On the Ratio of Two Correlated Normal Random Variables
Author(s): D. V. Hinkley
Source: *Biometrika*, Vol. 56, No. 3 (Dec., 1969), pp. 635-639
Published by: Oxford University Press on behalf of Biometrika Trust
Stable URL: http://www.jstor.org/stable/2334671
Accessed: 14-04-2017 13:32 UTC

Biometrika (1969), **56**, 3, *p.* 635
Printed in Great Britain

On the ratio of two correlated normal random variables

By D. V. HINKLEY
Imperial College

Summary

The distribution of the ratio of two correlated normal random variables is discussed. The exact distribution and an approximation are compared. The comparison is illustrated numerically for the case of the normal least squares estimate of α/β in the linear model $E(y_i) = \alpha + \beta u_i$ $(i = 1, \ldots, n)$ with uncorrelated normal error terms.

1. Introduction

In a regression analysis of bivariate data it is sometimes of interest to estimate the ratio of two population parameters. Two examples are:

(i) The analysis of the simple linear model $y_i = \alpha + \beta u_i + \epsilon_i$ $(i = 1, \ldots, n)$, where $\epsilon_1, \ldots, \epsilon_n$ are independently normally distributed with zero mean and variance σ^2; then $-\alpha/\beta$ is the intercept of the regression line with the u-axis.

(ii) The analysis of the two-line linear model

$$y_i = \begin{cases} \alpha_1 + \beta_1 u_i + \epsilon_i & (i = 1, \ldots, n_1), \\ \alpha_2 + \beta_2 u_i + \epsilon_i & (i = n_1 + 1, \ldots, n_1 + n_2), \end{cases}$$

with notation as in (i); the ratio $(\alpha_1 - \alpha_2)/(\beta_2 - \beta_1)$ is the abscissa of the intersection of the two regression lines.

In each example the maximum likelihood estimate of the ratio is the ratio of two correlated normally distributed variables, themselves estimates. In this paper the general distribution of ratios of this type is derived and compared with the approximation obtained by assuming the denominator random variable to be of constant sign. This has particular relevance to the examples above, and a numerical comparison is given for (i) above.

2. Theoretical results

Let X_1 and X_2 be normally distributed random variables with means θ_i, variances σ_i^2 $(i = 1, 2)$ and correlation coefficient ρ, and let $W = X_1/X_2$. The exact distribution of W and the standard approximation based on assuming $X_2 > 0$ are examined in some detail.

2·1. *Exact distribution of W*

The distribution of W when $\theta_1 = \theta_2 = 0$ was given by Geary (1930). Fieller (1932) and more recently Marsaglia (1965) considered the general problem with non-zero means. The latter studied the ratio

$$Z = \frac{a + Y_1}{b + Y_2},$$

in our notation, where Y_1 and Y_2 are independent $N(0, 1)$ variables. The connexion between Z and W is said (incorrectly) to be that 'It suffices to study Z; translations and changes of scale will provide the general ratio W.' This assumes there are four parameters, whereas there are five: θ_1, θ_2, σ_1^2, σ_2^2 and ρ. In fact Z has no great advantage over W, since the distributions of both involve the bivariate normal distribution.

If the joint density of (X_1, X_2) is $g(x, y)$ and the p.d.f. of W is $f(w)$, then

$$f(w) = \int_{-\infty}^{\infty} |y|\, g(wy, y)\, dy.$$

On substituting the bivariate normal density for $g(x, y)$, a simple integration gives

$$f(w) = \frac{b(w)\, d(w)}{\sqrt{(2\pi)}\, \sigma_1 \sigma_2 a^3(w)} \left[\Phi\left\{\frac{b(w)}{\sqrt{(1-\rho^2)}\, a(w)}\right\} - \Phi\left\{-\frac{b(w)}{\sqrt{(1-\rho^2)}\, a(w)}\right\}\right] + \frac{\sqrt{(1-\rho^2)}}{\pi \sigma_1 \sigma_2 a^2(w)} \exp\left\{-\frac{c}{2(1-\rho^2)}\right\}, \quad (1)$$

