

京都大学防災研究所 正員 ○長尾正志

京都大学防災研究所 正員 角屋 睦

1. 概要

最近、合理的な治水計画の樹立のために、洪水波の諸要因を同時に勘案した、いわゆる多変量統計の重要性が強く認識されつつあり、二変量統計はその基礎的役割を果たすものとして重要であることはいうまでもない。ところで従来この種の研究はほとんど正規分布に立脚したものであり、現実の水文諸要因が普通非対称な分布に従うものが多いために、種々の複雑な変換を必要とする。そこで、われわれは水文量の分布として広く適応性をもつと考えられる二変数ガンマ分布の理論の開発とその適用に関する基礎的研究に着手しているが、今回はその特別な場合として、計画の基礎となる日降水量や日流出量にみられるような指数分布についての二変数理論の誘導と水文量の模擬発生への応用についての基礎的考察を行なった結果を述べる。

2. 二変数指数分布の基礎理論

まず、周辺分布が

$$f_1(x_1) = (1/\sigma_1) \exp(-x_1/\sigma_1), \quad f_2(x_2) = (1/\sigma_2) \exp(-x_2/\sigma_2) \quad (1)$$

である二変数 x_1, x_2 に関する密度分布関数は次式で与えられる。

$$f(x_1, x_2) = \frac{1}{\sigma_1 \sigma_2 (1-\rho)} \exp\left\{-\frac{x_1}{\sigma_1(1-\rho)} - \frac{x_2}{\sigma_2(1-\rho)}\right\} \times I_0\left(\frac{2\sqrt{\rho}}{1-\rho} \sqrt{\frac{x_1 x_2}{\sigma_1 \sigma_2}}\right) \quad (2)$$

上式中の σ_1, σ_2 は尺度母数といわれる定数、 ρ は相関に關係し $0 < \rho < 1$ の定数であり、 $I_0(z)$ は 0 次の変形ベッセル関数である。また、標準化変量 $\xi = x_1/\sigma_1, \eta = x_2/\sigma_2$ を用いると、 ρ を固定した場合の ξ の条件付分布 $f(\xi|\eta)$ および条件付非超過確率 $F_P(\xi|\eta)$ はそれぞれ次式で与えられる。

$$f(\xi|\eta) = \frac{1}{1-\rho} \exp\left(-\frac{\xi + \rho\eta}{1-\rho}\right) \times I_0\left(\frac{2\sqrt{\rho}}{1-\rho} \sqrt{\xi\eta}\right) \quad (3), \quad F_P(\xi|\eta) = \int_0^\xi f(\xi|\eta) d\xi \quad (4)$$

なお、この $F_P(\xi|\eta)$ は、 ξ, η および ρ の関数であり、われわれは各種の ρ, ξ および η に対する数表をすでに用意している。

3. 分布母数の推定

上述の理論を実際問題に適用するには、密度分布関数中の未知母数 σ_1, σ_2 および ρ を推定しなければならない。そこで、これを全標本が既知の場合と、上位の部分標本が既知の場合について誘導してみた。このうち前者についての詳細は参考文献に譲り結果のみを記すにしておく。

1) 全標本の場合

$$\left. \begin{aligned} \text{a) 最尤解} \quad \hat{\sigma}_1 = \bar{x}_1, \quad \hat{\sigma}_2 = \bar{x}_2, \quad \sqrt{\hat{\rho}} &= \frac{1}{N} \sum_{i=1}^N \frac{I_1\left(\frac{2\sqrt{\hat{\rho}} \xi_i \eta_i}{1-\hat{\rho}}\right)}{I_0\left(\frac{2\sqrt{\hat{\rho}} \xi_i \eta_i}{1-\hat{\rho}}\right)} \sqrt{\xi_i \eta_i} \end{aligned} \right\} \quad (5)$$

ただし $\xi_i = \frac{x_{1i}}{\hat{\sigma}_1}, \quad \eta_i = \frac{x_{2i}}{\hat{\sigma}_2} \quad (N: \text{全標本数})$

b) 積率解

$$\hat{\sigma}_1 = \bar{x}_1, \quad \hat{\sigma}_2 = \bar{x}_2; \quad \hat{\sigma}_1^2 = \frac{\overline{x_1^2}}{2}, \quad \hat{\sigma}_2^2 = \frac{\overline{x_2^2}}{2}; \quad \hat{\rho} = \frac{\overline{x_1 x_2}}{\bar{x}_1 \bar{x}_2} - 1 \quad (6)$$

この結果をみると、 σ については最尤解と積率解が完全に一致すること、 p については積率解が通常の標本相関係数と一致するが、最尤解ではそうはならないことが分る。

