

Estimating the ratio of means

EDWARD OMEY
KULeuven@CampusBrussels
BELGIUM

MEITNER CADENA
DECE, Universidad de las Fuerzas Armadas
ECUADOR

Abstract: Confidence intervals for ratio of means for large paired and unpaired samples with finite variance, obtained by applying the central limit theorem and Cramér-Wold device, are given. Also, these intervals for ratios are obtained under infinite variance and when considering independent populations, by using stable distributions. Numerical illustrations by considering problems typically presented in practice are given.

Key-Words: Ratio of means, confidence interval, large sample, stable distribution, Cramér-Wold device

Received: February 2, 2020. Revised: August 1, 2020. Accepted: August 17, 2020. Published: September 13, 2020.

1 Introduction

It is often of interest to study the ratio of means in the place of the difference of the means. Researchers are often interested in the ratio of two measured quantities. An important example is the body mass index (BMI). The BMI or Quetelet index is a value derived from the mass (weight) and height of an individual. It is defined as the body mass (in kilograms) divided by the square of the body height (in meters). See for example [4]. In finance, an important ratio is the so-called Sharpe ratio. It is the ratio of the excess expected return of an investment to its return volatility or standard deviation. This ratio introduced in 1966 [22] and revised in 1994 [23] by Sharpe has become popular to adjust return rates of investments for risk, allowing the understanding of how the return of investment compares to its risk. In this way, higher Sharpe ratios would mean investments would generate better risk-adjusted returns. See e.g. [13, 16, 17].

In accountancy, accountants use a lot of financial ratios to compare companies with a benchmark, for instance for predicting failures. See e.g. [2]. In an earlier paper [19], authors studied central limit theorems for the coefficient of variation and for the ratio of variances for dependent and for independent samples.

More applications of ratios of means are recently found in a diversity of fields. For instance, in environment we have the called Arctic amplification which is most often expressed as the quotient between the mean change in Arctic temperatures and the global change over a period of analysis [21]; and, in occupa-

tional hygiene we have that a way for evaluating the effectiveness of implementing a work practice in reducing the concentration of an analyte in the urine of exposed workers is through the ratio between analyte concentrations obtained before and after implementing the work practice [15].

Regarding finite variance and large samples, typical approaches for building confidence intervals for ratios of means of independent populations have been based on methods as Fieller's Theorem [9, 10, 11], Taylor series [20] and bootstrap [6]. We tackle independent as well as dependent populations by applying the central limit theorem and Cramér-Wold device [3].

Regarding infinite variance and large samples, this fact matters because such a situation often occurs in many fields like meteorology, health care, finance and insurance. However, for the best of our knowledge, there is no proposals for building confidence intervals when ratios of means are required. We tackle this problem by using stable distributions if populations are independent.

In this paper we study ratios of the mean in the case of large paired (dependent) or large unpaired (independent) samples. In the paper we assume that moments exist whenever needed. In the third section we give some more comments in the case that the variance of the underlying distribution(s) is not finite. The last section presents concluding remarks, including future applications.

2 The ratio of two means

In what follows $N(\mu, \sigma^2)$ denotes the normal distribution with mean $\mu \in \mathbb{R}$ and variance $\sigma^2 (> 0)$.

2.1 Unpaired samples

Let X, X_1, X_2, \dots, X_n denote independent identically distributed (i.i.d.) random variables (r.v.s) with $EX = \mu_1$ and $Var(X) = \sigma_1^2$. Independent of the X , let Y, Y_1, Y_2, \dots, Y_m denote i.i.d. r.v. with $EY = \mu_2 \neq 0$ and $Var(Y) = \sigma_2^2$. This case was treated, among others, by Mahmoudi et al. [18], see also [24]. Let \bar{X} and \bar{Y} denote the sample means.

As $n, m \rightarrow \infty$, by the usual Central Limit Theorem (CLT) we have

$$\begin{aligned} \sqrt{n} \frac{\bar{X} - \mu_1}{\sigma_1} &\implies Z_1 \sim N(0, 1), \\ \sqrt{m} \frac{\bar{Y} - \mu_2}{\sigma_2} &\implies Z_2 \sim N(0, 1). \end{aligned}$$

Assume now that $n \sim \lambda m$ for some $\lambda > 0$. In this case we have

$$\sqrt{n} \frac{\bar{Y} - \mu_2}{\sigma_2} \implies Z_3 := \sqrt{\lambda} Z_2 \sim N(0, \lambda),$$

where Z_1, Z_2 are independent. Because of the independence, for all $(a, b) \neq (0, 0)$ we have

$$\sqrt{n} \left(\frac{a\bar{X} - a\mu_1}{\sigma_1} + \frac{b\bar{Y} - b\mu_2}{\sigma_2} \right) \implies aZ_1 + bZ_3,$$

or

$$\sqrt{n} \left(\frac{a\sigma_2\bar{X} - a\sigma_2\mu_1 + b\sigma_1\bar{Y} - b\sigma_1\mu_2}{\sigma_2\sigma_1} \right) \implies aZ_1 + bZ_3.$$

