Rapporto di ricerca n. 204

Some characteristics of the estimator of the ratio of two means

Carlotta GALEONE - Angiola POLLASTRI

Marzo 2011

Dipartimento di Metodi Quantitativi per le Scienze Economiche ed Aziendali
Università degli Studi di Milano Bicocca
Segreteria di redazione: Andrea Bertolini
Via Bicocca degli Arcimboldi, 8 - 20126 Milano - Italia
Tel +39/02/64483102/3 - Fax +39/02/64483105

SOME CHARACTERISTICS OF THE ESTIMATOR OF THE RATIO OF TWO MEANS

CARLOTTA GALEONE
Dipartimento di Medicina del Lavoro "Clinica del Lavoro Luigi Devoto", Università degli Studi di Milano
Dipartimento di Epidemiologia, Istituto di Ricerche Farmacologiche "Mario Negri", Milano
carlotta.galeone@unimi.it

ANGIOLA POLLASTRI
Dipartimento di Metodi Quantitativi per le Scienze Economico-Aziendali, Università degli Studi di Milano-Bicocca.
angiola.pollastri@unimib.it

Abstract

Generally only the asymptotic expectation and the asymptotic variance of the estimator of the ratio of two means are considered. In the present paper, the estimate of the asymptotic density function of the estimator of the ratio is presented. In fact, in different situations having the same asymptotic expectation and variance of the estimator, the density function may differ in a considerably way especially with respect to the skewness.
Confidence intervals for the ratio of two means which keep in consideration the shape of the distribution, proposed by Galeone, are reported.

Key words: ratio of two means, distribution of the estimator of the ratio of two correlated Normals, skewness, confidence intervals for the ratio.

1. Introduction

Often it is required to estimate the ratio of two means or the ratio of two totals. The estimator is a ratio of two random variables normally or asymptotically normally distributed. So the characteristics of the estimator of a ratio is connected with the ratio of two normal random variables. The distribution of the ratio of two normal random variables was studied by Geary (1930), Fieller (1932), Marsaglia (1965, 2006), Frosini (1970) and many others. In 1986 two papers due to Aroian and Oksoy et Aroyan proposed the Distribution Function of the ratio of two normal random variables.

Taking into account the result of Aroian and Oksoy, in the present paper we consider an estimate of the density function of the estimator of a ratio of two means. Then we present some examples in which the asyntotic variance of the estimator of the ratio is constant but the distributions has quite different shape especially with respect to the skewness.

Then in section 4 a proposal due to Galeone (2007) of confidence intervals for the ratio of two means which keep in account of the characteristics of the distribution of the estimator of a ratio is reported.

2. The distribution of the estimator of the ratio of two means

Let us consider a bivariate random variable $(r.v.)$ (X_1, X_2) having

$$E(X_1)=\mu_1 \;;\; E(X_2)=\mu_2 \;;\; Var(X_1)=\sigma^2{}_1 \;;\; Var(X_2)=\sigma^2{}_2 \;;\; Corr(X_1,X_2)=\rho\,.$$

Let us suppose we have drawn a simple random sample of n elements and we have obtained the observations (x_{1i}, x_{2i}) *(i=1,…,n).*

If (X_1, X_2) is a Bivariate Correlated Normal (B.C.N.) or if n is large, the r.v. $(\bar{X}_1, \bar{X}_2)$ tends to a B.C.N. having parameters

$$E(\bar{X}_1)=\mu_1 \;;\; E(\bar{X}_2)=\mu_2 \;;\; Var(\bar{X}_1)=\frac{\sigma^2{}_1}{n} \;;\; Var(\bar{X}_2)=\frac{\sigma^2{}_2}{n} \;;\; Corr(\bar{X}_1,\bar{X}_2)=\rho\,.$$

We can compute the maximum likelihood (M.L.) estimates of the means μ_1 and μ_2 indicated respectively by $\bar{x}_1$ and $\bar{x}_2$. The M.L. estimates of σ_1^2, σ_2^2 and ρ are respectively given by

$$s_1^2=\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)^2/n \qquad s_2^2=\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)^2/n$$

$$r=\sum_{i=1}^{n}\;(x_{1i}-\bar{x}_1)(x_{2i}-\bar{x}_2)\Big/\sqrt{\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)^2\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)^2}$$

Keeping into account the result obtained by Aroian (1986) and Oksoy et Aroyan (1986) about the Distribution Function (DF) of the ratio of two correlated Normals, we can obtain the distribution of the r.v. $W_n=\dfrac{\bar{X}_1}{\bar{X}_2}$, indicated by $F_{W_n}(w)$. Often $W_n=\dfrac{\bar{X}_1}{\bar{X}_2}$ is used as the estimator of $R=\dfrac{\mu_1}{\mu_2}$.

