

二変数ガンマ分布とその適用に関する研究 (5)

——二変数指数分布の母数推定に関する数値実験的検討——

長尾 正志・角屋 睦・白波瀬正道

STUDY ON TWO-VARIATE GAMMA DISTRIBUTION AND ITS ENGINEERING APPLICATION (5)

——Numerical Examination on Estimating Method of Population Parameters
of Two-variate Exponential Distribution ——

By Masashi NAGAO, Mutsumi KADOYA and Masamichi SHIRAHASE

Synopsis

Few methods for estimating population parameters of two-variate gamma distribution were already proposed by the authors in the previous papers, but the judgement for their merits and demerits was suspended because of difficulty for its theoretical discussion. In this paper, the technique of the Monte Carlo simulation is applied to generate many data regarded as samples from a population of the two-variate exponential distribution which is a special type of two-variate gamma distributions, and the effectiveness of the methods for estimating population parameters of the distribution is discussed from the viewpoints of unbiasedness and efficiency. The results in outline are as follows:

- 1) For the case that a sample from a total population is available; the estimates based on the maximum likelihood method give the most desirable results, but the ones on the moments method are not except the case using the correlation coefficient.
- 2) For the case that only a sample from a sub-population in the one-dimensional sense is available, in which the domain of independent variable is defined as upper tail over a given value; the estimates based on the maximum likelihood method give the good results, but the ones on the moments method are not.
- 3) For the case that only a censored sample in the two-dimensional sense is available, in which the domains of both variables are defined as upper tails over given values; no good means are found for estimating population parameters.

1. ま え が き

われわれはここ2, 3年来多変数非対称分布の土工計画への適用を目的として, 二変数ガンマ分布やその特別な場合に相当する二変数指数分布について基礎的・応用的研究を行なってきた。前報¹⁾において二変数ガンマ分布に含まれる未知の母数を理論的に推定する方法を提案したが, その推定結果がどの程度信頼でき, 実用に供し得るかという問題が残されていた。

上記問題の解決には, もちろん, 推定手法の理論的誘導の過程を逐次吟味することによってもある程度可能ではあるが, 解析的な手段だけでは必ずしも十分とはいえないようである。たとえば, 積率解と最尤解ではどのような相違があり, どの場合にどちらを使うべきかなどという問題は, 現在のところ解析的な考察だけで明確にしがたい面を多分に含んでいる。

このように解析的な解決の困難な問題に対処する技術的手段として, 近年コンピューター・シミュレーション

ヨンの技法が注目されてきている。とくに、いまの場合、議論の対象となる標本の不確定性を勘案すれば、多数の乱数発生を基礎とした統計の数値実験、いわゆるモンテカルロ・シミュレーションが有効と考えられるので、それを利用することにする。

ところで、正規分布以外の二変数分布で、推定母数の信頼性を数値実験的に検討した研究はあまり見当たらないが、その初期段階として Campbell の二変数ポアソン分布に関する研究²⁾がある。これは、二変数ポアソン分布について所与の平均値(したがって標準偏差)、相関係数をもつと予期される標本発生の物理モデルを作製し、実験によって観測された標本値と理論値との間の相違を検定し有意な差が認められないことを示したもので、それによって理論分布の実在性を証明されたわけである。この研究では、標本発生に物理的なモデルを採用した関係上、検討は 1 例しか行っていないが、考え方は示唆的であるといえよう。

2. 数値実験の方法

2.1 方針

さて、二変数分布において、相関に関係しない母数、すなわち周辺分布に関する母数は、いずれも一変数としての取扱いが可能であるから、その信頼性などについても従来からかなり研究されてきている。³⁾そこで、ここでは二変数問題における中心課題である相関母数の推定に重点を置いて検討することにする。さらに、ここでは、二変数ガンマ分布の特別な場合である二変数指数分布の母数推定のみを対象とし、議論を簡単にするために、原点母数は 0 と仮定しておく。

数値実験の方針を概括するとつぎのとおりである。まず、相関母数以外の母数は既知とし、すなわち、確率変数としては標準化変数 (standardized variable) を使用する。さらに、母集団の相関係数を仮定して、二変量の確率変数からなる標本を多数作り出す。つぎに、発生された標本から、各推定法によって標本相関係数などの推定値を算出しもとの母相関係数などの母数と比較することによって、各推定法の適用性さらに二変数標本の選び方などの検討を行なう。

その主な内容は、全標本および部分標本による推定値の変動特性、部分標本を全標本とみなした回帰特性の検討、さらに指数変量を正規化した場合の相関特性の検討である。

2.2 標本発生の概要

以後、標本発生とは、二変数指数分布に含まれる母数を与え、分布の理論に従って二変数指数標本の組 (XE, YE) を乱数を用いて逐次作り出していく過程をいうことにする。乱数の発生には電子計算機を利用する。

その手順は Fig. 1 のとおりである。まず区間 (0, 1) の一様乱数を発生させ、逆変換法によって、それを一変数の指数乱数 XE に変換する。ついで、XE およびある相関母数 ρ を与えた場合の条件付非超過確率 $F_\rho(YE|XE)$ として別に発生させた区間 (0, 1) の一様乱数を用いて、XE に対応する指数乱数 YE を作り出す。さらに、詳細な計算手順をつぎに列挙しておく。

2.3 計算手順

1) 標本個数 N , 組数 M を選定する。2) 擬似一様乱数 XZ', Z' を発生させる。3) 擬似一様乱数について一様性の検定 (χ^2 検定) と異常値の検定を行ない、合格したものを一様乱数 XZ, Z として採用する。4) $XE = -\sigma_1 \log XZ$ により XZ を尺度母数 σ_1 の指数乱数 XE 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 $\xi = XZ/\sigma_1$ に対応する標準化指数乱数 $\eta = YE/\sigma_2$ を、条件付非超過確率 $F = F_\rho(\eta|\xi)$ に一様乱数 Z を与えて算出する。その際にすでに計算済みの $F_\rho(\eta|\xi)$ の数表⁴⁾を用いる。6) $YE = \sigma_2 \eta$ で、 η を尺度母数 σ_2 の指数乱数に変換するが、ここでは $\sigma_2 = 1$ を採用する。7) $YZ = \exp(-YE/\sigma_2)$ で指数乱数 YE より一様乱数 YZ に変換する。さらに、一様乱数の組 (XZ, YZ) から次式で標準化正規乱数の組 (XN, YN) に変換する。

