II-134 二変数ガンマ分布の母数推定法

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

ガンマ分布は，その母数の選択に応じて，正規分布に近い形から指数分布のような単調減少する非対称分布に至る広い範囲の形状を表現できるので，多くの分野で応用面の広い分布と考えられているが，とくに水文統計では，降水量分布や降水期間・無降水期間の分布などの基礎的な水文量の分布によく適合することが報告されている。したがって，相関を考慮した二変数ないし多変数ガンマ分布の研究が実用上重要になるが，この方面の研究は，純粋数学的な若干のもの以外は，ほとんど見当らない。そこで，われわれは，その特別な場合である二変数指数分布についての理論の誘導と応用上の考察を，昨年度の講演会で発表したが，今回は，さらに一般的に，二変数ガンマ分布の実際上の適用を考え，その母数推定の方法を理論的に考察した結果を報告する。

2. 二変数ガンマ分布の基礎理論

一変数のガンマ分布は，もっとも一般的に，次式で定義される。

$$f(x)\,dx = \frac{1}{\Gamma(\nu)}\left(\frac{x-\upsilon}{\sigma}\right)^{\nu-1} exp\left(-\frac{x-\upsilon}{\sigma}\right) d\left(\frac{x-\upsilon}{\sigma}\right), \quad (x \geq \upsilon) \qquad (1)$$

ここに，νは形状母数，σは尺度母数，υは原点母数といわれる定数で，とくに$\nu=1$の場合には指数分布となる。なお，原点母数υについては普通0とおいても一般性を失わないので，とくに断らない限り以後$\upsilon=0$として議論しておく。

さて，周辺分布が形状母数$\nu_1 = n+m$，$\nu_2 = n$ $(n, m \geq 0)$のガンマ分布に従う二変数x_1，x_2の密度分布関数は次式で与えられる。

$$f(x_1, x_2) = \frac{1}{\Gamma(n)\Gamma(m)(\sigma_1\sigma_2)^{\frac{n+1}{2}}\sigma_1^{m}(1-\rho)\rho^{\frac{n-1}{2}}} \times (x_1x_2)^{\frac{n-1}{2}} x_1^{m} exp\left\{-\frac{x_1}{\sigma_1(1-\rho)} - \frac{x_2}{\sigma_2(1-\rho)}\right\}$$
$$\times\int_0^1 (1-t)^{\frac{n-1}{2}} t^{m-1} exp\left\{\frac{\rho x_1 t}{\sigma_1(1-\rho)}\right\} I_{n-1}\left(\frac{2\sqrt{\rho}}{1-\rho}\sqrt{\frac{x_1x_2}{\sigma_1\sigma_2}(1-t)}\right)dt \qquad (2)$$

ここに，$\Gamma(\nu)$；ガンマ関数，$I_\nu(z)$；ν次の変形ベッセル関数，σ_1，σ_2，ρ；定数でとくに$0 \leq \rho \leq 1$でなければならない。これより，x_i $(i=1, 2)$の周辺分布$f_i(x_i)$は

$$f_i(x_i) = \frac{1}{\Gamma(\nu_i)\sigma_i^{\nu_i}} x_i^{\nu_i-1} exp\left(-\frac{x_i}{\sigma_i}\right) \qquad (x_i \geq 0) \qquad (3)$$

また，x_2を与えたときのx_1の条件付分布$f(x_1|x_2)$は次式のようになる。

$$f(x_1|x_2) = \frac{1}{\Gamma(m)\sigma_1^{\frac{n+1}{2}+m}\sigma_2^{-\frac{n-1}{2}}(1-\rho)\rho^{\frac{n-1}{2}}} x_1^{\frac{n-1}{2}+m} x_2^{-\frac{n-1}{2}} exp\left\{-\frac{x_1}{\sigma_1(1-\rho)} - \frac{\rho x_2}{\sigma_2(1-\rho)}\right\}$$
$$\times\int_0^1 (1-t)^{\frac{n-1}{2}} t^{m-1} exp\left\{\frac{\rho x_1 t}{\sigma_1(1-\rho)}\right\} I_{n-1}\left(\frac{2\sqrt{\rho}}{1-\rho}\sqrt{\frac{x_1x_2}{\sigma_1\sigma_2}(1-t)}\right) dt \qquad (4)$$

