ISSN 1995-0802, Lobachevskii Journal of Mathematics, 2021, Vol. 42, No. 2, pp. 415–425. © Pleiades Publishing, Ltd., 2021.

Point Estimation for the Ratio of Medians of Two Independent Log-Normal Distributions

Lapasrada Singhasomboon[1*], Wararit Panichkitkosolkul[1], and Andrei Volodin[2***]**

(Submitted by Winai Bodhisuwan)

[1]*Department of Mathematics and Statistics, Thammasat University, Phathum Thani, Thailand*

[2]*Department of Mathematics and Statistics, University of Regina, Regina, Saskatchewan S4S 0A2, Canada*

Received May 10, 2020; revised June 03, 2020; accepted June 30, 2020

Abstract—We focus on the normal approximation for point estimation of the ratios of medians of two independent, log-normal distributions. We investigate its performance in terms of bias, variance, and mean square error, using Monte Carlo simulations. The results show that the normal approximation, which is relatively simple, provides a reliable result. The normal approximation approaches could be recommended on the basis of the specific values of the parameters and/or sample sizes. The point estimation is illustrated using real data examples.

DOI: 10.1134/S1995080221020177

Keywords and phrases: *Point estimation, Central tendency, Skewed distribution, Normal approximation, Simulation*

1. INTRODUCTION

In many applications, such as medicine, biology, exposure to pollution, economics, finance, reliability, survival and meteorology data analysis, measurements are often right-skewed. In these data, analyses are usually assumed by log-normal models. The log-normal distribution is common in many application areas. Estimating the parameters of the log-normal distribution is an interesting problem. There are many studies about approaches for interval estimation for log-normal distributions; for example, the interval for a single log-normal mean has been addressed multiple times in the literature [1–7]. The interval estimation for the ratio, or difference of two log-normal means is addressed in [8–16], and the interval estimation for the mean of several log-normal distributions is discussed in [17–22]. However, log-normal distributions that follow right-skewed data typically have extremely low measurements, which affect the median less than the mean. Thus, in this situation, the median is a more meaningful central tendency measure than the mean.

Some authors have considered the median of the log-normal distribution. Zellner [23] proposed a Bayesian and non-Bayesian estimator for the parameters of the mean and median of the log-normal distribution. Rao and D'Cunha [24] proposed the Bayes credible interval for the median of the log-normal distribution, and compared it with confidence intervals based on the MLE. The conclusion was that the Bayes credible interval has a shorter average length compared to the MLE interval. To our knowledge, there is no research paper on the point estimation for medians of two log-normal distributions; thus in this article, we investigate this question. We propose the normal approximation to construct the point estimator for the ratio of medians for two independent, log-normal distributions. We assess the accuracy of the estimator using the main precision characteristics, bias, variance, and mean square error. Typically, we prefer a point estimator with small mean square error.

*E-mail: lapasrada.shsb@hotmail.com

**E-mail: wararit@mathstat.sci.tu.ac.th

***E-mail: andrei@uregina.ca

In Section 2, we provide a description of the notation and log-normal model, then construct and discuss the point estimation of the ratio medians of two independent, log-normal distributions and its normal approximation. We present a simulation study to evaluate the properties of the normal approximation presented in Section 3, with the discussion and results located in Subsection 3.3.1. In Section 4, we use medical fees to illustrate the proposed point estimation, and give concluding remarks in Section 5.

2. POINT ESTIMATOR

The probability density function of a log-normal distribution with parameters μ and σ is

$$f(x) = \frac{1}{x}\sqrt{2\pi\sigma^2}e^{-\frac{1}{2}(\frac{\ln x-\mu}{\sigma})^2},$$

where $x > 0, -\infty < \mu < \infty, \sigma > 0$.

Let $\{X_i, 1 \leq i \leq n\}$ be a sample from the log-normal distribution with parameters μ and σ (denoted $X \sim LN(\mu, \sigma^2)$. Using the fact that the log-transformation of X_i, namely $Y_i = \ln(X_i)$ has the normal distribution $N(\mu, \sigma^2)$, we obtain the unbiased estimator (and the maximum likelihood estimator) for μ and $sigma^2$ as

$$\hat{\mu} = \bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i \quad \text{and} \quad \hat{\sigma^2} = S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2. \tag{1}$$

It is well-known that the median of the log-normal distribution can be calculated as follows:

$$m = e^{\mu} \tag{2}$$

Let $X_{ij};\ i = 1, 2,\ j = 1, \ldots, n_i$ be samples from two independent log-normal populations with the classical parameters μ_i and σ_i^2, respectively ($X_{ij} \sim LN(\mu_i, \sigma_i^2)$, $i = 1, 2$).

In this article, we are interested in the point estimation for the ratio ψ of medians of two independent log-normal distributions. Therefore, by (2), inferences are made on $\psi = m_1/m_2 = e^{\mu_1-\mu_2}$.

By using the plug-in estimator, according to (1), the unbiased (and maximum likelihood) point estimator of ψ is $\hat{\psi} = e^{\bar{Y}_1-\bar{Y}_2}$, where $\bar{Y}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} Y_{ij}$, $i = 1, 2$.

In the rest of this section, we address construction of the Normal approximation for the estimator $\hat{\psi}$.

2.1. The Normal Approximation

In the Normal Approximation approach, we use the Delta method to obtain a normality as the limiting distribution of an estimator $\hat{\psi}$ is the famous Delta method. It can be explained briefly in the following way.

Let $g(v_1, v_2)$ be a differentiable scalar function of two variables. Consider an estimator $T = g(V_1, V_2)$, which is a function of two basic statistics V_1 and V_2. Usually, statistics V_1 and V_2 have a simple form, and it is known that they are jointly asymptotically normal. The asymptotic distribution of an estimator, T, can be found by the Delta method, which is a procedure of stochastic representation of T.

We apply the Delta method to prove the asymptotically normality of statistic T as sample sizes tend to infinity and to find the asymptotic mean and variance for the estimator ψ.