$$XN = \Phi^{-1}(XZ), \quad YN = \Phi^{-1}(YZ)$$

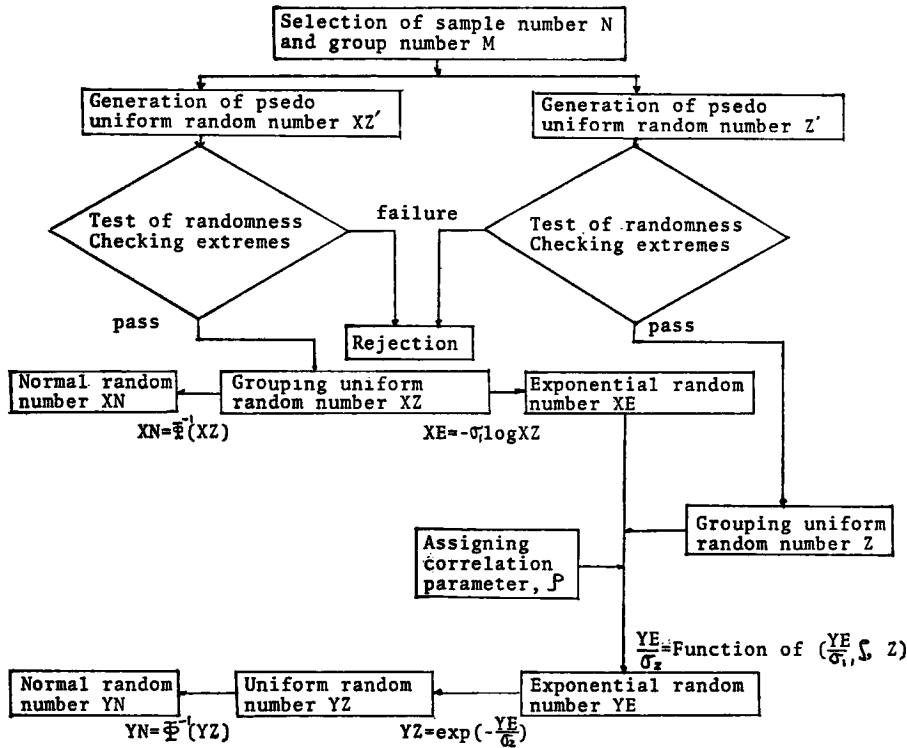


Fig. 1 Procedure for generating simulated data.

ここに、 $\Phi(x) = \int_{-\infty}^x \exp(-t^2/2) dt / \sqrt{2\pi}$ で、 Φ^{-1} は Φ の逆関数である。8) このようにして得られた二変数指数分布にしたがう標本から、各推定法によって相関母数を推定し、仮定した相関母数との比較検討を行なう。また、二変数正規乱数からも同様に相関母数を推定し、二変数指数乱数との比較、検討を行なう。

2.4 計算の細目

実際上の具体的な計算の内容はつぎのようである。

(1) 乱数発生プログラム；各 400 個の擬似一様乱数 $XZ(I)'$, $Z(I)'$ を FACOM 230-60 の基本外部関数 RANDOM を用いて発生させる。乱数発生の基本式は、次式の混合型合同法である。

$$X_i = aX_{i-1} + c \pmod{2^{35}}, \quad X_0 = b$$

定数 a , b , c としは $XZ(I)'$, $Z(I)'$ に対してそれぞれつぎのように設定する。

$XZ(I)'$	$Z(I)'$
a 257(=2 ⁸ +1)	129(=2 ⁷ +1)
b 12137516145	12137516145
c 12345678901	12345678901

(2) 乱数の一様性と異常値の検討；発生された乱数系列が、このシミュレーションの使用に適当か否かを一様性および異常値の面から検討する。検定は、乱数系列を200個ごとに20組に分けた各組について行なう。

一様性の検定には、 χ^2 検定を用い、危険率5%，自由度19に対する $\chi_0^2 = 30.14$ を棄却限界とした。また異常値の検定としては、危険率を5%とした棄却限界 $\varepsilon_0 = 1 - (1 - 0.05)^{1/200}$ を用い、0.999744 以上あるいは 0.000256 以下の乱数を異常値として棄却する。その結果 $K=1 \sim 20$ のうち、 $K=3, 4, 8, 12$ および 15

の計5組の乱数を不採用としている。

(3) 指数乱数対の模擬発生；指数乱数 ξ に対する所与の相関母数 ρ の下での η の計算には、あらかじめ二変数指数分布の条件付非超過確率の数表を計算機内に記憶させておき、与えた $\xi, F_\rho(\eta|\xi)$ の近傍の数値に対する η の値を数表から読み、それを2次元の Lagrange 補間によって内挿して求める。補間には普通4点×4点を用いる。

(4) 一様乱数の正規変換；あらかじめ標準正規分布の累積分布の表を計算機内に記憶させておき、与えられた XZ, YZ に対する XN, YN を Lagrange 補間で内挿して求める。補間には普通4点を用いる。

3. 全標本による母数推定値の検討

3.1 設定標本

以後の計算では、1組200個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個の発生標本のうち1000個を一様性・異常値の検定で棄却し、残った3000個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。

3.2 周辺分布の母数推定値の変動

まず相関母数の推定値の検討を行なうに先立って、周辺分布で仮定した母数が標本ではたして的確に推定できるかどうかをその推定値の変動に基づいて検討しておく。

原点母数を0と仮定しているの、周辺分布に関係する母数は σ_1, σ_2 である。その推定値

$$\hat{\sigma}_1 = \overline{XE}, \quad \hat{\sigma}_2 = \overline{YE}$$

を標本から計算した結果を、相関母数 $\rho=0.8$ について、各標本数 N に対して図示したのが、Fig. 2 (a), (b) である。図の上側に計算された推定値の出現割合を、0.05の階級に分けた頻度分布のヒストグラムで示す。ヒストグラムの面積が N の増加に伴って減少するのは、組数 M が N に反比例して採られているからである。同時に、平均値の位置を横線で、また下側に変動係数を示す。

その結果として、推定値は平均的にみて仮定値 $\sigma_1=\sigma_2=1$ よりわずかに小さいが、全体として1とみなしてよいといえよう。この傾向は相関母数 ρ が0.6, 0.4, 0.2と変わっても同様に認められる。

3.3 相関母数の推定値の変動

さて、相関母数の推定値の変動を各推定法について比較するが、先述のように、場合によって尺度母数が当初仮定した1という値より若干異なる場合もあることを勘案して、つぎの5種の統計量を用いることにする。

$$(1) \quad \rho_m(X, Y) = (\overline{XE \cdot YE} - \overline{XE} \cdot \overline{YE}) / (\overline{XE \cdot YE}) \dots\dots\dots (1)$$

XE, YE がそれぞれ尺度母数 σ_1, σ_2 の指数分布に従う場合の積率解である。

$$(2) \quad \rho_i(X, Y)$$

XE, YE がそれぞれ尺度母数 σ_1, σ_2 の指数分布に従う場合の最尤解、すなわち次式を満足する $\hat{\rho}$ である。

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

ただし i ; 任意の標本番号, N ; 標本数

ここに

$$\xi_i = (XE)_i / \hat{\sigma}_1, \quad \eta_i = (YE)_i / \hat{\sigma}_2 \dots\dots\dots (3)$$

$$(3) \quad \rho_m(\xi, \eta)$$

XE, YE が完全に標準化された ($\sigma_1=\sigma_2=1$ が成立つ) 指数分布に従うと仮定した場合の積率解である。

$$(4) \quad \rho_i(\xi, \eta)$$

XE, YE が完全に標準化された指数分布に従うと仮定した場合の最尤解である。

$$(5) \quad r = (\overline{XE \cdot YE} - \overline{XE} \cdot \overline{YE}) / (S_{XE} \cdot S_{YE}) \dots\dots\dots (4)$$

