水工学論文集　第37巻　1993年２月

多項分布モデルによる日降水量頻度分布の同一性の検討とグループ分け

Checking Homogeneity of Frequency-distributions of Daily Precipitation and Their Practical Grouping by Multinomial Distribution Model and AIC

鈴木正人*・長尾正志**
By Masato SUZUKI and Masashi NAGAO

Steadiness and homogeneity of hydrological quantities are mainly assumed in a customary stochastic analysis. This study gives reconsideration to this assumption, through checking homogeneity of shape of frequency-distributions. Judgement of homogeneity is given by AIC on the basis of multinomial-distribution model. If every distributions are judged not homogenous, practical grouping of those distributions is proposed by the same method.

Numerical calculation is carried out with daily precipitation in Gifu city from 1891 to 1990. It is shown that six-parts dividing model in order of maximum daily precipitation is most suitable.

Keywords: frequency-distribution, homogeneity, multinomial-distribution, AIC

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・均質性を前提とし，資料の母集団は一つと考えて，確率水文量などの計算が行われてきた。もちろん異常値検定に代表される異質な資料を除外する試みはなされてきたが，あくまでも，主体となる母集団は唯一であるという視点に立脚したものといえよう。そこで，本研究は，日降水量の頻度分布を対象として，従来”前提とする”の一言のもとに片づけられていた水文資料の均質性・定常性の仮定を見直し，均質でないと判断された場合は，実用的に同一とみなせるもの同士にグループ分けすることで，均質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ＡＩＣ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，ここでは周辺・同時の両分布についてその要点を記述しておく。

* 正会員　工博　岐阜工業高等専門学校助手　土木工学科
（〒５０１－０４　岐阜県本巣郡真正町）
** 正会員　工博　名古屋工業大学教授　工学部社会開発工学科
（〒４６６　名古屋市昭和区御器所町）

2．1　周辺分布の同一性の判定手法[2)]

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E1がk1 回，2番目の事象E2がk2回，・・・，c番目の事象がkc回起こる確率は次式の多項分布で与えられる。

$$m(k1,\cdots,kc\mid p1,\cdots\cdots,pc)=\frac{n!}{k1!\cdot\cdots\cdot kc!}\,p1^{k1}\cdot\cdots pc^{kc} \tag{1}$$

パラメータp1，・・・，pc 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。

$$pi=ni/n\ (i=0,\ 1,\ \cdot\cdot,\ c),\quad ni:\text{事象}i\text{の生起度数},\ n=\Sigma ni \tag{2}$$

以下の記号を用い，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。

n：全資料数（計算対象期間の全日数）

n（i1，i2）：各年毎の水文量の頻度分布（i1：年，i2：頻度分布の階級，i1＝1，2・・，c1；i2＝1，2，・・，c2）

n（i1）：i1年の資料数（365 または 366）

p（i2｜i1）：i1年にi2階級の水文量が生起する確率（$\sum_{i2} p(i2\mid i1)=1$）

いま，ある資料｛n（i1，i2）｝が得られる確率（＝P〔{n(i1,i2)}|{p(i2|i1)}〕）は，次式のようにc2個の項を持つc1個の多項分布の積で与えられる。

$$P[\{n(i1,i2)\}|\{p(i2|i1)\}]=\prod_{i1=1}^{c1}\left\{\frac{n(i1)!}{\prod_{i2=1}^{c2}n(i1,i2)!}\prod_{i2=1}^{c2}p(i2\mid i1)^{n(i1,i2)}\right\} \tag{3}$$

ここで，p（i2｜i1）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。

$$L[p(i2\mid i1)]=K4+\sum_{i1}\sum_{i2}n(i1,i2)\ln p(i2\mid i1)$$
$$K4=\ln\left\{\prod_{i1=1}^{c1}n(i1)!\Big/\prod_{i1=1}^{c1}\prod_{i2=1}^{c2}n(i1,i2)!\right\} \tag{4}$$

以下ではパラメータp（i2｜i1）の与えかたを変えて，それぞれ異なるモデルを表現し，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（c1），全体がM個（1≦M≦c1）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。

