wjert, 2021, Vol. 7, Issue 6, 202-210. Research Article ISSN 2454-695X

World Journal of Engineering Research and Technology

WJERT

www.wjert.org SJIF Impact Factor: 5.924

THE LOGNORMAL DISTRIBUTION'S GRAPHICALLY ANALYSIS AND ITSPOWER

B. Pratikno*, A. A. Pratama and E. L. Wulandari

Department of Mathematics, Faculty of Mathematics and Natural Science Jenderal Soedirman University, Purwokerto, Indonesia, Jl. Dr. Soeparno 61 Karangwangkal, Purwokerto, Banyumas, 53123, Central of Java, Indonesia.

Article Received on 05/10/2021 Article Revised on 26/10/2021 Article Accepted on 16/11/2021

***Corresponding Author**
B. Pratikno
Department of Mathematics, Faculty of Mathematics and Natural Science Jenderal Soedirman University, Purwokerto, Indonesia, Jl. Dr. Soeparno 61 Karangwangkal, Purwokerto, Banyumas, 53123, Central of Java, Indonesia.

ABSTRACT

The research discussed the graphically analysis of the probability density function (pdf), cumulative distribution function (cdf), and power and size of hypothesis testing (minimum) of the parameter shape on the Lognormal distribution. In this research, we derived the formula of the power and figured their curve using *R* code. The result showed that the power Lognormal distribution depended on the degree of freedom *n* and bound of the rejection area, and parameter shape (σ). The curves of the power are sigmoid and they increase and more faster to be one on the small parameter shape (σ) and large *n*.

KEYWORDS AND PHRASES: Lognormal distribution, the power function of the hypothesis testing, *R-code.*

1. INTRODUCTION

Following Wackerly, et al.[5], there are three important concepts of the hypothesis testing in rejecting or accepting null hypothesis (H_0), namely (1) probability error type I (α), (2) a probability error type II (β) and (3) a power. Here, the power is a significant method to test the testing on the parameter hypothesis testing. Therefore, we then studied more detail the power of the hypothesis testing on some various continuous distributions. One of them is Lognormal distribution. Note that, the power is defined as a probability to reject H_0

under H_1 in testing hypothesis H:$\theta = \theta$ versus H:$\theta \neq \theta$, for parameter θ (Wackerly, et al.[5]). From the previous research, we noted many authors, such as Pratikno[2], Khan and Pratikno[24] and Khan[14], already studied the power in testing intercept with non-sample prior information (NSPI). They used the probability integral of the cumulative distribution function (cdf) of the continuous distributions to calculate the power. Moreover, Pratikno[2] and Khan et al.[13] used the power and size to compute the cdf of the bivariate noncentral *F* (BNCF) distribution in multivariate and multiple regression models. Here, we also noted that many authors, such as Khan[14, 15, 16], Khan and Saleh[17,18,19, 22, 23], Khan and Hoque[21], Saleh[1], Yunus[6], and Yunus and Khan[9, 10, 11, 12], have contributed to the research of thepower in the context of the hypothesis area. In the context of the hypothesis testing with NSPI on multivariate and multiple regression models, Pratikno[2] and Khan et al.[13] used the BNCF distribution to compute the power using *R-code*. This is due to the computational of the probability integral of the probability distribution function (pdf) and cdf of the BNCF distribution are very complicated and hard (see Pratikno[2] and Khan et al.[20]), so the *R* code is used. Here, we noted that the probability density function (pdf) and cumulative distribution function (cdf) really significant contributed to analysis the power and size function as well as on continuous lognormal distribution.

From the previous research, we noted that many papers already discussed the power of the hypothesis testing on the continuous distribution, but here we focused on lognormal distributions. To investigate the power, we have to follow some steps, namely : (1) we must determine the sufficiently statistics, (2) we then create the rejection area using uniformly most powerful test (UMPT), (3) we then derive the formula of the power of the lognormal distribution, and (4) finally we simulate graphically analysed of the power using generate data.

