

II-135 二変数指数分布の母数推定に関する2, 3の数値実験的検討

京都大学防災研究所 正員 〇長尾正志
京都大学防災研究所 正員 角屋 睦
建設省土木研究所 正員 白波瀬正道

1. 概説

昨年度の報告で二変数指数分布の母数推定さらに本年度別稿でその二変数がガンマ分布への拡張を理論的に行なったが、はたしてその推定値がどの程度信頼でき実用に供し得るかがつぎの向題となる。このためには、もちろん、推定理論の誘導過程を逐次吟味していくことによってある程度の解決はできようが、解析的に未解明な点も多いし、また、たとえば積序解と最尤解ではどのような差違がありどのような場合にどちらを使った方が良いといった向題では、それぞれの解法にそれ相応の理論的根拠があり、現在のところ解析的手法だけでは完全に解明し難い面を多分に含んでいる。さて、このような向題を究明する有力な手段として、電子計算機上で乱数発生およびその統計処理を行ない現象を再現しようとする、いわゆる統計的数値実験の手法が開発されつつある。本研究は、このような手法を用いて二変数指数分布の母数推定の理論結果を実証的に検討しようとしたものである。すなわち、原変母数、尺度母数および相関母数が既知の二変数指数分布に属する二変数指数型標本を多数模擬発生させ、各推定法により母数を推定し、その結果を比較・考察したものである。なお、原変母数はとくに0と仮定しても一般性を失わないので、 $\nu=0$ と仮定して議論しておく。

2. 全標本による母数推定値の検討

1) 標本発生法

a. 標本発生手法の手順；まず、二変数指数分布の理論に従って、母数をすべて与えた場合の確率変数の組 (XE, YE) を、電子計算機内で発生させた乱数を使って逐次発生させていく。その手順は、図-1に示すとおりであるが、概要は以下のようなものである。最初に区間 $(0, 1)$ の一般乱数を発生させ、それを指数変換することによって一変数の指数乱数列 XE を得る。つぎに指数乱数 XE を与えたときの条件付非超過確率 $F(YE|XE)$ として別に発生させた区間 $(0, 1)$ の一般乱数を用いて、対応する指数乱数列 YE を算出する。以下に、具体的な計算手法を示そう。

b. 具体的手法； 1) 標本コ数 N 、組数 M を選定する。 2) 擬似一般乱数 XZ, Z' を発生させる。 3) 擬似一般乱数について一般性の検定(X^2 検定)と異常値の検定を行ない、合格したものを一般乱数 XZ, Z' とする。 4) $XE = -\sigma_1 \log XZ$ で、 XZ を尺度母数 σ_1 の指数変数に変換する。ただしここでは $\sigma_1 = 1$ を採用する。 5) 相関母数 ρ を仮定し、標準化指数乱数 $\xi = XE/\sigma_1$ を与えた場合に対応する標準化指数乱数 $\eta = YE/\sigma_2$ を、すでに計算済みの条件付非超過確率 $F = F_\rho(\eta|\xi)$ に一般乱数 Z' を与えて算出する。 6) $YE = \sigma_2 \eta$ で、 η を尺度母数 σ_2 の指数乱数 YE に変換する。ただしここでは $\sigma_2 = 1$ を採用している。

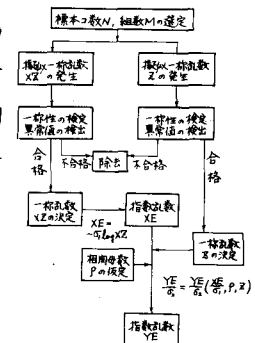


図-1 標本発生の手順

C. 設定標本；計算上の標本として、各組200コの標本系列を分割して標本数200, 100, 50, 25の標本を作製した。すなわち、4000コの擬似乱数資料のうち1000コを一様性・異常値の検定で棄却し、残った3000コを一様乱数標本として採用し、標本数 $N = 200, 100, 50, 25$ の標本をそれぞれ組数 $M = 15, 30, 60, 120$ 組用意した。

2) 周辺分布の母数推定値の変動

いま原点母数を0と仮定しているので、周辺分布に關係する母数は σ_1, σ_2 である。この推定値 $\hat{\sigma}_1 = \overline{XE}, \hat{\sigma}_2 = \overline{YE}$ の計算結果を、 $p = 0.8$ で標本数 N について示したものが図-2(a), (b)である。図の上側に実際の推定値の変動を0.05の階級に分けた頻度分布のヒストグラムで、その平均値の位置を横線で、また下側にその変動係数を示す。これを見ると、 $\hat{\sigma}_1, \hat{\sigma}_2$ の平均値は、全般に、仮定値 $\sigma_1 = \sigma_2 = 1$ より僅かに小さくなっているが、全体としてこのような推定によって尺度母数はほぼ正しく推定し得るといえる。なお、 $p = 0.2, 0.4, 0.6$ についても同じ傾向がみられ、相関母数による差異は認められない。

