

2変数異方性離散分布による渇水期水量の確率モデル化
Anisotropic Bivariate Discrete Stochastic Model
for Hydrological Series in Dry Season

長尾正志*, 小西宏和**, 鈴木正人***

By Masashi Nagao, Hirokazu Konishi and Masato Suzuki

This study presents an extension of the discrete stochastic model used to justify anisotropic bivariate hydrological quantities such as precipitation or runoff discharge in dry season. The theory of discrete distribution is obtained in the form of bivariate binomial distribution and bivariate negative binomial distribution. In those two type of distribution, bivariate have different parameters and correlation. The estimation of parameters is discussed by separating marginal and joint distribution in a maximum likelihood method. The results applied by negative binomial model prove to fit well for the data of runoff discharge series.

Keywords: bivariate distribution, discrete distribution, parameter estimation

1. 研究の目的

水資源施設の機能評価や操作法の研究をする場合、入力である河川流量や降水量の確率分布を理論モデルで表現する必要がよくある。しかも、計算上の要求からそれには離散分布が使われることが多い。著者らは利水用貯水池の取水機能の評価や合理的放流に関連して、2変数分布としての流入量時系列へ等方性の離散分布の適用を提案してきた¹⁾が、最近異方性をもつ離散分布が利用可能であることが分かってきたので、それら理論モデルとその母数推定について報告し、2つの隣接流域からの流出流量時系列の間に相互相関を考慮した2変数モデルを適用した結果を示す。

さて、渇水期の流況の確率分布としての特性としては、1) 各種の正の歪に対応する柔軟性がある 2) 顕著な持続性を有する 3) 一般に量の小さい領域の確率表現が重視される ことが上げられよう。したがって、特に3)との関連において、実用的な上限を設定することが有効である場合もある。以下2変数の離散型確率分布を取り扱う。

2. 異方性2変数離散分布モデル

著者らは、上限が有限で共通し、流況に応じて分布形状が変化させうる離散型モデルとして異方性の二項分布モデル²⁾や上限はなく共通母数は有するが、形状の変化に対応可能な異方性の負の二項分布モデル³⁾の提案を行ってきた。この場合、二項分布では平均>分散の関係が、負の二項分布では、平均<分散の関係が要求される。

以下では、理論の導出の詳細については文献を参照して頂くことにして、異方性モデルについてこれら2変数離散分布の概要を記述する。

* 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)

** 学生員 名古屋工業大学大学院 工学研究科社会開発工学専攻
(〒466 名古屋市昭和区御器所町)

*** 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)

2. 1 二項分布モデル

この分布では、隣接した2流域間の流出を問題にするときのように、各流域での共通の雨量上限や比流量上限の存在を想定する場合に有用なものである。

1) 条件付き分布

まず共通の上限 r 、形状母数の異なる2変数二項分布に従う変数 X_1 、 X_2 の確率母関数は次式で与えられる⁴⁾。

$$G_{X_1, X_2}(z_1, z_2) = [A + B_1 z_1 + B_2 z_2 + C z_1 z_2]^r, \quad A + B_1 + B_2 + C = 1 \quad (1)$$

上式を基礎として、形状母数 a_1 、 a_2 を持つ変数 X_1 、 X_2 の条件付き分布は

$$p_{ij} = P_r[X_1 = j | X_2 = i]$$

$$= a_1^j (1 - a_1)^{r-j} \cdot I_1 r^{i-j} \cdot I_3^i \cdot I_4^j \cdot \sum_{s=0}^{\min(i, j)} {}_i C_s \cdot r^{-i} C_{j-s} \left[\frac{I_1 I_2}{I_3 I_4} \right]^s \quad (2)$$

ただし

$$I_1 = 1 + \rho \sqrt{a_1 a_2} / \{ (1 - a_1) (1 - a_2) \}, \quad I_3 = 1 - \rho \sqrt{a_1 (1 - a_2)} / \{ a_2 (1 - a_1) \} \\ I_2 = 1 + \rho \sqrt{(1 - a_1) (1 - a_2)} / \{ a_1 a_2 \}, \quad I_4 = 1 - \rho \sqrt{a_2 (1 - a_1)} / \{ a_1 (1 - a_2) \} \quad (3)$$

