

## Test of significance of correlation coefficient without taking sum of squares and product

木村, 茂  
東北大学教育教養部

<https://doi.org/10.15017/12722>

---

出版情報 : 統計科学研究. 1 (3), pp.6-14, 1956-09. Research Association of Statistical Sciences  
バージョン :  
権利関係 :

# Test of significance of correlation coefficient without taking sum of squares and product.

木村 茂 (東北大教育教養部)

## § 1. 概 要

$x_1, x_2, \dots, x_n$  を  $N(m, \sigma^2)$  からとられた  $n$  の標本とするととき標本平均値  $\bar{x}$  より大きい  $\nu$  の標本を得る確率  $p(n, \nu)$  を求める。

二次元正規母集団からの標本  $(x_i, y_i)$  ( $i=1, 2, \dots, n$ ) とし  $x_i, y_i$  それぞれの平均値を  $\bar{x}, \bar{y}$  とするとき平面を  $x = \bar{x}, y = \bar{y}$  で区割される四区分の第一、第三(第二、第四)の区分内に  $m$  の標本を得る確率  $Q(n, m)$  を求める。

## § 2. $P(n, \nu)$ について

$$x'_1 \leq x'_2 \leq \dots \leq x'_n$$

を  $N(m, \sigma^2)$  からとられた順序づけられた標本とする。標本平均値  $\bar{x}$  より大きい  $\nu$  の標本を得る確率を  $P(n, \nu)$  とすれば

$$P(n, \nu) = \int \dots \int n! f(x'_1) f(x'_2) \dots f(x'_n) dx'_1 \dots dx'_n \quad \dots (1)$$

$$-\infty < x'_1 \leq x'_2 \leq \dots \leq x'_{n-\nu} \leq \bar{x} \leq x'_{n-\nu+1} \leq \dots \leq x'_n < \infty$$

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-m)^2}$$

(1) は塩谷氏 [1] によって次のように与えられた。

$$P_i(n, \nu) = \binom{n}{\nu} \left(\frac{n}{n-\nu}\right)^{\frac{1}{2}} \int_0^{\infty} F_{\nu} \left(\frac{1}{\sqrt{\nu}} x\right) F_{n-\nu} \left(\frac{\sqrt{\nu}}{n-\nu} x\right) \frac{1}{\sqrt{2\pi}} e^{-\frac{n}{2(n-\nu)} x^2} dx \quad \dots (2)$$

こゝに  $F_i(u)$  は F. E. Grubbs によって与えられた函数で  $N(0, 1)$  からとられた  $i$  の標本中の最大値  $x'_i$  と平均値  $\bar{x}$  との差が  $u$  以下となる確率を示す。

即ち

$$F_i(u) = P_n(x'_i - \bar{x} \leq u)$$

(2) は  $F_i(u)$  の性質を直接利用して次のようにも簡単に求められる。

$N(0,1)$  (一般性を失うことなく  $N(m, \sigma^2)$  において  $m=0, \sigma^2=1$  とできる) からとられた  $n$  の標本中特定の  $\nu$  と残りの  $n-\nu$  の標本平均値をそれぞれ  $\bar{x}_\nu, \bar{x}_{n-\nu}$  とするとき, 両変数の同時分布は

$$g(\bar{x}_\nu, \bar{x}_{n-\nu}) d\bar{x}_\nu d\bar{x}_{n-\nu} = \frac{\sqrt{\nu(n-\nu)}}{2\pi} \exp\left[-\frac{1}{2}\left\{\nu\bar{x}_\nu^2 + (n-\nu)\bar{x}_{n-\nu}^2\right\}\right] d\bar{x}_\nu d\bar{x}_{n-\nu} \quad \dots (a)$$

総計  $n$  の標本平均値を  $\bar{x}$  とし,  $\bar{x}_\nu = \bar{x} + u$  とすれば

$$\nu\bar{x} + (n-\nu)\bar{x}_{n-\nu} = n\bar{x} \quad \text{より} \quad \bar{x}_{n-\nu} = \bar{x} - \frac{\nu}{n-\nu} u$$