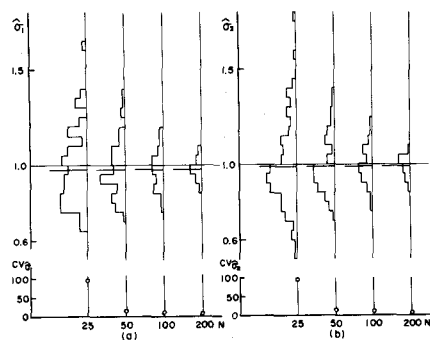


図-2 尺度母数の推定値の分布、平均値、変動係数 ($p = 0.8$)

3) 相関母数の推定値の変動

このように発生させた標本の尺度母数が、当初仮定した $\sigma_1 = \sigma_2 = 1$ より若干異なることを予想してつぎの5つの統計量を算出し、相関母数の推定精度の議論対象とした。

- $p_m(X, Y) = (\overline{XE \cdot YE} - \overline{XE} \cdot \overline{YE}) / (\overline{XE} \cdot \overline{YE})$; XE, YE がそれぞれ σ_1, σ_2 の尺度母数をもつ周辺分布に従うとした場合の積率解。
- $p_L(X, Y)$; $\xi = XE / \hat{\sigma}_1, \eta = YE / \hat{\sigma}_2$ を用いて規準化し直した場合の最尤解。
- $p_m(\xi, \eta)$; 発生された指数乱数が完全に規準化されている (すなわち $\sigma_1 = \sigma_2 = 1$ が成立) とした場合の積率解。
- $p_L(\xi, \eta)$; 発生された指数乱数が完全に規準化されているとした最尤解。
- $r = (\overline{XE \cdot YE} - \overline{XE} \cdot \overline{YE}) / (S_X \cdot S_Y)$; 標本相関係数、ただし S_X, S_Y は XE, YE の標準偏差。

以上の統計量による相関母数の推定結果を、推定値の頻度分布、平均値、変動係数を用いて、 $p = 0.8, N = 200$ について示したのが図-3である。最尤解は、他の推定値より平均的にもまた変動の少ないことから良好な推定結果を与え、尺度母数の相違の影響もほとんどないといえる。なお、発生された XE, YE が完全に指数分布に従っていれば、たとえば XE について、

$$S_X = \sqrt{\overline{XE^2} - (\overline{XE})^2} = \sqrt{2\sigma_1^2 - \sigma_1^2} = \sigma_1 = \overline{XE}$$

であるから、積率解と標本相関係数は完全に一致するはずであるが、

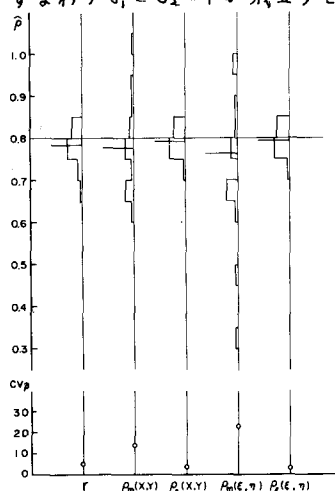


図-3 各推定法による相関母数の推定値の分布、平均値、変動係数 ($p = 0.8, N = 200$)

ここでは、両者は若干異なった結果を与えている。

なお、上記 iii), iv) の計算に必要な最尤解は次式

$$\hat{p} = \left[\frac{\frac{1}{N} \sum_{i=1}^N \frac{I_1 \left(\frac{2\sqrt{p} \xi_i \eta_i}{1-\hat{p}} \right)}{I_0 \left(\frac{2\sqrt{p} \xi_i \eta_i}{1-\hat{p}} \right)} \sqrt{\xi_i \eta_i} \right]^2 \quad (1)$$

ここに

$$\xi_i = (XE)_i / \hat{\sigma}_1, \quad \eta_i = (YE)_i / \hat{\sigma}_2 \quad (2)$$

で試算的に求めざるを得ないので以下の方法によ

った。すなわち、まず p の初期値 $p = p_1$ を仮定し、図-4 相関母数の最尤解 $p_2(\xi_i, \eta_i)$ の計算例 ($p=0.8$)