さらに，以後の計算に必要な条件付平均値 $E(x_2|x_1)$ および p，q 次の積率 ν_{pq} $(p, q = 0, 1)$ を示すとつぎのようになる。

$$E(x_2|x_1) = \nu\sigma_2 + \rho\frac{\sigma_2}{\sigma_1}(x_1 - \nu\sigma_1) \tag{5}$$

$$\nu_{10} = \nu_1\sigma_1, \quad \nu_{01} = \nu_2\sigma_2; \quad \nu_{20} = \nu_1(\nu_1+1)\sigma_1^2, \quad \nu_{02} = \nu_2(\nu_2+1)\sigma_2^2; \quad \nu_{11} = \nu_2(\nu_1+\rho)\sigma_1\sigma_2 \tag{6}$$

さて，上述の分布を実際に使用するには，分布関数中の未知の母数 ν_1，ν_2；σ_1，σ_2 および ρ を推定しなければならない。そこで，これを全標本が推定に利用できる場合と一変数の存在域が上位に限定された部分標本しか利用できない場合について考察した結果を説明する。

3. 全標本による母数推定

1) 形状母数の異なる二変数ガンマ分布の場合 $(\nu_1 \geq \nu_2)$

(2) 式で表わされる母集団分布からN組の標本 (x_{1i}, x_{2i}) $(i = 1, 2, \cdots, N)$ を取出し，これから算出される 1 次，2 次の積率に関して，(6) 式より次式が成立つ。

$$\frac{1}{N}\sum_{i=1}^{N} x_{1i} \equiv \overline{x_1} = \nu_1\sigma_1, \qquad \frac{1}{N}\sum_{i=1}^{N} x_{2i} \equiv \overline{x_2} = \nu_2\sigma_2$$

$$\frac{1}{N}\sum_{i=1}^{N} x_{1i}^2 \equiv \overline{x_1^2} = \nu_1(\nu_1+1)\sigma_1^2, \quad \frac{1}{N}\sum_{i=1}^{N} x_{2i}^2 \equiv \overline{x_2^2} = \nu_2(\nu_2+1)\sigma_2^2$$

$$\frac{1}{N}\sum_{i=1}^{N} x_{1i}x_{2i} \equiv \overline{x_1x_2} = \nu_2(\nu_1+\rho)\sigma_1\sigma_2$$

したがって，一般の積率解は (7) 式，とくに ν_1，ν_2 が既知の場合の解は (7)′ 式のようになる。

$$\left.\begin{aligned} \hat{\nu}_1 &= \frac{(\overline{x}_1)^2}{\overline{x_1^2} - (\overline{x}_1)^2}, \quad \hat{\nu}_2 = \frac{(\overline{x}_2)^2}{\overline{x_2^2} - (\overline{x}_2)^2} \\ \hat{\sigma}_1 &= \frac{\overline{x_1^2} - (\overline{x}_1)^2}{\overline{x}_1}, \quad \hat{\sigma}_2 = \frac{\overline{x_2^2} - (\overline{x}_2)^2}{\overline{x}_2} \\ \hat{\rho} &= \frac{\overline{x}_1(\overline{x_1x_2} - \overline{x}_1\overline{x}_2)}{\overline{x}_2\{\overline{x_1^2} - (\overline{x}_1)^2\}} = \sqrt{\frac{\hat{\nu}_1}{\hat{\nu}_2}}\, r \end{aligned}\right\} (7) \qquad \left.\begin{aligned} \hat{\sigma}_1 &= \frac{\overline{x}_1}{\nu_1}, \quad \hat{\sigma}_2 = \frac{\overline{x}_2}{\nu_2} \\ \hat{\sigma}_1^2 &= \frac{\overline{x_1^2}}{\nu_1(\nu_1+1)}, \quad \hat{\sigma}_2^2 = \frac{\overline{x_2^2}}{\nu_2(\nu_2+1)} \\ \hat{\rho} &= \nu_1\frac{\overline{x_1x_2}}{\overline{x}_1\overline{x}_2} - \nu_1 = \sqrt{\frac{\nu_1}{\nu_2}}\, r \end{aligned}\right\} (7)'$$