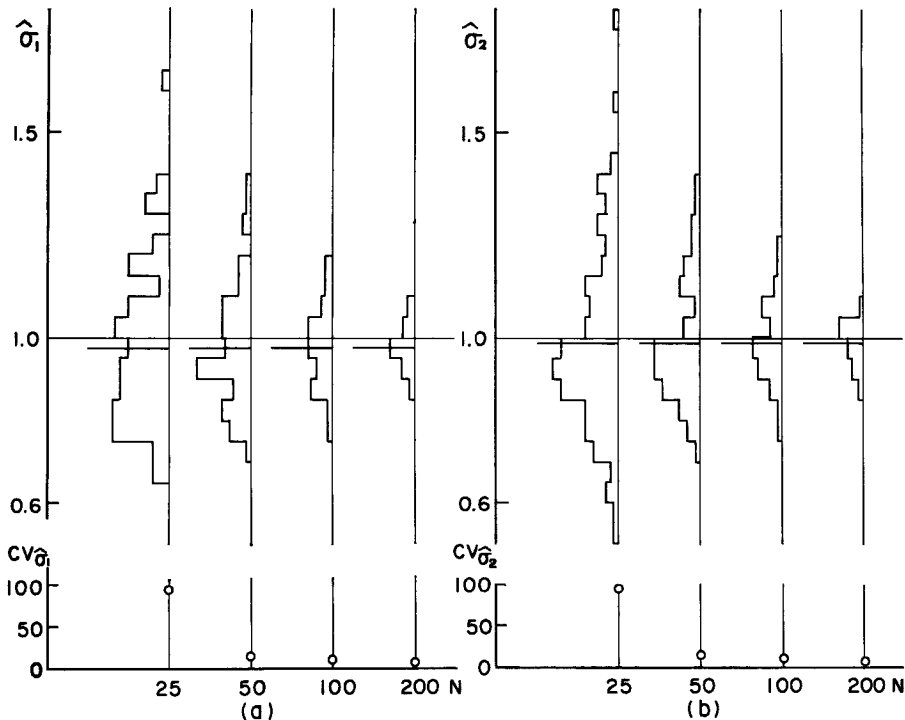


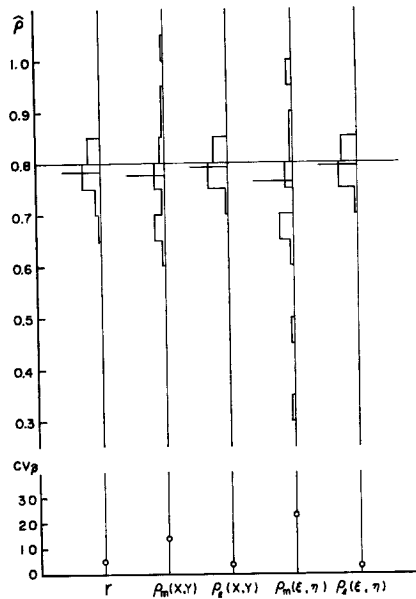
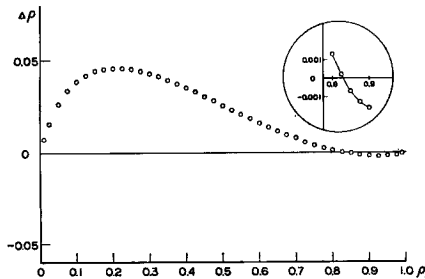
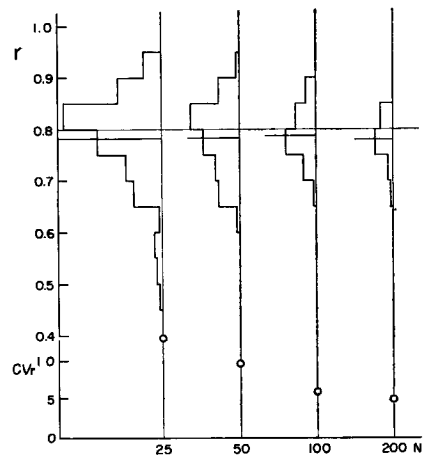
Fig. 2 Variation of estimates for scale parameters.

通常の定義に従う標本相関係数で、 S_{XE} , S_{YE} は XE , YE の標準偏差である。

以上の統計量による相関係数の推定結果を、推定値の頻度分布のヒストグラム、平均値および変動係数を用いて、 $\rho=0.8$, $N=200$ について示したのが Fig. 3 である。これによると、最尤解は、他の推定法より、平均的にもまた変動係数の小さいことから良好な結果を与える。すなわち不偏性、有効性において優れている。さらに尺度母数の推定値の相異による影響もほとんどみられない。それに比して、積率解はあまり良好な結果を与えるとはいえないようである。

なお、最尤解は試算的に求めざるを得ないが、実用的な簡便さを考えてつぎの方法によった。すなわち、まず ρ の初期値 $\rho=\rho_1$ を仮定し、(2) 式の右辺を計算して求めた $\rho=\rho_2$ との差 $\Delta\rho=\rho_2-\rho_1$ を ρ_1 を 0.025 ごとに 0 から 1 まで変えて算出しプロットした例が Fig. 4 である。最尤解は $\Delta\rho=0$ となる ρ であるから、 ρ の増加に伴って $\Delta\rho$ の符号が変わる ρ_1 の前後の値からその間の変化を直線近似によって、 $\Delta\rho=0$ に対する ρ の値として解を求めている。さらに、発生された二変数標本 XE , YE が完全に指数分布に従えば、 $S_{XE}=\overline{XE}$, $S_{YE}=\overline{YE}$ であるから、積率解と標本相関係数とは完全に一致するはずであるが、ここでは両者はかなり異なる結果となり、標本相関係数が最尤解に次いで良好な推定値を与えるのに比して積率解はかなり劣るようである。

つぎに推定結果への標本数の影響を調べる。標本相関係数による推定値の変動を、母相関係数が 0.8 について示したのが Fig. 5 である。一般に標本数の増加に伴って、平均値が母相関係数に近づくとともに変動係数も小さくなり、さらに、母相関係数が大きいほど変動係数は減少する傾向が認められるが、ここでは標本数が 100 のときに平均値が最も母数に近くなっている。その理由は、いまの場合、標本数の増加に反比例して組数が減少するように標本を選んでいるからであろう。

Fig. 3 Variation of estimates for correlation parameter. ($\rho=0.8$)Fig. 4 An example of calculation for obtaining a maximum likelihood estimate for correlation parameter [Estimate of $\rho_i(\xi, \eta)$ for $\rho=0.8$ and $K=1$].Fig. 5 Variation of sample correlation coefficients ($\rho=0.8$).