In this paper, the introduction and objective are given in Section 1 and 2. The literature review and research methods are then presented in Section 3 and 4, respectively. The result and discussion is obtained in Section 5. The conclusion is provided in Section 6.

2. OBJECTIVES

The research focused to graphically analyses of the probability density function (pdf), cumulative distribution function (cdf), and derived the formula of the power function and size ofthe hypothesis testing of the parameter shape on the Lognormal distribution.

3. Literature Review

To illustrate the power function, we follow Pratikno[2], Khan[14, 15, 16], Khan and Saleh[17,18,19, 22, 23], Khan and Hoque[21], Saleh[1], Yunus[6], and Yunus and Khan.[9,10,11, 12] Here, we noted that the maximum power and minimum size of the tests are used to test the eligible testing (among tests). Here, the power is defined as a probability to reject H_0 under H_1 in testing hypothesis, and the size is a probability to reject H_0 under H_0(see Wackerly, et al.[5] and Pratikno[2]). Detail of the power and size on several continuous distributions and testing coefficient parameters on the regression model are found Pratikno et al.[2, 3, 4] To illustrate the formula of the power function and their graph, we used the power function on case of the Binomial distribution (see Pratikno[2]), in testing H_0: $p=p_0=0.6$versus H: $p>0.6$, with rejection area $R=\{(x_1,x_2,...,x_i):Y\leq 4\}$ for several n=7,9,20 and 30. The formula of the power function of this distribution is given as $\pi(p)=P(\text{reject } H_0 | \text{Under } H_1: p=p_1)=\sum_{y=0}^{4}\binom{n}{y}p_1^y(1-p_1)^{9-y},\ p_1>0.6$, and their simulation graphs are then given at Figure 1.

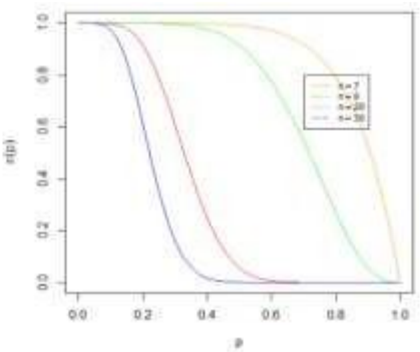


Figure 1. The power of Binomial distribution on several *n*.

From Figure 1, we see that the curve is sigmoid and tend to be zero more faster for large *n*. From the previous research, we then choose the blue (*n=30*) and red (*n=20*) curves are more suitable curve then others. They are quickly to be one for small *p*, we therefore recommend them as the significant curves of the power function of the Binomial distribution.

4. Research Methods

Step 1. We studied and simulated the pdf and cdf curves of the Lognormal Distributions.

Step 2. Following to the previous research, we then derived the formula of the power andsize of the Lognormal Distribution. Firstly, we find the sufficiently statistics. We then checked the monotonic maximum likelihood ratio (MLR) of the sufficientlystatistics. Finnaly, we use the *uniformly most powerful test* (UMPT) to get the rejection area (RR).

Step 3. Based on step 2, we then produced graphically analysis of the power function andsize

Step 4. The conclusion is drawn by choosing the maximum power and minimum size.

5. RESULTS

5.1. The Graphs of the pdf and cdf of the Lognormal Distribution

Following Hines, et al.[25], Bain and Engelhardt[7] and Balakrishnan and Lai[8], the pdf formula of the random variable *X* of the Lognormal distribution (as bell-shape curve) is given as

$$f(x;\mu,\sigma)=\frac{1}{\sqrt{2\pi}\sigma x}e^{-\frac{1}{2\sigma^2}[\ln(x)-\mu]^2},x\geq 0 \qquad \textbf{(1)}$$

with parameter $-\infty<\mu<\infty$, and $0<\sigma<\infty$, and the *X* is then notated as $X\sim LOGN(\mu,\sigma2)$. Taking the equation (1) into *ln*, we then get $\ln X\sim N(\mu,\sigma2)$. The cdf of the lognormal distribution of $X\sim LOGN(\mu,\sigma2)$,is then written as

$$F_X(x)=P[X\leq x]=P[\ln X\leq \ln x]=\Phi\left(\frac{\ln x-\mu}{\sigma}\right) \qquad (2)$$

where Φ is the cdf of the normal standard, *X* is *random* variable of the lognormal distribution, and ln*X* is random variable of the normal distribution. Using the equation (1), (2) and *R-code*, the graphs (curves) of their pdf and cdf are then presented in Figure 2.