このように、条件付き分布が上限 r 、形状母数 a_1 、 a_2 、および相関母数 ρ を用いて直接的に表記できるので、確率値の計算や各種シミュレーションなどに有効であると考えている。

2) 周辺分布と基本統計量

また周辺分布および平均、分散、歪係数は以下になる。($i = 1, 2$)

$$P_{xi}(X_i) = {}_r C_{xi} (1 - a_i)^{r - X_i} (a_i)^{X_i}, \quad E(X_i) = r a_i \\ V(X_i) = r a_i (1 - a_i), \quad C_s = (1 - 2 a_i) / [r a_i (1 - a_i)]^{1/2} \quad (4)$$

さらに、 X_1 と X_2 との間の相互相関係数 ρ は、 C を非負の値として

$$\rho = (C - a_1 a_2) / \sqrt{a_1 a_2 (1 - a_1) (1 - a_2)} \quad (5)$$

なお、 ρ には以下のように上限、下限の存在範囲があることが導ける。

$$\max[-\sqrt{a_1 a_2} / \{ (1 - a_1) (1 - a_2) \}, -\sqrt{(1 - a_1) (1 - a_2)} / \{ a_1 a_2 \}] \\ < \rho <$$

$$\min[\sqrt{a_1 (1 - a_2)} / \{ a_2 (1 - a_1) \}, \sqrt{a_2 (1 - a_1)} / \{ a_1 (1 - a_2) \}] \quad (6)$$

すなわち、範囲の上限(正值)は $a_1 = a_2$ に関して対称であり、両者の差が大きいほど小さくなる。なお $a_1 = a_2$ では上限は最大値1である。つぎに、下限(負値)は、 $a_1 a_2$ と $(1 - a_1) (1 - a_2)$ の差が大きいほど、つまり a_1 、 a_2 面において $a_1 + a_2 = 1$ の直線から離れるほど、大きくなり、 $a_1 + a_2 = 1$ では最小値-1となる。

3) 条件付き変量の平均、分散

式(2)の条件付き分布より、 $X_2 = s$ を指定した場合の条件付き変量 X_1 の平均および分散が得られる。結果のみを記述しておく。まず、条件付き平均 $\mu_1'(X_1 | X_2 = s)$ は指定変量 $X_2 = s$ に関して線形である。また、条件付き分散 $\mu_2(X_1 | X_2 = s)$ も指定変数に関して線形的な関係にあることが示される。

なお、以上の関係を用いると、条件付き平均値は、次式の通常の回帰直線の式に書き直せることが誘導される。なお、この関係は後述の負の二項分布についても成立する。したがって回帰推定には利用しやすい分布族であるといえる。

$$\mu_1'(X_1 | X_2) = \bar{X}_1 + \rho (\sigma_1 / \sigma_2) (X_2 - \bar{X}_2) \quad (\bar{X}, \sigma: \text{平均、標準偏差}) \quad (7)$$

4) 積率解としての母数推定

$$(4) \text{ 式より、まず形状母数 } a_1, a_2 \text{ を、} a_i = 1 - V(X_i) / E(X_i) \quad (i = 1, 2)$$

として推定した後、上限 r を、次式で求めるが、

$$r = 0.5 \times \{ \{E(X_1)\} / a_1 + \{E(X_2)\} / a_2 \} \quad (8)$$

上限 r は整数であるから、求めた値に最も近い整数として求め直したものを共通上限 $\hat{r}$ とする。そこで、再度形状母数を次式で計算する。

$$\hat{a}_i = E(X_i) / \hat{r} \quad (i = 1, 2) \quad (9)$$

つぎに定数 A 、 B_i ($i = 1, 2$)、 C が次のようにして求められる。

まず (5) 式より、相関母数 ρ に標本相関係数を使うと、次式で C が分かる。

$$\hat{C} = \hat{a}_1 \hat{a}_2 + \hat{\rho} \{ \hat{a}_1 \hat{a}_2 (1 - \hat{a}_1) (1 - \hat{a}_2) \}^{1/2} \quad (10)$$