MODEL（1）：$p(i2\mid i1)=\theta(i2)$　　自由パラメータ数：c2－1個

MODEL（c1）：$p(i2\mid i1)=\theta(i2\mid i1)$　　〃　　：c1×（c2－1）個

MODEL（M）：$p(i2\mid i1)=\theta(i2\mid mj)\ (j=1,2,\cdot\cdot,M)$　　〃　　：M×（c2－1）個

$$\theta(i2\mid mj)=\sum_{i1\in mj}n(i1,i2)\Big/\sum_{i1\in mj}n(i1),\quad mj:\text{分割されたグループ} \tag{5}$$

上式のMODEL（M）でM＝1，c1とすると，それぞれMODEL（1），MODEL（c1）となる。

各モデルのＡＩＣは次式で求められる。

$$AIC=-2\times L[p(i2\mid i1)]+2\times\text{自由パラメータ数} \tag{6}$$

対数尤度L〔p（i2｜i1）〕は，基本的には（4）式で求められるが，各モデルともK4の項は共通なので，実際の計算にはK4を減じた右辺第２項を用いた。

各モデルのＡＩＣ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，対応した資料分割数，および分割方法が得られる。

2．2　同時分布の同一性の判定手法[3)]

日降水量を計算対象として用いる場合，その生起量の持続性が無視できない場合がある。そこで一次の自己相関を考慮した同時頻度分布に対しても，その同一性の判定手法を記述しておく。以下の記号を採用する。

n（i1，i2，i3）：i1年にi2，i3階級の日降水量が継続して生起した頻度

$$(i1=1,2,\cdots,c1:\ i2,i3=1,2,\cdots,c2)$$

$p(i2,i3|i1)$ ：i1年という条件のもとで（i2，i3）階級の日降水量が継続して生起する確率

$$\sum_{i2}\sum_{i3} p(i2,i3|i1)=1$$

$p(i1,i2,i3)$ ：i1年，（i2，i3）階級の同時生起確率

$$\sum_{i1}\sum_{i2}\sum_{i3} p(i2,i3,i1)=1$$

いま，ある資料｛n（i1，i2，i3）｝が得られる確率（＝P〔{n(i1,i2,i3)}|{p(i1,i2,i3}〕）はつぎの多項分布で表される。

$$P\{\{n(i1,i2,i3)\}|\{p(i1,i2,i3\}\}=\frac{n!}{\prod n(i1,i2,i3)!}\prod p(i2,i3,i1)^{n(i1,i2,i3)} \quad (7)$$

ここで，Πはi1，i2，i3についての総乗を意味する。（7）式より，p（i2，i3|i1）をパラメータとみなしたときの対数尤度を，p（i2，i3|i1）に着目して記述すると，次式となる。

$$L\{p(i2,i3|i1)\}=K6+\sum_{i1}\sum_{i2}\sum_{i3} n(i1,i2,i3)\ln p(i2,i3|i1)$$

$$K6=\ln\{\prod_{i1=1}^{c1} n(i1)!\ /\ \prod_{i1=1}^{c1}\prod_{i2=1}^{c2}\prod_{i3=1}^{c2} n(i1,i2,i3)!\} \quad (8)$$

$$-\ln n!-\Sigma\ln n(i)!+\Sigma n(i)\ln p(i1)$$

周辺分布の場合と同様に，パラメータp（i2，i3|i1）の与えかたによりモデルの表現を行うが，その際，同時生起確率を独立と考える，すなわち同時分布を周辺分布の積で与えるか，または従属と考えるかでパラメータの表現は異なる。周辺分布の場合と同様に，一般的なＭＯＤＥＬ（Ｍ）について，両者を記述する。まず，独立とするモデルのパラメータは，

$$p(i2,i3|i1)=p(i2|i1)\times p(i3|i1)$$

$$=\theta(i2|mj)\times\theta(i3|mj)\quad \text{自由パラメータ数：M×（c2−1）個}$$

$$\theta(i2|mj)=\sum_{i1\in mj}\sum_{i3=1}^{c2} n(i1,i2,i3)\ /\ \sum_{i1\in mj} n(i1) \quad (9)$$