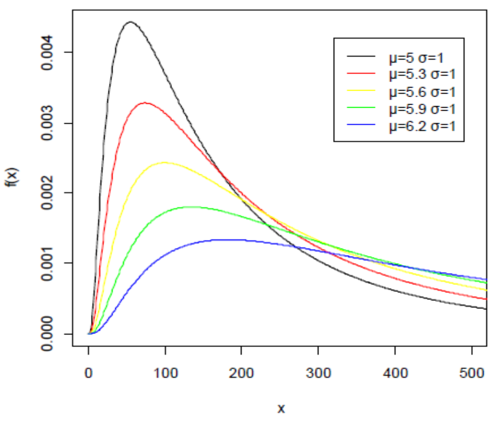


Figure 1.a. The pdf curve on several μ

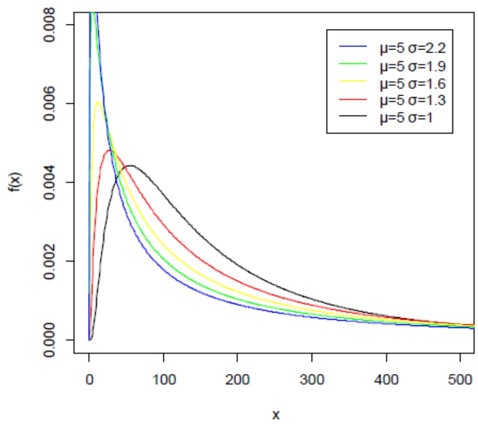
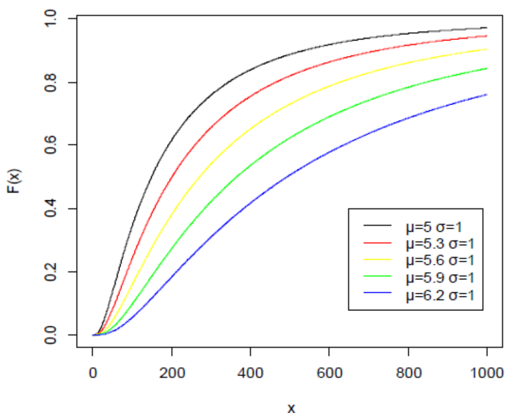


Figure 1.b. The pdf curve on several σ



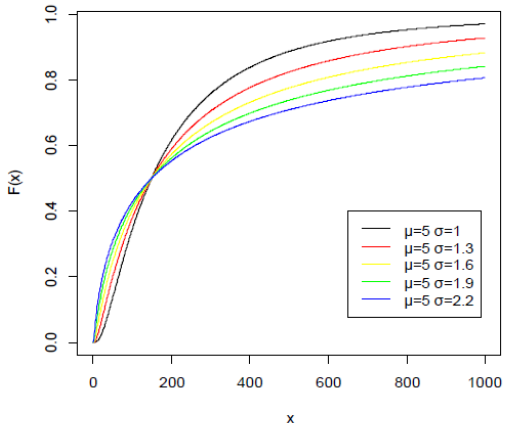


Figure 2.a. The cdf curve on several μ **Figure 2.b. The cdf curve on several σ**

5.2. The Power and Size of the Lognormal Distribution

Following the technique in deriving the formula of the power function the Section 2, with an example on the binomial distribution case, and also referring to many previous research have already discussed the power of the hypothesis testing on the continuous distribution (but not in lognormal distributions). We then focused to conduct graphically analysis of the pdf, cdf and its power-size on lognormal distribution. The procedure to derive the formula of the power are: (1) determine the sufficiently statistics, (2) create the rejection area using *uniformly most powerful test* (UMPT), (3) derive the formula of the power of the lognormal distribution, and (4) finally simulate graphically analysed of the power using generate data on *R-code*.