$(\bar{x}_\nu, \bar{x}_{n-\nu})$  を  $(\bar{x}, u)$  に変換すれば (a) は

$$h(\bar{x}, u) d\bar{x} du = \frac{n}{2\pi} \sqrt{\frac{\nu}{n-\nu}} \exp\left[-\frac{1}{2}\left(n\bar{x}^2 + \frac{n\nu}{n-\nu} u^2\right)\right] d\bar{x} du \quad \dots (b)$$

$$-\infty < \bar{x} < \infty, \quad -\infty < u < \infty$$

$\bar{x}$  を積分して  $u$  の分布は

$$l(u) du = \frac{1}{\sqrt{2\pi}} \sqrt{\frac{n\nu}{n-\nu}} \exp\left[-\frac{1}{2} \frac{n\nu}{n-\nu} u^2\right] du \quad \dots (c)$$

$\bar{x}_\nu - \bar{x}$  が  $(u, u+du)$  間にあり,  $\nu$  の標本がすべて  $\bar{x}$  より大きく  $n-\nu$  の標本がすべて  $\bar{x}$  より小さい確率は,  $F_i(u)$  の対称的互性質及び (c) から

$$F_\nu(u) \cdot F_{n-\nu}\left(\frac{\nu}{n-\nu} u\right) l(u) du$$

$$\infty > u \geq 0$$

一般に  $n$  の標本中  $\nu$  が  $\bar{x}$  より大きく  $n-\nu$  が  $\bar{x}$  より小さい確率は

$$\binom{n}{\nu} \int_0^\infty F_\nu(u) F_{n-\nu}\left(\frac{\nu}{n-\nu} u\right) l(u) du \quad \dots (d)$$

これは (2) と同様の結果である。明かに  $p(n, \nu) = p(n, n-\nu)$  の対称性をもつ。

表 I  $p(n, \nu)$  について

○  $n=20$  までは塩谷氏によって計算された。

○  $21 < n \leq 35$  の計算については F. E. Grubbs の数表<sup>[2]</sup> と (d) とによって数値積分された。



11	12	13	14	15	16	17	18	19	20
.0006	.0001								
.0170	.0056	.0017	.0004	.0001					
.1323	.0622	.0260	.0099	.0034	.0010	.0003	.0001		
.3501	.2431	.1440	.0748	.0348	.0148	.0056	.0021	.0006	.0001
.3501	.3779	.3283	.2396	.1516	.0850	.0431	.0197	.0084	.0035
		.3283	.3505	.3101	.2351	.1565	.0933	.0502	.0249
				.3101	.3282	.2945	.2300	.1595	.0996
						.2945	.3097	.2813	.2249
							.2813	.2813	.2940

30	31	32	33	34	35
.001					
.012	.005	.002	.001		
.047	.027	.013	.005	.002	.001
.117	.079	.051	.029	.015	.007
.202	.161	.119	.083	.052	.032
.241	.227	.198	.160	.122	.084
	.227	.234	.222	.194	.159
			.222	.227	.216
					.216





§3.  $Q(n, m)$  について

二次元正規母集団  $f(x, y; a, b, \sigma_1^2, \sigma_2^2, \rho, \sigma_1, \sigma_2)$  において特に  $\rho=0$  の場合、これからとられた標本を  $(x_i, y_i)$  とするとき  $x_i, y_i$  はそれぞれ独立に

$$f(x_i, y_i) dx_i dy_i = \left( \frac{1}{\sqrt{2\pi}} e^{-\frac{(x_i-a)^2}{2\sigma_1^2}} \right) \left( \frac{1}{\sqrt{2\pi}} e^{-\frac{(y_i-b)^2}{2\sigma_2^2}} \right) dx_i dy_i$$