また，従属とするモデルのパラメータは，

$$p(i2,i3|i1)=\theta(i2,i3|mj)\quad \text{自由パラメータ数：M×（c2×c2−1）個}$$

$$\theta(i2,i3|mj)=\sum_{i1\in mj} n(i1,i2,i3)\ /\ \sum_{i1\in mj} n(i1) \quad (10)$$

と表現される。周辺分布の場合と同様に，対数尤度のうち，共通する部分であるK6は除外し，ＡＩＣによりモデルを採択する。従属モデルは，独立モデルに比べて，パラメータ数が階級数の２乗の大きさで増加するので採択されにくくなる。

３．　分布の同一性の計算例

周辺分布モデルを例として，分布の同一性の判定の計算例を示し，判定にＡＩＣを用いる意義を明確にしておく。表－１に示す３種の頻度分布を用いる。頻度分布Ａは，岐阜市の1986年の日降水量を離散化幅１５ｍｍで離散化したもの，また頻度分布Ｂは同じく1971年のものである。また，頻度分布Ｃは頻度分布Ａの度数を各階級とも５倍したもので，相対頻度は頻度分布Ａと同一である。

まず，頻度分布Ａ，Ｂの同一性を判定する。二つの分布を同一とするＭＯＤＥＬ（１）の場合，対数尤度は次式で求められる。

表－１　計算例に用いる頻度分布（相対頻度）

階級	頻度（相対頻度）		
	A	B	C
1	330(0.904)	321(0.880)	1650(0.904)
2	19(0.052)	27(0.074)	95(0.052)
3	5(0.014)	7(0.019)	25(0.014)
4	6(0.016)	2(0.006)	30(0.016)
5	4(0.011)	1(0.003)	20(0.011)
6	0(0)	1(0.003)	0(0)
7	1(0.011)	3(0.008)	5(0.011)
8	0(0)	1(0.003)	0(0)
9	0(0)	2(0.006)	0(0)

LL=(330+321)×ln{(330+321)/730}+(19+27)×ln{(19+27)/730}+･･+(0+2)×ln{(0+2)/730}=-357.862

自由パラメータ数は階級数から１を減じた８だからこのモデルのＡＩＣ，ＡＩＣ（１）＝731.722　となる。また，二つの分布が異なるとするＭＯＤＥＬ（２）の対数尤度は，次式で求められる。

LL=330×ln{330/365}+19×ln{19/365}+･･･+321×ln{321/365}+･･･+2×ln{2/365}=-351.627

自由パラメータ数は，分布を２つと仮定しているのでＭＯＤＥＬ（１）の２倍の１６だから，ＡＩＣ（２）＝735.254となり，ＡＩＣ（１）の方が小さいことから，ＭＯＤＥＬ（１）が採択され，分布は同一と判定される。

つぎに，頻度分布Ｂ，Ｃの同一性を検討する。先述と同じ手順により，分布を同一と仮定した場合は，LL=-1002.283で，ＡＩＣ（１）＝2020.568，分布が異なるとするモデルでは，LL=-975.122でＡＩＣ（２）＝1982.244となり，ＭＯＤＥＬ（２）が採択され，分布は同一でないと判定される。

このように，相対頻度が等しい分布Ａ，Ｃでも，比較の対象となる分布により同一性の判定は異なる。これは，分布Ａ，Ｂの比較では，対数尤度はＭＯＤＥＬ（２）の方が大きく，適したモデルと考えられるが，自由パラメータ数が２倍になるので，対数尤度の差よりも自由パラメータ数の差の方が大きくなり，ＡＩＣではＭＯＤＥＬ（１）が採択される。一方，分布Ｂ，Ｃの比較では，やはり対数尤度はＭＯＤＥＬ（２）の方が大きく，自由パラメータ数もＭＯＤＥＬ（２）の方が多いのだが，頻度分布の度数が多いため，対数尤度の差が自由パラメータ数の差よりも大きく，ＡＩＣではＭＯＤＥＬ（２）が採択された。