By substituting in it the estimates of μ_1, μ_2, σ_1, σ_2 and ρ, we can estimate the DF of W_n

$$\hat{F}_{W_n}(w)=L\left(\frac{a_n-b_nt_w}{\sqrt{1+t_w^2}},-b_n,\frac{t_w}{\sqrt{1+t_w^2}}\right)+L\left(\frac{b_nt_w-a_n}{\sqrt{1+t_w^2}},b_n,\frac{t_w}{\sqrt{1+t_w^2}}\right)$$

where

$$a_n=\sqrt{\frac{n}{1-r^2}}\left(\frac{\bar{x}_1}{s_1}-r\frac{\bar{x}_2}{s_2}\right) \qquad b_n=\sqrt{n}\left(\frac{\bar{x}_2}{s_2}\right) \qquad t_w=\sqrt{\frac{1}{1-r^2}}\left(\frac{s_2}{s_1}w-r\right)$$

and where, according to the indication of Kotz et al (2000),

$$L(h,k,\rho)=\frac{1}{2\pi\sqrt{1-\rho^2}}\int_h^\infty\int_k^\infty \exp\left\{-\frac{1}{2\left(1-\rho^2\right)}(x_1^2-2\rho x_1x_2+x_2^2)\right\}dx_1dx_2$$

The estimate of the DF of W_n can be computed as a function of *w* using Fortran+IMSL library or MATLAB or other libraries, especially those containing routines regarding the DF of a B.C.N. or other functions which can give the DF of the B.C.N.

The density function is the following:

$$f(w)=\frac{s_2}{s_1}\sqrt{\frac{1}{1-r^2}}\,f(t)$$

where

$$f(t)=\frac{1}{\pi}e^{-\frac{1}{2}(a_n{}^2+b_n{}^2)}\frac{1}{\left(1+t^2\right)}\left\{1+c\int_0^q\varphi(u)du\right\}$$

$$q=\frac{b_n+a_nt}{\sqrt{1+t_w^2}}, \qquad c=q\{\varphi(q)\}^{-1}, \qquad \varphi(u)=(2\pi)^{-\frac{1}{2}}e^{-\frac{u^2}{2}}$$

Remembering that the variance of the estimator $W_n=\frac{\bar{X}_1}{\bar{X}_2}$ (see, f.i. Cochran, 1977), when *n* is not too small, is approximately given by:

$$Var(W_n)\cong\frac{R^2}{n}\left\{CV_1^2+CV_2^2-2\rho CV_1CV_2\right\}$$

where $CV_1=\frac{\sigma_1}{\mu_1}, CV_2=\frac{\sigma_2}{\mu_2}$

Cochran suggest to use the above approximate variance to build confidence intervals based on the Normal distribution.

We wish to present some proves in order to understand if this approximation is acceptable.

Now we compute

$$\Pr(W_n<0.8)=F_{W_n}(0.8) \text{ and } 1\text{-}\Pr(W_n<1.2)=1-F_{W_n}(1.2)\,.$$

And we compare corresponding probabilities in a Normal distribution in different situations having the same $Var(W_n)$.

