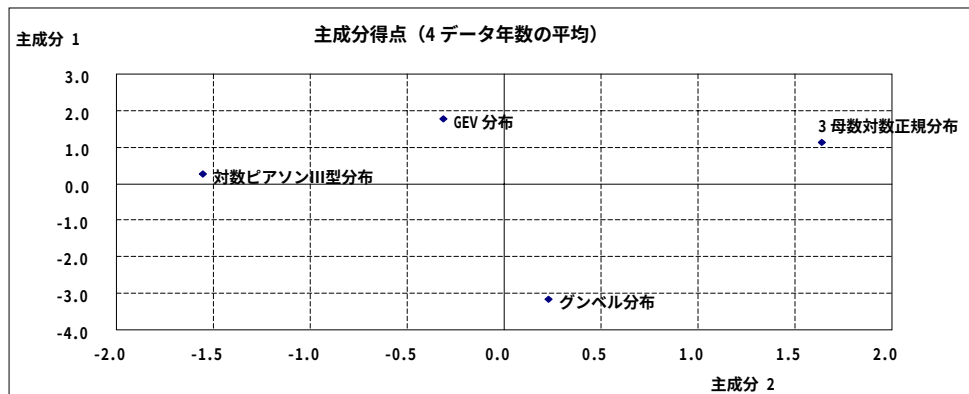


図 9-10 各確率分布に対する主成分得点(4 データ年数の平均)

10. 考察

本研究から得られた知見を列記する。

- 1) 検討対象の確率分布としては、対数正規分布の代表として 3 母数対数正規分布積率解、極値分布としてグンベル分布及び GEV 分布の L 積率解、ガンマ分布の代表として対数ピアソンIII型分布の積率解の 4 分布を用いた。対象とした年最大日雨量データで 4 分布の確率値を計算すると、確率値の小さい方から 3 母数対数正規分布、GEV 分布、対数ピアソンIII型分布、グンベル分布の順になった。
- 2) 確率分布を評価する指標として近年主に用いられているものに SLSC がある。これは原データと確率分布上で対応する推定値の「適合度」を表すもので、同様な指標として、重相関係数 (X-COR, P-COR) がある。これらの指標を各確率分布について計算すると、上記の適合度 3 指標の傾向はほぼ等しく、分布の適合度の順位では GEV 分布と 3 母数対数正規分布、対数ピアソン分布の差はほとんど無く、グンベル分布の適合度が悪いという結果になった。
- 3) 近年議論されている確率分布の判断基準として分布の「安定性」がある。実際の計画では新たにデータが加わっても確率値が大きく変動しないことが望まれる、その判断の指標としてリサンプリング法が提唱されている。ここではジャックナイフ誤差分散を用いて分布の比較を行なったところ、グンベル分布と 3 母数対数正規分布が小さく安定性が良く、対数ピアソンIII型分布、GEV 分布の順となり、2) の適合度とは異なる傾向であった。
- 4) 現在ほとんど用いられていないが、別の判断指標として確率値の「信頼性」がある。これは確率値の誤差分散あるいは標準偏差を指標とするもので、分布の適合度だけでなく、確率値そのものの信頼性を求める。ジャックナイフ誤差分散は安定性のほかに信頼性の指標とも考えられている。

本研究では「信頼性」をもう一つの判断基準と位置づけ、新たに回帰直線の信頼限界の考え方から求められる「予測値の誤差標準偏差 CI 」を水文確率値に適用する方法を考案し検討した。回帰直線の誤差分散 Se を用いると $CI = Se \times h_0$ と表され、その h_0 を信頼係数と定義した。 h_0 は対数正規分布やグンベル分布ではデータの値に左右されない独特な係数である。しかし CI に対する Se と h_0 の影響を調べたところ Se のほうがかなり大きいことが分かった。

- 5) CI および h_0 の算出にはプロットングポジションを使用する。Hazen, Weibull, Cunnane の 3 つのプロットによる CI および h_0 を比較すると、確率分布ごとの指標の大きさの傾向には殆ど差はなく、 CI は信頼性の高い（値が小さい）順に、3 母数対数正規分布、GEV 分布であり、対数ピアソン分布とグンベル分布は同程度であった。また h_0 は 3 母数対数正規分布、対数ピアソン分布、GEV 分布、グンベル分布の順であり、総合的に見て CI および h_0 については 3 母数対数正規分布が良好である。

- 6) 誤差分散は「確率分布の母数の誤差分散 S_T^2 」によっても算出することが出来る。この方法による推定誤差 S_T を各確率分布に適用して算出すると、GEV 分布が小さくて、グンベル分布が最も悪い。

CI 同様に「信頼性の指標」と位置づけたいが、傾向は「適合度」の指標に近く、検討が必要である。

- 7) 全体的な指標間の関係を知るために指標間で単相関を取ってみると、相関の強い組み合わせとして Se と CI 、SLSC と X-COR および P-COR、X-COR と P-COR などが示された。しかし多くの指標については個々の相関ではあまり傾向が明確にならなかった。