$$f(x_1,\dots,x_2|\mu)=\prod_{i=1}^{n} f(x_i|\mu)=\prod_{i=1}^{n}\left(\frac{1}{\sqrt{2\pi}\sigma x_i}e^{-\frac{1}{2\sigma^2}[\ln(x_i)-\mu]^2}\right)=\frac{1}{\prod_{i=1}^{n} x_i\left(\sqrt{2\pi}\sigma\right)^n}e^{\frac{-\sum_{i=1}^{n}(\ln x_i)^2}{2\sigma^2}}e^{\frac{\sum_{i=1}^{n}\ln x_i\mu}{\sigma^2}-\frac{n\mu^2}{2\sigma^2}}. \quad (3)$$

Let, $s=\sum_{i=1}^{n}\ln X_i$, we then get $g(s|\mu)=e^{\frac{s\mu}{\sigma^2}-\frac{n\mu^2}{2\sigma^2}}$, where $h(x)=\frac{1}{\prod_{i=1}^{n} x_i\left(\sqrt{2\pi}\sigma\right)^n}e^{\frac{-\sum_{i=1}^{n}(\ln x_i)^2}{2\sigma^2}}$ and

$s=\sum_{i=1}^{n}\ln X_i \sim N(n\mu, n\sigma^2)$, is called sufficient statistics, with $\lambda(s)$ is monotone likelihood ratio (MLR). Using UMP test in testing $H_0:\mu=\mu_0$ versus $H_1:\mu>\mu_0$, we then reject H_0 when $\sum_{i=1}^{n}\ln x_i \geq Z_\alpha\sigma\sqrt{n}+n\mu_0$. Based on this rejection area, we then derive the formula of the power as follows

$$\pi(\mu)=P(\text{Reject } H_0|\mu)=P\left(\sum_{i=1}^{n}\ln x_i \geq Z_\alpha\sigma\sqrt{n}+n\mu_0\middle|\mu\right)=1-P\left(\frac{\sum_{i=1}^{n}\ln x_i-n\mu}{\sigma\sqrt{n}}<Z_\alpha+\frac{\sqrt{n}(\mu_0-\mu)}{\sigma}\right). \quad (4)$$

Using the equation (4), the graphs of the power function of the lognormal distribution are presented in Figure 3.