2) 部分標本の場合

治水計画の基礎になる日降水量や日流出量は指数分布とみなせる場合が多いが、資料としては一部のものしか得られないことがあり、とくに上位標本による推定が必要とされることがよくある。そこで上位標本による母数推定を、まず周辺分布の母数推定ついで相関母数の推定の順で考えてみた。

a) 周辺分布の母数の推定 まず分布型を(1)式をさらに一般化した、原変母数 ν および尺度母数 σ をもつ指数分布で近似することにしよう。

$$g(x) = (1/\sigma) \exp\{-(x-\nu)/\sigma\} \quad (x \geq \nu); \quad 0 \quad (x < \nu) \quad (7)$$

上式の密度分布に含まれる母数 ν 、 σ を順序統計的な手法で最尤解として推定しよう。標本総数を N とし、その中から上位 n コの標本を抽出したとする。大きい方から i 番目(小さい方から $N-i+1$ 番目)の順序統計量 x_{N-i+1} の分布は(7)式の x に x_{N-i+1} を代入すればよく、 n コの標本のもつ尤度 L は、標本 $x_{N-n+1}, x_{N-n+2}, \dots, x_{N-1}, x_N$ の同時生起確率として次式で与えられる。

$$L = \binom{N}{N-n} \left[\int_{\nu}^{x_{N-n+1}} g(x) dx \right]^{N-n} \cdot n! \prod_{j=N-n+1}^N \{g(x_j)\} = \frac{N!}{(N-n)!} \sigma^{-n} [1 - \exp\{-(x_{N-n+1}-\nu)/\sigma\}]^{N-n} \exp\left\{-\frac{1}{\sigma} \sum_{j=N-n+1}^N (x_j - \nu)\right\} \quad (8)$$

最尤解 $\hat{\sigma}$ 、 $\hat{\nu}$ は $\partial \log L / \partial \sigma = 0$ 、 $\partial \log L / \partial \nu = 0$ とおき整理することにより次式で与えられる。

$$\hat{\sigma} = \frac{1}{n} \sum_{j=N-n+1}^N x_j - x_{N-n+1}, \quad \hat{\nu} = x_{N-n+1} - \left(\log \frac{N}{n}\right) \left(\frac{1}{n} \sum_{j=N-n+1}^N x_j - x_{N-n+1}\right) \quad (9)$$

このようにして σ 、 ν が判ったから、標準化変量としては $\xi = (x - \hat{\nu}) / \hat{\sigma}$ とおけばよい。

b) 相関母数の推定 以上のようにして変量 x_1, x_2 が標準化変量 ξ, η に変換されたので、その ξ, η を使って最尤解の手法を用いて p を推定しよう。いま、標本 (ξ_i, η_i) の組数を m とすると、この標本のもつ尤度 P は次式で与えられる。

$$P = \prod_{i=1}^m g(\xi_i, \eta_i) = (1-p)^{-m} \exp\left\{-\frac{1}{1-p} \left(\sum_{i=1}^m \xi_i + \sum_{i=1}^m \eta_i\right)\right\} \times \prod_{i=1}^m I_0\left(\frac{z\sqrt{p}}{1-p} \sqrt{\xi_i \eta_i}\right) \quad (10)$$

したがって P に対する最尤解としては $\partial \log P / \partial p = 0$ より次式を満足するように p を選べばよい。

$$1 - \frac{1}{1-p} (\bar{\xi} + \bar{\eta}) + \frac{1+p}{2p} \frac{\overline{G(z)}}{G(z)} = 0 \quad (11)$$

ただし

$$\left. \begin{aligned} G(z) &= \frac{z I_1(z)}{I_0(z)}, \quad z_i = \frac{z\sqrt{p}}{1-p} \sqrt{\xi_i \eta_i} \quad (I_1(z): 1\text{次の変形ベッセル関数}) \\ \bar{\xi} &= \frac{1}{m} \sum_{i=1}^m \xi_i, \quad \bar{\eta} = \frac{1}{m} \sum_{i=1}^m \eta_i, \quad \overline{G(z)} = \frac{1}{m} \sum_{i=1}^m G(z_i) \end{aligned} \right\} \quad (12)$$

ところで、(11)、(12)式において p は陰に含まれているから、上の関係式を満足するように初期値 $p = p^{(0)}$ と仮定した場合に(12)式で $\overline{G(z)}$ を計算し、 $p > 0$ であるから(11)式より補正值

$$p^{(1)} = \frac{-(\bar{\xi} + \bar{\eta} - 1) + \sqrt{(\bar{\xi} + \bar{\eta} - 1)^2 + \overline{G(z)} (\overline{G(z)} + 2)}}{\overline{G(z)} + 2} \quad (13)$$

