

②④ 二つの Quantiles の比による
変異係数の推定法に就て

奈良医大 山本純恭

§ 1. 二つの Quantiles の同時分布

母集団を連続型としての分布函数を $F(x)$, 密度函数を $f(x) = F'(x)$ とする. ξ_1, ξ_2 を夫々 p_1, p_2 -quantile ($0 < p_1 < p_2 < 1$) 即ち

$$(1) \quad F(\xi_1) = p_1, \quad F(\xi_2) = p_2$$

とする.

更に $f(\xi_i) \neq 0$ ($i = 1, 2$) 且つ $x = \xi_i$ の近傍で $f(x)$ は連続で且つ連続な導函数 $f'(x)$ を持つとする.

大きい n の Order Statistics を

$$(2) \quad x(1) < x(2) < \dots < x(n),$$

標本の p_1, p_2 -quantile を夫々

$$(3) \quad Z_1 \equiv x([\!np_1\!] + 1), \quad Z_2 \equiv x([\!np_2\!] + 1)$$

とする (但し $[\]$ は Gauss 記号)

このとき次の結果が成立する⁽¹⁾

Z_1, Z_2 の同時分布は漸近的に二次元正規分布であつて, その平均は ξ_1, ξ_2 , 二次積率の漸近的な形は

$$(4) \quad \mu_{20}(Z_1) = \frac{p_1 q_1}{n f^2(\xi_1)}, \quad \mu_{11}(Z_1, Z_2) = \frac{p_1 q_2}{n f(\xi_1) f(\xi_2)}, \quad \mu_{02}(Z_2) = \frac{p_2 q_2}{n f^2(\xi_2)}$$

である. 但し $q_i = 1 - p_i$ ($i = 1, 2$)

証明は省略する。

特に母集団が平均 m , 分散 σ^2 なる正規分布であるとき

$$(5) \quad \xi_i \equiv m + \sigma u_i \quad (i = 1, 2)$$

$$(6) \quad g_i \equiv \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u_i^2} \quad (i = 1, 2)$$

とすれば

$$(7) \quad p_i = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{u_i} e^{-\frac{t^2}{2}} dt, \quad f(\xi_i) = \frac{1}{\sigma} g_i, \quad (i = 1, 2)$$

n が十分大きいとき Z_1, Z_2 の同時分布の密度函数は漸近的に

$$(8) \quad \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2(1-\rho^2)} \left\{ \left(\frac{Z_1-m-\sigma u_1}{\sigma_1}\right)^2 - 2\rho \left(\frac{Z_1-m-\sigma u_1}{\sigma_1}\right) \left(\frac{Z_2-m-\sigma u_2}{\sigma_2}\right) + \left(\frac{Z_2-m-\sigma u_2}{\sigma_2}\right)^2 \right\}\right]$$

但し

$$(9) \quad \sigma_1^2 = \frac{p_1 q_1}{n g_1^2} \sigma^2, \quad \rho \sigma_1 \sigma_2 = \frac{p_1 q_2}{n g_1 g_2} \sigma^2, \quad \sigma_2^2 = \frac{p_2 q_2}{n g_2^2} \sigma^2$$

§ 2. 比の分布とその近似

Z_1, Z_2 の同時分布の密度が (8) であるとき

$$(10) \quad \xi \equiv \frac{Z_2}{Z_1}$$

の分布は所謂 "Index" の分布であつて種々の研究がある。(2)

Exact な分布函数は mixture of distribution⁽³⁾ の形で

$$(11) \quad F(\xi) = \sum_{v=0}^{\infty} e^{-\frac{h^2}{2}} \frac{\left(\frac{h^2}{2}\right)^v}{v!} F_v(\xi)$$

但し $F_\nu(\xi)$ は分布函数で

$$F_\nu(\xi) = \frac{1}{B\left(\frac{1}{2}, \nu + \frac{1}{2}\right)} \int_{-\frac{\pi}{2} + \alpha}^{\theta + \alpha} \sin^{2\nu} \theta d\theta,$$

$$(12) \quad h^2 = \frac{1}{1-p^2} \left\{ \left(\frac{\xi_1}{\sigma_1}\right)^2 - 2\rho \frac{\xi_1 \xi_2}{\sigma_1 \sigma_2} + \left(\frac{\xi_2}{\sigma_2}\right)^2 \right\},$$

$$\theta \equiv \tan^{-1} \frac{\sigma_1 \xi - \rho \sigma_2}{\sigma_1 \sqrt{1-p^2}},$$

$$\sin \alpha \equiv \frac{\xi_2}{h \sigma_1}, \quad \cos \alpha \equiv \frac{1}{h \sqrt{1-p^2}} \left(\frac{\xi_2}{\sigma_2} - \rho \frac{\xi_1}{\sigma_1} \right),$$