Tab.1: Tails of some distributions of the ratio with having the same $Var(W_n)$

1)n=20 $Var(W_n) \cong 0.025$

$\sigma^2{}_1$	$\sigma^2{}_2$	ρ	Norm	$F_{W_n}(0.8)$	1- $F_{W_n}(1.2)$
0.1	0.4	0.0	0.1030	0.0678	0.1273
0. 25	0.25	0.0	0.1030	0.0818	0.1203
0.4	0.1	0.0	0.1030	0.0943	0.1106

2) *n=20* $Var(W_n) \cong 0.015$

$\sigma^2{}_1$	$\sigma^2{}_2$	ρ	Norm	$F_{W_n}(0.8)$	1- $F_{W_n}(1.2)$
0.1	0.4	0.5	0.0512	0.0218	0.0843
0.25	0.25	0.4	0.0512	0.0369	0.0699
0.4	0.1	0.5	0.0512	0.0524	0.0524

3) *n=20* $Var(W_n) \cong 0.07$

$\sigma^2{}_1$	$\sigma^2{}_2$	ρ	Norm	$F_{W_n}(0.8)$	1- $F_{W_n}(1.2)$
0.1	0.4	0.9	0.0084	0.0003	0.0346
0.25	0.25	0.72	0.0084	0.0052	0.0170
0.4	0.1	0.9	0.0084	0.0165	0.0038

4) *n=30* $Var(W_n) \cong 0.0167$

$\sigma^2{}_1$	$\sigma^2{}_2$	ρ	Norm	$F_{W_n}(0.8)$	1- $F_{W_n}(1.2)$
0.1	0.4	0.0	0.1226	0.0942	0.1475
0.25	0.25	0.0	0.1226	0.1054	0.1395
0.4	0.1	0.0	0.1226	0.1160	0.1298

5) *n*=30 $Var(W_n) \cong 0.01$

$\sigma^2{}_1$	$\sigma^2{}_2$	ρ	Norm	$F_{W_n}(0.8)$	1- $F_{W_n}(1.2)$
0.1	0.4	0.5	0.0668	0.0397	0.0481
0.25	0.25	0.4	0.0668	0.0538	0.0826
0.4	0.1	0.5	0.0676	0.0668	0.0676

6) *n=30* $Var(W_n) \cong 0.0047$

$\sigma^2{}_1$	$\sigma^2{}_2$	ρ	Norm	$F_{W_n}(0.8)$	1- $F_{W_n}(1.2)$
0.1	0.4	0.9	0.0141	0.0022	0.0382
0.25	0.25	0.72	0.0141	0.01	0.0221
0.4	0.1	0.9	0.0141	0.0220	0.0085

3. Some considerations about the skewness of the distribution of the estimator of *R*

The analysis of the shape of a distribution is usually based on the traditional statistics $\sqrt{\beta_1}$, i.e. index of skewness, and β_2, i.e. index of kurtosis. These indices are both based on the moments of the distribution. In particular $\sqrt{\beta_1}$ is the third standardized moment and β_2 is the fourth standardized moment of the distribution. Unfortunately, this approach can not be used to study the shape of the ratio distribution because no moment of the r.v. W_n exists.

For these reasons, we studied the shape by means of the median and percentiles of the ratio distribution, that can not be obtained analytically, but numerical calculations have to be done for each particular case.

First of all, we consider the index proposed by Bowley in 1901 (Brentari, 1990)

$$sk\ (0.25) = \frac{Q_3 + Q_1 - 2Me}{Q_3 - Q_1}$$

where *Me* is the median, Q_1 and Q_3 are respectively the first and the third quartile. The index *sk(0.25)* varies in the interval *[-1,1]* and so it is easily interpretable.

Then we use a more analytic index based on the asymmetry of points suggested by David F.N. and Johnson in 1956 (see, f.i., Brentari,1990); for a continuous r.v. it is defined as

$$sk(p) = \frac{x(1-p) + x(p) - 2Me}{x(1-p) - x(p)} \qquad 0 \le p < \frac{1}{2}$$

where $x(p) = F^{-1}(p)$ is the p^{th} percentile.

The index *sk(p)* is the difference of the distance of the $(1\text{-}p)^{th}$ percentile from the median and the distance of the median from the p^{th} percentile divided by the distance from the $(1\ \text{-}p)^{th}$ percentile and the p^{th} percentile. So, having $-1 \le sk(p) \le 1$, the index $sk(p)$ is normalized.

Particular density functions and the relatives function $sk(p)$ of the r.v. W_n with different parameters are given in three sets of situation having the same variance of the estimators of the ratio of two means.