4. 部分標本による母数推定値の検討

4.1 部分標本の選び方

以上は標本の全部が利用できる場合の推定であるが、以下では部分的な標本による推定について述べる。それには前報の理論成果¹⁾を用い、全標本数 N のうちの上位 n 個の部分標本で母数を推定し、仮定した母数との比較を行なう。

さて、部分標本としてはつぎの2ケースを考えることにする。まず一つは、 XC をある定数とすると、 $XE \geq XC$ のような一変量の上位標本の場合で、他は、 $XE \geq XC$, $YE \geq YC$ のような二変量の上位標本の場合

合である。

4.2 一変量の上位標本の場合

XEがある定数XC以上の上位標本である場合に、その周辺分布に関する母数（一般に、それは原点母数 v_1 と尺度母数 σ_1 である）がどのように推定されるかを検討する。いまの場合、前提として $v_1=0$ および $\sigma_1=1$ の標本を発生しているはずであるが、さらに一般的な推定は、 $XE=x$ と略記して、次式で求められる¹⁾。

$$\left. \begin{aligned} \hat{v}_1 &= x_{(N-n+1)} - \left\{ \frac{1}{n} \sum_{j=N-n+1}^N x_{(j)} - x_{(N-n+1)} \right\} \log \frac{N}{n} \\ \hat{\sigma}_1 &= \frac{1}{n} \sum_{j=N-n+1}^N x_{(j)} - x_{(N-n+1)} \end{aligned} \right\} \dots\dots\dots (5)$$

ただし、 $x_{(j)}$ は x を小さい方から並べたときの j 番目順序統計量である。その $\rho=0.8$, $N=200$ に対する計算結果をFig. 6に、推定値の分布、平均値、標準偏差で示す。予想されるように、部分標本数 n が全標本数 N に比して小になると、平均値は母数より離れ、また標準偏差も大になるが、いまの場合、ほぼ $n/N \geq 1/2$ であれば結果は全標本の場合と大差ないようである。

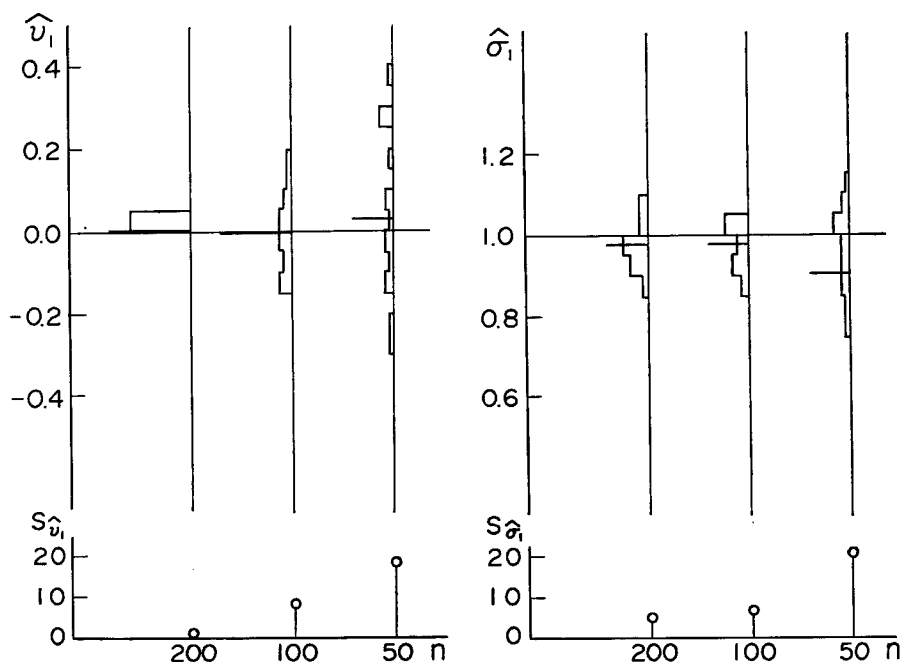


Fig. 6 Variation of estimates based on the censored samples in the one-dimensional sense ($\rho=0.8$, $N=200$).

さて、つぎの二変量の上位標本の推定値との対比を考え、相関母数の推定を最尤解と通常の標本相関係数で行なうことにする。なお最尤解 $\rho_i \equiv \rho_i(\xi, \eta)$ は次式を満足する ρ_i で、 $\hat{v}_1 = \hat{v}_2 = 1$ とおいて、求められる。

$$\left. \begin{aligned} \frac{2\rho_i}{1-\rho_i} \bar{\xi} &= \bar{G}, \quad G = G(Z_i) = Z_i I_1(Z_i) / I_0(Z_i) \\ Z_i &= \frac{2\sqrt{\rho_i(\rho_i \bar{\xi} + 1 - \rho_i)}}{(1-\rho_i)\bar{\eta}} \sqrt{\bar{\xi}_i \eta_i}; \quad \bar{\xi} = \frac{XE}{\hat{\sigma}_1}, \eta = \frac{YE}{\hat{\sigma}_2} \end{aligned} \right\} \dots\dots\dots (6)$$

Fig. 7は計算結果を示したものである。すなわち、部分標本数 n に対して、上側には、最尤解の変動を右に、標本相関係数の変動を左に、また下側にはそれぞれの変動係数を示す。これをみると、部分標本比 n/N が

小になるにつれて当然推定誤差は大になるが、最尤解は平均的にも変動係数からもほぼ良好な推定結果を与えるといえる。これに対して、標本相関係数では標本比が小になるにつれて、平均的に仮定した相関母数から推定値が離れてくる。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適当といえよう。

4.3 二変量の上位標本の場合

ここでは、 $XE \geq XC$ および $YE \geq YC$ (XC, YC ; 定数) の標本による相関母数の推定結果を述べる。既出の(6)式により、 $\hat{\sigma}_1 = \hat{\sigma}_2 = 1$ と仮定した場合の相関母数の最尤解 $\rho_i(\xi, \eta)$ を求め、これを通常の方法で求めた標本相関係数および相関母数と対比しながら、 $\rho = 0.8$ について、前と同様な表示をしたのが Fig. 8 である。

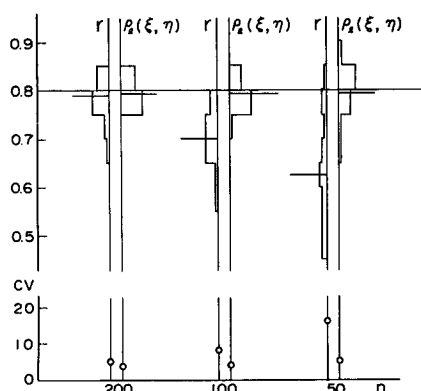


Fig. 7 Variation of estimates for correlation parameter based on the censored samples in the one-dimensional sense ($\rho = 0.8, N = 200$).