に従う。(以下一般性を失うことなく  $a=b=0$  とする) 各標本点を平面上に plot

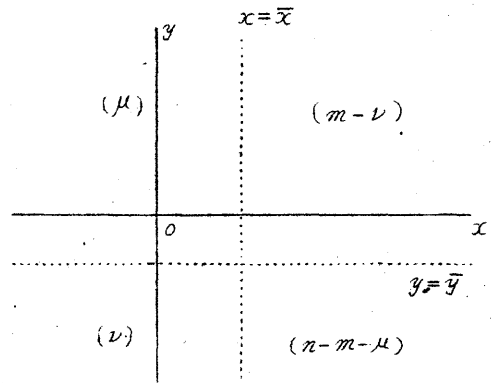
するとき,  $x=\bar{x}, y=\bar{y}$  で分けられる四区分の右上, 左上, 左下, 右下の各区分にそれぞれ  $m-\nu$ ,  $\mu$ ,  $\nu$ ,  $n-m-\mu$  の標本を得る確率を考える。

$x=\bar{x}$  の左側に  $\nu+\mu$  の  $x_i$  標本を得る確率は

$$P(n, \mu+\nu) \quad (4)$$

$y=\bar{y}$  の下側に  $n-m-\mu+\nu$  の  $y_i$  標本を得る確率は

$$P(n, n-m-\mu+\nu) \quad (5)$$



(4), (5) の条件のもとに  $x_i, y_j$  ( $i, j = 1, 2, \dots, n$ ) の組合せから各区分に上記の数値を持つ確率は

$$\begin{aligned} & \frac{1}{n!} \binom{m-\nu+\mu}{\mu} \binom{n-m-\mu+\nu}{\nu} \binom{n-\mu-\nu}{m-\nu} \binom{\mu+\nu}{\nu} (m-\nu)! \mu! \nu! (n-m-\mu)! \\ & = \binom{n-\mu-\nu}{m-\nu} \binom{\mu+\nu}{\nu} / \binom{n}{n-m-\mu+\nu} \quad (6) \end{aligned}$$

(4), (5), (6) から求める確率は

$$P(n, \mu+\nu) P(n, n-m-\mu+\nu) \binom{n-\mu-\nu}{m-\nu} \binom{\mu+\nu}{\nu} \binom{n}{n-m-\mu+\nu} \dots (7)$$

故に右上及び左下の部分に  $m$  の標本を得る確率を  $Q(n, m)$  とすれば,

$$Q(n, m) = \sum_{\nu=0}^m \sum_{\mu=0}^{\max(n-\nu-1, n-m+\nu-1)} P(n, \mu+\nu) P(n, n-m-\mu+\nu) \binom{n-\mu-\nu}{m-\nu} \binom{\mu+\nu}{\nu} \binom{n}{n-m-\mu+\nu} \dots (8)$$

明かに  $Q(n, m) = Q(n, n-m)$  の対称性をもつ。



表Ⅱ  $Q(n, m)$  は表Ⅰ  $P(n, \nu)$  の数値により (8) を用いて計算された。

平均値  $\bar{x}, \bar{y}$  の一致性から確率的に  $\bar{x} \rightarrow 0, \bar{y} \rightarrow 0$  ( $n \rightarrow \infty$ ), 故に  $Q(n, m)$  は漸近的に次の分布をなす。

$$Q(n, m) \sim \binom{n}{m} \left( \int_{-\infty}^0 \int_{-\infty}^0 f(x, y) dx dy + \int_0^{\infty} \int_0^{\infty} f(x, y) dx dy \right)^m \left( \int_{-\infty}^0 \int_{-\infty}^0 f(x, y) dx dy + \int_0^{\infty} \int_0^{\infty} f(x, y) dx dy \right)^{n-m}$$

$$= \binom{n}{m} \left( \frac{1}{2} \right)^m \left( \frac{1}{2} \right)^{n-m} = \binom{n}{m} \left( \frac{1}{2} \right)^n \quad \dots \dots (9)$$