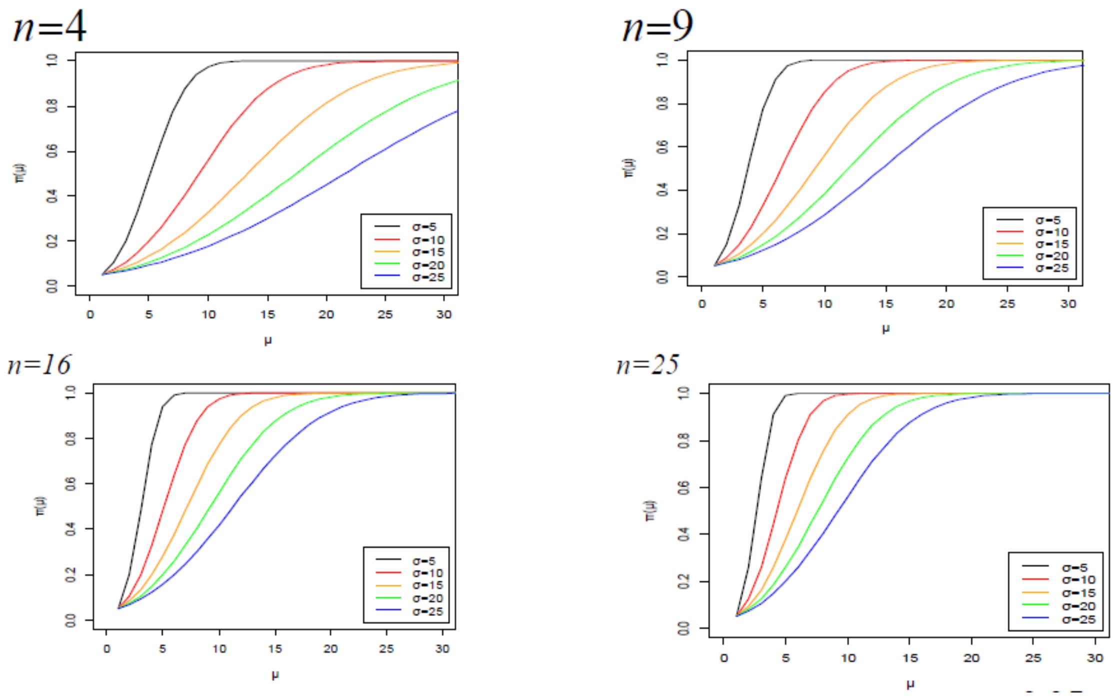
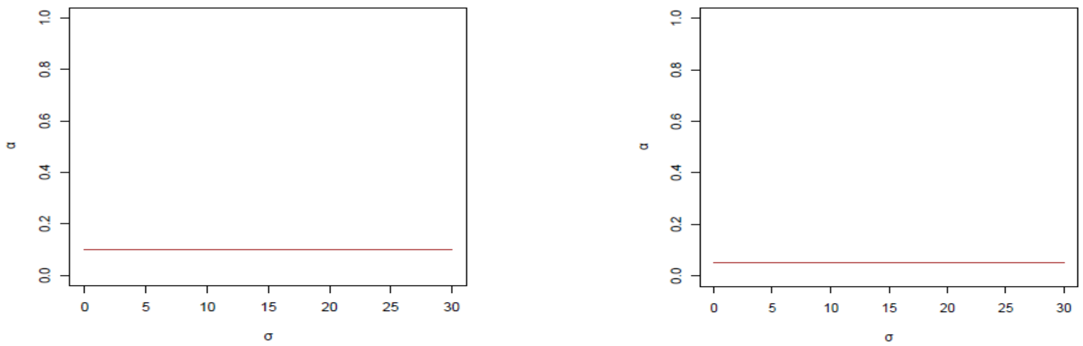


Figure 3: The power curve of the lognormal distribution at α=0.05.

We see from Figure 3. that the curves decrease as the σ increase, but they are more faster to be one when the *n* increases. Therefore, we note that both *n* and σ significantly affect to the skew-ness of their curves. Here, we also compute (and plot) the value of the size with different α*, α=0.01 *and* 0.05.Both size are constant, and they are different values, namely 0.10 and 0.049.Following the rule of the previous research, we must choose the lower value of the size (usually close to the level of the significance,α).The graphs of the size are then given in Figure 4.



(a) α=0.01 (b) α= 0.05

Figure 4: The size of the tests on α=0.01 & 0.05.

From Figure 4., it is clear that the lower (small) size is around 0.049 and occurred on *alpha* is 0.05, but not on *alpha* 0.01, the size is around 0.10. It means that we prefer to choose the minimum size when *alpha* is 0.05.

6. CONCLUSION

To derive the power of the lognormal distribution, we must consider sufficient statistics and *UMP test* for getting the rejection area. The result showed that the curves decrease as the increases, but they are more faster to be one when the *n* increases. The size is always constant, and the eligible size is 0.049 close to level of significance, and it is occurred when α=0.05.

ACKNOWLEDGMENT

I thankfully to the LPPM UNSOED for providing me granting of research. I also thank to colleague for providing me the ideas and data.