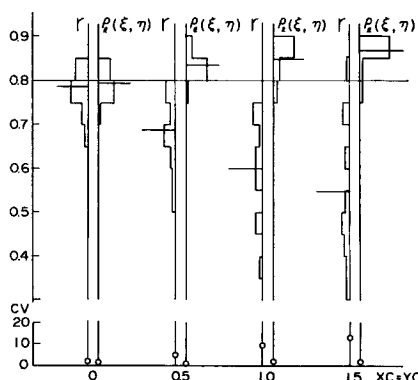


Fig. 8 Variation of estimates for correlation parameter based on the censored samples in the two-dimensional sense ($\rho = 0.8, N = 200$).

これによると、標本相関係数のみならず最尤解でも、標本比の減少につれて推定値は相関母数から一方的に離れてくる。したがって、いまのような推定方法をとる限り、このような部分標本の選び方は、推定値の不偏性や有効性からみて好ましくないと結論できよう。

5. 部分標本を全標本とみなした回帰特性の検討

利用し得る標本として上位の部分標本しかなく、しかも全標本数も不明の場合を想定する。そのような場合、独立変数 X より従属変数 Y を推定するため、通常、二変数正規標本としての回帰特性、

$$\hat{Y} = \bar{Y} + r(S_Y/S_X)(X - \bar{X})$$

が利用されるが、二変数指数分布に従う上位の部分標本に対してはこのような回帰特性の推定が可能か否かを検討してみる。

5.1 一変量の上位標本の場合

a. 理論的検討

ここでは、上位の部分標本による回帰特性を理論的に検討しておく。標本 (X, Y) は原点母数がともに 0、尺度母数が σ_1, σ_2 の二変数指数分布に従い、総数 N の標本のうちで X について XC 以上の n 個が既知とする。

さて、部分標本に対する X, Y の平均値および回帰係数をそれぞれ $\bar{X}_C, \bar{Y}_C$ および β_C と記し、これを求める。

まず平均値は、 $g(X)$ を X の周辺分布、 $f(Y|X)$ を X を与えたときの Y の条件付分布として、

$$\left. \begin{aligned} \overline{XC} &= \int_{XC}^{\infty} X \cdot \frac{N}{n} g(X) dX = XC + \sigma_1 \\ \overline{YC} &= \frac{N}{n} \int_{XC}^{\infty} g(X) \int_0^{\infty} Y f(Y|X) dY dX = \rho \frac{\sigma_2}{\sigma_1} (\overline{XC} - \sigma_1) + \sigma_2 \end{aligned} \right\} \dots\dots\dots (7)$$

となる。つぎに回帰係数は、

$$\beta_c = r_c S_{YC} / S_{XC} = CV_c / S_{XC}^2 \dots\dots\dots (8)$$

(S_{XC}^2 , S_{YC}^2 , r_c , CV_c ; いずれも部分標本に対する X , Y の分散, 相関係数, 共分散) で表わされるから, 上式右辺の諸量を求めると次式のようなになる。

$$\left. \begin{aligned} S_{XC}^2 &= \int_{XC}^{\infty} (X - \overline{XC})^2 \frac{N}{n} g(X) dX = \sigma_1^2 \\ CV_c &= \frac{N}{n} \int_{XC}^{\infty} (X - \overline{XC}) g(X) \int_0^{\infty} (Y - \overline{YC}) f(Y|X) dY dX \\ &= \rho \frac{\sigma_2}{\sigma_1} \frac{N}{n} \int_{XC}^{\infty} (X - \overline{XC})^2 g(X) dX = \rho \sigma_1 \sigma_2 \end{aligned} \right\} \dots\dots\dots (9)$$

以上の結果を用いると, 一変量の上位標本に対する回帰直線の式 $Y - \overline{YC} = \beta_c (X - \overline{XC})$ は, 結局

$$Y - \sigma_2 = \rho \frac{\sigma_2}{\sigma_1} (X - \sigma_1) \dots\dots\dots (10)$$

となり, 全標本に対するものと同じになることが示される。

b. 数値実験による検討

先述の理論の検証を数値実験によって行なう。発生された標本について $\sigma_1 = \sigma_2 = 1$ とみなすと, たとえば $\rho = 0.8$, $N = 200$ の上位標本に対する X , Y の理論的な平均値は, (7) 式によりつぎのようなになる。

$$n = 100; \quad \overline{XC} = 1 + \log(200/100) = 1.693, \quad \overline{YC} = 1 + 0.8(1.693 - 1) = 1.556$$

$$n = 50; \quad \overline{XC} = 1 + \log(200/50) = 2.386, \quad \overline{YC} = 1 + 0.8(2.386 - 1) = 2.109$$

その場合の発生標本についての計算結果が **Fig. 9** で, 図の上側に, $\overline{XC}$, $\overline{YC}$ の分布および平均値がヒストグラムおよび横実線で, 下側に変動係数が示されている。同図で理論的な平均値は横点線で示される。これより, 実験値は理論値に比べて僅かに小さいが, 全般的にほとんど差がないといえる。

つぎに, 標本から計算される回帰係数 β_c と比較のために標本相関係数 r_c を **Fig. 10** に示す。これによると, 標本相関係数は, 部分標本数 n が小に (すなわち標本下限値 XC が大に) なるにつれて次第に小さくなるが, 回帰係数は, 平均的にみて n に無関係に, 全標本に対する回帰係数 $\rho\sigma_2/\sigma_1$ を推定し得ることがわかる。

さて, 以上の結果を総合した回帰直線について述べる。 X を与えたときの Y の回帰曲線の理論式は, (10) 式より, $\sigma_1 = \sigma_2 = 1$, $\rho = 0.8$ に対して, 次式で与えられる。

$$Y = 0.8(X - 1) + 1 = 0.8X + 0.2$$

一方, 標本における $\overline{XC}$, $\overline{YC}$, β_c の各 n についての平均値を使って求めた回帰直線はつぎのようなになる。

$$n = 200; \quad Y = 0.799X + 0.204$$

$$n = 100; \quad Y = 0.806X + 0.193$$

$$n = 50; \quad Y = 0.816X + 0.149$$

これらを表示したものが **Fig. 11** で, 一変量の上位標本から全標本に対する回帰直線を推定し得るという推論の妥当性を示すものといえよう。

5.2 二変量の上位標本の場合

ここでは, $X \geq XC$, $Y \geq YC$ の標本を扱うが, この場合, 回帰特性を理論的に誘導することはむずかしい。そこで標本から直接回帰直線を求め, 一変量の上位標本の場合と比較してみる。

いま, $XC = YC$, $\rho = 0.8$, $N = 200$ の場合, 前と同様に, $\overline{XC}$, $\overline{YC}$ および β_c の n ごとの平均値を使って計算される回帰直線はつぎのようなになる。

$$XC=YC=0; Y=0.799X+0.204$$

$$XC=YC=0.5; Y=0.711X+0.478$$

$$XC=YC=1.0; Y=0.689X+0.816$$

$$XC=YC=1.5; Y=0.598X+1.100$$

これを示したものが Fig. 12 であり, $XC=YC$ の増加とともに回帰直線の勾配は減少し, Y 切片は増加し, 回帰直線の推定は困難になることがわかる。