又その密度函数は

$$(13) \quad \frac{1}{\pi} e^{-\frac{1}{2(1-p^2)} \left\{ \left(\frac{\xi_1}{\sigma_1}\right)^2 - 2\rho \frac{\xi_1 \xi_2}{\sigma_1 \sigma_2} + \left(\frac{\xi_2}{\sigma_2}\right)^2 \right\}} \frac{\sigma_1 \sigma_2 \sqrt{1-p^2}}{\sigma_1^2 \xi^2 - 2\rho \sigma_1 \sigma_2 \xi + \sigma_2^2} \\ \times \sum_{\nu=0}^{\infty} \frac{2^\nu \nu!}{(2\nu)!} \left\{ \frac{\sigma_1 (\xi_2 \sigma_1 - \xi_1 \rho \sigma_2) \xi + \sigma_2 (\xi_1 \sigma_2 - \xi_2 \rho \sigma_1)}{\sigma_1 \sigma_2 \sqrt{1-p^2} \sqrt{\sigma_1^2 \xi^2 - 2\rho \sigma_1 \sigma_2 \xi + \sigma_2^2}} \right\}^{2\nu}.$$

或は Fieller の形では

$$(14) \quad \frac{1}{\pi} \frac{\sigma_1 \sigma_2 \sqrt{1-p^2}}{\sigma_1^2 \xi^2 - 2\rho \sigma_1 \sigma_2 \xi + \sigma_2^2} e^{-\frac{1}{2(1-p^2)} \left\{ \left(\frac{\xi_1}{\sigma_1}\right)^2 - 2\rho \frac{\xi_1 \xi_2}{\sigma_1 \sigma_2} + \left(\frac{\xi_2}{\sigma_2}\right)^2 \right\}} \\ + \frac{\sigma_1 (\xi_2 \sigma_1 - \xi_1 \rho \sigma_2) \xi + \sigma_2 (\xi_1 \sigma_2 - \xi_2 \rho \sigma_1)}{\pi (\sigma_1^2 \xi^2 - 2\rho \sigma_1 \sigma_2 \xi + \sigma_2^2)^{3/2}} \cdot e^{-\frac{1}{2} \frac{(\xi_1 \xi - \xi_2)^2}{\sigma_1^2 \xi^2 - 2\rho \sigma_1 \sigma_2 \xi + \sigma_2^2}} \\ \times \int_0^{\frac{t}{\sigma_2}} \frac{\sigma_1 (\xi_2 \sigma_1 - \xi_1 \rho \sigma_2) \xi + \sigma_2 (\xi_1 \sigma_2 - \xi_2 \rho \sigma_1)}{\sigma_1 \sigma_2 \sqrt{1-p^2} \sqrt{\sigma_1^2 \xi^2 - 2\rho \sigma_1 \sigma_2 \xi + \sigma_2^2}} e^{-\frac{t^2}{2}} dt.$$

以上の如く "Index" の分布は Exact には扱い難いが (12) の h^2 がかなり大きいときは,

$$(15) \quad \eta = \frac{\xi_1 \xi - \xi_2}{\sqrt{\sigma_1^2 \xi^2 - 2\rho \sigma_1 \sigma_2 \xi + \sigma_2^2}}$$

とすれば η は標準正規分布に従うとしてよい近似が得られる。

§ 3. 変異係数の信頼限界

母集団が平均 m , 分散 σ^2 なる正規分布であるとき, それから取られた大きさ n の *Order Statistics* を (2), その p_i -quantile を (3), 対応する母集団の p_i -quantile を (5) とすれば, Z_1, Z_2 の同時分布の密度関数は漸近的に (8) で與えられることを知つた。こゝでは変異係数 ($V \equiv \sigma/m$) を問題にするのであるが, 実用上は平均 m が正で変異係数約 30% 以下を考えれば十分のように思われる。従つてこゝでは $\xi_1 > 0, \xi_2 > 0$ と仮定する。又このとき σ_1, σ_2 は何れも $n^{-\frac{1}{2}}$ の order であるし, 二つの quantiles が一致しない限り ($p_1 \neq p_2$) $p < 1$ であるから n が十分大きいときは $n^{\frac{1}{2}}$ は十分大きいから (15) 即ち

$$(16) \quad \eta = \frac{(\xi_1 - 1)/\sqrt{1 + u_1} \xi_1 - u_2}{\sqrt{\frac{1}{n} \left(\frac{p_1 g_1}{g_1^2} \xi_1^2 - 2 \frac{p_1 g_2}{g_1 g_2} \xi_1 + \frac{p_2 g_2}{g_2^2} \right)}}$$