where

$$a(w) = \left(\frac{w^2}{\sigma_1^2} - \frac{2\rho w}{\sigma_1 \sigma_2} + \frac{1}{\sigma_2^2}\right)^{\frac{1}{2}},$$

$$b(w) = \frac{\theta_1 w}{\sigma_1^2} - \frac{\rho(\theta_1 + \theta_2 w)}{\sigma_1 \sigma_2} + \frac{\theta_2}{\sigma_2^2},$$

$$c = \frac{\theta_1^2}{\sigma_1^2} - \frac{2\rho\theta_1\theta_2}{\sigma_1\sigma_2} + \frac{\theta_2^2}{\sigma_2^2},$$

$$d(w) = \exp\left\{\frac{b^2(w) - c a^2(w)}{2(1-\rho^2)\, a^2(w)}\right\}. \quad (2)$$

Also

$$\Phi(y) = \int_{-\infty}^{y} \phi(u)\, du, \quad \text{where} \quad \phi(u) = \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}u^2}.$$

This result for the probability density function of W was given by Fieller (1932).

The cumulative distribution function $F(w)$ of W is found by direct calculation to be

$$F(w) = L\left\{\frac{\theta_1 - \theta_2 w}{\sigma_1 \sigma_2 a(w)}, \quad -\frac{\theta_2}{\sigma_2}; \quad \frac{\sigma_2 w - \rho\sigma_1}{\sigma_1 \sigma_2 a(w)}\right\} + L\left\{\frac{\theta_2 w - \theta_1}{\sigma_1 \sigma_2 a(w)}, \quad \frac{\theta_2}{\sigma_2}; \quad \frac{\sigma_2 w - \rho\sigma_1}{\sigma_1 \sigma_2 a(w)}\right\}, \quad (3)$$

where

$$L(h, k; \gamma) = \frac{1}{2\pi\sqrt{(1-\gamma^2)}} \int_h^{\infty} \int_k^{\infty} \exp\left\{-\frac{x^2 - 2\gamma xy + y^2}{2(1-\gamma^2)}\right\} dx\, dy$$

is the standard bivariate normal integral tabulated by the National Bureau of Standards (1959).

It is easy to see from (3) that as $\theta_2/\sigma_2 \to \infty$, i.e. as $\mathrm{pr}\,(X_2 > 0) \to 1$,

$$F(w) \to \Phi\left\{\frac{\theta_2 w - \theta_1}{\sigma_1 \sigma_2 a(w)}\right\}, \quad (4)$$

a fact which is used in approximating to $F(w)$.

2·2. *Comparison of $F(w)$ with its approximation*

If $0 < \sigma_2 \ll \theta_2$, the limit in (4) suggests that

$$F^*(w) = \Phi\left\{\frac{\theta_2 w - \theta_1}{\sigma_1 \sigma_2 a(w)}\right\} \quad (5)$$

will be a useful approximation to $F(w)$. It is then important to compare $F(w)$ and $F^*(w)$ in some detail.

Now $$F(w) = \mathrm{pr}\,(X_1 - wX_2 \leqslant 0,\, X_2 > 0) + \mathrm{pr}\,(X_1 - wX_2 \geqslant 0,\, X_2 < 0)$$
$$= F^*(w) + \mathrm{pr}\,(X_2 < 0)\{1 - 2\,\mathrm{pr}\,(X_1 - wX_2 \leqslant 0, X_2 < 0)\}, \tag{6}$$
so that $$|F(w) - F^*(w)| \leqslant \mathrm{pr}\,(X_2 < 0) = \Phi(-\theta_2/\sigma_2). \tag{7}$$

This bound on the difference is attained at $w = \pm\infty$, for, by the definition of $F^*(w)$ or by (6),

$$F^*(-\infty) = \Phi(-\theta_2/\sigma_2) \quad \text{and} \quad F^*(\infty) = 1 - \Phi(-\theta_2/\sigma_2), \tag{8}$$

whereas $F(w)$ is proper. Note that (7), a familiar result, implies $F^* \to F$ uniformly as $\sigma_2 \to 0$.