表Ⅲは両者の数値的な漸近度を示す。これによれば  $n \geq 35$  においては小数オ三位まで一致していることが見られる。

表Ⅲ

$n \backslash m$	5		15		25		35	
	$Q(n, m)$	Approx.	$Q(n, m)$	Approx.	$Q(n, m)$	Approx.	$Q(n, m)$	Approx.
0	0.042	0.031						
1	.157	.156						
2	.301	.313	.004	.003				
3	.301	.313	.015	.014				
4			.043	.042				
5			.092	.092	.001	.002		
6			.152	.153	.005	.005		
7			.194	.196	.015	.014		
8			.194	.196	.034	.032	.001	.001
9					.061	.061	.002	.002
10					.098	.097	.005	.005
11					.132	.133	.012	.012
12					.154	.155	.024	.024
13					.154	.155	.043	.043
14							.068	.068
15							.095	.095
16							.118	.118
17							.132	.132
18							.132	.132

右欄は  $p = \frac{1}{2}$  の二項分布を示す。

#### §4. $f$ の検定

(14)

$H_0; \rho=0$  に対しては表Ⅱをもちいて検定を行うことが出来る.  $n > 35$  の場合は  $\rho = \frac{1}{2}$  の二項分布表による.

$H_0; \rho = \gamma$  に対しては母集団は  $f(x, y; 0, 0, \sigma_1^2, \sigma_2^2, r\sigma_1\sigma_2)$  と仮定され,  $\sigma_1/\sigma_2 = R$  があらかじめ知られているときは,  $\theta = \frac{1}{2} \tan^{-1} \left( \frac{2rR}{R^2-1} \right)$  だけ座標軸を回転することにより旧座標  $(x, y)$  は新座標  $(\xi, \eta)$  に変換され

$$\begin{cases} x = \xi \cos \theta - \eta \sin \theta \\ y = \xi \sin \theta + \eta \cos \theta \end{cases}$$

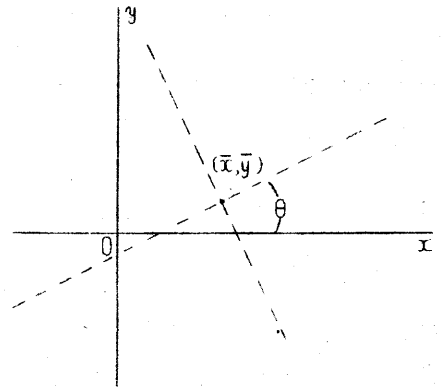
分布は

$$g(\xi, \eta) d\xi, d\eta = \frac{1}{2\pi \sigma_1 \sigma_2 \sqrt{1-\gamma^2}} \exp \left[ -\frac{1}{2(1-\gamma^2)} (\alpha \xi^2 + \beta \eta^2) \right] d\xi d\eta \quad (10)$$

( $\alpha, \beta$  は  $\sigma_1^2 \sigma_2^2 \lambda^2 - (\sigma_1^2 + \sigma_2^2) \lambda + 1 - \gamma^2 = 0$  の二根)

従って  $\xi, \eta$  はそれぞれ独立となり,  $(\bar{x}, \bar{y})$  点はそのまゝ  $(\bar{\xi}, \bar{\eta})$  なることを用いれば  $\rho=0$  の場合と同様に  $(\bar{x}, \bar{y})$  を通り旧座標軸と  $\theta$  角をなす直交二直線で分けられる四区分の第一, 第三(第二, 第四)の区分内に入る標本の箇数によって検定される.

通常の相関係数  $r$  による検定に対比してこの検定の効率については現在漸近的结果しか得られていない. この点の検討及び平均値を *mediau* に替えた場合の検定等に対する効率の検討等は今後を期したい.



### 参 考 文 献

- [1] 塩谷実: 正規母集団からとられた標本における正負の偏差の和の分布について, Inst. of Stat. Math., 6, No. 2.
- [2] F.E. Grubbs: Sample criteria for testing outlying observation, Ann. of Math. Stat., 21. (1950).