土木学会第44回年次学術講演会（平成元年10月）

II-78　ＡＩＣを用いた相関渇水時流況の離散モデル化の研究

名古屋工業大学　○学生員　鈴木正人　正員　長尾正志

１．　研究の概要

われわれは貯水池による渇水時利水補給という観点から、統計的貯水池理論を用いて解析を行なってきたが、簡便化などの理由により、時間的、量的に離散化された量を対象とする。したがって、入力としての渇水時流況の的確な離散モデル化が重要な課題となる。しかし、渇水時流況を離散化して経験分布を求めると、離散化単位量にもよるが、通常全体の数％以下に当たる個数の、量の大きなデータの存在が、全体の積率を引っ張りモデル化に大きな影響を及ぼすことがよくある。そこで、本研究は、量の多い範囲と少ない範囲とは異なった統計的特性を持つとし、それぞれ異なった形式でモデル化し、各モデルの混合を行ったうえで、最適なモデルをＡＩＣにより選択しようとするものである。

２．　対象とするデータ

量的、時間的に離散化された時点ｔに対する流量系列ｑｔ（ｔ＝１，・・，Ｎ）を所与とする。ｑtは０，１，２，‥といった無次元の整数値で、その上限をｃとする。モデル化には、

ｎ(i)：ｑｔ＝ｉの発生個数（ｉ＝０，１，・・，ｃ）ただし、　$\sum_{i=0}^{c} n(i) = N$

ｎij：｛（ｑｔ＝ｉ）∩（ｑt+1＝ｊ）｝(t=1,‥,N-1)の発生個数(i,j=0,1,‥,c)ただし、$\sum_{i,j=0}^{c} n_{ij} = N-1$

をもつ経験分布としての周辺分布、同時分布をその対象とする。

３．　渇水時流況の離散化モデル

３．１　モデルの概念　流量時系列を自己相関性の違いにより、低水、高水流量集団にそれぞれ分離する手法[1)]が提案されている。本研究でもそれに習い、自己相関性の有無により、流量データを二つの集団に分割する。すなわち、同時分布において、分割の境界をｉ，ｊ＝ｋ（ｋ＝０，１，・・，ｃ）とし、量の少ない範囲（ｉ，ｊ＝０，１，・・，ｋ）は一次の自己相関性を考慮した同時二項分布（ｐ1ij（ｒ1，ａ1，ρ））に、また、それ以外の範囲は、相関のない二項（ｐ2(i)（ｒ2，ａ2））、ポアソン（ｑ(i)（λ））のどちらかの分布に従うと仮定する（図－１参照）。このように別個に求めた分布を混合して、全体のモデルを構成し、それをMODELm（ｋ）と呼ぶことにする。ここで、ｍ＝１，２でそれぞれ二項分布、ポアソン分布との混合を意味する。例えば、MODEL2（０）は全てのデータが無相関のポアソン分布に従うとするモデルである。最適なモデルが選択できれば、母集団の分割とその推定とを同時に行うことができる。

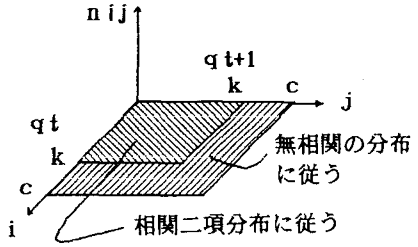
３．２　パラメータの推定　量の少ない範囲のモデル、つまり相関二項分布のパラメータ推定には、先述の仮定により、経験分布として、ｎij（ｉ、ｊ＝０，１，・・，ｋ）の同時分布を用いる。

また、量の多い範囲のモデル、つまり周辺二項、ポアソンの各分布のパラメータ推定には、相関がないという仮定にしたがい経験分布として、ｎ(i)（ｉ＝ｋ＋１，ｋ＋２，・・，ｃ）の周辺分布を用いる。パラメータ推定手法にはともに最尤法を用いた。

３．３　モデルの混合　前項で求めたパラメータをもつ相関二項分布と、独立な同時分布としてのポアソンまたは、二項分布の両分布を混合する。混合された分布をｐｑijとすれば、例えばポアソン分布との混合の場合、

$$pq_{ij} = \alpha \times p_{ij} + (1-\alpha) \times q(i) \times q(i)$$

と表される。ここで、αは混合の比率で、全データの内、相関二項分布に従うデータ数の割合を意味しており、本研究では次式で近似しておく。



図－１モデルの概念図

$$\alpha = NL / (N-1), \quad NL = \sum_{i,j=0}^{k} n_{ij} \quad \cdots\cdots (1)$$