The choice $a\sigma_2\mu_1 = -b\sigma_1\mu_2$ implies that

$$\sqrt{n} \left(\frac{a\sigma_2\bar{X} + b\sigma_1\bar{Y}}{\sigma_2\sigma_1} \right) \implies aZ_1 + bZ_3.$$

Choosing $a = \sigma_1$ we arrive at

$$\sqrt{n} \left(\bar{X} - \frac{\mu_1}{\mu_2} \bar{Y} \right) \implies \sigma_1 Z_1 - \sigma_2 \frac{\mu_1}{\mu_2} Z_3.$$

Setting $\mu = \mu_1/\mu_2$ we get at

$$\begin{aligned} \sqrt{n}(\bar{X} - \mu\bar{Y}) \\ \implies \sigma_1 Z_1 - \sigma_2 \mu Z_3 =: Z_4 \sim N(0, \sigma_1^2 + \sigma_2^2 \mu^2 \lambda), \end{aligned}$$

or equivalently

$$\sqrt{n} \frac{\bar{X} - \mu\bar{Y}}{\sqrt{\sigma_1^2 + \sigma_2^2 \mu^2 \lambda}} \implies Z_5 \sim N(0, 1),$$

or

$$\frac{\bar{X} - \mu\bar{Y}}{\sqrt{\frac{1}{n}\sigma_1^2 + \frac{1}{m}\sigma_2^2\mu^2}} \implies Z_5 \sim N(0, 1). \quad (1)$$

We can replace σ_1^2 by $s_1^2 = \sum_{i=1}^n (X_i - \bar{X})^2 / (n-1)$, σ_2^2 by $s_2^2 = \sum_{i=1}^m (Y_i - \bar{Y})^2 / (m-1)$, and μ by $\hat{\mu} = \bar{X}/\bar{Y}$ with $\bar{Y} \neq 0$ to obtain

$$A = \frac{\bar{X} - \mu\bar{Y}}{\sqrt{\frac{1}{n}s_1^2 + \frac{1}{m}s_2^2\hat{\mu}^2}} \implies Z_5 \sim N(0, 1). \quad (2)$$

Formula (2) leads to confidence statements for μ of the form, for a confidence level $0 < \alpha < 1$:

$$\mu = \frac{\bar{X} \pm z_{\nu/2} \sqrt{\frac{1}{n}s_1^2 + \frac{1}{m}s_2^2\hat{\mu}^2}}{\bar{Y}}, \quad (3)$$

where z_ν is the $\nu \times 100$ th percentile of the standard normal distribution. To test $H_0 : \mu = \mu_0$, we can use the test statistic

$$A_0 = \frac{\bar{X} - \mu_0\bar{Y}}{\sqrt{\frac{1}{n}s_1^2 + \frac{1}{m}s_2^2\mu_0^2}},$$

which under H_0 has an asymptotic standard normal distribution as pointed out in [2].

Remark 1

- 1) If $\sigma_1^2 = \sigma_2^2 = \sigma^2$ we can estimate σ^2 by the pooled variance s_p^2 and then [2], [3] can be replaced by

$$\begin{aligned} A &= \frac{\bar{X} - \mu\bar{Y}}{s_p \sqrt{\frac{1}{n} + \frac{1}{m}\hat{\mu}^2}} \implies Z_5 \sim N(0, 1), \\ \mu &= \frac{\bar{X} \pm z_{\nu/2} s_p \sqrt{\frac{1}{n} + \frac{1}{m}\hat{\mu}^2}}{\bar{Y}}. \end{aligned}$$

- 2) If also $n = m$, we find that

$$\begin{aligned} A &= \sqrt{n} \frac{\bar{X} - \mu\bar{Y}}{s_p \sqrt{1 + \hat{\mu}^2}} \implies Z_5 \sim N(0, 1), \\ \mu &= \frac{\bar{X} \pm z_{\nu/2} s_p n^{-1/2} \sqrt{1 + \hat{\mu}^2}}{\bar{Y}}. \end{aligned}$$

2.1.1 Application

Consider the blood pressure measurements (30 sec pulse) collected in the National Health and Nutrition Examination Survey, US, corresponding to 2017-2018. These measurements are taken for males (X)

and females (Y). These data without missing values concern 3301 observations for males and 3441 for females. Members of these samples are represented by X_i and Y_i , respectively. We assume that X_i and X follow the same distribution, and in a similar way for Y_i and Y . Also, we assume that X_i and X_j , $i \neq j$, are independent from each other, and the same assumption between Y_i and Y_j , $i \neq j$. Further, we assume that X and Y are independent from each other. Finally, for X as well as for Y , we admit that these r.v.s have at least moment 2.

We are interested on behavior of $\mu = \mu_m / \mu_f$ with μ_m the expected blood pressure for male and μ_f the expected blood pressure for female. To this aim, we formulate the null hypothesis $H_0: \mu = 1$ and go to build the confidence interval given in (3) at $\nu = 5\%$.

From those data, we get

$$\begin{aligned} \hat{\mu}_m &= 72.45198 \\ \hat{\mu}_f &= 74.99506 \\ s_m^2 &= 155.4878 \\ s_f^2 &= 148.0352. \end{aligned}$$

These inputs then give the confidence interval $[0.9582443; 0.9739359]$. This fact therefore means that H_0 is rejected at $\nu = 5\%$. Further, this interval would suggest that $\mu < 1$.