を算出し、最終的に補正量 $\Delta p \equiv p^{(1)} - p^{(0)}$ が0となるような p を試算で求めざるを得ないが、この計算は一般にかなりの労力を要する。そこでつぎのような実用的な解法を提示する。

まず、上記の計算に必要な関数 $G(z)$ は、図-1 のようであり、 z の極端に大きいまたは小さい値に対しては、変形ベッセル関数の性質によつて次式で近似できる。

$$z \geq 10 \text{ では } G(z) = z - 0.5 - 0.125/z \text{ (誤差約 } 1/7000 \text{ 以下)} \quad (14-1)$$

$$0 \leq z \leq 0.5 \quad G(z) = z^2/2 - z^4/16 - z^6/96 \text{ (" } 1/2000 \text{ 以下)} \quad (14-2)$$

ところで、いま対象としている治水計画では一般に z 、 η はかなり大きい。たとえば超過確率 $1 - F(z)$ が $1/20$ 、 $1/50$ 、 $1/100$ に対する z は 2.995、3.912、4.605 である。また、相関を考慮しなければならない場合には p はかなり大きく、しかも $R \equiv 2\sqrt{p}/(1-p)$ は p とともに単調に増加する。たとえば $p = 0.5$ では $R = 2.828$ である。結局二つの変量が大きくしかも両者の間にかかなりの相関がある場合には z はかなり大きく、 $G(z)$ としては (14-1) 式を使えばよく、さらに実用的には右辺の3項を省略した形、 $G(z) \approx z - 0.5$ で近似でき、そのとき $\overline{G(z)}$ は次式で簡単に求められる。

$$\overline{G(z)} = \bar{z} - 0.5 = \frac{2\sqrt{p}}{1-p} \sqrt{\xi\eta} - 0.5 \quad (15)$$

標本 (ξ_i, η_i) に関して $\bar{z}$ 、 $\bar{\eta}$ および $\sqrt{\xi\eta}$ は一義的に求められるから、(15) 図-2 相関母数の推定例

(13) 式より Δp が符号を変える p の値として p の近似値を求め、さらに厳密な解が必要ならば、その近傍で (14-1) 式によつて同様な試算を行なえばよい。この計算の1例を図-2 に示すが、(15) 式によつた近似解は、(14-1) 式によつた厳密解より一般に僅かに大きいが、実用的にはほとんど差がないといえよう。

4. 二変数指数分布に基づく模擬発生

以上によつて、規準化変量への変換形および相関母数がついたので、(4) 式の条件付非超過確率 $F_p(\xi|\eta)$ を用いれば、二変数指数分布の場合の水文資料の模擬発生は容易である。すなわち、長期間の水文資料の得られている基準地突水水量を η_0 とし、これを規準化変量 η に変換しておく。ついで区間 $(0, 1)$ の一様乱数と乱数表などを使って発生させ、それを (4) 式の $F_p(\xi|\eta)$ と考え、与えられた p および

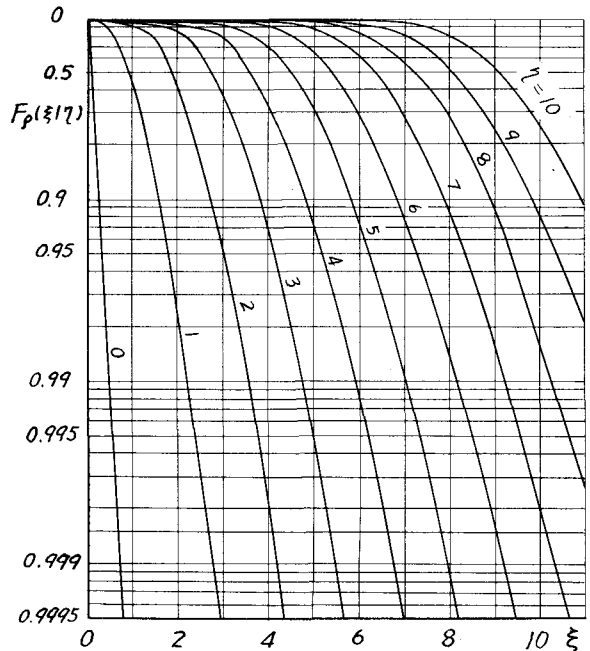
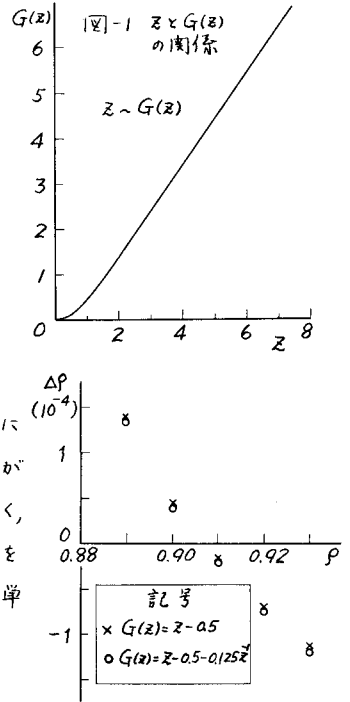


