Journal of Computational Science &Engineering 36 (2018) 975-978

Available at http://www.asocse.org ISSN 1710-4068

Confidence interval on the ratio of two means

Cuijuan Ge, Xinmin Li*

School of Mathematics and Statistics, Qingdao University, Qingdao 266061, China

Abstract: **In this paper, we considered interval estimations on the ratio of two normal means. This method for interval estimation was proposed based on bootstrap. The simulation results show that the bootstrap approach has better prosperities under the large sample size.**

Keywords: Bootstrap; mean ratios; coverage probability

Received 8 June 2018, Revised 12 July 2018, Accepted 15 July 2018

*Corresponding Author: Xinmin Li, xmli@qdu.edu.cn

1. Introduction

Nowadays, since the inference on the ratio of means in complex cases often have limitations, it has become one of the hot issues. Using the inference on the ratio of means, we may know how large one quantity is relative to another one, or estimate the position at which a regression line intersects the abscissa (Miller[1] (1986), Buonaccorsi [2] (2001)). And the result of the research can be applied in many important fields, such as, bioassay(Finney[3](1978)), biological application test(Ionut[4](2009)), bioequivalence(Chow and Liu[5] (1992); Berger and Hsu[6](1996)) and image marginal test. In the theory of sampling distribution, a series of accurate sampling distributions such as F distribution and t distribution have been proposed, but the distributions are difficult to drive in many cases. If the variance of the population with second-order moments is positive, according to the central limit theorem, the population follows asymptotic normality when the sample size is larger, then normal distribution which is used commonly is easy to derive. In this article, we construct the confidence interval on the ratio of means by generalized fiducial inference and bootstrap under the normal assumption.

Let the population distribution be the normal distribution $N(\mu,\Sigma)$, where

$$\mu=(\mu_1,\mu_2)', \Sigma=\begin{pmatrix}\sigma_1^2 & \rho\sigma_1\sigma_2\\ \rho\sigma_1\sigma_2 & \sigma_2^2\end{pmatrix},(\rho\neq\pm1),$$

then the ratio of means $\theta=\frac{\mu_1}{\mu_2}$ is the interesting parameter. The problem of constructing confidence interval on the ratio of normal means is usually called Fieller problem.

Let $X_j=(X_1,X_2,\cdots;X_{n_j})$ be the sample from $N(\mu_j,\sigma_j^2)$ (j = 1,2), and the sample mean and variance is denoted as

$$\overline{X_j}=\frac{1}{n_j}\sum_{i}^{n_j}X_i,\ S_j^2=\frac{\sum_i^{n_j}(X_i-\overline{X_j})^2}{n_j},\ j=1,2.$$

When $\rho=0$, Fieller[7]constructed a confidence interval on the ratio of means $\theta=\frac{\mu_1}{\mu_2}$ as follows. Under the assumption of identical variance, an important statistic was given as

$$T=\frac{(\overline{Y}-\theta\overline{X})}{\sqrt{\left(\frac{1}{n_1}+\frac{\theta^2}{n_2}\right)S^2}}$$

where $S^2=\frac{n_1S_1^2+n_2S_2^2}{n_1+n_2}$. It is obvious that $\sqrt{\frac{(n_1+n_2-2)}{(n_1+n_2)}}T$ follows the Student's t distribution with (n_1+n_2-2) degrees of freedom. Then the confidence interval of θ for the given level $1-\alpha$ could be given as

$$\left\{\theta:\sqrt{\frac{(n_1+n_2-2)}{(n_1+n_2)}}|T|<t_{\frac{1-\alpha}{2}}\right\}.$$

In this paper, we propose the bootstrap method for constructing confidence intervals. The rest of the article is organized as follows. Two existing methods by Jack C. Lee[8](2004) and Guimei Zhao[9] (2014) are introduce in Section 2. In Section 3, a new method of constructing confidence intervals is proposed. In Section 4, Monte Carlo simulation study is used to test the advantages and disadvantages of methods. Some conclusions are given in Section 5.

2. Preliminary

Jack C. Lee[8](2004) developed the method of estimation on the ratio of means under heteroscedastic conditions based on Fieller[7]. For the given statistics X_j and S_j^2, we have

$$Z_j=\frac{\sqrt{n_j}\left(\overline{X_j}-\mu_j\right)}{\sigma_j}\sim N(\mathbf{0},1),\ U_j=\frac{n_jS_j^2}{\sigma_j^2}\sim\chi^2_{n_j-1},\ j=(1,2)$$

According to the generalized inference, they provided the generalized pivotal quantity

$$T_\theta=\frac{\overline{x_1}-Z_1s_1/\sqrt{U_1}}{\overline{x_2}-Z_2s_2/\sqrt{U_2}}=\frac{\overline{x_1}-T_1s_1/\sqrt{n_1-1}}{\overline{x_2}-T_2s_2/\sqrt{n_2-1}},\quad(1)$$

where $\overline{x_1}$, $\overline{x_2}$, s_1, s_2 are the observed values of $\overline{X_1}$, $\overline{X_2}$, S_1, S_2. And $T_J\sim t(n_j-1)$, j = (1,2). According to the numerical result, the confidence interval constructed by Jack C. Lee[8]not only has good coverage under the same variance, but also has good properties under the condition of heteroscedasticity.