つまり，頻度の度数が大きくなると，対数尤度の絶対値は増加するが，一方自由パラメータ数は変わらないため，ＡＩＣにおいて対数尤度が占める割合が増加し，分布形状の相違が同一性の判定に支配的に働く。逆に，分布形状にある程度の相違はあっても，度数が少ないと対数尤度の差は小さくなり，分布は同一と判定されやすくなる。したがって，分布の同一性の判定は，分布形状の相違とその生起度数とを総合的に判断して行われているといえよう。

４．適用計算

４．１　計算方法

1891年～1990年の岐阜市日降水量を資料として，各年の日降水量の頻度分布（周辺分布，および同時分布）を同一性の検討対象とした。また，全資料数からスタージェスの式により，頻度分布の階級数は１８個，階級幅は最大値，最小値より１５ｍｍとした。

分布の同一性の判定には，２．で述べたように各年の降水量頻度分布の資料についてグループ数を仮定し，幾つかのグループに分割する必要がある。その方法として，無作為に100年分の資料を分割するとすれば，全体を２組に分割するだけでも，（2^{100} －２）／２通りの分割の方法があり膨大な計算量になる。そこで，本研究では，計算の便宜を計り，基礎統計量を各年の頻度分布の特性を表す指標として用い，資料を年を一括した単位として昇順に順位づけし，その順位にしたがって２組，および３組に分割することにした。２組に分割する場合は，まず１位～100位までを順位づけしておいて，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。また，３組に分割する場合は，同様に4851（$={}_{99}C_2$）通りの分割の仕方がある。２または３分割されたグループをさらに分割することで，８分割モデルまでを検討の対象にした。

各分割モデルにおけるＡＩＣを比較すれば，その指標における最適な分割が求まる。また各指標における最適分割モデルのＡＩＣを比較することから，どの指標による分割が適当かが判断できる。

４．３　結果および考察

（ａ）　周辺分布

順位づけの指標には，２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分割には２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。各指標による最適

分割モデルのＡＩＣを表－２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を同一とみなすよりも６つの分布で表現したほうが妥当だと思われる。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より値が大きいことから，経年的にみて日降水量の頻度分布形状にトレンドのような一方的な変化は認められない。

表－２ 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	――	――
６分割 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

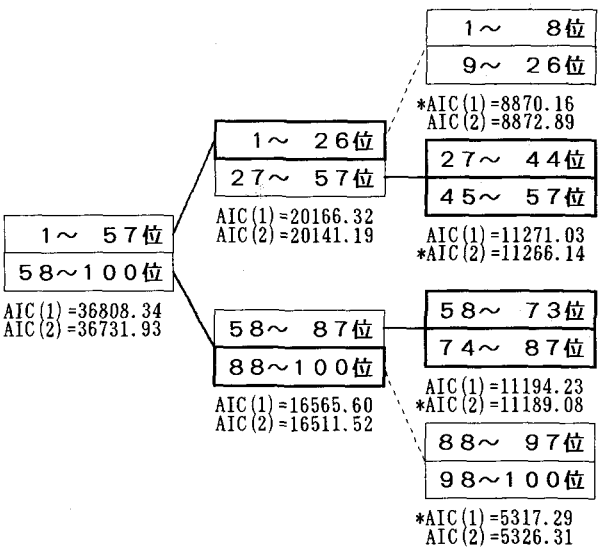
※６分割で3-2は、３分割をそれぞれ２分割、2-3は２分割を３分割したもの

表－３ 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)

つぎに，年最大日降水量を指標とした場合の８分割の手順を図－１に示す。まず全体を２分割する場合，１～57位，58～100位の分割がAIC(2)=36731.93で最も適当な分割となる。このAIC(2)をMODEL(1)のAIC(1)（=36808.34）と比較すると，AIC(2)の方が小さいので２分割モデルの方が適当と判断され，つぎの４分割へと計算を進める。以後，AIC(1)，AIC(2)をそれぞれ比較し，分割するか否かの判定をしていくことで，この手順における最も適した分割が得られる。図中で太線で囲まれた順位が最終的に得られた分割の結果（１～26，27～44，45～57，58～73，74～87，88～100位の６分割）で，４分割から８分割へのモデルでＡＩＣの値の小さい方（図中で＊を付したもの）を総計したものが，この分割モデルのＡＩＣになる。