6. 正規化変数の相関特性との比較

最後に, このような二変指数分布を直接使った統計資料と, 従来のようにそれを正規変数に変換したもの

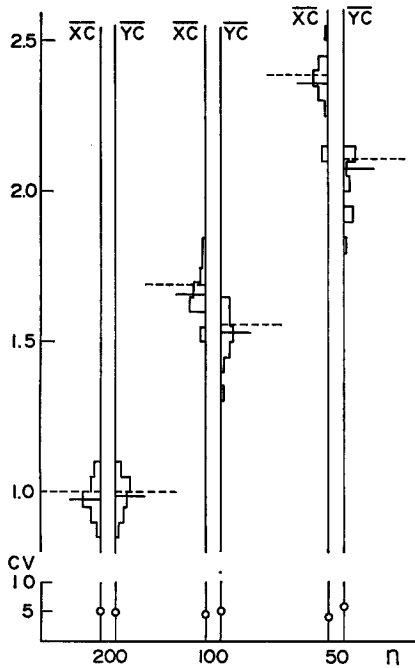


Fig. 9 Variation of $\bar{X}_c$ and $\bar{Y}_c$ estimated based on the censored samples in the one-dimensional sense ($\rho=0.8$, $N=200$).

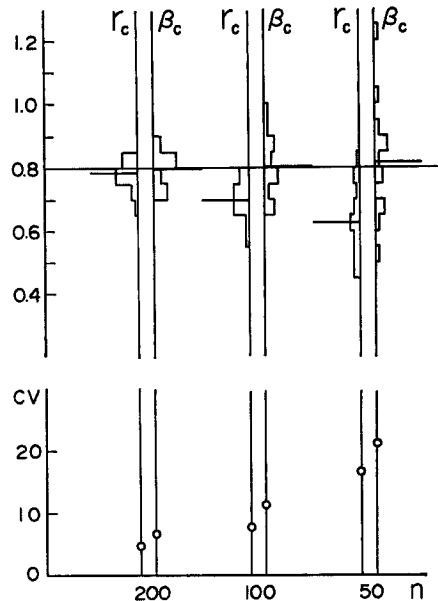


Fig. 10 Variation of β_c and r_c estimated based on the censored samples in the one-dimensional sense ($\rho=0.8$, $N=200$).

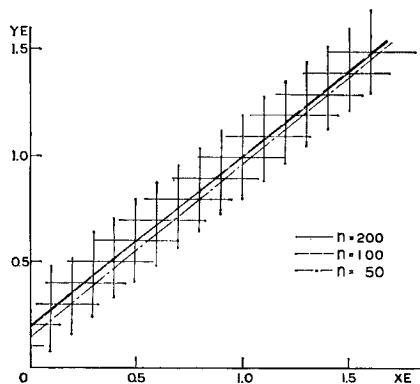


Fig. 11 Regression lines estimated based on the censored samples in the one-dimensional sense ($\rho=0.8$, $N=200$).

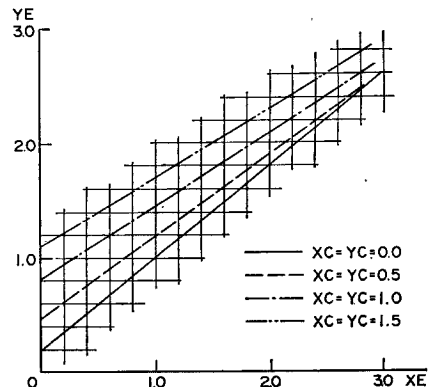


Fig. 12 Regression lines estimated based on the censored samples in the two-dimensional sense ($\rho=0.8$, $N=200$).

との相関特性の相違を知るために、既出の Fig. 1 のようにして発生された二変数指数標本と二変数正規標本の各々から相関係数を求め比較する。

6.1 理論的考察

いま、二変数指数変量 (XE, YE) から二変数正規変量 (XN, YN) へ変換する場合を考え、それぞれを次式の規準化変量 (ξ_e, η_e) および (ξ_n, η_n) で表わすことにする。

$$\left. \begin{aligned} \xi_e &= XE/\sigma_{XE}, & \eta_e &= YE/\sigma_{YE} \\ \xi_n &= (XN - m_{XN})/\sigma_{XN}, & \eta_n &= (YN - m_{YN})/\sigma_{YN} \end{aligned} \right\} \dots\dots\dots (11)$$

上式で、 m は平均値、 σ は標準偏差を意味する。

さて、(ξ_e, η_e) の同時生起を、指定変数 ξ_e の生起および ξ_e の生起下における条件付変量 ($\eta_e|\xi_e$) の生起と考えると、(ξ, η_e) より (ξ_n, η_n) への変換は、 ξ_e より ξ_n への変換および ($\eta_e|\xi_e$) より ($\eta_n|\xi_n$) への変換に分けて取扱うことができる。

ところで、変換の前後において変量の起りやすさは同一でなければならない。したがって、まず指定変数の変換において、次式の ξ_e, ξ_n の非超過確率

$$\begin{aligned} F(\xi_e) &= 1 - \exp(-\xi_e) \\ F(\xi_n) &= \int_{-\infty}^{\xi_n} (1/\sqrt{2\pi}) \exp(-u^2/2) du = \Phi(\xi_n) \end{aligned}$$

が等しくなければならない。すなわち、次式

$$\xi_n = \Phi^{-1}[1 - \exp(-\xi_e)] \quad (\Phi^{-1}; \Phi \text{ の逆関数}) \dots\dots\dots (12)$$

が成立しなければならない。この関数を Fig. 13 に示す。

さらに、条件付変数の変換においても、次式の ($\eta_e|\xi_e$), ($\eta_n|\xi_n$) の非超過確率

$$\begin{aligned} F(\eta_e|\xi_e) &= \int_0^{\eta_e} \frac{1}{1-\rho_e} \exp\left(-\frac{\eta + \rho_e \xi_e}{1-\rho_e}\right) \cdot I_0\left(\frac{2\sqrt{\rho_e \xi_e}}{1-\rho_e} \sqrt{\eta}\right) d\eta \\ &\equiv F(\rho_e, \eta_e|\xi_e) \\ F(\eta_n|\xi_n) &= \Phi((\eta_n - \rho_n \xi_n)/\sqrt{1-\rho_n^2}) \end{aligned}$$

が等しくなければならないから、

$$t_F \equiv \Phi^{-1}[F(\rho_e, \eta_e|\xi_e)] \dots\dots\dots (13)$$

と記すと、

$$\eta_n = \rho_n \xi_n + \sqrt{1-\rho_n^2} \cdot t_F \dots\dots\dots (14)$$

でなければならない。

いまの場合、 ρ_e, ξ_e, η_e は既知であるから、(12), (13) 式より ξ_n, t_F は求まるが、(14) 式において ρ_n が未知なため η_n を求めることはできない。