Example 1

The first case has $CV_x^2 = 0.1$, $CV_y^2 = 0.4$, r=0.5

The second case has $CV_x^2 = 0.4$, $CV_y^2 = 0.1$, r=0.5

The third case has $CV_x^2 = 0.25$, $CV_y^2 = 0.25$, r=0.4

The sample size is *n=30*. The asymptotic variance in the three situations is $Var(\hat{R}) \cong 0.01$

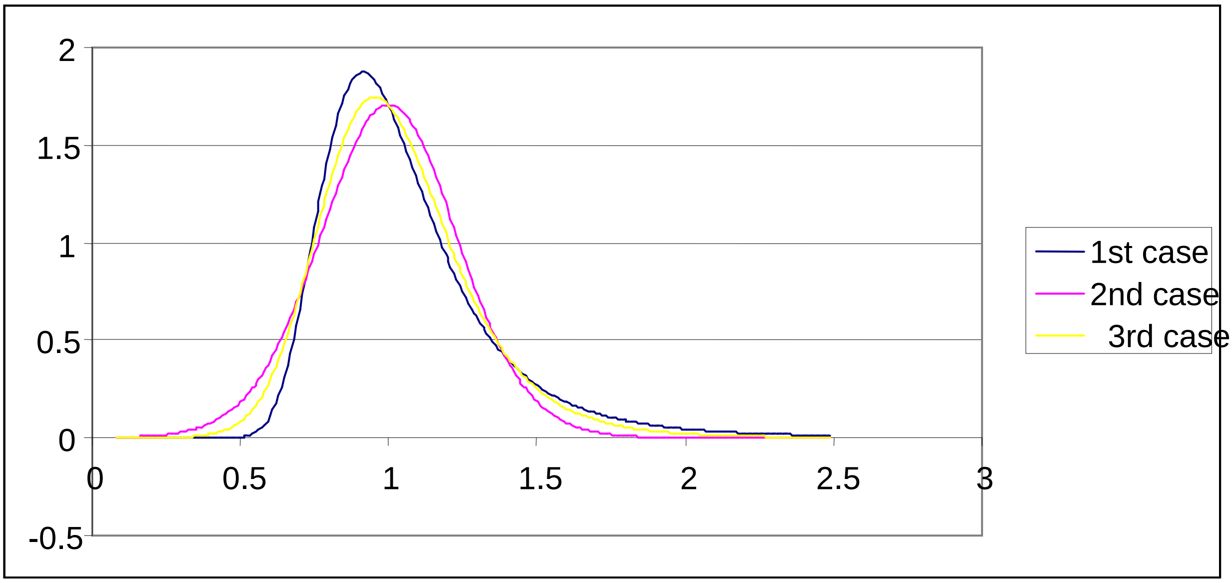


Fig.1: Density function in three situations having $Var(\hat{R}) \cong 0.01$