ただし r は標本相関係数である。なお，この場合，現在のところ最尤解は得られていない。

2) 形状母数の等しい二変数ガンマ分布の場合 $(\nu_1 = \nu_2 = \nu)$

a. 積率解

上記の (7)，(7)′ 式に対応して，一般の場合は (8) 式，ν が既知の場合は (8)′ 式が解となる。

$$\left.\begin{aligned} \hat{\nu} &= \frac{(\overline{x}_1)^2}{\overline{x_1^2} - (\overline{x}_1)^2} = \frac{(\overline{x}_2)^2}{\overline{x_2^2} - (\overline{x}_2)^2} \\ \hat{\sigma}_1 &= \frac{\overline{x_1^2} - (\overline{x}_1)^2}{\overline{x}_1}, \quad \hat{\sigma}_2 = \frac{\overline{x_2^2} - (\overline{x}_2)^2}{\overline{x}_2} \\ \hat{\rho} &= \frac{\overline{x_1x_2} - \overline{x}_1\overline{x}_2}{\sqrt{\overline{x_1^2} - (\overline{x}_1)^2}\sqrt{\overline{x_2^2} - (\overline{x}_2)^2}} = r \end{aligned}\right\} (8) \qquad \left.\begin{aligned} \hat{\sigma}_1 &= \frac{\overline{x}_1}{\nu}, \quad \hat{\sigma}_2 = \frac{\overline{x}_2}{\nu} \\ \hat{\sigma}_1^2 &= \frac{\overline{x_1^2}}{\nu(\nu+1)}, \quad \hat{\sigma}_2^2 = \frac{\overline{x_2^2}}{\nu(\nu+1)} \\ \hat{\rho} &= \nu\frac{\overline{x_1x_2}}{\overline{x}_1\overline{x}_2} - \nu = r \end{aligned}\right\} (8)'$$

b. 最尤解

一般の場合の解は誘導が困難であり，形状母数νが既知の場合の結果のみを記すと次式が得られる。

$$\left.\begin{aligned} &\hat{\sigma}_1 = \frac{\bar{x}_1}{\nu}, \quad \hat{\sigma}_2 = \frac{\bar{x}_2}{\nu}, \qquad \nu\sqrt{\hat{\rho}} = \frac{1}{N}\sum_{i=1}^{N} \frac{I_{\nu}\left(\dfrac{2\sqrt{\hat{\rho}\,\xi_i\eta_i}}{1-\hat{\rho}}\right)}{I_{\nu-1}\left(\dfrac{2\sqrt{\hat{\rho}\,\xi_i\eta_i}}{1-\hat{\rho}}\right)}\sqrt{\xi_i\eta_i} \\ &\text{ただし}\quad \xi_i = x_{1i}/\hat{\sigma}_1, \quad \eta_i = x_{2i}/\hat{\sigma}_2 \end{aligned}\right\} \qquad (9)$$

ところで，上式で相関母数ρは陰に含まれているから試算的に求めざるを得ないので，実用上の計算には前述の積率解を使った方が便利である。

3)　二変数指数分布の場合　（$\nu_1 = \nu_2 = 1$）

二変数指数分布の場合には，前記(8)′，(9)式において$\nu = 1$とおけば，結果は簡単に求められる。

4.　部分標本による母数推定

以上は母集団から全標本が何の制限もなく抽出できた場合の推定法であったが，実際には何らかの制限がつくことがある。たとえば治水計画では高水のピーク流量の分布が問題となるが，高水観測として警戒水位あるいは指定水位などの一定の水位以上で実施されることが多い。もちろん，実用上からもなるべく少数の標本で推定できれば好都合である。

