

II-310 A I C を用いた水文量頻度分布の同一性の検討

岐阜工業高等専門学校 正員 鈴木 正人 名古屋工業大学 正員 長尾 正志

1. はじめに

従来、水文量を統計的に取り扱う際には、おもに資料の定常性、均質性を仮定し、確率水文量などの計算が行われてきた。本研究は、各年の水文量頻度分布の同一性の検討をA I Cを用いて行う。さらに、分布が同一でないと判断された場合には、実用的に同一とみなせるもの同士にグループ分けし、同一性を満たすような水文資料の分割を試みたものである。

2. 多項分布モデルによる分布の同一性の判定

複数の頻度分布が与えられた場合、多項分布モデルを基礎にそれらの頻度分布の同一性をA I Cにより判定する手法が提案されている¹⁾。その詳細は文献を参考にしていただくとして要点を記述しておく。

n 回の試行が独立で、各回の試行の結果が c 個の排反な事象に分けられる場合、1番目の事象 E_1 が k_1 回、2番目の事象 E_2 が k_2 回、 $\dots$ 、 c 番目の事象が k_c 回起こる確率は次式で多項分布で与えられる。

$$m(k_1, \dots, k_c | p_1, \dots, p_c) = \frac{n!}{k_1! \dots k_c!} p_1^{k_1} \dots p_c^{k_c} \quad (1)$$

パラメータ $p_1, \dots, p_c$ はそれぞれ対応する事象の生起確率であり、その最尤解は次式で与えられる。

$$p_i = n_i / n \quad (i = 0, 1, \dots, c), \quad n_i: \text{事象 } i \text{ が起こった度数} \quad (2)$$

これより、各年の水文量頻度分布を多項分布でモデル化することで分布の同一性の検討を行う。以下の記号を採用する。

$n(i_1, i_2)$: 各年毎の水文量の頻度分布 (i_1 : 年、 i_2 : 頻度分布の階級, $i_1 = 1, 2, \dots, c_1$;
 $i_2 = 1, 2, \dots, c_2$)

$n(i_1)$: i_1 年の資料数

$p(i_2 | i_1)$: i_1 年に i_2 階級の水文量が生起する確率

いま、ある資料 $\{n(i_1, i_2)\}$ が得られる確率($= P[\{n(i_1, i_2)\} | \{p(i_2 | i_1)\}]$)は、 c_2 個の項を持つ、 c_1 個の多項分布の積で与えられ、 $p(i_2 | i_1)$ をパラメータとみなしたときの対数尤度は次式で表現される。

$$L(p(i_2 | i_1)) = K_4 + \sum_{i_1=1}^{c_1} \sum_{i_2=1}^{c_2} n(i_1, i_2) \ln p(i_2 | i_1) \quad (3)$$

$$K_4 = \ln \left\{ \frac{c_1!}{\prod_{i_1=1}^{c_1} n(i_1)!} \prod_{i_1=1}^{c_1} \frac{c_2!}{\prod_{i_2=1}^{c_2} n(i_1, i_2)} \right\}$$

パラメータ $p(i_2 | i_1)$ の与えかたによりモデルを表現し、分布の同一性の判定を行う。すなわち、各年が同一の分布に従うとするモデルMODEL(1)、各年がそれぞれ異なる分布に従うとするモデルMODEL(c_1)、全体が M 個($1 \leq M \leq c_1$)の分布に従うとするモデルMODEL(M)のパラメータをそれぞれつぎのように表現する。

$$\begin{aligned} \text{MODEL}(1) : p(i_2 | i_1) &= \theta(i_2) && \text{自由パラメータ数: } c_2 - 1 \text{ 個} \\ \text{MODEL}(c_1) : p(i_2 | i_1) &= \theta(i_2 | i_1) && \text{〃 : } c_1 \times (c_2 - 1) \text{ 個} \\ \text{MODEL}(M) : p(i_2 | i_1) &= \theta(i_2 | m_j) \quad (j=1, 2, \dots, M) && \text{〃 : } M \times (c_2 - 1) \text{ 個} \\ \theta(i_2 | m_j) &= \sum_{i_1 \in m_j} n(i_1, i_2) / \sum_{i_1 \in m_j} n(i_1), \quad m_j: \text{分割されたグループ} && (4) \end{aligned}$$

上式のMODEL(M)で $M=1$ 、 c_1 とすると、それぞれMODEL(1)、MODEL(c_1)となる。

各モデルのA I Cを計算し、MODEL(1)が採択されれば、分布は同一であると判断される。また、MODEL(M)が採択されれば、分布は同一でないと判断されるとともに、対応した資料分割が得られる。

3. 適用計算

3.1 計算方法 1891年～1990年までの岐阜市日降水量を資料として用い、各年の日降水量の頻度分布を同一性の検討対象とした。階級数は全資料数からスタージェスの式により18個、階級幅は最大、最小値より