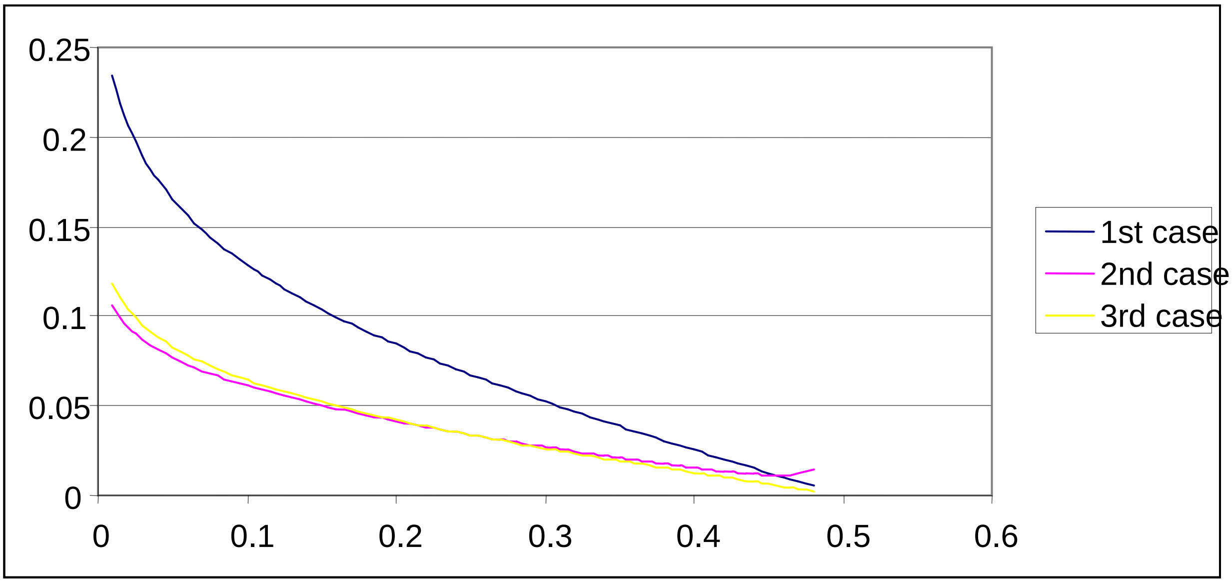


Fig. 2: Asymmetry indexes $sk(p)$

Example 2

The first case has CV_x^2 =0.1 e. $CV_y^2 = 0.4$ r=0.9

The second case has $CV_x^2 = 0.4$ e $CV_y^2 = 0.1$ r=0.9

The third case has CV_x^2 =0.25 e $CV_y^2 =$ 0.25 r=0.72

The asymptotic variance in the three situations (sample size *n=30*) is $Var(\hat{R}) \cong 0.0047$.

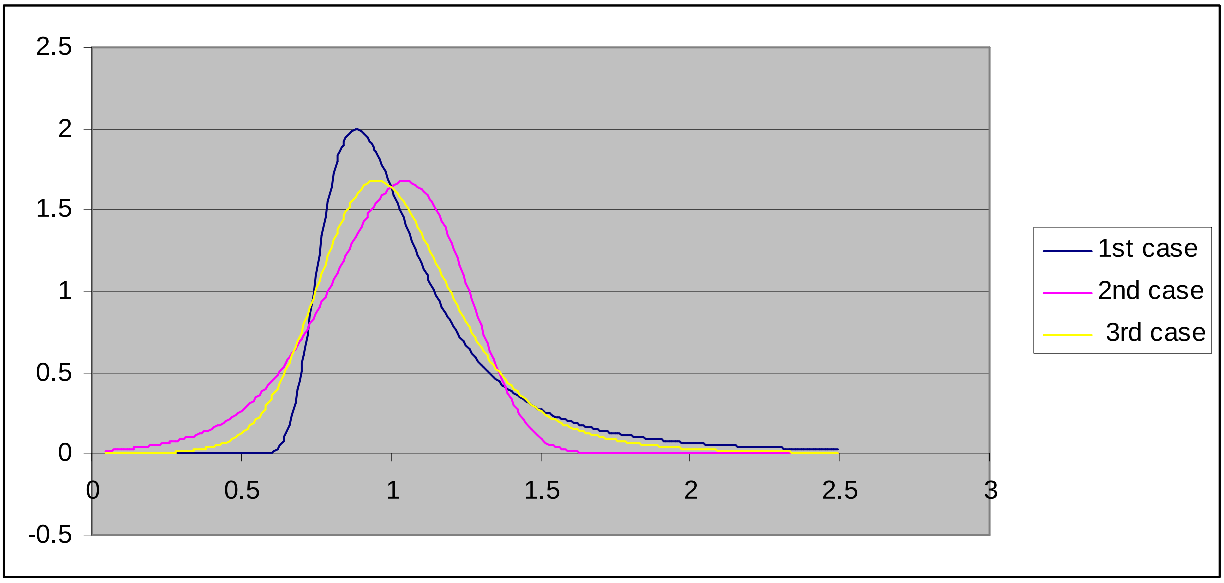
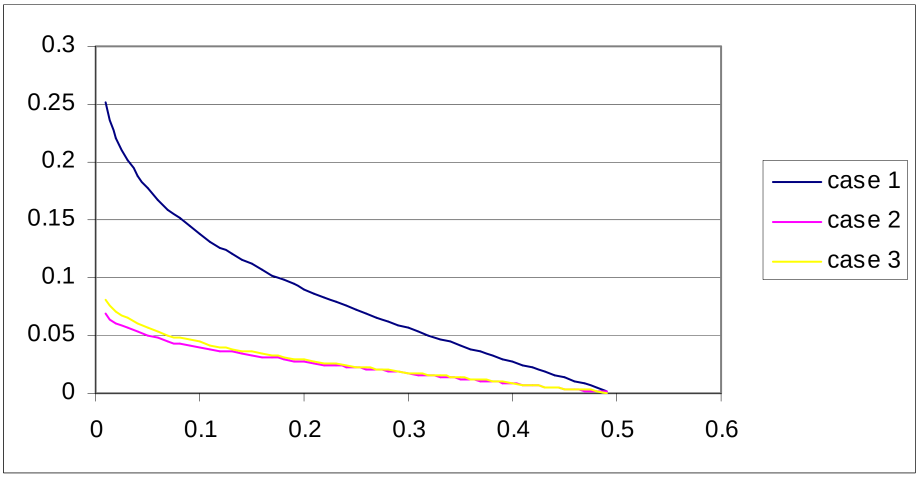