そこで，ここでは部分標本として変量x_1に関してある限界x_{1c}以上の標本が得られる場合，換言すると，制限なしに得られたNコの全標本のうち，x_1に関して上位nコの同時標本（x_1，x_2）しか分らない場合の母数推定を考える。説明の便宜上，x_1を制約付変量，x_2を制約外変量と呼ぶと，まず制約付変量の周辺分布の母数の推定をした後，制約外変量の周辺分布の母数および相関母数の推定を行なうことにする。ただし標本数N，nは十分に大きいものとし，形状母数が等しくνである二変数ガンマ分布の場合についてのみ述べる。もちろん，二変数指数分布に対する解は最終的な結果に$\nu = 1$を代入すれば容易に求められる。

a.　周辺分布の母数推定

(3)式で$i = 1$とおいた周辺分布$f_1(x_1)$に従う大きさNのx_1についての標本中より大きい方からnコの上位標本を抽出する。標本には小さい方から$x_{1(1)}$，$x_{1(2)}$，… のように添字を付けると，大きい方からi番目の順序統計量$x_{1(N-i+1)}$　（$i = 1, 2, \cdots, n$）の密度分布は，(3)式のx_1に$x_{1(N-i+1)}$を代入することによって求められる。以後簡単に制約付変量を意味する添字1を略記すると，nコの上位標本のもっと尤度Lは次式で表わすことができる。

$$L = \frac{N!}{(N-n)!}\{\Gamma(\nu)\}^{-N}\sigma^{-\nu n}\left[\gamma\left(\nu, \frac{x_{(N-n+1)}}{\sigma}\right)\right]^{N-n} \times \prod_{j=N-n+1}^{N}\{x_{(j)}\}^{\nu-1} \times \exp\left(-\frac{1}{\sigma}\sum_{j=N-n+1}^{N} x_{(j)}\right) \qquad (10)$$

ただし$\gamma(\nu, x)$は第1種不完全ガンマ関数である。

さらに，表現を簡単にするためにつぎの記号を用いることにする。

$$S = \frac{1}{x_c}\prod_{j=N-n+1}^{N}\{x_{(j)}\}^{1/n}, \quad T = \frac{1}{x_c}\cdot\frac{1}{n}\sum_{j=N-n+1}^{N} x_{(j)}, \quad x_c = x_{(N-n+1)} \qquad (11)$$

これより未知母数ν，σの推定に関する最尤方程式$\partial \log L/\partial\nu = \partial \log L/\partial\sigma = 0$を求めると次式が得られる。

$$S = \exp\left[\frac{d}{d\nu}\left\{\log\frac{\Gamma(\nu,\xi_c)}{\Gamma(\nu)}\right\} + \psi(\nu)\right]\cdot\xi_c^{-1},\quad T = \frac{\xi_c^{\nu-1}e^{-\xi_c}}{\Gamma(\nu,\xi_c)} + \frac{\nu}{\xi_c} \tag{12}$$

ここに，$\psi(\nu) = \Gamma'(\nu)/\Gamma(\nu)$，$\Gamma(\nu,\xi) = \Gamma(\nu) - \gamma(\nu,\xi)$，$\xi_c = x_c/\sigma$ (13)

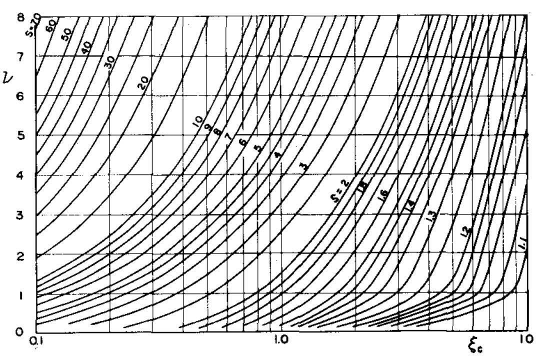
ただし，(12)式の誘導には，標本数が大きいとしての次式の近似関係を用いている。

$$\frac{n}{N} = \int_{x_c}^{\infty} f_1(x)\,dx = \int_{\xi_c}^{\infty}\frac{1}{\Gamma(\nu)}\xi^{\nu-1}e^{-\xi}\,d\xi = \frac{\Gamma(\nu,\xi_c)}{\Gamma(\nu)} \tag{14}$$