他方 B_i は、 $\hat{B}_i = \hat{a}_i - \hat{C}$ で、最後に A は $\hat{A} = 1 - \hat{B}_1 - \hat{B}_2 - \hat{C}$ で計算できる。

2. 2 負の二項分布

先述の結果で、

$$r = -k \quad (k > 0), \quad - (B_i + C) = a_i \quad (> 0) \quad (i = 1, 2) \quad (11)$$

とおけば、負の二項分布が得られる。これについて簡単に触れておく。なお、この分布では上限は無限大である。

1) 周辺分布と基本統計量

まずその周辺分布を通常の表現に改めると、次式のようになる。

$$P_{x_i}(x_i) = x_i! k^{-1} C_{x_i} \cdot (p_i)^k (q_i)^{x_i} \quad (x_i = 0, 1, 2, \dots) \quad (12)$$

ただし、 $p_i \equiv (1 + a_i)^{-1}$, $q_i \equiv a_i (1 + a_i)^{-1}$

また、各種統計量は以下のものである。

$$\text{平均 } E(X_i) = k a_i, \quad \text{分散 } V(X_i) = k a_i (1 + a_i) \quad (13)$$

$$\text{歪係数 } C_s = (1 + 2 a_i) / [k a_i (1 + a_i)]^{1/2} \quad (14)$$

$$\text{相関係数 } \rho = (a_1 a_2 - C) / \sqrt{a_1 a_2 (1 + a_1) (1 + a_2)} \quad (15)$$

2) 条件付き分布

さらに、条件付き分布は、

$$p_{i,j} = a_1^j (1 + a_1)^{-k-j} \frac{I_1^{-k-i-j} I_3^i I_4^j}{\Gamma(k+i+j-s)} \times \sum_{s=0}^{\min(i,j)} (-1)^s \frac{\Gamma(k+i)}{\Gamma(k+i+j-s)} \frac{i!}{s! (i-s)! (j-s)!} \left[\frac{I_1 I_2}{I_3 I_4} \right]^s \quad (16)$$

ただし

$$I_1 = 1 - \rho \sqrt{a_1 a_2} / \{ (1 + a_1) (1 + a_2) \}, \quad I_3 = 1 - \rho \sqrt{a_1 (1 + a_2)} / \{ a_2 (1 + a_1) \}, \\ I_2 = 1 - \rho \sqrt{(1 + a_1) (1 + a_2)} / \{ a_1 a_2 \}, \quad I_4 = 1 - \rho \sqrt{a_2 (1 + a_1)} / \{ a_1 (1 + a_2) \}$$

なお、この相関係数には以下のように上限のみの制限が存在する。

$$\rho < \min \left[\sqrt{a_1 (1 + a_2)} / \{ a_2 (1 + a_1) \}, \sqrt{a_2 (1 + a_1)} / \{ a_1 (1 + a_2) \} \right] \quad (17)$$

この上限は、 $a_1 = a_2$ に関して対称で、 a_1 、 a_2 の差が大きいほど上限の値は小さくなり、 $a_1 = a_2$ において最大 1 となる。なお、上式の [] の中の項を L_1 および L_2 とすると、 $a_1 > a_2$ のとき $L_1 > L_2$ したがって上限は L_2 、逆の場合は上限は L_1 となる。

3) 積率解としての母数推定

平均と分散より、まず形状母数 a_i を、 $\hat{a}_i = V(X_i) / E(X_i) - 1$ として推定した後、 k を次式の平均値で求める。 $\hat{k} = 0.5 \times (E(X_1) / \hat{a}_1 + E(X_2) / \hat{a}_2)$ 、さらに形状母数を次式