この意味から条件付変数の完全な変換はむずかしいと考えざるを得ないので、以後は従属変数についても指定変数と同様な変換、すなわち次式

$$\eta_n = \Phi^{-1}[F(\eta_e)] = \Phi^{-1}[1 - \exp(-\eta_e)] \dots\dots\dots (15)$$

を用いることにする。

6.2 数値実験の結果

各相関母数 ρ_e について、二変数指数変量の相関係数 RE と、(12), (15) 式で正規化した変量の相関係数 RN について平均値と変動係数を Fig. 14 に示す。ただし、標本数 200 の右側のプロットは、発生乱数の妥当性を検証するために、区間 (0, 1) の一称乱数の代りに区間 (0, 1) を等分割した超過確率の値を用いて計算したものである。いずれの場合にも正規変換後の相関係数はもとの相関係数に比べて小さくなっている。Fig. 14 は平均値としての結果であるが、個々の標本に対しても、たとえば $\rho_e = 0.8$, $N=200$ における RE, RN の比較は Fig. 15 のようであり、同様な結果が認められる。

この正規化変換に伴う相関係数の減少の原因を考察してみる。これは、本質的には、前述のように、各変

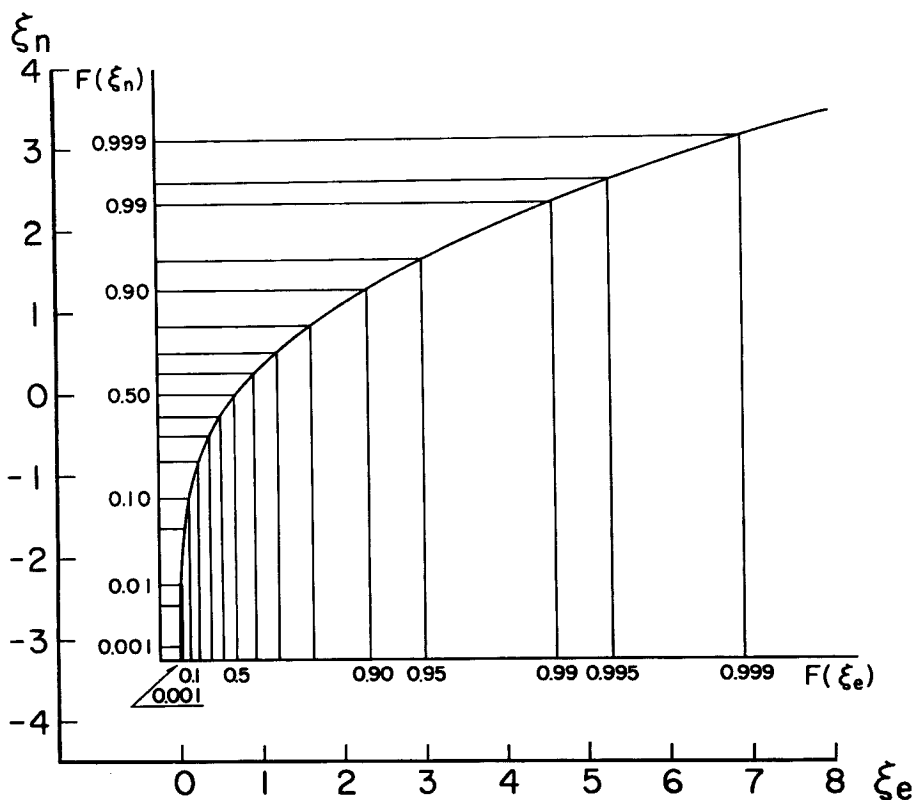


Fig. 13 Relation between standardized normal variate and standardized exponential variate.

量の周辺分布については正規化されているが、条件付変量については完全な正規化が困難であることによるものであろう。すなわち、いまのようにして発生された標本においては、通常の変数における二変数正規変量としての回帰特性、

1) 指定変数にかかわらず回帰曲線が一定の直線となり、標準化変量による表示でその勾配は相関係数に相当する。

2) 指定変数にかかわらず条件付変量が等分散性を示す。

は成立しない。これを $\rho_e = 0.8$ について具体的に述べる。

まず、Fig. 16 の実線は、変換後の回帰曲線(条件付平均値)を示したもので、破線 $\eta_n = \rho_e \xi_n$ は二変数指数標本から二変数正規標本へ完全に変換されかつ $\rho_e = \rho_n$ が成立つと仮定した場合の回帰曲線である。図によれば、変換後の回帰線は曲線となり、相関係数に相当する勾配は、 ξ_n が大になれば ρ_e より若干大きくなるが、 ξ_n が負で絶対値が増すと急激に減少する。したがって、 ξ_n の対称性を勘案すると、 ξ_n の全域における平均的な勾配として、 ρ_e は ρ_n より小さい値となることが予想される。さらに、条件付中央値の変換後の軌跡を点線で併記したが、上記の傾向は全く同じである。また、 ρ_e を変えてもこの傾向は同様に認められる。

したがって、部分的な標本抽出によって相関係数が異なって計算されるはずである。たとえば、 $\rho_e = 0.8$ 、 $N=200$ の標本に対して、指定変数 ξ_n の正の領域と負の領域に分けて部分的に相関係数を計算した結果が Table 1 であり、負の領域の相関係数が小さいことが、全領域に対する相関係数を小さくしている原因であることがわかる。

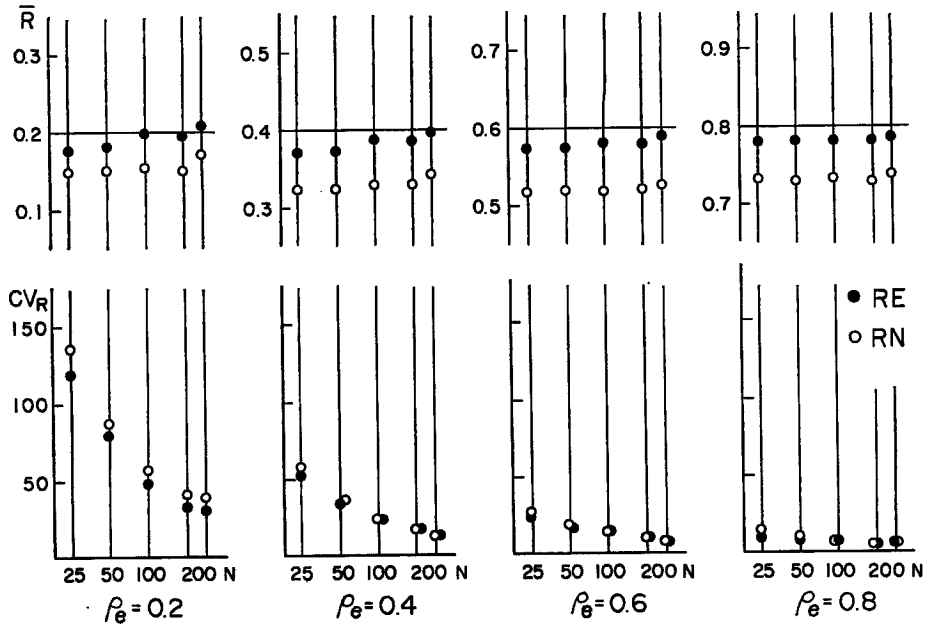


Fig. 14 Average and coefficient of variation for sample correlation coefficients, RE and RN, ($N=200$). In these figures, right-hand plots for $N=200$ show the results obtained based on samples such ordered as $P=i/(N+1)$, in which i means order of a sample.