In the Delta method, the function g is used to expand into the Taylor series at the point $\mu_1 = E(V_1)$ and $\mu_2 = E(V_2)$:

$$g(V_1, V_2) = g(\mu_1, \mu_2) + \frac{\partial g(\mu_1, \mu_2)}{\partial v_1}(V_1 - \mu_1) + \frac{\partial g(\mu_1, \mu_2)}{\partial v_2}(V_2 - \mu_2) + \text{Remainder}.$$

Note that it is possible to prove that Remainder $\to 0$ in probability as the sample sizes go to infinity.

Table 1. Bias of $\hat{\psi}$

(σ_1^2, σ_2^2)	(n_1, n_2)	(25,25)					(50,50)				
		μ_2									
	μ_1	1.0	1.5	2.0	2.5	3.0	1.0	1.5	2.0	2.5	3.0
	1.0	0.0085	−0.3850	−0.6290	−0.7740	−0.8630	0.0052	−0.391	−0.631	−0.0776	−0.8640
	2.0	1.7480	0.6659	0.0102	−0.3860	−0.6294	1.7352	0.6579	0.0044	−0.3894	−0.6304
	3.0	6.4628	3.5220	1.7437	0.6660	0.0110	6.4289	3.4936	1.7296	0.6600	0.0045
	(n_1, n_2)	(100,100)					(25,50)				
		μ_2									
(0.2,0.3)	μ_1	1.0	1.5	2.0	2.5	3.0	1.0	1.5	2.0	2.5	3.0
	1.0	0.0019	−0.3920	−0.6320	−0.7770	−0.8645	0.0080	−0.3880	−0.6290	−0.7753	−0.8640
	2.0	1.7271	0.6534	0.0021	−0.3920	−0.6312	1.7423	0.6571	0.0084	−0.3889	−0.6297
	3.0	6.4085	3.4968	1.7250	0.6520	0.0031	6.4424	3.5242	1.7394	0.6605	0.0074
	(n_1, n_2)	(25,100)					(50,100)				
		μ_2									
	μ_1	1.0	1.5	2.0	2.5	3.0	1.0	1.5	2.0	2.5	3.0
	1.0	0.0056	−0.3900	−0.6300	−0.7760	−0.8636	0.0033	−0.3900	−0.6310	−0.7760	−0.8640
	2.0	1.7347	0.6618	0.0063	−0.3890	−0.6296	1.7284	0.6521	0.0024	−0.3905	−0.6304
	3.0	6.4266	3.5099	1.7339	0.6589	0.0060	6.4091	3.4943	1.7292	0.6535	0 .0031
	(n_1, n_2)	(25,25)					(50,50)				
		μ_2									
	μ_1	1.0	1.5	2.0	2.5	3.0	1.0	1.5	2.0	2.5	3.0
	1.0	0.1091	−0.329	−0.591	−0.752	−0.85	0.0540	−0.3570	−0.6140	−0.7650	−0.8584
	2.0	2.0021	0.8241	0.1019	−0.331	−0.595	1.8630	0.7298	0.0560	−0.3640	−0.6129
	3.0	7.1771	3.9656	2.0000	0.8158	0.1050	6.7910	3.6920	1.8377	0.7270	0.04410
	(n_1, n_2)	(100,100)					(25,50)				
		μ_2									
(2.0,3.0)	μ_1	1.0	1.5	2.0	2.5	3.0	1.0	1.5	2.0	2.5	3.0
	1.0	0.0250	−0.3780	−0.6220	−0.7720	−0.8610	0.0814	−0.3500	−0.6050	−0.7600	−0.8540
	2.0	1.7760	0.6947	0.0311	−0.3760	−0.6250	1.9262	0.7633	0.0748	−0.3470	−0.6050
	3.0	6.5920	3.5957	1.7808	0.6850	0.0243	6.9011	3.8087	1.9248	0.7756	0.0784
	(n_1, n_2)	(25,100)					(50,100)				
		μ_2									
	μ_1	1.0	1.5	2.0	2.5	3.0	1.0	1.5	2.0	2.5	3.0
	1.0	0.0590	−0.3620	−0.6110	−0.7630	−0.8568	0.0330	−0.3730	−0.6190	−0.7690	−0.8600
	2.0	1.8880	0.7461	0.0582	−0.3550	−0.6108	1.8220	0.7043	0.0377	−0.3770	−0.6170
	3.0	6.7730	3.7519	1.8754	0.7324	0.0541	6.6780	3.6106	1.8091	0.7029	0.0332

Table 1. Продолжение

	(n_1, n_2)	(25,25)					(50,50)				
		μ_2									
	μ_1	1.0	1.5	2.0	2.5	3.0	1.0	1.5	2.0	2.5	3.0
	1.0	0.5356	−0.0720	−0.4340	−0.6530	−0.7910	0.2350	−0.2380	−0.5430	−0.7210	−0.8310
	2.0	3.2081	1.6010	0.5468	−0.0680	−0.4270	2.4240	1.0740	0.2474	−0.2450	−0.5412
	3.0	10.6300	6.0744	3.2921	1.5175	0.5478	8.2190	4.6179	2.3748	1.0256	0.2360
	(n_1, n_2)	(100,100)					(25,50)				
		μ_2									
(10.0,12.0)	μ_1	1.0	1.5	2.0	2.5	3.0	1.0	1.5	2.0	2.5	3.0
	1.0	0.1170	−0.3260	−0.5930	−0.7520	−0.8490	0.3833	−0.1630	−0.5010	−0.6950	−0.8130
	2.0	2.0500	0.8588	0.1126	−0.3210	−0.5870	2.7281	1.2776	0.3605	−0.1670	−0.4970
	3.0	7.194	3.9671	2.0429	0.8482	0.1138	9.3186	5.2541	2.7623	1.2548	0.3862
	(n_1, n_2)	(25,100)					(50,100)				
		μ_2									
	μ_1	1.0	1.5	2.0	2.5	3.0	1.0	1.5	2.0	2.5	3.0
	1.0	0.3010	−0.2210	−0.5190	−0.7150	−0.8230	0.1610	−0.2880	−0.5700	−0.7380	−0.8420
	2.0	2.5070	1.1446	0.2939	−0.2190	−0.5230	2.1970	0.9282	0.1803	−0.2940	−0.5680
	3.0	8.4660	4.7735	2.4950	1.1788	0.2923	7.7730	4.2055	2.1941	0.9609	0.1569

In our case, $g(v_1, v_2) = e^{v_1 - v_2}, V_1 = \bar{Y}_1$ and $V_2 = \bar{Y}_2$. The statistic $V_1 = \bar{Y}_1$ is normally distributed as $N(\mu_1, \sigma_1^2)$ and the statistic $V_2 = \bar{Y}_2$ is normally distributed as $N(\mu_2, \sigma_2^2)$. They are jointly normally distributed by their independence.