$a_i = E(X_i) / \hat{k}$ で計算し直す。これを標本相関係数を ρ とした (15) 式に用いると、 C が判明する。

$$\hat{C} = \hat{a}_1 \hat{a}_2 - \rho \{ \hat{a}_1 \hat{a}_2 (1 + \hat{a}_1) (1 + \hat{a}_2) \}^{1/2}$$

最後に、 B_1 および A は、次式から推定できる。

$$\hat{B}_i = -\hat{a}_i - \hat{C} \quad (i = 1, 2), \quad \hat{A} = 1 - \hat{B}_1 - \hat{B}_2 - \hat{C}$$

さらに、この負の二項分布で $k = 1$ とおけば、異方性の2変数幾何分布が得られる。

最後に、これら異方性分布において、各母数に等方性（たとえば $a_1 = a_2$ ）を仮定すれば、従来得られていた結果と完全に合致することが確認できた。

3. 最尤解による母数推定

積率解による母数推定については先述したが、一般に積率解より最尤解のほうが統計的に好ましい結果が得られることが知られている。そこで、最尤解による母数推定の手順を記述する。ところで、既述のように条件付き分布が分かっているの、それに周辺分布を掛けた同時分布から対数尤度を求め、これが最大となるように、各種母数を推定すれば結果が得られるはずである。しかし、相関母数に制限が存在することからこれを繰り返し演算で求めることは困難であった。そこで以下のように、まず周辺分布の母数を求めた後、相関母数を推定するという2段階の手順で計算することにした。これを負の二項分布の場合について記述しておく。

3.1 周辺分布の母数推定

この場合の未知母数は k, a_1, a_2 の3個であるが、まず k が異なる母数 k_1, k_2 を持つとみなし、これを求めるように考える。それには(12)式の周辺分布から n 個の標本組 ($X_{ij}; i = 1, 2, \dots, n; j = 1, 2$) に対する対数尤度 LL を求め、 $\partial LL / \partial k_j = 0, \partial LL / \partial a_j = 0$ を各 j ごとに連立させた、いわゆる尤度方程式を資料総数 n で割った形式の次式から k_j を求める⁵⁾。

$$\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^2 \frac{x_{ij}-1}{k_j + s} - \ln \left[1 + \frac{X_j}{k_j} \right] \equiv D(k_j) = 0 \quad (j = 1, 2) \quad (18)$$

この式は資料1個当たりの対数尤度の k の増加に伴う変化率と解釈できるものである。実用的には、 k_j を順次変えていき上式の関数 $D(k_j)$ の値がある微小な許容範囲 ε 内にあれば、それが解であると判別できよう。つぎに、これら k_1, k_2 の平均で共通の母数 k とする。さらに(13)式の関係から標本平均を用いて a_i を求める。

3.2 相関母数の推定

(12)式の周辺分布と(16)式の条件分布の積で同時分布 $f(i, j) = P_i \times p_{ij}$ が得られる。この同時分布は唯一の未知母数 ρ を含む。この (i, j) は同時に取った標本の組であるから、この意味を分かりやすくするために、これを n 個の標本の i 番目の組 (x_i, y_i) に対する値 $f(x_i, y_i)$ のように表記する。したがって対数尤度関数 $LL = \sum \log f(x_i, y_i) \quad (i = 1, 2, \dots, n)$ をデータに対して求め、この値を最大する母数 ρ が最尤推定値を与える。

4. モデルの適用

4.1 渇水期の設定

渇水の状態は、本来自然現象である供給と、社会現象である需要との相対関係で決まるものであるから、単純ではない。しかし自然の流況との関連でいえば、平均、変動係数、歪係数、自己相関性などの考慮が主体となるであろう。

4.2 離散化

連続量を離散化するのであるから、期間とその間の平均量が現象記述に適切でなければならない。貯水池による流水管理の場合では、この期間には貯水池の操作単位期間が対応する。現実には適当な単位量を用い