の分布は標準正規分布でよく近似出来る。

η は p_1, p_2 を定めれば g_1, g_2 が定まり u_1, u_2 も定まるから parameter として $V (= \sigma/m)$ を含むだけである。従つて η を 0 ならしめる値として V の推定値

$$(17) \quad \hat{V} = \frac{\xi_1 - 1}{u_2 - u_1 \xi_1} = \frac{Z_2 - Z_1}{u_2 Z_1 - u_1 Z_2}$$

が得られる。又 *Critical region* として正規分布の 100 α % point t_α を用いて

$$(18) \quad |\eta| \geq t_\alpha$$

を選べば仮説 $V = V_0$ を検定することが出来る。次に不等式

$$(19) \quad |\eta| \leq t_\alpha$$

を解けば

$$\left\{ \frac{(\xi - 1) / \bar{V} - (u_2 - u_1 \xi)}{\sqrt{\frac{1}{n} \left(\frac{p_1 q_1}{g_1^2} \xi^2 - 2 \frac{p_1 q_2}{g_1 g_2} \xi + \frac{p_2 q_2}{g_2^2} \right)}} \right\}^2 \leq t_\alpha^2$$

即ち

$$\left\{ (u_2 - u_1 \xi)^2 - t_\alpha^2 \frac{1}{n} \left(\frac{p_1 q_1}{g_1^2} \xi^2 - 2 \frac{p_1 q_2}{g_1 g_2} \xi + \frac{p_2 q_2}{g_2^2} \right) \right\} \cdot \bar{V}^2 - 2(\xi - 1)(u_2 - u_1 \xi) \bar{V} + (\xi - 1)^2 \leq 0$$

となり

$$(20) \quad \underline{V} \leq V \leq \bar{V}$$

但し

$$\underline{V} = \frac{\xi - 1}{u_2 - u_1 \xi + \frac{t_\alpha}{\sqrt{n}} \Delta}, \quad \bar{V} = \frac{\xi - 1}{u_2 - u_1 \xi - \frac{t_\alpha}{\sqrt{n}} \Delta}$$

(21)

$$\Delta = \sqrt{\frac{p_1 q_1}{g_1^2} \xi^2 - 2 \frac{p_1 q_2}{g_1 g_2} \xi + \frac{p_2 q_2}{g_2^2}}$$

として信頼係数 $100(1 - \alpha)\%$ の変異係数の信頼区間が得られる。

§ 4. 最良配置と効率

Random interval (20) の中は

$$(22) \quad \bar{V} - \underline{V} = \frac{2 \frac{t_\alpha}{\sqrt{n}} \frac{A}{\xi - 1}}{\left(\frac{u_2 - u_1 \xi}{\xi - 1} \right)^2 - \frac{t_\alpha^2}{n} \left(\frac{A}{\xi - 1} \right)^2}$$

であるから、ここでは以下述べる意味に於ける最短区間を與える Spacing (p_1, p_2 を定めること) を求める。

母変異係数 V が V_0 のとき ξ が $\eta = 0$ を與える値 ξ_0 (η の平均で ξ の漸進的な median) は

$$(23) \quad \xi_0 = \frac{1 + u_2 V_0}{1 + u_1 V_0} \quad \text{or} \quad V_0 = \frac{\xi_0 - 1}{u_2 - u_1 \xi_0}$$

このとき (22) を最小にする spacing p_1, p_2 を求めよう。そのためには

$$(24) \quad \psi(p_1, p_2, V_0) = \left(\frac{\Delta}{\xi_0 - 1} \right)^2$$

$$\frac{\frac{p_1 g_1}{g_1^2} (1 + u_2 V_0)^2 - 2 \frac{p_1 g_2}{g_1 g_2} (1 + u_2 V_0)(1 + u_1 V_0) + \frac{p_2 g_2}{g_2^2} (1 + u_1 V_0)^2}{(u_2 - u_1)^2 V_0^2}$$