2.2 Paired samples

In this section, let (X, Y) , (X_1, Y_1) , (X_2, Y_2) , ..., (X_n, Y_n) denote i.i.d. random vectors with $EX = \mu_1$, $EY = \mu_2$, $Var(X) = \sigma_1^2$, $Var(Y) = \sigma_2^2$ and with $\rho = \rho(X, Y)$. Now consider linear combinations of the form $L = aX + bY$. Clearly we have

$$\begin{aligned} \mu_L &= EL = a\mu_1 + b\mu_2, \\ \sigma_L^2 &= Var(L) = a^2\sigma_1^2 + b^2\sigma_2^2 + 2ab\rho\sigma_1\sigma_2. \end{aligned}$$

The CLT for L shows that

$$\sqrt{n} \frac{\bar{L} - \mu_L}{\sigma_L} \implies Z \sim N(0, 1),$$

or equivalently that

$$\begin{aligned} \sqrt{n}(a\bar{X} + b\bar{Y} - a\mu_1 - b\mu_2) \\ \implies \sigma_L Z \\ \sim N(0, a^2\sigma_1^2 + b^2\sigma_2^2 + 2ab\rho\sigma_1\sigma_2), \end{aligned} \quad (4)$$

for all $(a, b) \neq (0, 0)$. By the Cramér-Wold device, it follows that

$$\sqrt{n}(\bar{X} - \mu_1, \bar{Y} - \mu_2) \implies (U, V), \quad (5)$$

where (U, V) has a bivariate normal law with means $EU = EV = 0$ and with variance-covariance matrix

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}.$$

Taking $a = \mu_2$ and $b = -\mu_1$ relations (4), (5) show that

$$\begin{aligned} \sqrt{n}(\mu_2\bar{X} - \mu_1\bar{Y}) \\ \implies \mu_2U - \mu_1V := Q \\ \sim N(0, \mu_2^2\sigma_1^2 + \mu_1^2\sigma_2^2 - 2\mu_1\mu_2\rho\sigma_1\sigma_2). \end{aligned}$$

Note that $Var(Q) = Var(\mu_2U - \mu_1V)$. Setting $\mu = \mu_1/\mu_2$, we also have

$$\sqrt{n}(\bar{X} - \mu\bar{Y}) \implies R \sim N(0, \sigma_1^2 + \mu^2\sigma_2^2 - 2\mu\rho\sigma_1\sigma_2),$$

or

$$\sqrt{n} \frac{\bar{X} - \mu\bar{Y}}{\sqrt{\sigma_1^2 + \mu^2\sigma_2^2 - 2\mu\rho\sigma_1\sigma_2}} \implies Z \sim N(0, 1).$$

Replacing the unknown quantities by their respective estimates, we obtain that

$$\sqrt{n} \frac{\bar{X} - \mu\bar{Y}}{\sqrt{s_1^2 + \hat{\mu}^2 s_2^2 - 2\hat{\mu}r s_1 s_2}} \implies Z \sim N(0, 1).$$

Now we get confidence statements of the form

$$\mu = \frac{\bar{X} \pm z_{\nu/2} \frac{1}{\sqrt{n}} \sqrt{s_1^2 + \hat{\mu}^2 s_2^2 - 2\hat{\mu}r s_1 s_2}}{\bar{Y}}. \quad (6)$$

Remark 2

1) If $\sigma_1 = \sigma_2 = \sigma$ we can simplify and find

$$\sqrt{n}(\bar{X} - \mu\bar{Y}) \implies N(0, \sigma^2(1 + \mu^2 - 2\mu\rho)).$$

2) If also $\rho = 0$, then we find

$$\sqrt{n}(\bar{X} - \mu\bar{Y}) \implies N(0, \sigma^2(1 + \mu^2))$$

2.2.1 Application

Consider the height (X) and weight (Y) measurements (cm and kg) collected in the National Health and Nutrition Examination Survey, US, corresponding to 2017-2018. For females, these data without missing values concern 4108 observations (X, Y) . We assume that (X_i, Y_i) and (X_j, Y_j) , $i \neq j$, are independent. Also, we assume X_i follows the same distribution than X , and for Y_i a similar hypothesis.

We are interested on to evaluate behavior of the Body Mass Index (BMI) which is calculated as weight

in kilograms divided by height in meters squared. To this aim, we formulate the null hypothesis H_0 : $BMI < 30$, because indexes greater or equal than 30 are a concern of health. To evaluate such a hypothesis, we go to build the confidence interval given in (6) at $\nu = 5\%$, adapting it for testing inequalities.

From those data, we get

$$\begin{aligned} \bar{X} &= 65.28055 \\ \bar{Y} &= 2.335582 \\ s_1^2 &= 808.1152 \\ s_2^2 &= 0.273935 \\ r &= 0.728730. \end{aligned}$$

These inputs then give the confidence interval $[0.00000; 28.17132]$. This fact therefore implies that H_0 is not rejected at $\nu = 5\%$.

3 Domains of attraction

In this section we briefly discuss the case where the variances are possibly not finite.