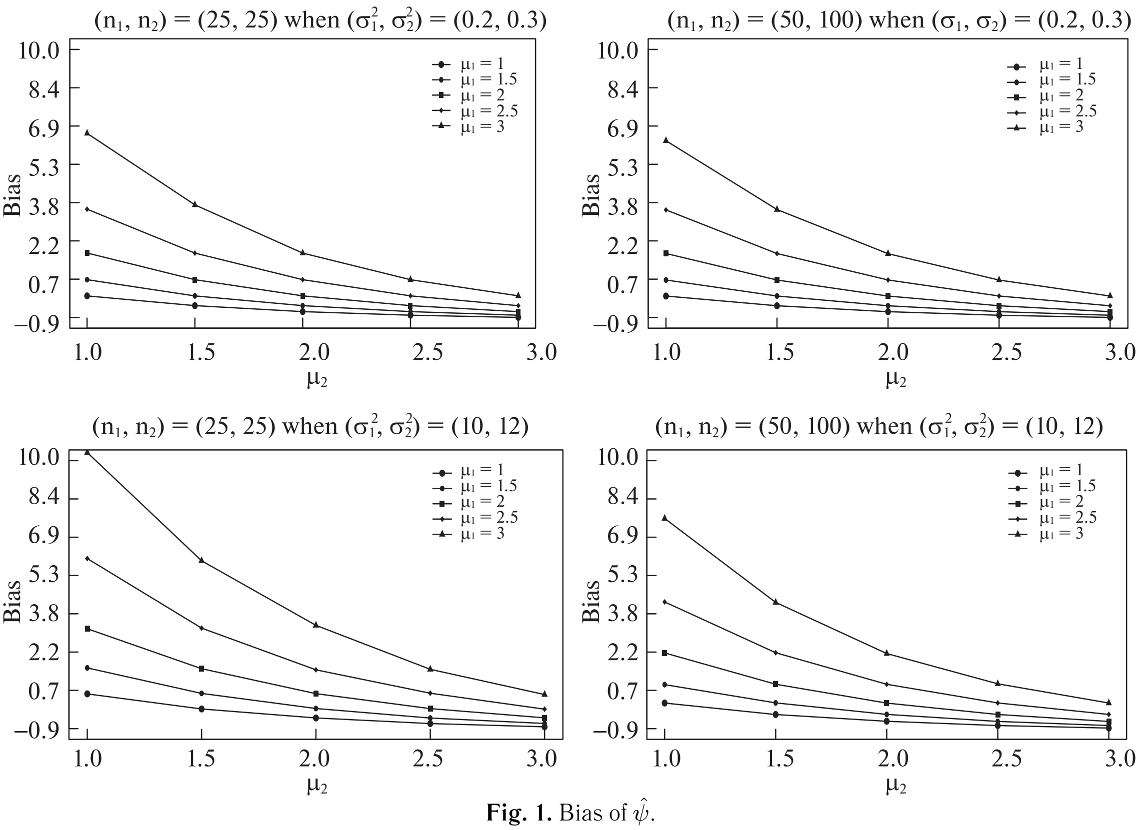

Fig. 1. Bias of $\hat{\psi}$.

Table 2. Mean Square Error of $\hat{\psi}$

(σ_1^2, σ_2^2)	(n_1, n_2)	(25,25)					(50,50)				
(0.2,0.3)		μ_2									
	μ_1	1.0	1.5	2.0	2.5	3.0	1.0	1.5	2.0	2.5	3.0
	1.0	0.0202	0.1561	0.3983	0.5999	0.7453	0.0103	0.1565	0.3989	0.6027	0.7467
	2.0	3.2061	0.4987	0.0206	0.1569	0.3991	3.0867	0.4610	0.0100	0.1553	0.3988
	3.0	42.8932	12.8212	3.1950	0.4995	0.0207	41.8864	12.4106	3.068	0.4635	0.0104
	(n_1, n_2)	(100,100)					(25,50)				
		μ_2									
	μ_1	1.0	1.5	2.0	2.5	3.0	1.0	1.5	2.0	2.5	3.0
	1.0	0.0050	0.1556	0.3996	0.6032	0.7475	0.0149	0.1559	0.3979	0.6019	0.7461
	2.0	3.0206	0.4405	0.0051	0.1558	0.3991	3.1428	0.4703	0.0146	0.1566	0.3985
	3.0	41.3505	12.3282	3.0134	0.4386	0.0050	12.7071	3.1304	0.4754	0.0141	0.0113
	(n_1, n_2)	(25,100)					(50,100)				
		μ_2									
	μ_1	1.0	1.5	2.0	2.5	3.0	1.0	1.5	2.0	2.5	3.0
	1.0	0.0113	0.1559	0.3985	0.6025	0.7462	0.0071	0.1551	0.3991	0.6026	0.7468
	2.0	3.0927	0.4680	0.0113	0.1555	0.3980	3.0409	0.4446	0.0070	0.1551	0.3984
	3.0	41.9021	12.5487	3.0898	0.4651	0.0114	441.4712	12.3519	3.0423	0.4461	0.0071
(2.0,3.0)	(n_1, n_2)	(25,25)					(50,50)				
		μ_2									
	μ_1	1.0	1.5	2.0	2.5	3.0	1.0	1.5	2.0	2.5	3.0
	1.0	0.2887	0.2064	0.3872	0.9753	0.7270	0.1148	0.1704	0.3932	0.5910	0.7390
	2.0	5.9484	1.4370	0.2864	0.2079	0.3914	4.3248	0.8372	0.1201	0.1761	0.3914
	3.0	66.4982	21.0610	5.9309	1.3838	0.2865	52.4088	15.9385	4.1938	0.8440	0.1160
	(n_1, n_2)	(100,100)					(25,50)				
		μ_2									
	μ_1	1.0	1.5	2.0	2.5	3.0	1.0	1.5	2.0	2.5	3.0
	1.0	0.0546	0.1627	0.3948	0.5979	0.7431	0.1896	0.1862	0.3894	0.5861	0.7333
	2.0	3.5482	0.6311	0.0545	0.1616	0.3974	5.0461	1.0591	0.1763	0.1840	0.3896
	3.0	46.3822	14.0246	3.5615	0.6105	0.0548	56.8428	18.0231	5.0043	1.0646	0.1809
	(n_1, n_2)	(25,100)					(50,100)				
		μ_2									
	μ_1	1.0	1.5	2.0	2.5	3.0	1.0	1.5	2.0	2.5	3.0
	1.0	0.1341	0.1788	0.3912	0.5884	0.7366	0.0784	0.1681	0.3938	0.5947	0.7409
	2.0	4.5680	0.9111	0.1301	0.1738	0.3908	3.9067	0.7122	0.0792	0.1663	0.3908
	3.0	52.8729	16.7045	4.4718	0.8835	0.1327	48.9102	14.5976	3.8610	0.7007	0.0792