を最小にすればよい。(24) の最小値を與える p_1, p_2 は当然 V_0 に depend する。又特定の V_0 に対しても一般には解けていない。然し (24) の形の対称性から symmetric spacing 即ち $p_1 + p_2 = 1$ (又は $u_1 + u_2 = 0$) がよいように思はれる。

Spacing の対称性を假定して

$$u = u_2 = -u_1, \quad p = p_2 = 1 - p_1,$$

$$g = g_1 = g_2,$$

とおくと (24) は

$$(25) \quad \psi(p, V_0) = \frac{\frac{p g}{g^2} (1 + u^2 V_0^2) + \frac{(1-p)^2}{g^2} (1 - u^2 V_0^2)}{2 u^2 V_0^2}$$

$$= \frac{1}{2 V_0^2} \psi_1(p) + \frac{1}{2} \psi_2(p)$$

但し

$$\psi_1(p) = \frac{1}{u^2} \frac{(1-p)(2p+1)}{g^2}, \quad \psi_2(p) = \frac{1-p}{g^2}$$

数表⁽⁴⁾を用いて計算すれば $\psi_1(p)$ は $\frac{1}{2} < p < 1$ の範囲で

$$(26) \quad p = 0.930 \quad u = 1.476$$

に於て最小になる。 $\psi_1(p)$, $\psi_2(p)$ の $\frac{1}{2} < p < 1$ に於ける変化を示したのが第1図である。 V_0 を考慮に入れると $V_0 < 0.1$ に於ては $\psi_2(p)$ の影響は (26) の精度に於ては無視出来て (25) の最小値を興える spacing は (26) である。 然し V_0 が漸次大きくなり 0.2 附近では $\psi_2(p)$ の影響により (25) の最小値を興える spacing は

$$(27) \quad p = 0.922, \quad u = 1.419$$

となる。

実用上種々の Sample size に対する信頼帯を例えは第2図の如く図示しておくことは便利であるが、変異係数 10% 以下のことが多いように思われるので対称配置で (26) 即ち

$$(28) \quad \begin{array}{ll} p_1 = 0.070 & p_2 = 0.930 \\ u_1 = -1.476 & u_2 = 1.476 \end{array}$$

を採用するがよいであろう。 又このとき信頼限界は

$$(29) \quad \begin{array}{l} \bar{V} = \frac{\xi - 1}{1.476(\xi + 1) + \frac{t_{\alpha}}{\sqrt{n}} \Delta} \\ \underline{\bar{V}} = \frac{\xi - 1}{1.476(\xi + 1) - \frac{t_{\alpha}}{\sqrt{n}} \Delta} \end{array}$$

但し

$$\Delta = \sqrt{3.611\xi^2 - 0.544\xi + 3.611}$$

である。

最後に通常用いられる変異係数の推定値

$$(30) \quad v = \frac{s}{\bar{x}},$$

但し

$$\bar{x} = \frac{1}{n} \sum_1^n x_i, \quad s^2 = \frac{1}{n} \sum_1^n (x_i - \bar{x})^2,$$

に対しては $m > 0$ 且つ V が小さいとき近似的にその平均及分散を興える式として

$$(31) \quad E(v) = V + O\left(\frac{1}{n}\right), \quad D^2(v) = \frac{V^2}{2n}(1+2V) + O\left(n^{-\frac{3}{2}}\right)$$

よめる。勿論 n が十分大きいときである。この Standard error $D(v)$ を用いて信頼区間を作成したときと吾々の方法による場合とを比較するために次の計量をとるのが漸近的な効率として自然であろう。

$$(32) \quad e = \lim_{n \rightarrow \infty} \frac{z_{\alpha} \frac{V}{\sqrt{2n}} \sqrt{1+2V}}{z_{\alpha} \frac{t_{\alpha}}{\sqrt{n}} \sqrt{\psi(p,V)}} \cdot \frac{1}{\frac{1}{V^2} - \frac{t_{\alpha}^2}{n} \psi(p,V)}$$