Fig.3: Density function in three situations having $Var(\hat{R}) = 0.0047$



Example 3

The first case has CV_x^2 =1 e. $CV_y^2 = 2$ r=0.9

The second case has $CV_x^2 = 2$ e $CV_y^2 = 1$ r=0.9

The third case has CV_x^2 =1.5 e CV_y^2 = 1.5 r=0.8485

The asymptotic variance, in the above situations having sample size equal to 30, is $Var(\hat{R}) = 0.01515$

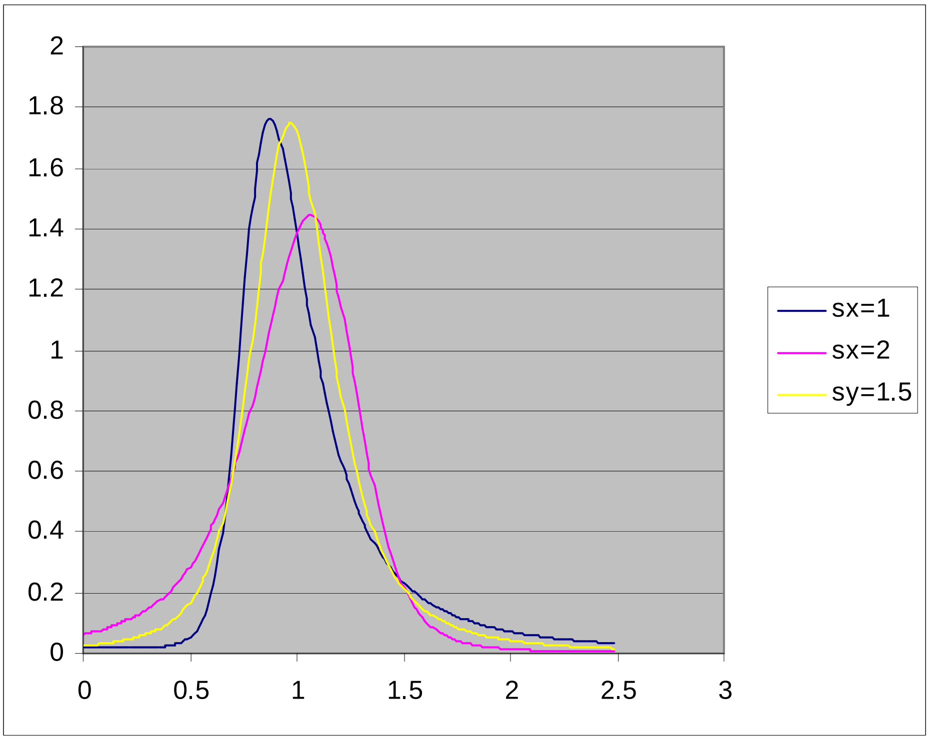


Fig.5: Density function in three situations having $Var(\hat{R}) = 0.01515$

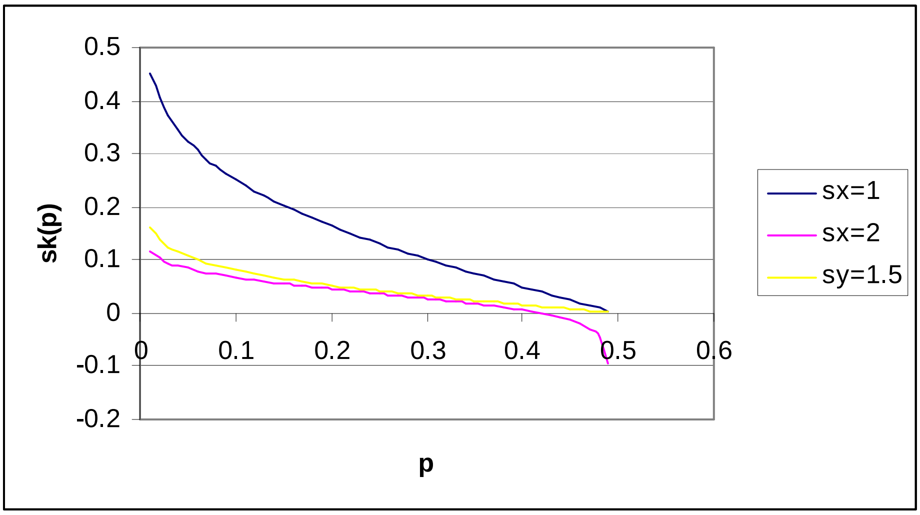


Fig. 6: Asymmetry indexes $sk(p)$

The examples above considered show the importance of considering the exact distribution of the estimator of R and not only the variance of the estimator.

4. The Confidence interval for *R* based on the exact distribution or the estimator

As discussed previously, none of the moments of W_n exists, and thus it is impossible to infer from the mean value $E(W_n)$ and variance $Var(W_n)$.
Cochran (1977), in order to construct the confidence intervals for R, uses a Normal distribution having asymptotic expected value R and Var (W_n).

The Fieller method refers to a general approach to obtain the confidence region for the ratio of the means in a Bivariate Normal distribution. Fieller (1940) introduced this method for the first time in his paper on the standardization of insulin. Subsequently, in another paper, the problem was described by the same author with more details (1954), and a new expression for the ratio of the means as a linear combination of rvs made relatively simple the computation of confidence