て放流量、貯水池容量を表現しなければならないので、計算の簡便さと現象表現の正確さを勘案しながら選定していくことになる。

4. 3 モデル選定の基準

さて、以上のような前段階を経て、分布母数を推定しモデル選定を行なっていくが、その際に渇水期の流量分布のモデル作成に、渇水期の全ての資料をそのまま利用するのが、妥当であろうかという問題がある。特に議論の出るのは、渇水期とはいっても、ごく稀にかなり大きな値が介入し、それが渇水期を代表していると考えられる分布特性を非常に曖昧で、不安定なものにしないかという懸念である。したがって、本研究では、この欠点を除くために、資料中の最大値から小さいほうに第2位、第3位と順に資料を除外していきながら、残された部分標本での母数推定を実施し、適合を計っていくことにした。また、周辺分布の適合性の判定には、(18)式の $D(k_j)$ の他に、 X^2 - 検定を併用した。

4. 4 実資料の当てはめ

宮崎県大淀川綾北ダム（流域面積148.3 km^2 ）、綾南ダム（101.0 km^2 ）の1966-1986年の日流量を基礎に、平均などの特性から渇水期を11月2日からの120日間、20年分とした。また単位期間を5日とし、これを比流量（ $\text{m}^3/\text{s}/100\text{km}^2$ ）の形で離散化し、対応する量をそれぞれ、 X_1 、 X_2 ととる。なお常時確保される流量分は除外している。したがって、相互相関係数を勘案した両者の流量時系列に以上の異方性離散分布を当てはめてみた。

表-1は、その手順と結果の概要を示したものである。

まず周辺分布の母数推定では、全標本より上位標本を除外していくと、平均、標準偏差はともに減少するが、特に標準偏差の減少は著しい。したがって、全標本付近で平均<分散から負の二項分布が適合するが、除外標本が増すにつれて平均>分散に移り、二項分布が適合することになる。この適合性を尤度方程式の誤差および X^2 - 検定としてみたのがその上の2行である。すなわち、除外標本が少ないうちは、 X^2 - 検定（1%、5%）は不合格であったものが、次第に合格に変わっていく。また尤度方程式誤差 $D(k)$ は次第に小さくなる。誤差の許容限界を $\epsilon_a=7\times 10^{-3}$ にとると、それに対応する係数 k_1 、 k_2 が定まる。後に平均した近似値を採用することを考えて、これらを整数のうちから選ぶことにする。ここでは $k_1=2$ 、 $k_2=5$ から共通母数には $k=4$ を採用する。

つぎに形状母数 a_1 、 a_2 はそれぞれの平均を k で割って、 $a_1=0.212$ 、 $a_2=0.229$ となる。なお、この場合、標本数は全体では480、部分標本では X_1 、 X_2 のそれぞれで460、474となる。最後に相関母数を求めるために以上の採用域の共通部分の459個について最尤解、および積率解（標本相関係数）で求めた結果は、 $\rho=0.719$ および0.743である。なお、全標本での結果は、それぞれ0.777 および0.829となり、最尤解が積率解より若干小さな値となっているが、それほど大きな差異ではない。

つぎに適合結果を経験分布と理論分布の比較として示そう。図-1は周辺分布の比較で、経験分布では若干全体から突出した部分が見受けられるが、全体的には満足できる適合結果といえよう。なお χ^2 検定では X_1 ではほぼ1%程度、 X_2 では5%の有位水準で適合性が認められる。ついで図-2は同時分布の比較で

相 関 母 数		0.719 0.743	0.777 0.829	最尤解 積率解 相互相関係数 ρ
	大	$a_1=0.212$ $a_2=0.229$	小	形状母数 a_1, a_2
	大	4	小	共通母数 k
	小	$\epsilon_a=7\times 10^{-3}$	大	尤度方程式誤差 $D(k)$
	OK	$\Leftrightarrow$	NO	X^2 - 検定
	二項分布 平均>分散 小	$\Leftrightarrow$	負の二項分布 平均<分散 大	分散
	小		大	平均
	部分		全体	採用標本
		$X_1:460, X_2:474$ (X_1, X_2):459	480	標本個数