4．ＡＩＣの計算

MODELm（ｋ）のＡＩＣ、ＡＩＣm（ｋ）を計算する。ＡＩＣm（ｋ）は、次式で定義される[2)]。

$$AICm(k) = -2 \times LLm(k) + 2 \times \text{自由パラメータ数}, \quad LLm(k) = \sum_{i,j=0}^{c} n_{ij} \times \ln(p\,q_{ij}) \quad \cdots (2)$$

また、自由パラメータ数は、それぞれ、相関二項分布のパラメータが３個、周辺二項分布が２個、ポアソン分布が１個、混合の比率が１個だから、一般的に、MODEL1（ｋ）は６個、MODEL2（ｋ）は５個、特別な場合として、MODELm（ｃ）は３個、MODEL1（０）は２個、MODEL2（０）は１個である。

各モデルの内、ＡＩＣの最小のモデルを最適なモデルとして採択する。複雑な混合をするほど、自由パラメータ数、つまりＡＩＣが大きくなり、そのモデルは不採択となる傾向がある。しかし、そのようなモデルがあえて採択されたときには、母集団全体がそれだけ異なった性質を持つ複数の集団から形成されていたといえるだろう。

5．適用計算

表１　最適モデル策定例（基礎の分布：r1＝5，a1=0.4，ρ＝0.6，r2＝10，a2=0.8，α＝0.9）

No	α	r1	a1	Corr.	MODEL1			MODEL2	
					r2	a2	AIC1	λ	AIC2
0	0.	−	−	−	10	0.2588	8945.791016	2.5878	8087.584961
1	0.3039	1	0.7077	0.3060	10	0.3382	8709.198242	3.3818	8012.357422
2	0.6158	2	0.6590	0.3900	10	0.4504	8152.145020	4.5039	7589.318848
3	0.8235	3	0.5840	0.4790	23	0.2727	6916.266113	6.2727	6871.724609
4	0.8927	4	0.4890	0.5550	10	0.7738	6386.554688	7.7383	6452.198730
5	0.9047	5	0.3999	0.6031	10	0.8084	◎6246.466797	8.0842	○6417.455566
6	0.9137	6	0.3340	0.6341	10	0.8302	6271.068848	8.3023	6435.738770
7	0.9338	8	0.2570	0.6760	10	0.8697	6362.094727	8.6970	6487.827637
8	0.9639	25	0.0889	0.7460	10	0.9278	6592.485840	9.2778	6624.443848
9	0.9900	25	0.0984	0.7613	10	1.0000	6965.612305	10.0000	6799.995605
10	1.0000	25	0.1035	0.7654	−	−	6954.753906	−	6954.753906

最適なモデルの策定例として、r1＝5，a1＝0.4，ρ＝0.6の相関二項分布に、r2＝10，a2＝0.8の無相関二項分布をα＝0.9で混合したデータ数1000個の離散分布に対し、MODEL1，MODEL2でモデル化した結果を表１に示す。表中のNoとは、分割の境界、つまり（２）式におけるｋに対応する。AICを比較すると、MODEL1のNo.5が最小で、理論分布のパラメータも正確に推定されている。２番目に小さいMODEL1のNo.6とのAICの差も、約25と大きいことから、このモデルの採択が合理的なことがわかる。ポアソン分布との混合であるMODEL2のなかに限って比較すると、やはりNo.5が最適になるから、主体になる分布（今の場合相関二項分布）の設定が適切であれば、無相関の混合する分布がそれほど適切でなくても、分布の境界はほぼ正確に推定できるように思える。

つぎに、データ数500個の場合で、混合率、および量の多い範囲の分布パラメータが、パラメータ推定精度に与える影響を比較したのが表２である。表中で、×は両方の分布でパラメータ推定に失敗したことを、△は量の少ない範囲のパラメータだけ推定できたことを、○は両分布のパラメータ推定ができたことを意味する。αが一定の場合、ａ2が大きく、つまり平均が大きくなるほど推定精度は良くなっている。これは、両分布の統計的性質の相違が大きいほど推定精度が良いことを意味する。また、同一のパラメータでは、ａ2が小さい場合はαが大きいほど推定精度がよく、逆にａ2が大きい場合には推定精度が悪くなる。これは、両分布の平均が近い場合は、混合率の増加は量の少ない範囲のパラメータ推定に雑音の減少として好影響を及ぼすが、平均がある程度離れている場合は、量の多い範囲のパラメータ推定にデータ個数の減少として悪影響を及ぼすためだと思われる。

表２　パラメータ推定精度の比較(N=500)

α＼a2	0.5	0.6	0.7	0.8	0.9
0.90	×	△	△	○	○
0.95	×	△	△	○	○
0.99	△	△	△	△	○

参考文献　１）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，水理講演会講演集，1969
２）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