Hereafter RV_α denotes the set of regularly varying functions f with index $\alpha \in \mathbb{R}$, i.e. $f(tx)/f(x) \rightarrow t^\alpha$ as $x \rightarrow \infty$ for all $t > 0$. If $\alpha = 0$, such functions are the so-called slowly varying functions.

3.1 Univariate case

Suppose that the real r.v. X is in the domain of attraction of a stable law $U(\alpha)$ with exponent α where $1 < \alpha \leq 2$. Notation: $X \in DA(U(\alpha))$. This means that we can find a sequence of positive numbers (a_n) so that

$$n \frac{\bar{X} - EX}{a_n} \Rightarrow U(\alpha). \quad (7)$$

In the case where $\alpha = 2$, $U(2)$ has a normal distribution. To characterize (7), we consider the truncated moment function $V_X(x) = \int_{-x}^x y^2 dF_X(y)$. We have the following result.

Theorem 1 (Chapter XVII.5 in [8])

(i) Suppose that $\alpha = 2$ and that $F_X(x)$ is not concentrated at one point. We have $X \in DA(U(2))$ if and only if $V_X(x) \in RV_0$.

(ii) Suppose that $1 < \alpha < 2$ and that

$$\frac{P(X > x)}{P(|X| > x)} \rightarrow p, \quad \frac{P(X < -x)}{P(|X| > x)} \rightarrow q.$$

Then $X \in DA(U(\alpha))$ if and only if $V_X(x) \in RV_{2-\alpha}$.

(iii) In both (i) and (ii) we have

$$\frac{x^2 P(|X| > x)}{V_X(x)} \rightarrow \frac{2 - \alpha}{\alpha}.$$

Feller's theorem also shows that in (7) we can take a_n in such a way that $nV_X(a_n) \sim a_n^2$. Now define $W(x) = x^2/V_X(x) \in RV_\alpha$. Since $\alpha > 0$, $W(x) \sim W^\circ(x)$ where $W^\circ(x) \in RV_{1/\alpha}$ is increasing. Then we find that $n \sim W(a_n)$ and $a_n \sim W^\circ(n)$. Among others it follows that (a_n) is regularly varying with index $1/\alpha$. The finite variance case corresponds to the case where $V_X(x) \uparrow EX^2 < \infty$, and then we can choose a_n so that $a_n^2 \sim nEX^2$.

3.2 Independent random variables

Now suppose that $X \in DA(U(\alpha))$ and $Y \in DA(V(\beta))$ where $1 < \alpha, \beta \leq 2$, and where X and Y are independent. This means that there exist sequences $(a_n), (b_n)$ so that

$$n \frac{\bar{X} - \mu_1}{a_n} \Rightarrow U(\alpha), \quad m \frac{\bar{Y} - \mu_2}{b_m} \Rightarrow V(\beta).$$

If $m \sim \lambda n, \lambda > 0$, we have

$$n \frac{\bar{X} - \mu_1}{a_n} \Rightarrow U(\alpha), \quad n \frac{\bar{Y} - \mu_2}{b_n} \Rightarrow c(\lambda)V(\beta),$$

where $c(\lambda)$ is a positive constant depending on λ .

If $b_n = o(a_n)$, then for all $(a, b) \neq (0, 0)$ we have

$$n \frac{a\bar{X} - a\mu_1 - b\bar{Y} + b\mu_2}{a_n} \Rightarrow aU(\alpha).$$

Choosing $a = 1, b = \mu = \mu_1/\mu_2$ we get that

$$n \frac{\bar{X} - \mu\bar{Y}}{a_n} \Rightarrow U(\alpha).$$

If $a_n = o(b_n)$ we have a similar result.

Next suppose that $V_X(x) \sim cV_Y(x)$ where $c > 0$. This implies that $\alpha = \beta$ and that we can take $a_n = b_n$. Now we find that for all $(a, b) \neq (0, 0)$ we have

$$n \frac{a\bar{X} - a\mu_1 - b\bar{Y} + b\mu_2}{a_n} \Rightarrow aU(\alpha) - bc(\lambda)V(\beta).$$

Choosing $a = 1, b = \mu = \mu_1/\mu_2$ we get that

$$n \frac{\bar{X} - \mu\bar{Y}}{a_n} \Rightarrow U(\alpha) - \mu c(\lambda)V(\beta).$$

From some of the previous results, we have the following one, which will be used later.

Proposition 1 Let X and Y be r.v.s belonging to domains of attraction of stable laws $U(\alpha)$ and $V(\beta)$ with $1 < \alpha, \beta < 2$, respectively. If $V_X(x) \sim cV_Y(x)$ for some $c > 0$, where $V_Z(x) = \int_{-x}^x y^2 dF_Z(y)$, then $c = 1$.

