71. *On a Confidence Interval of the Ratio of Population Means of a Bivariate Normal Distribution*

By Junjiro OGAWA.

(Comm. by K. KUNUGI, M.J.A., July 12, 1951.)

We frequently encounter the following problem: Given a random sample of size

$$(x_i, y_i),\ i = 1, 2, \cdots, n,$$

it is required to estimate the ratio of population means

$$\lambda = m_2/m_1$$

of a bivariate normal population of which the means m_1, m_2, variances σ_1^2, σ_2^2, and the coefficient of correlation ρ are all unknown.

The maximum likelihood estimate of λ is

$$\hat{\lambda} = \bar{y}/\bar{x},$$

where $\bar{x}$ and $\bar{y}$ are respective sample means. The sampling distribution of $\hat{\lambda}$ is very complicated depending all unknown population parameters. The frequency function of $\hat{\lambda}$ is seen to be

$$f(z) = \frac{\sigma_1\sigma_2\sqrt{1-\rho^2}}{\pi(\sigma_2^2-2\rho\sigma_1\sigma_2 z+\sigma_1^2 z^2)}\exp.\left\{-\frac{n}{2(1-\rho^2)}\left(\frac{m_1^2}{\sigma_1^2}-2\rho\frac{m_1 m_2}{\sigma_1\sigma_2}+\frac{n_2^2}{\sigma_2^2}\right)\right\}+$$

$$+\frac{\sqrt{n}\,\{\sigma_2(\sigma_2 m_1-\rho\sigma_1 m_2)+z\sigma_1(\sigma_1 m_2-\rho\sigma_2 m_1)\}}{\pi(\sigma_2^2-2\rho\sigma_1\sigma_2 z+\sigma_1^2 z^2)^{3/2}}\exp.\left\{-\frac{n(mz-m_2)^2}{\sigma_2^2-2\rho\sigma_1\sigma_2 z+\sigma_1^2 z^2}\right\}$$

$$\times\int_0^{\frac{\sqrt{n}\,\{\sigma_2(\sigma_2 m_1-\rho\sigma_1 m_2)+z\sigma_1(\sigma_1 m_2-\rho\sigma_2 m_1)\}}{\sigma_1\sigma_2\sqrt{(1-\rho^2)(\sigma_2^2-2\rho\sigma_1\sigma_2 z+\sigma_1^2 z^2)}}}\exp.-\frac{t^2}{2}\,dt.$$

The author has not been able to manage the above distribution any how. An approach given in the following will be useful in some situations.

Let the sample central moments of the second order be

$$l_{11}=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2,\ l_{12}=l_{21}=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y}),\ l_{22}=\frac{1}{n}\sum_{i=1}^{n}(y_i-\bar{y})^2,$$

and let S.S. Wilk's generalized variance be

$$L = l_{11}l_{22}-l_{12}^2,$$

then it will be found that the statistic

$$F=\frac{n-2}{2}\left[\frac{l_{22}}{L}(\bar{x}-m_1)^2-2\frac{l_{12}}{L}(\bar{x}-m_1)(\bar{y}-m_2)+\frac{l_{11}}{L}(\bar{y}-m_2)^2\right]$$

is distributed according to Snedecor's F-distribution with degrees of freedom $(2, n-2)$, what is easily derivable from H. Hotelling's generalized Student's ratio T.

If we denote the $100\,\alpha\%$ point of the F-distribution of degrees of freedom $(2, n-2)$ by $F^{\,2}_{n-2}\,(100\alpha)$, that is

$$P(F \geqq F^{\,2}_{n-2}\,(100\alpha)) = \alpha\,,$$

we have in consequence

$$P(F \leqq F^{\,2}_{n-2}\,(100\alpha)) = 1-\alpha\,.$$

This gives a confidence region of the true mean point (m_1, m_2) of confidence coefficient $100\,(1-\alpha)\%$. The region

$$F \leqq F^{\,2}_{n-2}\,(100\alpha)$$

is the interier of the ellipse

$$\frac{l_{22}}{L}(m_1-\bar{x})^2-2\frac{l_{12}}{L}(m_1-\bar{x})(m_2-\bar{y})+\frac{l_{11}}{L}(m_2-\bar{y})^2=\frac{2}{n-2}F^{\,2}_{n-2}\,(100\alpha).$$

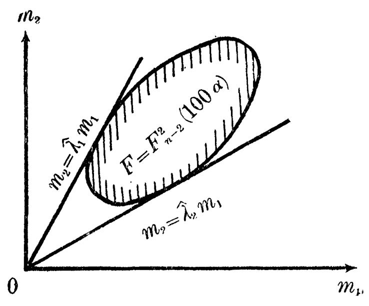
in the m_1, m_2 -plane.