(1) 式の右辺を計算して求めた $p = p_2$ との差 $\Delta p = p_2 - p_1$ を p_1 についてプロットすると、たとえば図-4 のようになる。最尤解は $\Delta p = 0$ を満す p であるから、 Δp が符号を変える p_1 の前後の値から直線近似によって $\Delta p = 0$ となる $p = p_1$ を最尤解とした。

つぎに仮定した p に対する標本相関係数 r による推定結果を述べる。図-5 は $p = 0.8$ に対する各標本数 N についての r の頻度分布、平均値を示したものである。このように全般に標本数 N が大きくなると r の平均値 $E(r)$ は仮定した p に近づき、その変動係数も小さくなる傾向が認められるが、いまの場合、標本数 N が増すにつれて組数 M が N に反比例して減少するように標本抽出を行なっているために、 $N = 100$ の場合に $E(r)$ が p にもっとも近いという結果が出ている。

3. 部分標本による母数推定値の検討

1) 部分標本の選定

ここでは、部分標本として、 N コの全標本 (XE, YE) の中から XE がある限界 XC 以上 n コの上位標本のみを使用して母数を推定し、仮定した母数との比較を行なっている。

2) 周辺分布の母数推定値の変動

以後、 XE, YE に関する母数にはそれぞれ添字 1, 2 を付けることにする。原変母数 v_1 、尺度母数 σ_1 の上位標本 $XE_{(j)}$ ($j = N-n+1, \dots, N$) による推定値は一般に次式で与えられる。

$$\begin{aligned} \hat{v}_1 &= XE_{(N-n+1)} - \left(\sum_{j=N-n+1}^N XE_{(j)} / n - XE_{(N-n+1)} \right) \log(N/n) \\ \hat{\sigma}_1 &= \sum_{j=N-n+1}^N XE_{(j)} / n - XE_{(N-n+1)}, \quad (XE_{(j)}; XE \text{ の } j \text{ 番目順序統計量}) \end{aligned} \quad (3)$$

$p = 0.8, N = 200, n = 200, 100, 50$ に対するこの計算結果を図-6 に、推定値の密度分布、平均値、標準偏差で示す。いまの場合、 $v_1 = 0, \sigma_1 = 1$ の標本が発生されているはずであるから、標本比 n/N が小さくなるにつれて、平均値が仮定値より離れ、かつ標準偏差も大きくなるが、この例ではほぼ $n/N \geq 1/2$ であれば推定精度は良好であるといえよう。

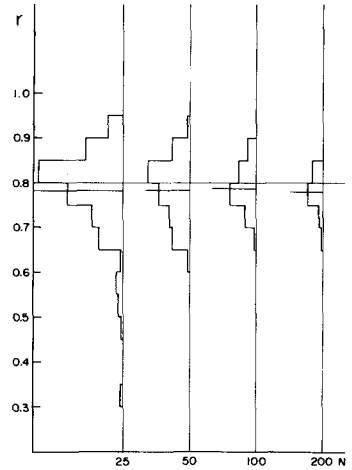
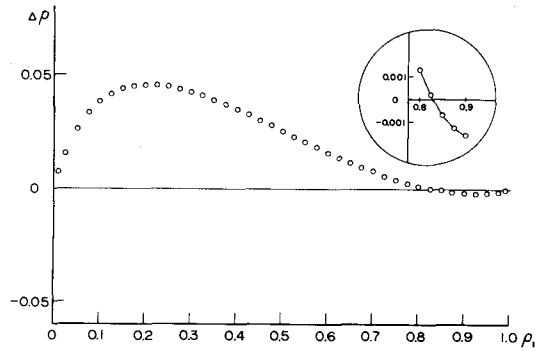


図-5 標本相関係数の頻度分布と平均値 ($p=0.8$)

3) 相関母数 推定値の変動

標本が完全に規
準化されていると
した場合の最尤解
見(ξ, η) および
上位標本に対して
通常の方法で求め
た標本相関係数r

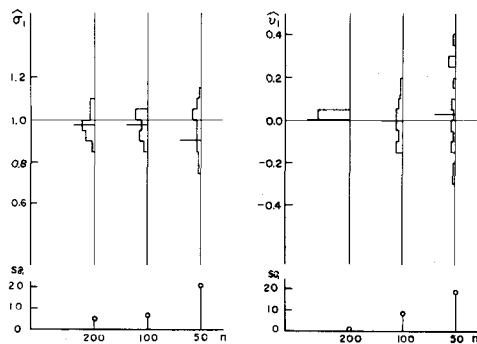


図-6 $\hat{G}_1, \hat{G}_2$ の分布, 平均値, 標準偏差
($\rho = 0.8, N = 200$)