intervals of the ratio. Fieller method has been used as a touchstone by several authors (Finney 1964, Rao 1965, Kendall and Stuart 1961), because of its importance in examining the general techniques for constructing confidence intervals using resampling techniques, such as the jackknife or bootstrapping. However, there are a few degenerate cases for which the confidence region is not an interval, so that practical interpretation of the results is impossible. In fact, the existence of a bounded $(1-\alpha)\%$ confidence interval for R with Fieller method is not always guaranteed. Gardiner at al. (2001) proved that the confidence interval is bounded if and only if the estimated $\bar{X}_2$ is significantly different from zero at level α.
In order to always obtain the existence of a bounded $(1-\alpha)\%$ confidence interval for R, Galeone (2007) proposed the following confidence intervals for R, that can be obtained by using the inverse cumulative density function of W_n, as followed:

$$P\{W_{\alpha/2} \le R \le W_{1-\alpha/2}\} = 1-\alpha \tag{2}$$

where $W_{\alpha/2}$ and $W_{1-\alpha/2}$ are the estimators (see Galeone et Pollastri, 2008) of $(\alpha/2)th$ and the $(1-\alpha/2)th$ quantile of the distribution of the r.v. W_n.

We consider a first simulation in which there are situations where Fieller method presents problems. Using the method proposed by Galeone is always possible to build confidence intervals for *R* We select *5000* samples from a Bivariate Normal r.v. having parameters $\mu_1 = 0.25$, $\mu_2 = 1.20$, $\sigma^2{}_1 = 9$, $\sigma_2^2 = 16$.