Further comparison is possible by examination of the derivative of $F^*(w)$, which by (5) and (2) is

$$f^*(w) = \frac{b(w)\,d(w)}{\sqrt{(2\pi)}\,\sigma_1\sigma_2 a^3(w)}. \tag{9}$$

Clearly $f^*(w) < 0$ when $b(w) < 0$, that is for w in the region

$$w < -\frac{\psi_1}{\psi_2} \quad (\psi_2 < 0),$$
or
$$w > -\frac{\psi_1}{\psi_2} \quad (\psi_2 > 0), \tag{10}$$

where $\psi_1 = \rho\theta_1\sigma_1\sigma_2 - \theta_2\sigma_1^2$ and $\psi_2 = \rho\theta_2\sigma_1\sigma_2 - \theta_1\sigma_2^2$. When $b(w) > 0$, it is easy to see that $f(w) > f^*(w)$. For by (1) and (9)

$$\frac{f(w)}{f^*(w)} = \Phi\left\{\frac{b(w)}{a(w)\sqrt{(1-\rho^2)}}\right\} - \Phi\left\{-\frac{b(w)}{a(w)\sqrt{(1-\rho^2)}}\right\} + \frac{2\sqrt{(1-\rho^2)}\,a(w)}{b(w)}\,\phi\left\{\frac{b(w)}{a(w)\sqrt{(1-\rho^2)}}\right\}.$$

The inequality $\Phi(-u) < u^{-1}\phi(u)$ $(u > 0)$ gives the desired result. Hence $f(w) > f^*(w)$ for all w. That is, $F(w) - F^*(w)$ increases monotonically from $-\Phi(-\theta_2/\sigma_2)$ at $w = -\infty$ to $\Phi(-\theta_2/\sigma_2)$ at $w = +\infty$.

The next step is to determine the point at which $F(w) = F^*(w)$. Now by (6)

$$F(w) = F^*(w) + \Phi\left(-\frac{\theta_2}{\sigma_2}\right) - 2\int_{-\infty}^{0}\int_{-\infty}^{wy} g(x,y)\,dx\,dy, \tag{11}$$

where $g(x,y)$ is the bivariate normal density. The double integral in (11) is easily reduced to

$$\int_{-\infty}^{-\theta_2/\sigma_2} \phi(u)\,\Phi\left\{\frac{(w\sigma_2 - \rho\sigma_1)\,u + (w\theta_2 - \theta_1)}{\sigma_1(1-\rho^2)}\right\} du, \tag{12}$$

so that the solution to $F^*(w) = F(w)$ is the solution to

$$\int_{-\infty}^{-\theta_2/\sigma_2} \phi(u)\,du = 2\int_{-\infty}^{-\theta_2/\sigma_2} \phi(u)\,\Phi\left\{\frac{(w\sigma_2 - \rho\sigma_1)\,u + (w\theta_2 - \theta_1)}{\sigma_1(1-\rho^2)}\right\} du.$$

When $\psi_2 = 0$, the solution is $w_0 = \rho\sigma_1/\sigma_2$, and in general the solution w_0 satisfies

$$w_0 < \frac{\rho\sigma_1}{\sigma_2} \quad \text{if} \quad \psi_2 < 0,$$
$$w_0 > \frac{\rho\sigma_1}{\sigma_2} \quad \text{if} \quad \psi_2 > 0;$$

in fact $|w_0 - (\rho\sigma_1/\sigma_2)|$ increases as $|\psi_2|$ increases.

Finally, more precise bounds than (7) can be obtained for finite w. Starting with (12) it is possible to put bounds on the double integral, first integrating by parts and then using the inequality $\Phi(u) < \phi(u)/(-u)$ for negative u. The resulting bounds on $F(w)-F^*(w)$ are

$$\begin{aligned} J \leqslant F(w)-F^*(w) \leqslant J+K(w) \quad (w \geqslant \rho\sigma_1/\sigma_2), \\ J+K(w) \leqslant F(w)-F^*(w) \leqslant J \quad (w < \rho\sigma_1/\sigma_2), \end{aligned} \tag{13}$$

where
$$J = \Phi\left(-\frac{\theta_2}{\sigma_2}\right)\left[1-2\Phi\left\{\frac{\psi_2}{\sigma_1\sigma_2^2\sqrt{(1-\rho^2)}}\right\}\right]$$

and
$$K(w) = \frac{2(w\sigma_2-\rho\sigma_1)}{\sigma_1\theta_2 a(w)}\,\phi\left\{\frac{w\theta_2-\theta_1}{\sigma_1\sigma_2 a(w)}\right\}\Phi\left\{-\frac{b(w)}{a(w)\sqrt{(1-\rho^2)}}\right\}.$$