$V \leq 0.1$ のときは $\psi_{\alpha}(p)$ は無視出来て

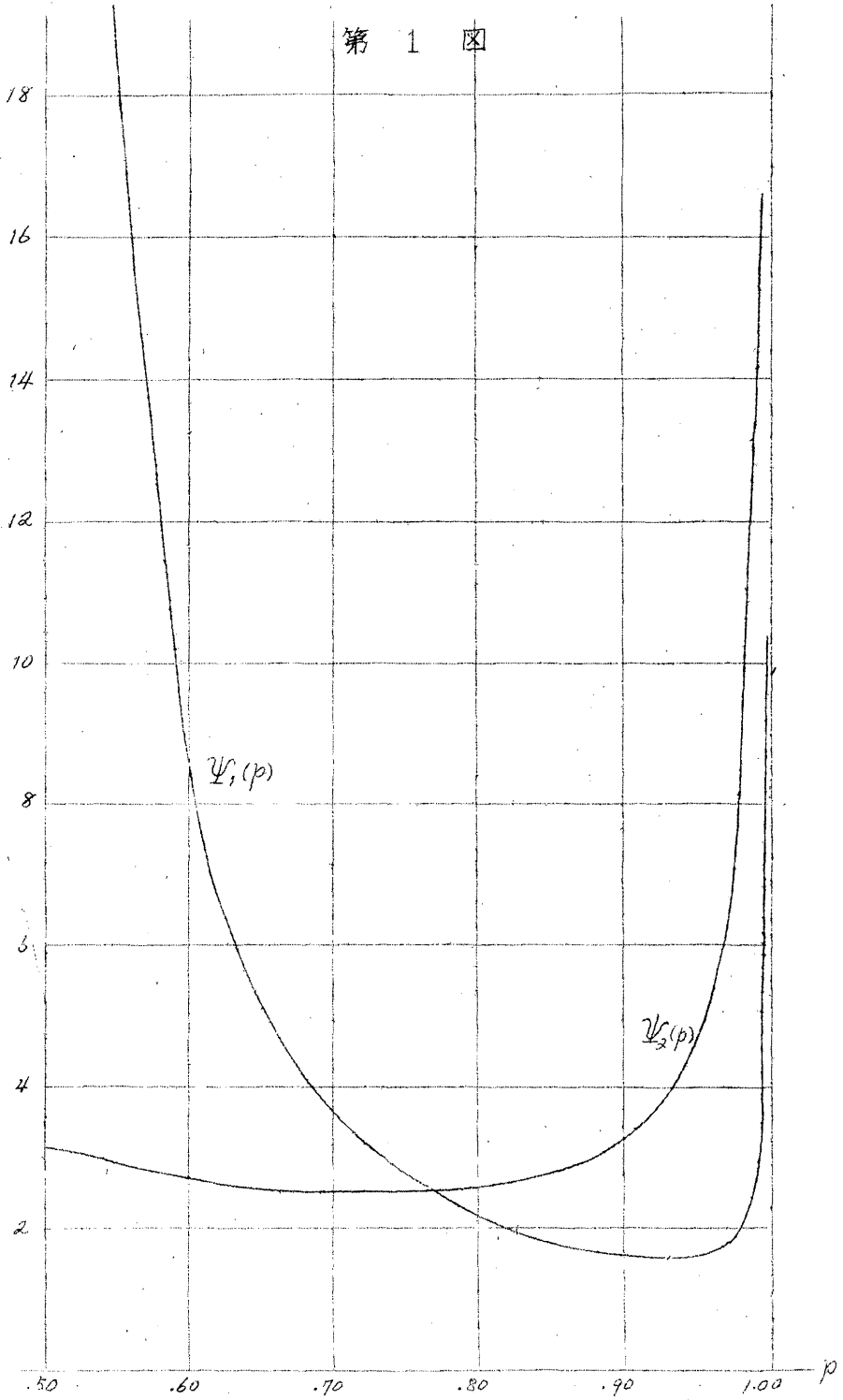
$$e = \frac{1}{\sqrt{\psi_1(p)}} \sqrt{1+2V^2} = 0.80$$