- 8) 指標間の関連、分布と指標の関連等をより明確にするために主成分分析を行なった。数値のオーダーや指標の良好性の方向を合わせるため、標準化を行ない、X-COR、P-COR 以外の指標の正負を変えて分析計算を行った。4 つのデータ年数で行なったが累積寄与率は第 2 主成分でいずれも 80% 以上であり良好であった。

一般的に総合力を表すとされる第一主成分において、影響度を表す固有ベクトルは比較的差は小さいが X-COR、 S_T 、P-COR、SLSC、 CI の順に大きく、この順に分布関数の総合力に対する影響度が大きいという解釈が出来る。またジャックナイフ誤差分散のみがマイナスで、逆の関係性を有している。なお第 1 主成分はプラスが適合度、マイナスがジャックナイフ誤差分散の安定性を示す指標という見方も可能である。

- 9) 主成分分析における第 2 主成分が大きいものはジャックナイフ誤差分散、 Se 、 CI 、 h_0 の順であり、その他の指標はマイナスである。これにより第 2 主成分はジャックナイフ誤差分散の安定性と CI の信頼性に関する指標がプラスで、適合度の指標がマイナスといえる。二つの主成分から 8 つの指標を分類すると、安定性を表すジャックナイフ誤差分散、適合度を中心にしたグループ、これらの中間の性質を持つ CI のグループの 3 つに分けることが出来る。

- 10) 主成分分析による各確率分布を主成分得点から見ると第 1 主成分は大きい順に GEV 分布、3 母数対数正規分布、対数ピアソン分布、グンベル分布であり、第 2 主成分は 3 母数対数正規分布、グンベル分布、GEV 分布、対数ピアソン III 型分布の順である。

固有ベクトルと対比させ 4 つの確率分布を評価すると、GEV 分布は X-COR などの適合度が高く総合的にも優れているが安定性が低い分布、3 母数対数正規分布は安定性と CI の持つ信頼性に優れた分布、ピアソン III 型分布は SLSC 等の適合度が最も高いが安定性が低い分布、グンベル分布は安定性が高いが適合度や CI の信頼性が低い分布である。

11. おわりに

水文確率値を算定する場合どのような確率分布が最適か、またそれを判断する指標は何が適当であり、指標相互の関係および確率分布との関連性はどのようなになっているかを調べるため、道内 8 箇所の日雨量データを用いて検討を行った。

本研究では、近年問題にされる適合度と安定性の両方を満たす確率分布を模索したが、検討結果から、それぞれの指標が良好な確率分布はかなり異なっていることが分かった。その中で 3 母数対数関分布が比較的両方の指標で良好である。また第 3 の指標として「信頼性」を考えて、誤差分散あるいは誤差標準偏差を 3 つの手法で求めたが、 S_T 、 CI 、ジャックナイフ誤差分散のそれぞれの指標が良好な確率分布には違いがみられ、誤差分散の指標の中でも持っている性質に違いがあることがわかった。

主成分分析の結果は採用するデータによって違いが出るため、確率分布や評価指標の特性把握のためには、さらにいろいろなデータを用いて他の分布関数や変数も組み合わせ、検討を行うことが求められる。また確率値はその母数推定方法によってかなり異なる値をとることがあるので、他の母数推定方法についても検討する必要がある。

参考文献

- 1) 星 清: 開発土木研究月報, No540, pp.35-48, 1998.
- 2) (財)国土開発技術研究センター: 高水計画の手引き検討の手引き(案) 参考資料, pp5-24, 2000.
- 3) 竹内 啓, 芳賀 敏郎, 野澤 昌弘, 岸本敦司: SAS による回帰分析, pp.197-203, 1996.
- 4) ラリー・ゴミック, ウルコット・スミス: 確率統計が驚異的によくわかる, pp187-206, 1995.
- 5) 馬場 敬之, 久池井 茂: 確率統計キャンパスゼミ, pp.168-172, pp.128-131, 2003.
- 6) 水文・水資源学会編集: 水文・水資源ハンドブック, pp239-247, 1997.
- 7) 田中 茂信, 宝 馨: 河川流量の頻度解析における適合度と安定性の評価, 水工学論文集第 43 巻, pp.127-132, 1999.
- 8) 宝 馨, 高棹 琢馬: 水文頻度解析における確率分布モデルの評価基準, 土木学会文集, 第 393 号/II -9, pp153-156, 1988.
- 9) David R. Maidment: HANDBOOK OF HYDROLOGY, Chapter18, pp.29-33, 1993.
- 10) G.W.Kite: FREQUENCY AND RISK ANALYSES IN HYDROLOGY, pp.33-127, 1977.
- 11) 池淵 周一, 椎葉充晴, 宝馨, 立川康人: エース水文学, pp.159-179, 2006.
- 12) 管 民朗: 図解 統計分析教室, pp252-296, 2006.
- 13) 管 民朗: 多変量解析の実践上, pp127-157, 1993.