Table 2. Продолжение

	(n_1,n_2)	(25,25)					(50,50)				
		μ_2									
	μ_1	1.0	1.5	2.0	2.5	3.0	1.0	1.5	2.0	2.5	3.0
	1.0	3.8394	1.3312	0.6128	0.5998	0.6825	0.9260	0.3859	0.4045	0.5593	0.7067
	2.0	37.0604	11.1480	3.9103	1.2807	0.6574	12.3118	3.3660	0.8767	0.3682	0.4015
	3.0	289.8737	111.9812	38.4053	12.8062	3.3571	114.6552	38.7245	12.1558	3.3681	0.9820
	(n_1,n_2)	(100,100)					(25,50)				
		μ_2									
(10.0,12.0)	μ_1	1.0	1.5	2.0	2.5	3.0	1.0	1.5	2.0	2.5	3.0
	1.0	0.3085	0.2176	0.3885	0.5801	0.7243	1.7895	0.7810	0.4656	0.5616	0.6904
	2.0	6.4317	1.5421	0.3294	0.2142	0.3889	19.5132	3.1603	1.7630	0.5989	0.4816
	3.0	66.6252	21.7559	6.3464	1.5349	0.3253	181.9149	63.8960	19.2874	6.1035	1.7686
	(n_1,n_2)										
		μ_2									
	μ_1	1.0	1.5	2.0	2.5	3.0	1.0	1.5	2.0	2.5	3.0
	1.0	1.2572	0.4712	0.4285	0.5592	0.6974	0.5439	0.2783	0.3905	0.5732	0.7181
	2.0	14.7542	4.5920	1.2021	0.4462	0.4320	8.6614	2.2528	0.5427	0.2658	0.3946
	3.0	140.6147	46.9427	14.3412	4.2729	1.2274	86.6303	28.4016	8.4672	2.3737	0.5607

To calculate partial derivatives,

$$\frac{\partial g(v_1,v_2)}{\partial v_1} = e^{v_1-v_2} \quad \text{and} \quad \frac{\partial g(v_1,v_2)}{\partial v_2} = -e^{v_1-v_2}.$$

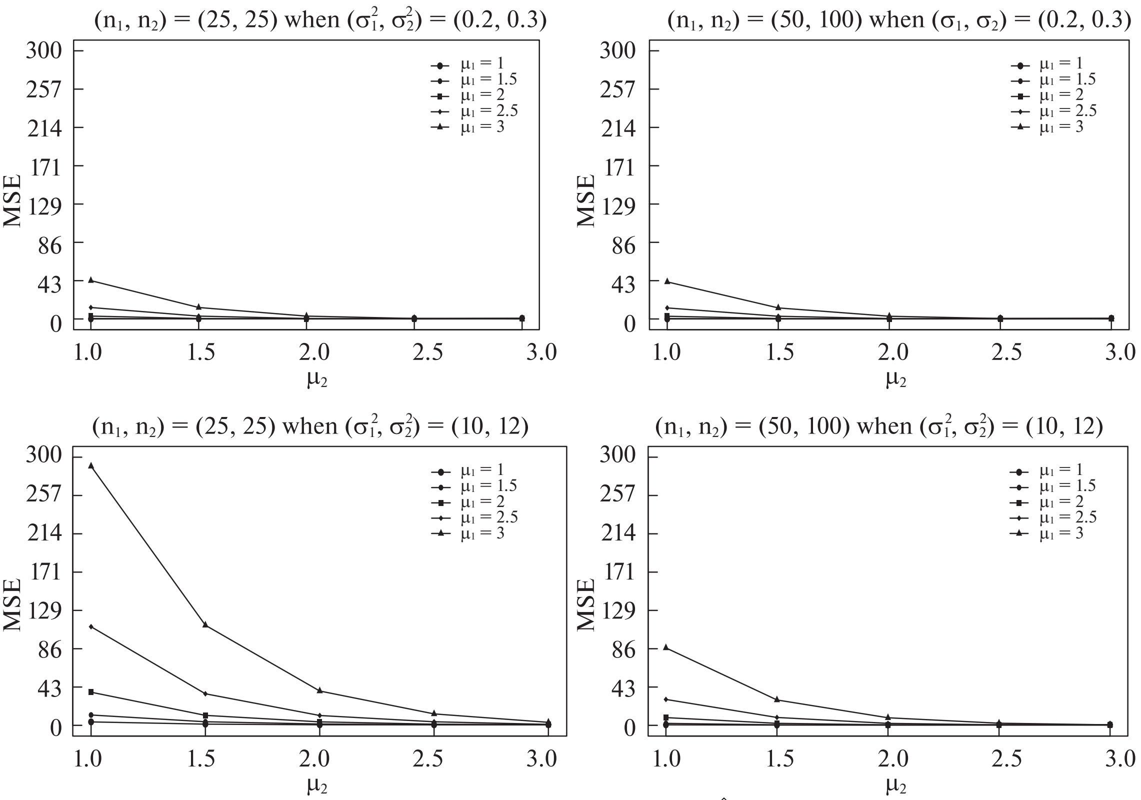

Fig. 2. Mean Square Errors of $\hat{\psi}$.