Tab. 2: Comparison between the Confidence intervals due to Fieller and the Confidence Intervals due to Galeone

$\rho = 0.3$			
n		Fieller method	Exact dist method
25	(1 – â)	0.394	0.926
	%ds	0.56	0.61
	Amp	-	5.047
50	(1 – â)	0.594	0.919
	%ds	0.34	0.54
	Amp	-	3.00
100	(1 – â)	0.819	0.910
	%ds	0.23	0.52
	Amp	-	1.569
200	(1 – â)	0.863	0.896
	%ds	0.63	0.53
	Amp	-	0.728
400	(1 – â)	0.897	0.897
	%ds	0.48	0.48
	Amp	0.433	0.432
800	(1 – â)	0.901	0.901
	%ds	0.51	0.51
	Amp	0.290	0.290

In table 3 are reported the results of a simulation in which we compare the confidence interval proposed by Fieller with the confidence intervals based on the exact distribution proposed by Galeone. The parameters of the Bivariate Normal are $\mu_1 = 12$; $\mu_2 = 10$; $\sigma^2{}_1 = 1$; $\sigma_2^2 = 1.5$. The confidence probability is fixed $(1-\alpha) = 0.90$.

Tab. 3: Comparison between the size of the Confidence intervals due to Fieller and the Confidence Intervals due to Galeone

ρ	Fieller method			Exact distr method		
	$(1-\hat{\alpha})$	Amp	%ds	$(1-\hat{\alpha})$	Amp	%ds
0	0.911	0.021	0.48	0.901	0.021	0.51
0.3	0.903	0.017	0.51	0.901	0.017	0.50
0.6	0.895	0.019	0.49	0.897	0.019	0.50
0.9	0.893	0.008	0.50	0.896	0.008	0.50
-0.3	0.911	0.023	0.52	0.911	0.023	0.52
-0.6	0.903	0.026	0.52	0.903	0.025	0.52
-0.9	0.907	0.011	0.43	0.905	0.011	0.42

$(1-\hat{\alpha})$: coverage probability
Amp: average width of the intervals
%ds: degree of symmetry of miscoverage

It is possible to observe that the averages width of the intervals of the two methods are very closed.

5. Conclusions

The present paper shows the importance of considering the real distribution of the estimator of *R.*

In fact the distribution of the estimator of the ratio of two means or of two totals is useful also to build the confidence intervals for the parameter *R*. The method proposed by Cochran based on the Normal distribution does not consider the exact distribution of the estimator but only the asymptotic variance of the estimator. Here we show that we can have the same asymptotic variance of the estimator of the ratio of two totals in situations where the estimator has different distribution.

The method, proposed by Galeone, here considered always exists while the method of Fieller, generally used in order to build confidence intervals, sometimes presents problems. When the two methods exist, they supply confidence intervals very closed.

REFERENCES

Aroian L.A. (1986) The distribution of the quotient of two correlated random variables. *Proceedings of the Am. Stat. Ass. Business and Economic Section.*

Brentari E. (1990), *Asimmetria e misure di Asimmetria*, Giappichelli Ed., Torino.

Cochran W.C. (1997), *Sampling Techniques*, John Wiley and Sons, New York.

.Fieller, E. C. (1932). "The distribution of the index in a Normal Bivariate population." *Biometrika* **24**(3/4): 428-440.

Fieller, E. C. (1954). "Some Problems in Interval Estimation*." Journal of the Royal Statistical Society*. Series B (Methodological) **16**(2): 175-185.

Frosini, B. V. (1970). "La stima di un quoziente nei grandi campioni." *Giornale degli economisti e annali di economia* : 381-400.

Galeone C. (2007) On the ratio of two Normal r.v., jointly distributed as a Bivariate Normal, *PhD Thesis, Università di Milano-Bicocca*.

Galeone C., Pollastri A. (2008) Estimation of the quantiles of the ratio of two correlated Normals, *XLIV Riunione scientifica della Società Italiana di Statistica.*

Geary, R. C. (1930). "The frequency distribution of the quotient of two Normal variates." *J Royal Stat Society*, **93**(3): 442-446.

Kotz S., Balakrishnan N., Johnson N.L. (2000) *Continuous Multivariate Distributions,* Wiley, New York.

Oksoy D., Aroian L.A. (1986) Computational Techniques and examples of the density and the distribution of the quotient of two correlated normal variables, *Proceeding of the American Statistical Association Business an Economic Section.*

Marsaglia, G. (1965). "Ratios of Normal variables and ratio of sums of Uniform variables." *J Am Statist Ass* **60**(309): 193-204.

Marsaglia, G. (2006). "Ratios of Normal variables." *Journal of Statistical Software* **16**(4).