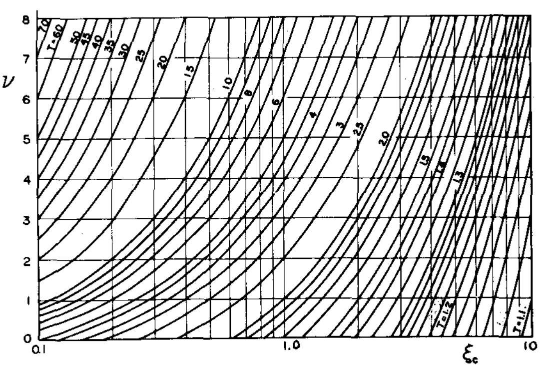
したがって，上位標本について(11)式でS，Tを求め，これを(12)式に使った連立式を解いて得たν，ξ_cの推定値$\hat{\nu}$，$\hat{\xi}_c$を求めれば，σの推定値$\hat{\sigma}$は，(13)式より次式で求められる。

$$\hat{\sigma} = x_{(N-n+1)}/\xi_c \tag{15}$$

さて，この計算には，あらかじめξ_c，ν面上にS，Tの曲線群を描いておき，二曲線$S=S$，$T=T$の交点の座標によって$\hat{\nu}$，$\hat{\xi}_c$は簡単に求められる。曲線群は以下の図－1，2のようになる。



図－1　$S = \exp\left[\frac{d}{d\nu}\left\{\log\frac{\Gamma(\nu,\xi_c)}{\Gamma(\nu)}\right\} + \psi(\nu)\right]\cdot\xi_c^{-1}$



図－2　$T = \frac{\xi_c^{\nu-1}e^{-\xi_c}}{\Gamma(\nu,\xi_c)} + \frac{\nu}{\xi_c}$

b. その他の母数の推定

以上によりσ_1，νは既知となったので，残された母数σ_2，ρを推定する。これにはx_1を与えた場合のx_2の条件付平均値が(5)式となることを用いる。すなわち，固定したx_1に対して$x_2 = E(x_2|x_1)$でx_2を推定しようとする際のx_1，x_2面内の軌跡が回帰曲線であるが，それは同時にx_1よりx_2を推定する際の最小自乗解であることが知られている。推定式を簡単に，

$$y = au + b,\quad a = \rho\sigma_2,\quad b = \nu\sigma_2,\quad u = x_1/\sigma_1 - \nu,\quad y = x_2 \tag{16}$$

と表現しておくと，上式が回帰曲線であることから定数a，bは，回帰曲線による推定の自乗誤差$I = \sum_{i=1}^{n}(y_i - au_i - b)^2$を最小にする条件$\partial I/\partial a = \partial I/\partial b = 0$よりもとめられ，したがって(16)式より，未知母数σ_2，ρは次式で推定される。ただし，たとえば$\bar{\xi}$はnコのξ_iの平均を意味する。

$$\hat{\sigma}_2 = \frac{\overline{\xi^2}\,\overline{x_2} - \bar{\xi}\,\overline{\xi x_2} + \nu(\overline{\xi x_2} - \bar{\xi}\,\overline{x_2})}{\nu\{\overline{\xi^2} - (\bar{\xi})^2\}},\quad \hat{\rho} = \frac{\nu(\overline{\xi x_2} - \bar{\xi}\,\overline{x_2})}{\overline{\xi^2}\,\overline{x_2} - \bar{\xi}\,\overline{\xi x_2} + \nu(\overline{\xi x_2} - \bar{\xi}\,\overline{x_2})},\quad \xi = x_1/\hat{\sigma}_1 \tag{17}$$

参考文献　長尾正志・角屋睦；二変数ガンマ分布とその適用に関する研究(1)，(2)，京大防災研年報，13号B，14号B，昭45，46．