Proof. Because X and Y belong to domains of attraction of stable laws $U(\alpha)$ and $V(\beta)$, respectively, there exist sequences a_n and b_m such that

$$n \frac{\bar{X} - \mu_1}{a_n} \implies U(\alpha), m \frac{\bar{Y} - \mu_2}{b_m} \implies V(\beta),$$

as $n \rightarrow \infty$ and $m \rightarrow \infty$, respectively. Hence, Theorem 1 implies that

$$nV_X(a_n) \sim a_n^2 \quad \text{and} \quad mV_Y(b_m) \sim b_m^2,$$

as $n \rightarrow \infty$ and $m \rightarrow \infty$, respectively. On the other hand, by the hypothesis $V_X(x) \sim cV_Y(x)$, we can take $a_n = b_n$. Hence, we deduce that

$$a_n^2 \sim nV_X(a_n) \sim ncV_Y(a_n) = ncV_Y(b_n) \sim cb_n^2,$$

i.e. $c = 1$. The proposition then follows. ■

Further, let F and G be the distributions functions associated to X and Y , respectively. Assume there exist a slowly varying function L_X at infinite and constants $c_{X,1}, c_{X,2} \geq 0, c_{X,1} + c_{X,2} > 0$, such that

$$\begin{aligned} x^\alpha(1 - F(x)) &= (c_{X,1} + o(1))L_X(x) \\ x^\alpha F(-x) &= (c_{X,2} + o(1))L_X(x), \end{aligned} \quad (8)$$

as $x \rightarrow \infty$, and there exist a slowly varying function L_Y at infinite and constants $c_{Y,1}, c_{Y,2} \geq 0, c_{Y,1} + c_{Y,2} > 0$, such that

$$\begin{aligned} x^\alpha(1 - G(x)) &= (c_{Y,1} + o(1))L_Y(x) \\ x^\alpha G(-x) &= (c_{Y,2} + o(1))L_Y(x), \end{aligned} \quad (9)$$

as $x \rightarrow \infty$.

Consider the following result proved by Ibragimov and Linnik [12], see also e.g. [1].

Theorem 2 (Theorem 2.6.5 in [12]) If F belongs to the domain of attraction of a stable law $U(\alpha)$ with $0 < \alpha < 2$, and satisfies (8) for some constants c_1 and c_2 and L being slowly varying function, then the logarithm of the characteristic function of $1 - F$ as $t \rightarrow 0$ is

$$it\delta - c|t|^\alpha L(|t|^{-1}) \left[1 - i\beta \frac{t}{|t|} \tan\left(\frac{\alpha\pi}{2}\right) \right] + o(|t|^\alpha L(|t|^{-1})),$$

where

$$\begin{aligned} \beta &= \frac{c_1 - c_2}{c_1 + c_2}, \\ c &= \Gamma(1 - \alpha)(c_1 + c_2) \cos\left(\frac{\alpha\pi}{2}\right), \\ \delta &= E[X]. \end{aligned}$$

A consequence of this result is the following one.

Proposition 2 Let X_1, \dots, X_n be a sample of r.v.s of a r.v. X , which follows a distribution function F . Let Y_1, \dots, Y_m be a sample of r.v.s of a r.v. Y , which follows a distribution function G . Assume that X and Y are independent and that they satisfy the hypothesis of Proposition 1 with $E(Y) \neq 0$. Assume that F and G satisfy (8) and (9) for some constants $c_{X,1}, c_{X,2}$ and $c_{Y,1}, c_{Y,2}$ and L_X and L_Y being slowly varying functions, respectively. If $n = \lambda m$ for some $\lambda > 0$ and taking $\mu = \mu_1/\mu_2$ with $\mu_1 = E(X)$ and $\mu_2 = E(Y)$, then we have, for some stable law W ,

$$n \frac{\bar{X} - \mu\bar{Y}}{a_n} \implies W(\alpha) = W(\alpha, \beta_W, c_W, \delta_W),$$

where

$$\beta_W = \frac{c_X \beta_X + c_Y \left|\frac{\mu}{\lambda}\right|^\alpha \beta_Y}{c_X + c_Y \left|\frac{\mu}{\lambda}\right|^\alpha},$$

$$c_W = \left(c_X + c_Y \left|\frac{\mu}{\lambda}\right|^\alpha \right)^{1/\alpha},$$

$$\delta_W = 0,$$

with $c_X = \Gamma(1 - \alpha)(c_{X,1} + c_{X,2}) \cos\left(\frac{\alpha\pi}{2}\right)$, $\beta_X = (c_{X,1} - c_{X,2})/(c_{X,1} + c_{X,2})$, $c_Y = \Gamma(1 - \alpha)(c_{Y,1} + c_{Y,2}) \cos\left(\frac{\alpha\pi}{2}\right)$ and $\beta_Y = (c_{Y,1} - c_{Y,2})/(c_{Y,1} + c_{Y,2})$, which describe the logarithm of the characteristic function of W given by, as $t \rightarrow 0$,

$$i\delta_W t - c_W^\alpha |t|^\alpha \left[1 - i\beta_W \frac{t}{|t|} \tan\left(\frac{\alpha\pi}{2}\right) \right] + o(|t|^\alpha).$$