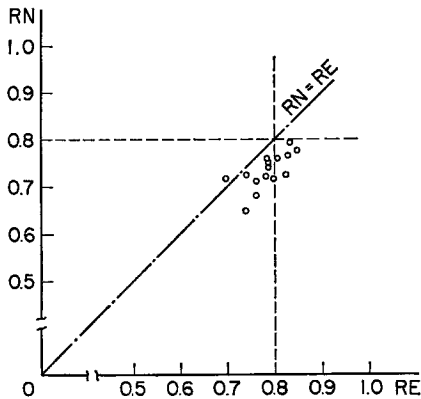


Fig. 15 Comparison of sample correlation coefficients obtained for the case of $\rho_e=0.8$ in two-variate exponential distribution ($N=200$).

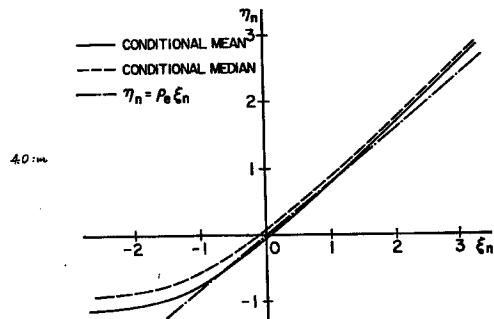


Fig. 16 Loci of the conditional mean and median obtained for the case of $\rho_e=0.8$ in the two-variate exponential distribution.

最後に、逆に二変数正規変量から二変数指数変量へ変換する場合について簡単に触れておく。この場合にも、(12), (15) 式を使って変換することになろう。

Fig. 17 は、 $\rho_n=0.8$, $N=200$ に対する RN, RE の対比を示したもので、この場合には、RN に比べて RE が小さくなっていることがわかる。その原因として、 $\rho_n=0.8$ に対する変換後の回帰線（これは同時に条件付中央値の軌跡でもある）を計算したものが Fig. 18 の実線である。なお、図において、破線 $\eta_e = \rho_n(\xi_e - 1) + 1$ は、変換が完全でかつ $\rho_e = \rho_n$ が成立つと仮定した際の回帰直線である。

Table 1 Comparison of the correlation coefficients obtained based on the normalized samples in positive and negative regions of normals.

($\rho_e = 0.8$, $N = 200$)

Data Number K	Correlation Coefficient R_N	
	$0 < \xi_n < +\infty$	$-\infty < \xi_n < 0$
1	0.6636	0.3878
2	0.6914	0.3459
5	0.6438	0.4782
6	0.6087	0.2293
7	0.6497	0.4469
9	0.5234	0.3269
10	0.5760	0.2646
11	0.5883	0.2826
13	0.6853	0.4216
14	0.6867	0.3183
16	0.5750	0.4491
17	0.6649	0.4256
18	0.5624	0.2428
19	0.6466	0.2318
20	0.5317	0.5003
Average	0.6198	0.3569

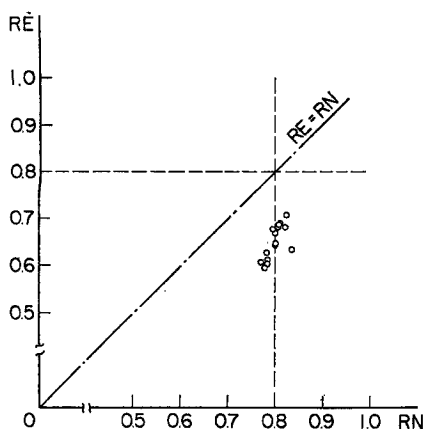


Fig. 17 Comparison of sample correlation coefficients obtained for the case of $\rho_n = 0.8$ in two-variate normal distribution ($N = 200$).

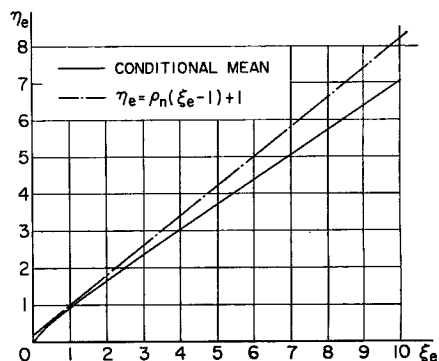


Fig. 18 Loci of the conditional mean and median obtained for the case of $\rho_n = 0.8$ in two-variate normal distribution.

この場合も、変換後の回帰線は一般に曲線となり、相関係数 ρ_e に対応する勾配は、 ξ_e が 0 のごく近傍以外では ρ_n より小さく、その相違は ξ_e が大きくなるにつれて漸増する。相関係数の値には ξ_e の大きい値が支配的に影響するから、標本全体として RE が RN より小さくなるといえよう。

以上、各変量についてそれぞれの周辺分布は非超過確率を変えないという条件で、二変数指数標本から二変数正規標本へ、あるいはその逆の変換を行なうと、変換後の相関係数は変換前に比べて小さくなると結論できよう。

7. 結 言

以上二変数指数分布に含まれる母数、とくに相関母数、の推定法の良否を、数値実験的に検討したが、得られた主要な結果を要約すると以下のようである。

1) まず、標本全部を使った推定では、相関母数の推定には最尤解によるものが、解の不偏性や有効性から、最も良好で、標本相関係数による推定がこれに次ぎ、積率解による推定はあまり良好な結果を与えない。

2) つぎに、標本の一部分を使った推定では、一変量の上位標本からの推定においては、最尤解が一般に良好な結果を与えるのに反し、標本相関係数による推定にはかなり問題がある。さらに、二変量の上位標本による推定では、推定結果からみてこのような標本の選定は不適当であるといえよう。

3) 最後に、周辺分布の非超過確率を変えないという条件の下に、二変数指数標本から二変数正規標本へ、あるいはその逆の変換を行なうと、変換後の相関係数は元に比べて小さくなることがわかった。

参 考 文 献

- 1) 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(2) ——二変数ガンマ分布の母数推定法——，京大防災研究所年報，第14号B，昭46. 4，pp. 43-56.
- 2) J. T. Campbell：The Poisson correlation function, Proceeding of Edinburgh mathematical society, series 2, vol. 4, 1934, pp. 18-46.
- 3) 垣尾忠彦：確率水量推定諸法の適応性に関する研究，京大大学院工学研究科修士論文，昭45. 3，
- 4) 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(3) ——二変数指数分布の数値表——，京大防災研究所年報，第14号B，昭46. 4，pp. 57-75.