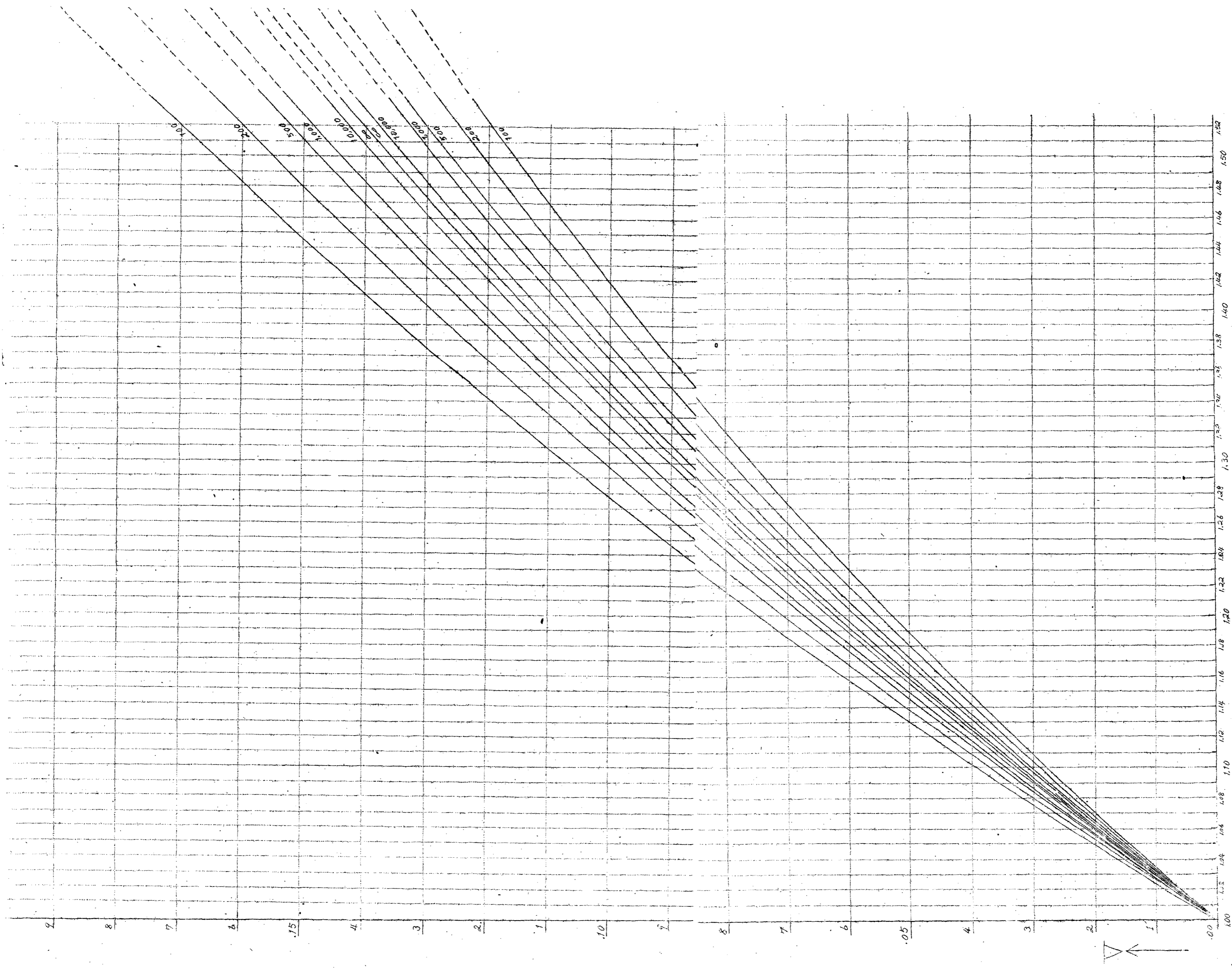
となる。勿論 V に depend するが吾々の方法の効率は、約 80% である。

参 考 文 献

- (1) Cramér, H. ; *Mathematical Methods of Statistics*. Princeton Univ. Press, (1946). pp. 369 - 370.
Smirnoff, N. : *Über die verteilung dess allgemeinen Gliedes in der Variationsche*.
Metron, Vol. 12 (1935) (未見)
- (2) Merrill, A.S. : *Frequency distribution of an Index when both components follows the normal law*. *Biom.*, 20 (1928)
Geary, R.C. : *The frequency distribution of the quotient of two normal variables*. *J.R.S.S.*, Vol. 9.3 (1932) (未見)
Fieller, E.C. : *The distribution of an Index in a normal bivariate population*. *Biom.*, Vol. 24 (1932)
Nicholson, C. : *A geometrical analysis of the frequency-distribution of the ratio between two variables*. *Biom.*, 32 (1941)
山本 純 恭 : 正規分布に従う二つの確率変数の比の分布に就て (未発表)
- (3) Robbins, H. : *Mixture of distributions*. *Ann. Math. Stat.*, Vol. 19 (1948)
- (4) 統計数値表 I. 河出書房 (1943) pp. 68-76
- (5) Cramér, H. *ibid* pp. 357-358

第 1 图





第 2 図
 変異係数の信頼区間 (信頼係数 95%)