Proof. Let X and Y be r.v.s satisfying $V_X(x) \sim cV_Y(x)$ and belonging to domains of attraction $DA(U(\alpha))$ and $DA(V(\alpha))$, respectively. Hence and by hypothesis, there exist sequences a_n and a_m such that

$$T_{X,n} = n \frac{\bar{X} - \mu_1}{a_n} \implies U(\alpha),$$

and

$$T_{Y,m} = m \frac{\bar{Y} - \mu_2}{a_m} \implies V(\alpha),$$

as $n \rightarrow \infty$ and $m \rightarrow \infty$. Hence, Theorem 2 indicates that when $t \rightarrow 0$, the logarithm of the characteristic function of $T_{X,n}$ is, after straightforward computations, as $n \rightarrow \infty$,

$$\begin{aligned} &\log \varphi_{T_{X,n}}(t) \\ &= -nc_X \left| \frac{t}{a_n} \right|^\alpha L_X \left(\left| \frac{t}{a_n} \right|^{-1} \right) \\ &\quad \times \left[1 - i\beta_X \frac{t}{|t|} \tan\left(\frac{\alpha\pi}{2}\right) \right] \\ &\quad + o \left(n \left| \frac{t}{a_n} \right|^\alpha L_X \left(\left| \frac{t}{a_n} \right|^{-1} \right) \right), \end{aligned}$$

where $c_X = \Gamma(1 - \alpha)(c_{X,1} + c_{X,2}) \cos\left(\frac{\alpha\pi}{2}\right)$ and $\beta_X = (c_{X,1} - c_{X,2}) / (c_{X,1} + c_{X,2})$. Because $T_{X,n}$ converges to $U(\alpha)$, then $\log \varphi_{T_{X,n}}(t)$ converges to the logarithm of the characteristic function of $U(\alpha)$ as $n \rightarrow \infty$. Hence we deduce, as $n \rightarrow \infty$,

$$\frac{nL_X\left(\left|\frac{t}{a_n}\right|^{-1}\right)}{a_n^\alpha} \rightarrow 1. \quad (10)$$

In a similar way, we deduce, as $m \rightarrow \infty$,

$$\frac{mL_Y\left(\left|\frac{t}{a_m}\right|^{-1}\right)}{a_m^\alpha} \rightarrow 1. \quad (11)$$

Next, we analyze $T_{W,n} = n(\bar{X} - \mu\bar{Y})/a_n$. After straightforward computations, we get that, as $t \rightarrow 0$, the logarithm of its characteristic function is

$$\begin{aligned} \log \varphi_{T_{W,n}}(t) &= -nc_X \left|\frac{t}{a_n}\right|^\alpha L_X\left(\left|\frac{t}{a_n}\right|^{-1}\right) \\ &\quad \times \left[1 - i\beta_X \frac{t}{|t|} \tan\left(\frac{\alpha\pi}{2}\right)\right] \\ &\quad + o\left(n \left|\frac{t}{a_n}\right|^\alpha L_X\left(\left|\frac{t}{a_n}\right|^{-1}\right)\right) \\ &\quad - nc_Y \left|\frac{t\mu}{\lambda a_n}\right|^\alpha L_Y\left(\left|\frac{t\mu}{\lambda a_n}\right|^{-1}\right) \\ &\quad \times \left[1 - i\beta_Y \frac{t}{|t|} \tan\left(\frac{\alpha\pi}{2}\right)\right] \\ &\quad + o\left(n \left|\frac{t\mu}{\lambda a_n}\right|^\alpha L_Y\left(\left|\frac{t\mu}{\lambda a_n}\right|^{-1}\right)\right). \end{aligned}$$

Taking $n \rightarrow \infty$ and then applying (10) and (11) give

$$\begin{aligned} \log \varphi_{T_{W,n}}(t) &\rightarrow -\left(c_X + c_Y \left|\frac{\mu}{\lambda}\right|^\alpha\right) |t|^\alpha \\ &\quad \times \left[1 - i \frac{c_X \beta_X + c_Y \left|\frac{\mu}{\lambda}\right|^\alpha \beta_Y}{c_X + c_Y \left|\frac{\mu}{\lambda}\right|^\alpha} \frac{t}{|t|} \tan\left(\frac{\alpha\pi}{2}\right)\right] \\ &\quad + o(|t|^\alpha). \end{aligned}$$

The proposition then follows. ■

3.2.1 Application

In this application we evaluate the null hypothesis that μ is equal to a given μ_0 when data would not have variance but do mean. To this aim, we consider logarithmic variations of exchange rates r_t between US and Indonesia (US-I) and US and Malaysia (US-M), from first January 1994 to 18th May 2020. These logarithmic variations are expressed as $\log(r_t/r_{t-1})$.