3. A NUMERICAL ILLUSTRATION

To illustrate the results in the previous section, in particular the new bounds (13) on $F(w)-F^*(w)$, several calculations were carried out for the simple linear regression case mentioned in § 1. The variables X_1 and X_2 are respectively $\hat{\alpha}$ and $\hat{\beta}$, and specific values $\alpha = 0$, $\beta = 0{\cdot}2$, $\sigma = 1$ are used. The covariance matrix of $(\hat{\alpha}, \hat{\beta})$ gives the formulae for σ_1, σ_2 and ρ. Lastly, $\theta_1 = \alpha$ and $\theta_2 = \beta$.

The short table given (Table 1) is for the case $n = 10$. Values of $F(w)$, $F^*(w)$, bounds (13) and uniform bound (7), are given for seven values of w. The values of $F(w)$ were obtained by a two-dimensional iterative Simpson's Rule, and are correct to the number of figures given.

Table 1. *Comparison of* $F(w)$ *and* $F^*(w)$ *in the regression case with* $\alpha = 0$, $\beta = 0{\cdot}2$, $\sigma = 1$ *and* $n = 10$

			Bounds (13)		
w	$F(w)$	$F^*(w)$	Lower	Upper	Bound (7)
−15	0·05939	0·09403	−0·03464	−0·03462	0·03464
−10	0·08995	0·12458	−0·03464	−0·03462	.
−5	0·16739	0·20203	−0·03464	−0·03462	.
0	0·46536	0·50000	−0·03464	−0·03462	.
5	0·96445	0·99908	−0·03463	−0·03462	.
10	0·96568	0·99967	−0·03462	−0·03352	.
15	0·96832	0·99698	−0·03462	−0·02495	.

The value of $-\psi_1/\psi_2$ is 7·0 and $\psi_2 > 0$, so that by (10) $F^*(w)$ is decreasing for $w > 7{\cdot}0$, albeit very slowly. The precision of the bounds (13) is quite striking, and it is interesting to see that the uniform bound is itself nearly attained over a large part of the range. The bounds do become slacker in the tails. This and other examples suggest as an improved approximation to $F(w)$

$$F^{**}(w) = F^*(w) - \Phi(-\theta_2/\sigma_2), \tag{14}$$

when $F^*(w) > F(w)$ over most of the range; the correction term in (14) would be added if $F^*(w) < F(w)$ over most of the range.

The relative simplicity of the approximations $F^*(w)$ and $F^{**}(w)$ facilitates extensive use of $F(w)$, particularly in the two regression examples of § 1, where approximation is very accurate for moderate sample sizes. For an application to the two-phase regression situation, see Hinkley (1969). Approximation may be useful even for computing a single value of $F(w)$ since computation of (3) can often involve trivariate interpolation from tables of $L(h, k; \gamma)$.

I am grateful to Mrs E. A. Chambers for the numerical computations.

References

Fieller, E. C. (1932). The distribution of the index in a normal bivariate population. *Biometrika* **24**, 428–40.

Geary, R. C. (1930). The frequency distribution of the quotient of two normal variates. *J. R. Statist. Soc.* **93**, 442–6.

Hinkley, D. V. (1969). Inference about the intersection in two-phase regression. *Biometrika* **56**, 495–504.

Marsaglia, G. (1965). Ratios of normal variables and ratios of sums of uniform variables. *J. Am. Statist. Ass.* **60**, 193–204.

National Bureau of Standards (1959). *Tables of the Bivariate Normal Distribution Function and Related Functions. Applied Mathematics Series*, no. 50. Washington: U.S. Government Printing Office.

[*Received January* 1969. *Revised May* 1969]