Table 3. Simulated and Theoretical Variances of $\hat{\psi}$. Theoretical Variance is reported in brackets

(σ_1^2, σ_2^2)	(n_1, n_2)	(25,25)					(50,50)				
		μ_2									
	μ_1	1.0	1.5	2.0	2.5	3.0	1.0	1.5	2.0	2.5	3.0
	1.0	0.0202	0.0076	0.0027	0.0010	0.0004	0.0102	0.0037	0.0014	0.0005	0.0002
		(0.0207)	(0.0077)	(0.0028)	(0.0010)	(0.0004)	(0.0102)	(0.0038)	(0.0014)	(0.0005)	(0.0002)
	2.0	0.1506	0.0552	0.0205	0.0076	0.0028	0.0756	0.0281	0.0100	0.0037	0.0014
		(0.1540)	(0.0567)	(0.0209)	(0.0077)	(0.0028)	(0.0755)	(0.0277)	(0.0102)	(0.0038)	(0.0014)
	3.0	1.1254	0.4163	0.1545	0.0559	0.0206	0.5552	0.2056	0.0764	0.0279	0.0104
		(1.1347)	(0.4161)	(0.1535)	(0.0566)	(0.0208)	(0.5592)	(0.2038)	(0.0753)	(0.0279)	(0.0102)
	(n_1, n_2)	(100,100)					(25,50)				
		μ_2									
(0.2,0.3)	μ_1	1.0	1.5	2.0	2.5	3.0	1.0	1.5	2.0	2.5	3.0
	1.0	0.0050	0.0019	0.0007	0.0003	0.0001	0.0149	0.0053	0.0019	0.0007	0.0003
		(0.0050)	(0.0019)	(0.0007)	(0.0003)	(0.0001)	(0.0144)	(0.0053)	(0.0019)	(0.0007)	(0.0003)
	2.0	0.0379	0.0135	0.0051	0.0019	0.0007	0.1074	0.0386	0.0145	0.0053	0.0020
		(0.0374)	(0.0137)	(0.0050)	(0.0019)	(0.0007)	(0.1069)	(0.0390)	(0.0145)	(0.0053)	(0.0019)
	3.0	0.2821	0.1005	0.0376	0.0136	0.0050	0.7937	0.2873	0.1049	0.0391	0.0141
		(0.2754)	(0.1017)	(0.0373)	(0.0137)	(0.0051)	(0.7884)	(0.2897)	(0.1066)	(0.0392)	(0.0144)
	(n_1, n_2)	(25,100)					(50,100)				
		μ_2									
	μ_1	1.0	1.5	2.0	2.5	3.0	1.0	1.5	2.0	2.5	3.0
	1.0	0.0113	0.0040	0.0015	0.0005	0.0002	0.0071	0.0026	0.0010	0.0004	0.0001
		(0.0112)	(0.0041)	(0.0015)	(0.0006)	(0.0002)	(0.0071)	(0.0026)	(0.0010)	(0.0003)	(0.0001)
	2.0	0.0836	0.0300	0.0113	0.0042	0.0015	0.0535	0.0193	0.0070	0.0026	0.0010
		(0.0829)	(0.0307)	(0.0113)	(0.0042)	(0.0015)	(0.0524)	(0.0192)	(0.0071)	(0.0026)	(0.0010)
	3.0	0.6013	0.2289	0.0832	0.0309	0.0114	0.1418	0.0521	0.0521	0.0190	0.0071
		(0.6129)	(0.2259)	(0.0830)	(0.0308)	(0.0112)	(0.1424)	(0.0525)	(0.0525)	(0.0192)	(0.0071)

Hence,

$$\frac{\partial g(\mu_1,\mu_2)}{\partial v_1} = e^{\mu_1-\mu_2} \quad \text{and} \quad \frac{\partial g(\mu_1,\mu_2)}{\partial v_2} = -e^{\mu_1-\mu_2}.$$

By the Taylor series expansion,

$$g(V_1,V_2) \approx e^{\mu_1-\mu_2} + e^{\mu_1-\mu_2}(\hat{\mu}_1-\mu_1) - e^{\mu_1-\mu_2}(\hat{\mu}_2-\mu_2) \approx e^{\mu_1-\mu_2}(\hat{\mu}_1-\hat{\mu}_2-\mu_1+\mu_2+1).$$

From this, we take the expected value and variance. The asymptotic mean and variance of ψ are $\mu_\psi = e^{\mu_1-\mu_2}$ and $\sigma_\psi^2 = e^{2(\mu_1-\mu_2)}\left(\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}\right)$, respectively.

To summarize, as sample sizes $n_1, n_2 \to \infty$, the estimator $\hat{\psi}$ is approximately normal $N(\mu_\psi, \sigma_\psi^2)$.

Obviously, values μ_1, μ_2, σ_1^2, and σ_2^2 are unknown when we estimate the parameter function ψ having only samples in our hands. In this case, we use the estimators (1) and obtain the following plug-in