REFERENCES

1. A. K. Md. E. Saleh. Theory of preliminary test and Stein-type estimation with applications. John Wiley andSons, Inc., New Jersey, 2006.
2. B. Pratikno. Test of Hypothesis for Linear Models with Non-Sample Prior Information. Unpublished PhD Thesis, University of Southern Queensland, Australia, 2012.
3. B. Pratikno. The noncentral *t* distribution and its application on the power of the tests. *Far East Journal ofMathematical Science (FJMS),* 2018; 106(2): 463-474.
4. B. Pratikno, Power of hypothesis testing parameters shape of the distributions. *Far East Journal ofMathematical Science (FJMS)*, 2019; 110(1): 15-22.
5. D. D. Wackerly, W.Mendenhall III, and R. L.Scheaffer. Mathematical statistics with application, 7th Ed.Thomson Learning, Inc., Belmont, CA, USA, 2008.
6. R. M. Yunus. Increasing power of M-test through pre-testing. Unpublished PhD Thesis, University of SouthernQueensland, Australia, 2010.
7. L.J. Bain and M. Engelhardt 1992. *Introduction to Probability and Mathematical Statisttics,* 2th ed. USA:Duxbury Press., Belmont, CA.
8. N. Balakrishnan and C.D. Lai. 2009. *Continous Bivariate Distributions*, 2th ed. USA: Springer.
9. R. M. Yunus and S. Khan. Test for intercept after pre-testing on slope a robust method.

In: 9th Islamic Countries Conference on Statistical Sciences (ICCS-IX): Statistics in the Contemporary World - Theories, Methods and Applications, 2007.

10. R. M. Yunus and S. Khan. Increasing power of the test through pre-test a robust method. Communications in Statistics-Theory and Methods, 2011a; 40: 581-597.
11. R. M. Yunus and S. Khan. M-tests for multivariate regression model. Journal of Nonparamatric Statistics, 2011b; 23: 201-218.
12. R. M. Yunus and S. Khan. The bivariate noncentral chi-square distribution - Acompound distribution approach. Applied Mathematics and Computation, 2011c; 217: 6237-6247.
13. S. Khan, B. Pratikno, A.I.N. Ibrahim and R.M Yunus, The correlated bivariate noncentral F distribution and Its application. Communications in Statistics—Simulation and Computation, 2016; 45: 3491–3507.
14. S. Khan. Estimation of the Parameters of two Parallel Regression Lines Under Uncertain Prior Information. Biometrical Journal, 2003; 44: 73-90.
15. S. Khan. Estimation of parameters of the multivariate regression model with uncertain prior information and Student-t errors. Journal of Statistical Research, 2005; 39(2): 79-94.
16. S. Khan. Shrinkage estimators of intercept parameters of two simple regression models with suspected equal slopes. Communications in Statistics - Theory and Methods, 2008; 37: 247-260.
17. S. Khan. and A. K. Md. E. Saleh. Preliminary test estimators of the mean based on p-samples from multivariate Student-t populations. Bulletin of the International Statistical Institute. 50th Session of ISI, Beijing, 1995; 599-600.
18. S. Khan. and A. K. Md. E. Saleh. Shrinkage pre-test estimator of the intercept parameter for a regression model with multivariate Student-t errors. Biometrical Journal, 1997; 39: 1-17.
19. S. Khan. and A. K. Md. E. Saleh. On the comparison of the pre-test and shrinkage estimators for the univariate normal mean. Statistical Papers, 2001; 42(4): 451-473.
20. S. Khan. , Z. Hoque and A. K. Md. E. Saleh. Improved estimation of the slope parameter for linear regression model with normal errors and uncertain prior information. Journal of Statistical Research, 2002; 31(1): 51-72.
21. S. Khan. and Z. Hoque. Preliminary test estimators for the multivariate normal mean based on the modified W, LR and LM tests. Journal of Statistical Research, 2003; 37: 43-55.
22. S. Khan. and A. K. Md. E. Saleh. Estimation of intercept parameter for linear regression with uncertain non- sample prior information. Statistical Papers, 2005; 46: 379-394.

23. S. Khan. and A. K. Md. E. Saleh. Estimation of slope for linear regression model with uncertain priorinformation and Student-*t* error. Communications in Statistics - Theory and Methods, 2008; 37(16): 2564-258.
24. S. Khan and B. Pratikno, Testing Base Load with Non-Sample Prior Information on Process Load. Statistical Papers, 2013; 54(3): 605-617,.
25. W.W. Hines, D.C. Montgomery, D. M. Goldsman, and C. M. Borror. 2003. *Probability and Statistics in Engineering's. USA*: John Willey & Sons, Inc.