ちなみに，この手順で得られた分割方法，ＡＩＣは，表－２に示した年最大日降水量を指標とした６分割（２－３分割）の結果と一致している。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－３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



図－１ 最大値を指標とした分割の手順（８分割）

れないので，本研究で用いたモデルでは頻度分布の階級の最大値，すなわち年最大日降水量が頻度分布の同一性の判定に支配的に関与していると思われる。

（ｂ）　同時分布

各年日降水量の同時頻度分布に対し，周辺分布への適用で有効だった年最大日降水量と，自己相関係数の２つの指標で順位づけし，２分割を繰り返すことで８分割モデルまでの検討を行った。なお，２．２で述べたように，同時分布の場合は，自己相関性を考慮するか否かで２通りのモデルが作れるので，２分割モデルの場合，分割と自己相関の有無の組み合わせで，計４通りのモデルが検討対象となる。

自己相関係数を指標とした場合は，１～20位（独立），21～29位（独立），30～100位（従属）の３分割がAIC=73865.22で最適なモデルという結果が出た。一方，年最大日降水量を指標とした場合は１～８位，９～25位，26～44位，45～56位，57～73位，74～87位，88～92位，93～100位（いずれも独立）の８分割がAIC=73509.76で最適なモデルとなり，自己相関係数を指標として用いた場合よりもＡＩＣが小さくなった。自己相関性の有無は，１～100位を一まとめに取り扱ったモデルでは，従属モデルが採用されたが，他の場合はいずれも独立モデルが採用された。これは，資料を分割することで，頻度分布の度数が減少し，ＡＩＣのうち自由パラメータ数の増加が対数尤度の増加を上回ってしまい，従属モデルが採択されなかったのであろう。なお，この分割結果を周辺分布の分割結果である図－１と比較すると，最終的な分割数に差異はあるものの分割の境界の順位はほぼ類似していることがわかる。

（ｃ）　多雨期降水量の周辺分布

以上の各年日降水量の頻度分布に対する検討では，周辺・同時分布ともに年最大日降水量を指標とした分割が適していることがわかった。そこで，年間の頻度分布ではなく，年最大日降水量が生起した期間（便宜上多雨期と呼ぶ）の周辺分布に対して，同様の適用を行った。期間の決め方としては，100年の日降水量資料に対し，最大値が生起した月日を調べ，その最早日，最遅日ではさまれた期間を多雨期とした。具体的には，２月１６日（1922年に生起）～１０月２１日（1980年に生起）の２３７日間である。順位づけの指標には，年最大日降水量，分散，平均の三種を用いたが，最終的に得られた最適モデルは，年最大日降水量を指標とした場合の，１～25位，26～44位，45～72位，73～87位，88～100位の５分割で，AIC=29982.87であった。この結果を各年単位の周辺分布に対する結果と比べると，多雨期の45～72位に相当するグループが，年単位では45～57位，58～72位に２分割されている他は完全に一致した結果となっている。これは，今回用いたモデルでは，年単位の頻度分布も，多雨期の頻度分布も，共にその同一性の判定は年最大日降水量という最大階級により左右されているからであろう。

５．　まとめ

頻度分布の同一性の検討という観点から，本研究では，一地点の日降水量頻度分布に対して検討を行った。その結果，周辺分布，同時分布，ともに年最大日降水量を指標にして順位づけした資料の分割が有効であることがわかった。多項分布モデルという比較的パラメータ数の多い分布をモデルの基礎として用いたので，同時分布では自己相関性を考慮したモデルは採択されにくかったが，二項分布，指数分布などのパラメータ数の少ない分布を用いることで，自己相関性の表現も可能になるであろう。今後は，本手法を異なる地点間の頻度分布の同一性の判定にも適用・検討していく予定である。

６．　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，1983,pp.74～77

２）鈴木正人・長尾正志：ＡＩＣによる水文量頻度分布の同一性の検討，第４７回土木学会年次学術講演会概要集，1992,第II部門，pp.670～671

３）前掲１），pp.96～105