In 2014, Zhao et al.[9]gave the generalized p-value

$$p(x,y)=P\left(\frac{\overline{x_1}-T_1s_1/\sqrt{n_1-1}}{\overline{x_2}-T_2s_2/\sqrt{n_2-1}}\leq\theta|\theta=\theta_0\right),$$

for the one side hypothesis

2018@ The American Computational Science Society. All rights reserved.

$H_0:\theta\leq\theta_0$ vs $H_1:\theta>\theta_0$,

According to the posterior predictive p-value (Meng[10]), the one side posterior predictive p-value was constructed to solve the Fieller problem.

Table 1 Confidence interval coverage probabilities comparison

n	(σ_1^2,σ_1^2)	method	(μ_1,μ_2) (1,1)	(1,3)	(1,7)	(1,10)
10	(1,3)	Lee	0.9824	0.9674	0.9600	0.9580
			(18.2806)	(0.7623)	(0.2179)	(0.1454)
		BCI	0.9522	0.9202*	0.9176*	0.9170*
			(16.9004)	(0.6830)	(0.1818)	(0.1233)
	(10,1)	Lee	0.9796	0.9586	0.9604	0.9540
			(13.5387)	(1.6415)	(0.6788)	(0.4543)
		BCI	0.9522	0.9202*	0.9176*	0.9170*
			(10.2890)	(1.3435)	(0.5487)	(0.3856)
50	(1,3)	Lee	0.9676	0.9476	0.9530	0.9476
			(2.5163)	(0.2214)	(0.0842)	(0.0574)
		BCI	0.9628	0.9534	0.9448	0.9486
			(2.4186)	(0.2195)	(0.0814)	(0.0560)
	(10,1)	Lee	0.9502	0.9154*	0.9564	0.9578
			(2.0600)	(0.6041)	(0.2637)	(0.1864)
		BCI	0.9404*	0.9466	0.9434*	0.9442
			(1.9982)	(0.5907)	(0.2497)	(0.1750)
100	(1,3)	Lee	0.9506	0.9532	0.9478	0.9504
			(1.0407)	(0.1574)	(0.0578)	(0.0406)
		BCI	0.9514	0.9492	0.9450	0.9474
			(0.9794)	(0.1524)	(0.0575)	(0.0397)
	(10,1)	Lee	0.9518	0.9472	0.9470	0.9492
			(1.3646)	(0.4229)	(0.1796)	(0.1247)
		BCI	0.9502	0.9448	0.9428*	0.9486
			(1.3480)	(0.4164)	(0.1767)	(0.1237)

3. Bootstrap Method

Recently, some authors considered the Fieller problem when $\rho=0$. Cheng[11](1995) applied the bootstrap method in computer simulation. Hwang[12] (1995) applied the bootstrap method to the Fieller problem. Choquet et al. [13](1998) considered the shortcomings of the method, and gave the calculation methods of one-sided and two-sided confidence. They also consider the application of bootstrap method to solve the problem. Next, we will show how to use bootstrap method to estimate the confidence interval on the ratio of means.

Let $X_1,X_2,\cdots;X_n$ be the sample from$N(\mu,\Sigma)$,then the unbiased estimation of parameter of (μ,Σ) are $\bar{X}$ and S, respectively, where $\bar{X}^{'}=\frac{1}{n}\sum_{h=1}^{n}X_h$, and $S=(s_{ij})=\frac{1}{n-1}\sum_{h=1}^{n}(X_h-\bar{X})(X_h-\bar{X})^{'}$, i=1,2, j=1,2.It is easy to know that

$$\bar{X}\sim N\left(\mu,\frac{1}{n}\Sigma\right),\ S\sim W\left(n-1,\frac{1}{n-1}\Sigma\right).$$

Then we could get the confidence interval by bootstrap method. The algorithm of bootstrap method shows as follows.

Step 1. For the samples from the populations. Calculate$\bar{X}$and S. Set $k:=1$ and$k_{max}=M$.

Step 2. Generate bootstrap samples$X^*=(X_1^*,X_2^*)\sim N(\bar{X},S)$.

Step 3. Calculate$\bar{X}_1^*,\bar{X}_2^*$, and compute $\theta_{B;k}=\frac{\bar{X}_1^*}{\bar{X}_2^*}$.