Table 3. Продолжение

	(n_1,n_2)	(25,25)					(50,50)				
		μ_2									
	μ_1	1.0	1.5	2.0	2.5	3.0	1.0	1.5	2.0	2.5	3.0
	1.0	0.2768	0.0984	0.0376	0.0134	0.0050	0.1119	0.1119	0.0429	0.0060	0.0021
		(0.3030)	(0.1105)	(0.0406)	(0.0149)	(0.0055)	(0.2197)	(0.1224)	(0.0456)	(0.0061)	(0.0022)
	2.0	1.9398	0.7578	0.2760	0.0984	0.0369	0.8533	0.3047	0.1170	0.0433	0.0157
		(2.1903)	(0.8188)	(0.2992)	(0.1088)	(0.0400)	(0.9060)	(0.3291)	(0.1233)	(0.0447)	(0.0166)
	3.0	14.9866	5.3351	1.9709	0.7183	0.2755	6.2894	2.3077	0.8167	0.3155	0.1141
		(16.3561)	(6.0286)	(2.1783)	(0.8021)	(0.2985)	(6.6821)	(2.4317)	(0.8880)	(0.3302)	(0.1200)
	(n_1,n_2)	(100,100)					(25,50)				
		μ_2									
(2.0,3.0)	μ_1	1.0	1.5	2.0	2.5	3.0	1.0	1.5	2.0	2.5	3.0
	1.0	0.0540	0.0199	0.0073	0.0026	0.0010	0.1830	0.0635	0.0239	0.0089	0.0033
		(0.0551)	(0.0204)	(0.0074)	(0.0027)	(0.0010)	(0.1905)	(0.0682)	(0.0251)	(0.0093)	(0.0034)
	2.0	0.3925	0.1485	0.0535	0.0199	0.0073	1.3359	0.4764	0.1707	0.4764	0.1707
		(0.4049)	(0.1511)	(0.0558)	(0.0205)	(0.0074)	(1.3884)	(0.5037)	(0.1854)	(0.5037)	(0.1854)
	3.0	2.9287	1.0951	0.3902	0.1412	0.0542	9.2175	3.5167	1.2992	0.4631	0.1747
		(3.0255)	(1.1125)	(0.4057)	(0.1489)	(0.0552)	(9.9915)	(3.7289)	(1.3817)	(0.5070)	(0.1866)
	(n_1,n_2)	(25,100)					(50,100)				
		μ_2									
	μ_1	1.0	1.5	2.0	2.5	3.0	1.0	1.5	2.0	2.5	3.0
	1.0	0.1307	0.0478	0.0179	0.0067	0.0024	0.0773	0.0288	0.0108	0.0038	0.0014
		(0.1377)	(0.0501)	(0.0186)	(0.0070)	(0.0025)	(0.0799)	(0.0295)	(0.0110)	(0.0040)	(0.0015)
	2.0	1.0045	0.3544	0.1267	0.0477	0.0177	0.5858	0.2162	0.0778	0.0288	0.0106
		(1.0279)	(0.3747)	(0.1376)	(0.0509)	(0.0187)	(0.6003)	(0.2187)	(0.0807)	(0.0296)	(0.0110)
	3.0	7.0052	2.6274	0.9548	0.3470	0.1298	4.3128	1.5614	0.5880	0.2066	0.0781
		(7.3842)	(2.7737)	(1.0142)	(0.3684)	(0.1358)	(4.4416)	(1.5958)	(0.5940)	(0.2171)	(0.0804)

estimators of $\hat{\mu}_\psi$ and $\hat{\sigma}^2_\psi$ as follows:

$$\hat{\mu}_\psi = e^{\bar{y}_1-\bar{y}_2}\text{and}\hat{\sigma}^2_\psi = e^{2(\bar{y}_1-\bar{y}_2)}\left(\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}\right),$$

where $\bar{y}_i$ and s_i^2 are the observed values of $\bar{Y}_i$ and S_i^2, respectively.

3. SIMULATION STUDY

We used simulation studies to evaluate the properties of the normal approximation. We estimated the Bias, Mean Square Error (MSE), Simulated Variance (Simuvar) and Theoretical Variance (TVar) through Monte Carlo simulation with the R statistical software. For the parameter configurations, we generated 10,000 random samples from two independent log-normal populations with parameter μ_i and $\sigma_i^2 i = 1, 2$. Numerical results on the values of the accuracy measurements, Bias, Mean Square

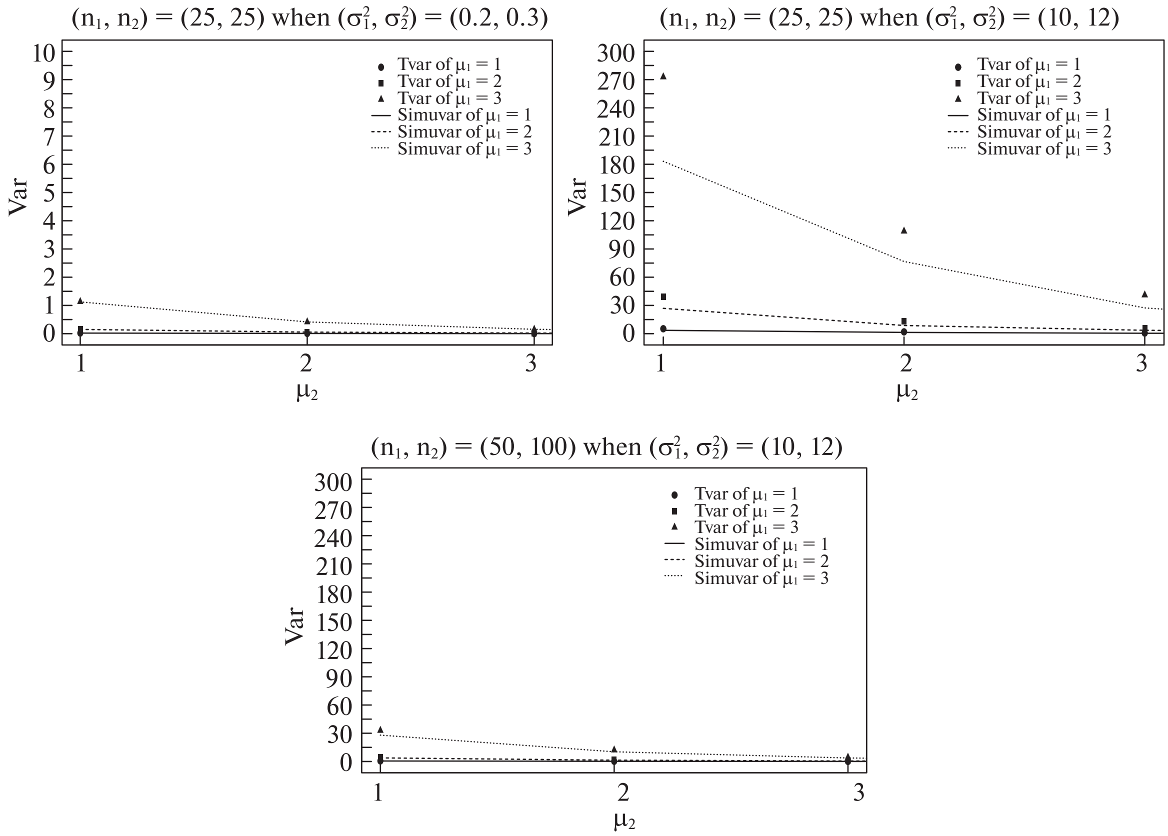

Fig. 3. Variances of $\hat{\psi}$.