In almost all circumstances of practical occurrence the means m_1 and m_2 are positive. If the level of significance α had been chosen sufficiently large, so that the above ellipse lies wholly in the first quadrant, then we can draw two tangents through the origin

$$m_2 = \hat{\lambda}_1\, m_1 \text{ and } m_2 = \hat{\lambda}_2\, m_1$$

to the ellipse (see Fig. 1).

Fig. 1



It is clear that the inequality

$$P(\hat{\lambda}_1 \leqq \lambda \leqq \hat{\lambda}_2) \geqq 1-\alpha$$

holds for all values λ independently of the other unknown parameters. Thus we get a confidence interval for λ in a certain sense.

After some easy calculations, it will be seen that

$$\hat{\lambda}_1=\left[\frac{2}{n-2}F^{\,2}_{n-2}\,(100\alpha)\,l_{12}-\bar{x}\,\bar{y}\pm\left(\frac{2}{n-2}F^{\,2}_{n-2}\,(100\alpha)L\cdot\left(\frac{l_{22}}{L}\bar{x}^2-2\frac{l_{12}}{L}\bar{x}\,\bar{y}+\frac{l_{11}}{L}\bar{y}^2-\frac{2}{n-2}F^{\,2}_{n-2}\,(100\alpha)\right)\right)^{\frac{1}{2}}\right]\Bigg/\left(\frac{2}{n-2}F^{\,2}_{n-2}\,(100\alpha)\,l_{11}-\bar{x}^2\right),$$

$$\hat{\lambda}_2 = \left[\frac{2}{n-2}F_{n-2}^{2}(100\alpha)\, l_{12} - \bar{x}\,\bar{y} \pm \left(\frac{2}{n-2}F_{n-2}^{2}(100\alpha)L\cdot\left(\frac{l_{22}}{L}\bar{x}^2 - 2\frac{l_{12}}{L}\bar{x}\,\bar{y} + \frac{l_{11}}{L}\bar{y}^2 - \frac{2}{n-2}F_{n-2}^{2}(100\alpha)\right)\right)^{\frac{1}{2}}\right] \Big/ \left(\frac{2}{n-2}F_{n-2}^{2}(100\alpha)\, l_{11} - \bar{x}^2\right),$$

where the signs of radicals should be chosen so as to $\hat{\lambda}_1 < \hat{\lambda}_2$. Whence, it follows that the level of significance α should have been chosen so as to be

$$\frac{l_{22}}{L}\bar{x}^2 - 2\frac{l_{12}}{L}\bar{x}\bar{y} + \frac{l_{11}}{L}\bar{y}^2 \geqq \frac{2}{n-2}F_{n-2}^{2}(100\alpha),$$

in other words, the null hypothesis H_0: $m_1 = m_2 = 0$ should have been rejected at the level α.

Illustrative Example: The following table shows the data of catches of the sand-eel (Ammodytes Personatus) which had been obtained by Mr. T. Miyazaki[1] at Osaka Bay in 1948, who had taken experiments with the purpose of comparing efficiencies of Komase-nets of different sizes, measured by their diameters and lengths.

Table.

Catches of sand-eel by two Komase-nets of different sizes (Unit in Kwan)

No. 1	Diameter 21m. Length 60m.	84 87 224 539 72 200 256 164 106 80	88 84 328 162 389 96 27	392 312 574 68 32
No. 2	Diameter 30m. Length 90m.	218 613 438 353 166 427 324 191 42 256	45 361 367 272 381 378 132	1110 585 134 228 101

If it is not so unreasonable to assume that the catches by Komase-nets of different sizes in almost uniform circumstances are distributed according to a multivariate normal distribution, then we shall be justified in estimating the ratio of true catches by two Komase-nets by means of the method explained above. The results are as follows:[2]

$$n = 22 \qquad \frac{2}{n-2} = 0.1$$

$$l_{11} = 24588, \qquad l_{12} = l_{21} = 16370, \qquad l_{22} = 61468,$$

1) The head of Sumoto Branch of the Inland Sea Regional Fisheries Research Laboratory, to whom I express my deep thanks for his kindness of putting the data at my disposal.

2) The computations are carried out by Mr. Y. Shiobara of the Institute of Statistical Mathematics and checked by Mr. Y. Miyamoto. The author confesses his deep thanks to them.

$$L = l_{11} l_{22} - l_{12}^2 = 1243380000$$
$$\bar{x} = 198.363, \qquad \bar{y} = 341.909$$
$$\bar{x}^2 = 39347, \qquad \bar{y}^2 = 116901$$

(1) $\alpha = 0.01 \qquad F_{n-2}^{2}(1) = F_{20}^{2}(1) = 5.85.$

$$\hat{\lambda}_1 = 0.85, \qquad \hat{\lambda}_2 = 3.82.$$

(2) $\alpha = 0.05, \qquad F_{n-2}^{2}(5) = F_{20}^{2}(5) = 3.49$

$$\hat{\lambda}_1 = 1.03, \qquad \hat{\lambda}_2 = 3.00$$