Go to Step 2, k=k+1.

Step 4. String M values of $\theta_{B;k}$,and sorting them in ascending order$\theta_{B;(1)},\ldots,\theta_{B;(M)}$.

For given $1-\alpha$, the $(\alpha/2)$th and $(1-\alpha/2)$th quantiles are $\theta_{B;\alpha/2*M}\theta_{B;(1-\alpha/2))M}$ respectively. Then $[\theta_{B;\alpha/2*M}\theta_{B;(1-\alpha/2))M}]$ is the $1-\alpha$ confidence interval.

4. Simulation Study

In this section, Monte Carlo simulation study is used to analyze the performance of the method described in Section.3. We compare bootstrap (BCI) method with Jack C. Lee[8]method. The criteria are coverage probabilities and average interval length. In the simulation study, 5000 Bootstrap samples to obtain the coverage probabilities of bootstrap text. Each sample

2018@ The American Computational Science Society. All rights reserved.

follows multivariable normal distribution. We chose the sample size (n) to be 10,50 and 100. The mean values were (1,1),(1,3), (1,7), and (1,10) ,respectively. The variances were (1,3) and (10,1) respectively. At the same time, the correlation coefficients were ±0.1, ±0.5, ±0.8 and 0 respectively.

Let $\alpha = 0.05$, the simulation study using the following steps:

Step 1. Generate sample from N(μ,Σ)

Step 2. Calculate the confidence interval by BCI and Lee, respectively. Where the simulation size is M=5000,

Step 3. Repeat steps 1-2 5000 times, calculate the coverages and mean length of confidence intervals.

Table 2 Confidence interval coverage probabilities($\sigma_1^2 = 1, \sigma_2^2 = 3$)

			ρ					
n	(μ_1, μ_2)	method	-0.8	-0.5	-0.1	0.1	0.5	0.8
10	(1,1)	Lee	0.9450	0.9520	0.9742	0.9880	0.9994	0.9998
		BCI	0.8962*	0.9584	0.9474	0.9360*	0.9782	0.9910
	(1,3)	Lee	0.9318*	0.9302*	0.9560	0.9802	0.9932	0.9894
		BCI	0.9208*	0.9432*	0.9330*	0.9420*	0.9770	0.9704
	(1,7)	Lee	0.9480	0.9480	0.9574	0.9662	0.9756	0.9702
		BCI	0.9318*	0.9604	0.9458	0.9500	0.9694	0.9540
	(1,10)	Lee	0.9536	0.9558	0.9598	0.9652	0.9648	0.9622
		BCI	0.9400*	0.9546	0.9464	0.9504	0.9656	0.9528
50	(1,1)	Lee	0.8838*	0.8932*	0.9544	0.9810	0.9986	0.9996
		BCI	0.9092*	0.9378*	0.9620	0.9378*	0.9726	0.9966
	(1,3)	Lee	0.9150*	0.9092*	0.9370*	0.9574	0.9738	0.9636
		BCI	0.9416*	0.9696	0.9542	0.9540	0.9868	0.9818
	(1,7)	Lee	0.9400*	0.9520	0.9558	0.9576	0.9644	0.9608
		BCI	0.9578	0.9740	0.9554	0.9476	0.9852	0.9762
	(1,10)	Lee	0.9462	0.9466	0.9540	0.9552	0.9548	0.9516
		BCI	0.9600	0.9578	0.9578	0.9558	0.9852	0.9716
100	(1,1)	Lee	0.9864*	0.8834*	0.9350*	0.9664	0.9990	0.9994
		BCI	0.9040*	0.9406*	0.9556	0.9296*	0.9724	0.9978
	(1,3)	Lee	0.9120*	0.8962*	0.9354*	0.9686	0.9876	0.9854
		BCI	0.9522	0.9692	0.9564	0.9456	0.9880	0.9852
	(1,7)	Lee	0.9314*	0.9210*	0.9390*	0.9640	0.9736	0.9660
		BCI	0.9606	0.9770	0.9544	0.9590	0.9882	0.9770
	(1,10)	Lee	0.9434*	0.9326*	0.9440	0.9536	0.9652	0.9548
		BCI	0.9644	0.9804	0.9592	0.9556	0.9882	0.9744

Table 3Confidence interval coverage probabilities($\sigma_1^2 = 10, \sigma_2^2 = 1$)