Error (MSE), Simulated Variance (Simuvar), and Theoretical Variance (TVar) for the estimator of ψ of two independent log-normal distributions when the variances (σ_1^2, σ_2^2) are (0.2,0.3), (2.0,3.0) and (10.0, 12.0) and both means, μ_1 and μ_2 values change from 1.0 to 3.0 by step, say, 0.5 for different group sizes (n_1, n_2) varying from small to large under equal and unequal sizes are reported in Tables 1–3, respectively.

3.1. Discussion

According to the results presented in Tables 1–3, we can conclude that the bias of our approach increases with the increase of variances σ_i^2. As expected, the bias decreases when the sample size increases for all value of the parameter μ_i. In cases of the value of μ_1 is small, the bias of our approach is small and decreases when the value of the parameter μ_2 becomes larger especially for moderate to large sample sizes.

We can say the same about mean square error (MSE) and simulated variance and theoretical variance. Moreover, they are very close to zero for small μ_1, ($\mu_1 = 1.0, 1.5$) and for moderate to large sample sizes.

In addition, to illustrate the overall performance of the normal approximation method clearly, the data in the tables are plotted as the graphs. The bias, MSE, variance values is plotted against μ_i for small and large value of the parameter variances σ_i^2, according to Tables 1–3 shown in Figures 1–3, respectively.

Figure 1 shows the bias against μ_i for small and large value of the parameter variances σ_i^2. As we see in Figure 1, in case of small variances $(\sigma_1^2, \sigma_2^2) = (0.2.0.3)$, the bias is smaller than for scenario with larger variances, $(\sigma_1^2, \sigma_2^2) = (10.0, 12.0)$. Moreover, the our estimator performs better as the values of μ_i become smaller for all values of the parameter variances σ_i^2 especially for moderate and large sample sizes.

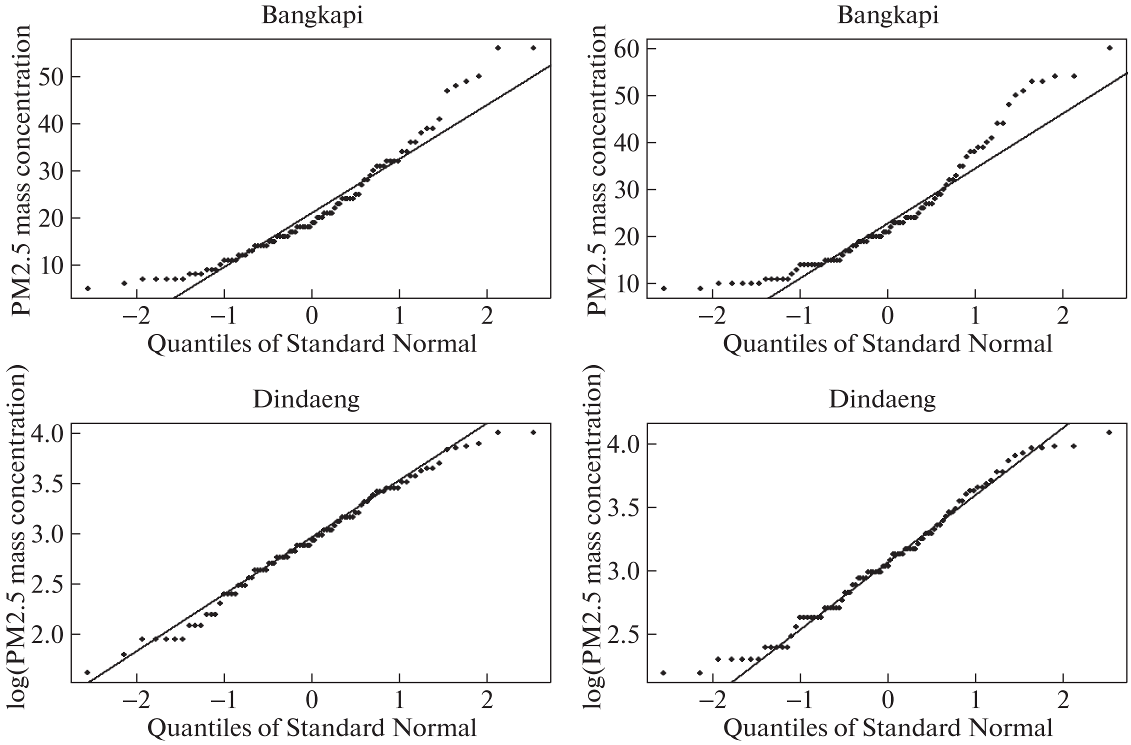

Fig. 4. Quantile plots of Medical Fees and log(Medical Fees) data.

Figure 2 shows the mean square error (MSE) against μ_i for small and large value of the parameter variances σ_i^2. From Figure 2 we can draw similar conclusions as for the bias. In case of small variances $(\sigma_1^2, \sigma_2^2) = (0.2.0.3)$, the MSE is smaller than for scenario with larger variances $(\sigma_1^2, \sigma_2^2) = (10.0, 12.0)$. The MSE are small for value of the parameter μ_i is small. Therefore, we can say that our approach are efficient for small value of the parameters μ_i, σ_i^2 and moderate to large sample sizes.

Figure 3 shows the simulated variance and theoretical variance against μ_i for small and large value of the variance parameters σ_i^2. From Figure 3, the theoretical and asymptotic variance of the estimator are accurate when μ_i is small and $(\sigma_1^2, \sigma_2^2) = (0.2.0.3)$ for moderate to large sample sizes.