To this aim, we assume the distribution function F for representing logarithmic variations of exchange rates related to US-I as

$$F(x) = \begin{cases} \frac{1}{2} \left(1 - \frac{x}{\sigma_X}\right)^{-\alpha} & \text{if } x < 0 \\ 1 - \frac{1}{2} \left(1 + \frac{x}{\sigma_X}\right)^{-\alpha} & \text{if } x \geq 0. \end{cases}$$

F is then continuous and

$$F'(x) = \begin{cases} \frac{\alpha}{\sigma_X} \left(1 - \frac{x}{\sigma_X}\right)^{-\alpha-1} & \text{if } x < 0 \\ \frac{\alpha}{\sigma_X} \left(1 + \frac{x}{\sigma_X}\right)^{-\alpha-1} & \text{if } x > 0, \end{cases}$$

and it is defined $F'(0) = \alpha/\sigma_X$. Then F is also continuous differentiable, i.e. the involved probability density function is continuous. Also, F is clearly regularly varying with tail indexes, see e.g. [14],

$$\begin{aligned} \lim_{x \rightarrow \infty} -\frac{\log(1 - F(x))}{\log x} &= \lim_{x \rightarrow \infty} -\frac{\log \frac{1}{2} \left(1 + \frac{x}{\sigma_X}\right)^{-\alpha}}{\log x} \\ &= \alpha \end{aligned}$$

and

$$\begin{aligned} \lim_{x \rightarrow \infty} -\frac{\log F(-x)}{\log x} &= \lim_{x \rightarrow \infty} -\frac{\log \frac{1}{2} \left(1 + \frac{x}{\sigma_X}\right)^{-\alpha}}{\log x} \\ &= \alpha \end{aligned}$$

On the other hand, we have

$$\begin{aligned} x^\alpha(1 - F(x)) &= \frac{1}{2} x^\alpha \left(1 + \frac{x}{\sigma_X}\right)^{-\alpha} \\ &= \frac{\sigma_X^\alpha}{2} [1 + o(1)] \end{aligned}$$

and

$$\begin{aligned} x^\alpha F(-x) &= \frac{1}{2} x^\alpha \left(1 + \frac{x}{\sigma_X}\right)^{-\alpha} \\ &= \frac{\sigma_X^\alpha}{2} [1 + o(1)] \end{aligned}$$

This means that $L_X(x) = \sigma_X^\alpha/2$ and $c_{X,1} = c_{X,2} = 1$. In particular, we get $\beta_X = 0$.

We define Y in a similar way to X , but taking σ_Y instead of σ_X .

Thus, W in Proposition 2 becomes

$$W\left(\alpha, 0, \kappa^{1/\alpha}, 0\right),$$

where

$$\kappa = \Gamma(1 - \alpha) \cos\left(\frac{\alpha\pi}{2}\right) \left[\sigma_X^\alpha + \sigma_Y^\alpha \left|\frac{1}{\mu}\right|^\alpha\right].$$

Note that this law W is symmetric. Hence, for a $\nu \times 100$ % of confidence, $W_{(1-\nu)/2}$ and $W_{1/2+\nu/2}$ satisfy $W_{(1-\nu)/2} = -W_{1/2+\nu/2}$. Then, a confidence interval for μ is

$$\frac{\bar{X} \pm W_{(1-\nu)/2} a_n / n}{\bar{Y}}.$$

In order to establish the parameters of W data are fitted, producing the estimates

$$\begin{aligned} \hat{\alpha}_{US-I} &= 1.740163811 \\ \hat{\sigma}_{US-I} &= 0.004037815 \\ \hat{\alpha}_{US-M} &= 1.827724618 \\ \hat{\sigma}_{US-M} &= 0.002430342 \end{aligned}$$

Hence, we take

$$\begin{aligned} \alpha &= 1.800000000 \\ \sigma_{US-I} &= 0.004037815 \\ \sigma_{US-M} &= 0.002430342 \end{aligned}$$

Under these assumptions, at $\nu = 5$ %, Table 1 presents hypothesis tests for some values μ_0 . These results imply that hypothesis like $\mu_0 = 1$ or $\mu_0 = 2$ may be not rejected, but the one $\mu_0 = 3$ may be rejected.

μ_0	Inferior limit of CI	Superior limit of CI
1	-2.371444	2.890368
2	-2.058195	2.577119
3	-1.988479	2.507403

Table 1: CI for some μ_0

4 Concluding remarks

- 1) In the paper we used a Central Limit Theorem to construct confidence statements for the ratio of means for large samples. We obtained a result for both the independent and the dependent case. Also in the infinite variance case, we proved a result. In the latter, more results are needed to construct attractive confidence statements.

- 2) In studying ratio's there are different approaches possible. Consider the following three econometric models:

$$\text{model 1: } Y = \tau X + \epsilon;$$

$$\text{model 2: } Y = (\tau + \epsilon)X;$$

$$\text{model 3: } Y = (\tau\sqrt{X} + \epsilon)\sqrt{X}.$$

When we do not take into account the error term ϵ , we find $Y/X = \tau$ in the three cases. However, when we use data and apply the least squares method, we find different estimates for τ :

$$\text{model 1: } \hat{\tau}_1 = \overline{XY}/\overline{X^2};$$

$$\text{model 2: } \hat{\tau}_2 = \overline{Y}/\overline{X};$$

$$\text{model 3: } \hat{\tau}_3 = \overline{Y}/\overline{X}.$$

We notice that the 3 models give a different estimate for τ which ideally represents the same ratio Y/X .

- 3) The case of dependency under infinite variance needs approaches for representing dependence structures. Copulas have been used for playing such a role, as for instance e.g. [5] and [7]. In a forthcoming paper we will present new alternatives for tackling confidence intervals for ratios of means under those conditions.
- 4) Further research is needed to estimate parameters in the dependent case/domain of attraction case. Also we plan to investigate ratio of variances, skewness and curtosis. Also we have a plan to apply our results in studying financial indicators like the RSI (relative strenght index).