			ρ					
n	(μ_1, μ_2)	method	-0.8	-0.5	-0.1	0.1	0.5	0.8
10	(1,1)	Lee	0.9432*	0.9320*	0.9800	0.9828	0.9938	0.9986
		BCI	0.9390*	0.9364*	0.9386*	0.9386*	0.9378*	0.9344*
	(1,3)	Lee	0.9210*	0.9436*	0.9634	0.9772	0.9874	0.9874
		BCI	0.9144*	0.9200*	0.9208*	0.9160*	0.9146*	0.9228*
	(1,7)	Lee	0.9354*	0.9470	0.9586	0.9600	0.9740	0.9768
		BCI	0.9254*	0.9232*	0.9306*	0.9246*	0.9204*	0.9144*
	(1,10)	Lee	0.9430*	0.9444	0.9596	0.9660	0.9740	0.9692
		BCI	0.9198*	0.9248*	0.9220*	0.9166*	0.9186*	0.9176*
50	(1,1)	Lee	0.8956*	0.9256*	0.9554	0.9756	0.9942	0.9988
		BCI	0.9392*	0.9380*	0.9386*	0.9488	0.9472	0.9488
	(1,3)	Lee	0.8950*	0.9206*	0.9462	0.9642	0.9874	0.9910
		BCI	0.9494	0.9406*	0.9444	0.9426*	0.9472	0.9494

2018@ The American Computational Science Society. All rights reserved.

	(1,7)	Lee	0.9280*	0.9368*	0.9560	0.9536	0.9692	0.9762
		BCI	0.9452	0.9444	0.9464	0.9408*	0.9444	0.9434*
	(1,10)	Lee	0.9360*	0.9480	0.9508	0.9574	0.9642	0.9640
		BCI	0.9410*	0.9392*	0.9404*	0.9488	0.9500	0.9464
100	(1,1)	Lee	0.8684*	0.8854*	0.9384*	0.9612	0.9920	0.9988
		BCI	0.9470	0.9420*	0.9490	0.9460	0.9480	0.9508
	(1,3)	Lee	0.9038*	0.9178*	0.9434*	0.9560	0.9864	0.9880
		BCI	0.9438*	0.9380*	0.9466	0.9526	0.9462	0.9464
	(1,7)	Lee	0.9316*	0.9266*	0.9518	0.9568	0.9722	0.9724
		BCI	0.9500	0.9484	0.9444	0.9504	0.9474	0.9476
	(1,10)	Lee	0.9296*	0.9404*	0.9450	0.9526	0.9600	0.9664
		BCI	0.9404*	0.9462	0.9516	0.9478	0.9460	0.9438*

For $\rho \neq 0$, the results showed in Table 2 and Table 3 are the interval coverage probabilities. The results are demonstrated in the following.

(a)The change of correlation coefficient has great influence on Jack C. Lee[8]method and bootstrap method (BCI), confidence interval coverage probabilities below lower bound of accuracy limit.

(b)Jack C. Lee[8]method and bootstrap method (BCI) have anticonservative coverage probabilities performance when the variance difference between X_1 and X_2 is large. They have good coverage probabilities performance in positive correlation, large sample sizes and the variance difference betweenX_1 and $X_2$is small.

In conclusion, bootstrap method (BCI) is better than the Jack C. Lee[8]method.

5. Conclusion

We have proposed the method for interval estimation based on bootstrap. For $\rho=0$, when the sample size (n) is large, bootstrap method (BCI)has short average interval length. For $\rho \neq 0$, when the variance difference betweenX_1 and $X_2$is small, the sample sizesare large and the correlations are positive correlation, bootstrap method (BCI) has conservative coverage probabilities performance. In addition, bootstrap method (BCI) takes less time to simulation. Therefore, bootstrap method (BCI) could be applied in most cases.

References

[1] R.G. Miller, Technometrics. 29 (1986) 387-387.
[2] J.P. Buonaccorsi, Encyclopedia of Environmetrics. Wiley.(2001) 773-775.
[3] D.J. Finney, Grin, London (1978).
[4] I.F. Bebu. T. Mathew, Biometrical Journal. 51 (2009) 1047-1058.
[5] S.C. Chow, J.P. Liu, Marcel Dekker, Inc, New York. (1992).
[6] R.L. Berger, J.C. Hsu, Statistical Science. 11 (1996) 283-302.
[7] E.C. Fieller, Journal of the Royal Statistical Society. 16 (1954) 175-185.
[8] J.C. Lee, S.H. Lin. Journal of Statistical Planning & Inference, 123 (2004) 49-60.
[9] G.M. Zhao, X.Z. Xu, Statistics & Probability Letters. 95 (2014) 57-62.
[10] XLi. Annals of Statistics, 22 (1994) 1142-1160.
[11] R.C.H. Cheng, Simulation Conference Proceedings IEEE. (1995) 171-177.
[12] J.T. Hwang, StatisticaSinica. 5 (1995) 161-171.
[13] D. Choquet, P. L'Ecuyer, Acm Transactions on Modeling & Computer Simulation. 9 (1998) 326-348.

2018@ The American Computational Science Society. All rights reserved.