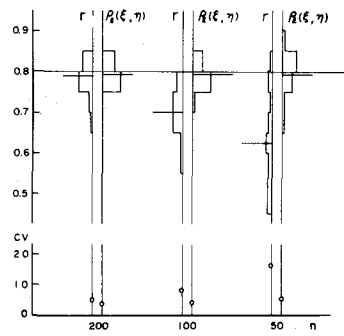


図-7 $\rho(\xi, \eta), r$ の分布, 平均値,
変動係数 ($\rho = 0.8, N = 200$)

の分布, 平均値, 変動係数を $\rho = 0.8, N = 200$ について図-7 に示す。当然 n/N が小さくなるにつ
れて変動係数は大きくなるが $\rho(\xi, \eta)$ は平均的にみてほぼ正しく ρ を推定しているといえる。これに
対して, r は n/N の減少につれて変動係数が急激に増加するのみならず, 平均的にも仮定した ρ か
ら大きく離れてくる。したがって, このような部分標本に対する相関母数の推定に標本相関係数を用
いることは不当であることが分る。

4. 部分標本を全標本とみなした回帰特性の検討

いま入手し得る標本として上位標本しかなく, 全標本数 N も不明である場合を想定する。このよう
な場合, 独立変量 X より従属変量 Y の推定には, 通常 $\hat{Y} = \bar{Y} + r(S_Y/S_X)(X - \bar{X})$ の回帰特性が利用
されるが, 二変数指数型変数の上位標本において, これが正しく推定されるかどうかを検討してみる。

a. 理論的検討; まず上位標本から回帰特性に關係する統計量を導いておく。上位標本に対する X_E ,
 Y_E の平均値をそれぞれ $\bar{X}_C, \bar{Y}_C$, 分散を $S_{X_C}^2, S_{Y_C}^2$, 共分散を CV_C と記すと, それらはそれぞれつ
ぎのように求められる。

$$\bar{X}_C = X_C + \sigma_1, \bar{Y}_C = \rho \frac{\sigma_2}{\sigma_1} X_C + \sigma_2, S_{X_C}^2 = \sigma_1^2, S_{Y_C}^2 = \sigma_2^2 + 2 \frac{\sigma_2^2}{\sigma_1^2} \rho(1-\rho) X_C, CV_C = \rho \sigma_1 \sigma_2 \quad (4)$$

したがって, 上位標本に対する相関係数 r_C , さらに回帰係数 β_C はそれぞれ次式のようにになる。

$$r_C = CV_C / (S_{X_C} S_{Y_C}) = \rho / \sqrt{2\rho(1-\rho)(X_C/\sigma_1) + 1}, \beta_C = r_C S_{Y_C} / S_{X_C} = \rho \sigma_2 / \sigma_1 \quad (5)$$

以上により, 上位標本を全標本とみなした回帰直線の式 $Y_E - \bar{Y}_C = \beta_C(X_E - \bar{X}_C)$ は, 結局全標本に対
する式 $Y_E - \sigma_2 = \rho(\sigma_2/\sigma_1)(X_E - \sigma_1)$ と全く同じ形式で表現で
示ることができる。

b. 数値実験による検討; 以上を発生標本について数値的に
検討したのが図-8 である。一方, $\sigma_1 = \sigma_2 = 1$ を仮定した理論
的な回帰曲線式は, $Y_E = 0.8X_E + 0.2$ であり, いずれも上
記の理論的結果を裏付けているものといえる。

参考文献 長尾正志・角屋睦; 二変数指数分布とその適用に
関する研究, 土木学会25回年次講演会概要集, 昭45.11,
Ⅱ-59, pp. 169-172.

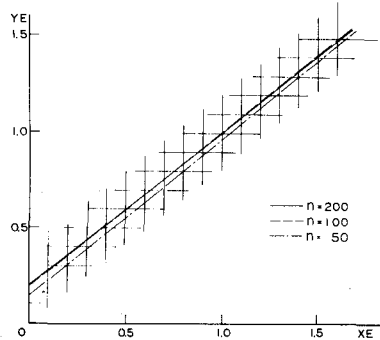


図-8 上位標本による回帰曲線の推定
($\rho = 0.8, N = 200$)