図-3 条件付非超過確率 ($p = 0.9$)

7に対する ξ を適当な内挿によって $F_p(\xi|q)$ の数表から算出する。 $F_p(\xi|q)$ の $p=0.9$ とした例を図-3に示しておく。最後に、 ξ をもとの変量 x_1 に変換し直せば、求める水文資料が発生できたことになる。

5. 指数分布による模擬発生と正規分布を仮定した模擬発生との比較

さて、このようにして指数分布を使って模擬発生した水文資料と、従来行なわれているように単純に正規分布を仮定して模擬発生したものとの統計的性格の相違を比較してみよう。それには実際に模擬発生した標本を用いた比較もできるが、ここではつぎのような理論的考察を行なっておく。

すなわち、正規分布、指数分布に従う標準化変量をそれぞれ $\xi_n = (x - m_n)/\sigma_n$, $\xi_e = (x - m_e)/\sigma_e$ (m, σ ; 平均値, 標準偏差, 添字 n ; 正規分布, e ; 指数分布) とすると、非超過確率 F はそれぞれつぎのようになる。

$$F(\xi_n) = \Phi(\xi_n) \equiv \int_{-\infty}^{\xi_n} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt, \quad F(\xi_e) = 1 - \exp\{-\xi_e + 1\} \quad (\because m_e = \sigma_e) \quad (16)$$

したがって非超過確率が等しくなるのは、

$$\Phi(\xi_n) = 1 - \exp\{-\xi_e + 1\} \quad \text{すなわち} \quad \xi_e = -\log\{1 - \Phi(\xi_n)\} + 1 \quad (17)$$

の曲線上である。この ξ_n と ξ_e の関係は図-4 のようになり超過確率 $1-F$ が減少するに従って標準指数変量 ξ_e は標準正規変量 ξ_n に比べて急激に増大していく傾向が分る。換言すれば、同じ乱数表を使った模擬発生では、超過確率の小さい場合に、正規分布より指数分布の方が一般に大きな値が出現することになる。

つぎに、指定変数 x_1 に対する x_2 の条件付分散をみると、正規分布では周知のとおり、

$$D_n^2(x_1|x_2) = \sigma_1^2(1-\rho^2)$$

であって x_2 に無関係であり、いわゆる等分散性を有するのに対し、指数分布の場合には、(3) 式より

$$D_e^2(x_1|x_2) = \sigma_1^2(1-\rho)^2 + 2 \frac{\sigma_1^2}{\sigma_2} \rho(1-\rho)x_2 \quad (18)$$

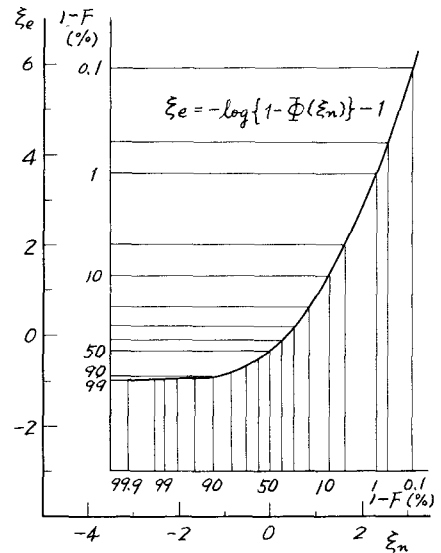


図-4 標準正規変量と標準指数変量の関係

となり x_2 に依存し、 x_2 が増加するにつれて線型的に増大する。すなわち、指定変数 x_1 を変えても、正規分布の場合には従属変数は回帰曲線の周囲に等しい程度でばらつくのに対し、指数分布の場合には x_1 が大きくなるにつれて従属変数 x_2 のばらつき方は幾らでも増していくことが分る。このように模擬発生に使用する分布関数によって模擬発生された資料の統計的性格は著しく異なるから、従来のように盲信的に正規分布としての模擬発生を乱用することは厳に慎まねばならないといえよう。

参考文献 長尾正志・角屋睦：二変数がガンマ分布とその適用に関する研究(1) — 二変数指数分布の基礎理論 — 京都大学防災研究所年報第13号B, 昭和45年3月, pp. 105 ~ 133.