15mmとした。また、資料分割の方法では、無作為に100組の資料を分割すると、2組に分割するだけで、 $(2^{100} - 2) / 2$ 通りの分割の方法があり計算量が膨大になる。そこで、頻度分布の特性を表す指標として、基礎統計量により各年の資料を昇順に順位づけし、その順位にしたがって2組、および3組に分割した。たとえば、2組に分割する場合は、1位~100位まで順位づけし、1位と2~100位、1~2位と3~100位、・・・、1~99位と100位というように99通りの分割の仕方がある。各分割モデルにおけるAICを比較することで、その指標における最適な分割が求まり、各指標における最適分割モデルのAICを比較することで、どの指標による分割が適当かが判断できる。具体的には、2組の分割には、平均、分散、変動係数、歪係数、無降雨日数、年最大日降水量を、3組の分割には2組分割の結果を参考にして、平均、分散、年最大日降水量を指標として用いた。さらに、2組に分割されたグループを2分割することで4分割、3分割することで6分割といったように、最大8分割までの計算を行った。

3. 2 結果および考察 各指標による最適分割モデルのAICを表-1に示す。表中で、指標に関わらず1分割がMODEL(1)に、100分割がMODEL(100)に対応している。また、6分割において3-2は3分割を2分割、2-3は2分割を3分割したことを意味する。AICが最小なのは、年最大日降水量を指標とした6分割モデルであり、各年の日降水量頻度分布が同一とみなすよりも6つの分布で表現したほうが妥当と思われる。指標別にみても、どの分割数に対しても年最大日降水量を指標としたモデルのAICが最小となる。また、年代順に並べた資料を2分割した場合は、MODEL(1)のAICより値が大きいくことから、経年的にみて日降水量の頻度分布に一方的な変化があるとは思えない。

年最大日降水量を指標とする最適な6分割(2-3分割)の結果は1~26, 27~44, 45~57, 58~73, 74~87, 88~100位であり、100年のうち13年分が年最大日降水量が大きいグループに属している。分割された各グループの相対頻度分布および頻度分布を表-2に示す。最大階級は段階的に減少しているが他の階級の相対頻度には明確な変化は認められないので、本研究で用いたモデル

表-1 各指標による分割モデルのAIC

	平均	分散	歪係数	最大	無降雨	年代順
1分割	36808.34	同左	同左	同左	同左	同左
2分割	36744.78	36742.25	36752.88	36731.93	36807.88	36813.64
3分割	36728.65	36719.78	36724.37	36677.84	——	——
4分割	36729.16	36713.78	36724.92	36652.72	36831.09	——
6分割	3-2	36759.00	36737.83	36745.25	36656.58	——
	2-3	36754.33	36733.12	36741.23	36642.67	——
8分割	36787.95	36770.18	36781.85	36654.42	——	——
100分割	39155.62	同左	同左	同左	同左	同左

※6分割で3-2は、3分割をそれぞれ2分割、2-3は2分割を3分割したもの

表-2 最大値を指標とした最適分割の相対頻度分布(頻度分布)

i	up	middle1	middle2	middle3	middle4	low
1	0.88(4161)	0.89(4541)	0.89(5194)	0.89(4193)	0.89(5840)	0.89(8497)
2	0.07(312)	0.05(277)	0.06(354)	0.06(305)	0.07(433)	0.06(537)
3	0.02(116)	0.03(145)	0.03(147)	0.03(131)	0.02(157)	0.03(256)
4	0.02(80)	0.01(61)	0.01(60)	0.01(58)	0.01(69)	0.01(115)
5	0.01(31)	0.01(39)	0.01(32)	0.00(23)	0.01(36)	0.00(60)
6	0.00(10)	0.00(15)	0.00(20)	0.00(18)	0.00(16)	0.00(34)
7	0.00(10)	0.00(12)	0.00(14)	0.00(4)	0.00(22)	0(0)
8	0.00(7)	0.00(1)	0.00(7)	0.00(15)	0(0)	0(0)
9	0.00(1)	0.00(5)	0.00(18)	0(0)	0(0)	0(0)
10	0(0)	0.00(10)	0(0)	0(0)	0(0)	0(0)
11	0(0)	0.00(5)	0(0)	0(0)	0(0)	0(0)
12	0.00(6)	0(0)	0(0)	0(0)	0(0)	0(0)
13	0.00(3)	0(0)	0(0)	0(0)	0(0)	0(0)
14	0.00(5)	0(0)	0(0)	0(0)	0(0)	0(0)
15	0.00(2)	0(0)	0(0)	0(0)	0(0)	0(0)
16	0.00(1)	0(0)	0(0)	0(0)	0(0)	0(0)
17	0.00(2)	0(0)	0(0)	0(0)	0(0)	0(0)
18	0.00(2)	0(0)	0(0)	0(0)	0(0)	0(0)

4. おわりに

今回は、日降水量の頻度分布を一変数の多項分布でモデル化したため、自己相関性は考慮されていないが、二変数の多項分布でモデル化することで、自己相関性を考慮した同一性の検討も可能であると考えている。

最後に、資料収集・整理等において前岐阜工業高等専門学校学生加藤昇君の助力をえたことを記し謝意を表する。

参考文献 1) 坂元慶行・石黒真木夫・北川源四郎：情報量統計学，共立出版 pp.74~77, 1983