Acknowledgements:

The authors thanks the reviewers for their valuable comments and suggestions which helped us to improve the paper.

References:

- [1] J. Aaronson and M. Denker, Characteristic functions of random variables attracted to 1-stable laws, *Ann. Probab.*, Vol. 26, No. 1, 1998, pp. 399–415.
- [2] W.H. Beaver, Financial Ratios As Predictors of Failure. *Journal of Accounting Research*, Vol. 4, 1966, pp. 71–111.
- [3] P. Billingsley, *Probability and Measure*, 2nd edition, Wiley, New York, 1986.
- [4] H. Blackburn and D. Jacobs, Commentary: Origins and evolution of body mass index (BMI): continuing saga, *International Journal of Epidemiology*, Vol. 43, No. 3, 2014, pp. 665–669.

- [5] A.C. Cebrian, M. Denuit and P. Lambert, Generalized Pareto Fit to the Society of Actuaries' Large Claims Database, *North American Actuarial Journal*, Vol. 7, No. 3, 2003, pp. 18–36.
- [6] B. Efron and R.J. Tibshirani, *An introduction to the bootstrap*, Chapman & Hall, Boca Raton, 1993.
- [7] H. Esmaeili and C. Klüppelberg, Parametric estimation of a bivariate stable Lévy process, *Journal of Multivariate Analysis*, Vol. 102, No. 5, 2011, pp. 918–930.
- [8] W. Feller, *An introduction to Probability Theory and Its Applications*, Vol. II, 2nd edition, Wiley, New York, 1971.
- [9] E.C. Fieller, The biological standardization of insulin, *Supplement to the Journal of the Royal Statistical Society*, Vol. 7, No. 1, 1940, pp. 1–64.
- [10] E.C. Fieller, A fundamental formula in the statistics of biological assays and some applications, *Quarterly Journal of Pharmacy and Pharmacology*, Vol. 17, 1944, pp. 117–123.
- [11] E.C. Fieller, Some problems in interval estimation, *J. R. Stat. Soc. B*, Vol. 16, No. 2, 1944, pp. 175–185.
- [12] I.A. Ibragimov and Y.V. Linnik, *Independent and stationary sequences of random variables*, English Trans, edited by J. F. C. Kingman, Wolters-Noordhoff Publishing, Groningen, The Netherlands, 1971.
- [13] C.L. Israelsen, A refinement to the Sharpe ratio and information ratio, *J. Asset Manag.*, Vol. 5, No. 6, 2005, pp. 423–427.
- [14] J. Karamata, Sur un mode de croissance régulière. Théorèmes fondamentaux, *Bull. Soc. Math. Fr.*, tome 61, 1933, pp. 55–62.
- [15] K. Krishnamoorthy, T. Mathew and Z. Xu, Comparison of Means of Two Lognormal Distributions Based on Samples with Multiple Detection Limits, *J. Occup. Environ. Hyg.*, Vol. 11, No. 8, 2014, pp. 538–546.
- [16] O. Ledoit and M. Wolf, Robust performance hypothesis testing with the Sharpe ratio, *J. Empir. Finance*, Vol. 15, No. 5, 2008, pp. 850–859.
- [17] A.W. Lo, The Statistics of Sharpe Ratios, *Financial Analysts Journal*, Vol. 58, No. 4, 2002, pp. 36–52.
- [18] M.R. Mahmoudi, J. Behdoodian and M. Maleki, Large sample inference about the ratio of means in two independent populations, *J. Stat. Theory and Appl.*, Vol. 16, No. 3, 2017, pp. 366–374.
- [19] E. Omeý and S. Van Gulck, Central limit theorems for variances and correlation coefficients, *Unpublished preprint HUBrussel*, 2008.
- [20] D. Polsky, H.A. Glick, R. Willke and K. Schulman, Confidence intervals for cost-effectiveness ratios: A comparison of four methods, *Health Economics*, Vol. 6, No. 3, 1997, pp. 243–252.
- [21] M.C. Serreze and R.G. Barry, Processes and impacts of Arctic amplification: A research synthesis, *Global Planet. Change*, Vol. 77, No. 1-2, 2011, pp. 85–96.
- [22] W.F. Sharpe, Mutual Fund Performance, *Journal of Business*, Vol. 39, No. 1, 1966, pp. 119–138.
- [23] W.F. Sharpe, The Sharpe Ratio, *J. Portf. Manag.*, Vol. 21, No. 1, 1994, pp. 49–58.
- [24] H. Wang and S.C. Chow, A practical approach for comparing means of two groups without equal variance assumption, *Stat. in Medicine*, Vol. 21, 2012, pp. 3137–3151.

Contribution of individual authors

The authors have equally contributed to the writing, editing and style of the paper. All authors have read and agreed to the published version of the manuscript.

Sources of funding for research

This research received no external funding.

Creative Commons Attribution License 4.0 (Attribution 4.0 International , CC BY 4.0)

This article is published under the terms of the Creative Commons Attribution License 4.0
https://creativecommons.org/licenses/by/4.0/deed.en_US