4. AN APPLICATIONS TO REAL DATA

Zhou et al. [4] gave a dataset on the effects of race on medical fees of African American and Caucasian patients with type I diabetes who had received inpatient or outpatient care on at least two occasions during the period from 1 January 1993 to 30 June 1994. For African American patients, the summary statistics are: $n_1 = 119$, $\bar{y}_1 = 9.067$, $s_1^2 = 1.824$ and Caucasian patients, the summary statistics are: $n_2 = 106$, $\bar{y}_2 = 8.693$, $s_2^2 = 2.692$.

Figure 4 shows the QQ plots for the original and log-transformed medical fee data. These plots show that the distribution of dataset concentrations are positively skewed, and the logarithmically transformed data are approximately symmetric. Here, we give a point estimator of the ratio of medians ψ of medical fees of African American and Caucasian patients. The Normal approximation for the estimator $\hat{\psi}$ is 1.4535. The results indicate that the median (average) medical fees of African American patients is greater than the Caucasian patients. The results are consistent with the our simulation study, the normal approximation is recommended for moderate to large sample size.

5. CONCLUDING REMARKS

In this paper, we focus on the normal approximation to the estimator for the ratio of medians of two independent, log-normal distributions. The performance of the estimator is evaluated in terms of its bias, variance, and the mean square error in the simulation study section. The results show that the estimator has reliable accuracy. The normal approximation is recommended for moderate to large sample sizes when μ_i and σ_i^2 are small.

ACKNOWLEDGEMENTS

The third author's research was funded by the subsidy allocated to Kazan Federal University for the state assignment in the sphere of scientific activities, project № 1.13556.2019/13.1.

REFERENCES

1. C. E. Land, "An evaluation of approximate confidence interval estimation methods for lognormal means," Technometrics **14**, 145–158 (1972).
2. J. E. Angus, "Inferences on the lognormal mean for complete samples," Commun. Stat.-Simul. Comput. **17**, 1307–1331 (1988).
3. J. E. Angus, "Bootstrap one-sided confidence intervals for the lognormal mean," Statistician **43**, 395–401 (1994).
4. X. H. Zhou and S. Gao, "Confidence intervals for the lognormal mean," Stat. Med. **16**, 783–790 (1997).
5. D. J. Taylor, L. L. Kupper, and K. E. Muller, "Improved approximate confidence intervals for the mean of a lognormal random variable," Stat. Med. **21**, 1443–1459 (2002).
6. J. Wu, A. C. M. Wong, and G. Jiang, "Likelihood-based confidence intervals for a log-normal mean," Stat. Med. **22**, 1849–1860 (2003).
7. N. Smithpreecha, S. Niwitpong, and S. Niwitpong, "Confidence intervals for common mean of lognormal distributions," in *Proceedings of the International Econometric Conference of Vietnam* (Springer, Cham, 2018), pp. 276–289.
8. X.-H. Zhou and W. Tu, "Interval estimation for the ratio in means of log-normally distributed medical costs with zero values," Comput. Stat. Data Anal. **35**, 201–210 (2000).
9. J. Wu, G. Jiang, A. C. M. Wong, and X. Sun, "Likelihood analysis for the ratio of means of two independent log-normal distributions," Biometrics **58**, 463–469 (2002).
10. K. Krishnamoorthy and T. Mathew, "Inferences on the means of lognormal distributions using generalized p-values and generalized confidence intervals," J. Stat. Plan. Inf. **115**, 103–120 (2003).
11. P. Gill, "Small-sample inference for the comparison of means of log-normal distributions," Biometrics **60**, 525–527 (2004).
12. Y.-H. Chen and X.-H. Zhou, "Interval estimates for the ratio and difference of two log-normal means," Stat. Med. **25**, 4099–4113 (2006).
13. K. Cimermanová, "Estimation of confidence intervals for the log-normal means and for the ratio and the difference of log-normal means with application of breath analysis," Meas. Sci. Rev. **7** (4), 31–36 (2007).
14. G. Y. Zou, J. Taleban, and C.-Y. Huo, "Confidence interval estimation for lognormal data with application to health economics," Comput. Stat. Data Anal. **53**, 3755–3764 (2009).
15. J. Poirier, G. Y. Zou, and J. Koval, "Confidence intervals for a difference between lognormal means in cluster randomization trials," Stat. Methods Med. Res. **26**, 598–614 (2017).
16. S. G. Kang, W. D. Lee, and Y. Kim, "Objective Bayesian analysis using modified profile likelihood for the ratio of two log-normal means," J. Korean Stat. Soc., 1–22 (2020).
17. A. Baklizi and M. Ebrahem, "Interval estimation of common lognormal mean of several populations," J. Probab. Stat. Sci. **3**(1), 1–16 (2005).
18. L. Tian and J. Wu, "Inferences on the common mean of several log-normal populations: The generalized variable approach," Biometr. J. **49**, 944–951 (2007).
19. J. Behboodian and A. Jafari, "Generalized inference for the common mean of several lognormal populations," J. Stat. Theory Appl. **5**, 240–259 (2006).
20. S. H. Lin and R. S. Wang, "Modified method on the means for several log-normal distributions," J. Appl. Stat. **40**, 194–208 (2013).
21. A. Malekzadeh and M. Kharrati-Kopaei, "Inferences on the common mean of several heterogeneous log-normal distributions," J. Appl. Stat. **46**, 1066–1083 (2019).
22. A. Zellner, "Bayesian and non-Bayesian analysis of the log-normal distribution and log-normal regression," J. Am. Stat. Assoc. **66** (334), 327–330 (1971).
23. K. A. Rao and J. G. D'Cunha, "Bayesian inference for median of the lognormal distribution," J. Mod. Appl. Stat. Methods **15**, 526–535 (2016).
24. G. Y. Zou, "On the estimation of additive interaction using the four-by-two table and beyond," Am. J. Epidemiol. **168**, 212–224 (2008).
25. G. Y. Zou and A. Donner, "Construction of confidence limits about effect measures: A general approach," Stat. Med. **27**, 1693–1702 (2008).
26. S. Weerahandi, "Generalized confidence intervals," J. Am. Stat. Assoc. **88**, 899–905 (1993).