表-1 母数推定の概念図

ある。ここでも周辺分布における場合と同様に、部分的にバラツキが見られるが、総合してほぼ妥当な適合結果であるといえよう。

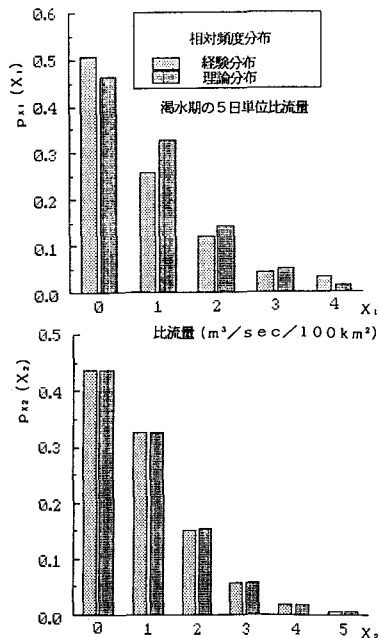


図-1 経験分布と理論分布の比較(周辺分布)

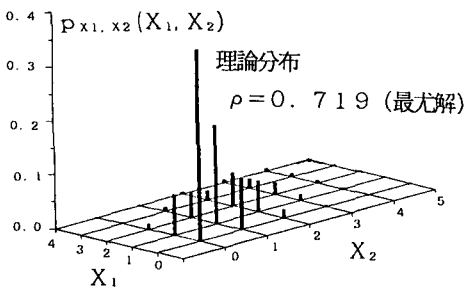
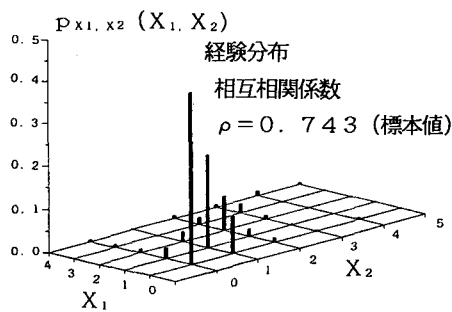


図-2 経験分布と理論分布の比較(同時分布)

従来相関を勘案した離散分布として、等方性の理論しか得られていなかったが、上記のように異方性が勘案できたので、採用される理論分布系の選択に柔軟性がでることになった。もちろん、以上の異方性分布で X_1 、 X_2 の母数を共通に取れば、等方性分布が得られる。利用分野としては、この例のように雨量や比流量などを対象とした相互相関としての利用や、時系列間の分布特性の違いを表現するには非等方性分布を、特性が共通として時系列的に扱う場合などには等方性分布を適用することが考えられる。

しかし、各種母数の一括した推定法や、合理的な判定基準の策定、2変数以上の多変数への拡張など残された課題も少なくないので、これらは今後の目標としていくつもりである。

参考文献

- 1) 長尾正志、池田吉隆：流量相関を考慮した利水用貯水池の機能評価に関する確率過程論の応用、第23回水理講演会論文集、1978.2、pp.247-255
- 2) 長尾正志、小西宏和：形状母数の異なる2変数二項分布の理論モデル、平成3年度土木学会中部支部研究発表会講演概要集、1992.3、pp.159-160
- 3) 長尾正志、小西宏和、鈴木正人：異方性2変数離散分布の理論—負の二項分布を中心として—、第47回土木学会学術講演会概要集、1992.9、第II部門、pp.668-669
- 4) C.B. Edwards and J. Gurlands: A Class of Distributions applicable to Accidents, Jour. Amer. Statist. Ass., 56, pp.503-517, 1961
- 5) Norman L. Johnson and Samuel Kotz: Discrete Distributions, Houghton Mifflin Co., 1969, pp.131-132