The Power of Hypothesis Parameters Testing on Geometric Distributions and Its Simulation

Budi Pratikno and Evita Luaria Wulandari
Department of Mathematics, Faculty of Mathematics and Natural Sciences,
Universitas Jenderal Soedirman, Purwokerto, Indonesia
budi.pratikno@unsoed.ac.id, evita.wulandari@mhs.unsoed.ac.id

Abdul Talib Bon
Department of Production and Operations,
University Tun Hussein Onn Malaysia, Malaysia
talibon@gmail.com

Abstract

The research studied power of hypothesis parameters testing on geometric distributions. To derive and compute the power of this distribution, we must determine the sufficiently statistics and compute the rejection area using *uniformly most powerful test* (UMPT). The *R-code* is then used to figure the graph of the power, we then did graphically analysis of the graph. The result showed that the curves of the power are depended on the value of the rejection area.

Keywords: *The power, parameter testing, rejection area, and R-code*

1. Introduction

Population inference is drawn by sample, and it could be improved using testing parameter on their distribution (Sirait et al., 2020a; Sirait et al., 2020b). Here, we should use the power and size to testing the hypothesis parameter. Note that in the hypothesis testing, a probability error type I (α), a probability error type II (β) and a **power** are really significant useful. In this case, the power is defined as a probability to reject H_0 *under* H_1 in testing hypothesis, $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$, for parameter θ (Wackerly, et al., 2008). We then used it as a statistical techniques to investigate the population inference. Following Wackerly, et al. (2008) and Casella and Berger (2002), we noted more detail the definition of the power and size, namely, (1) the power is the probability to reject H_0 under H_1 and (2) the size is the probability to reject H_0 under H_0.

Following the previous research, many authors studied the improving inference population using power and size such as Pratikno et al. (2019, 2020), Pratikno (2012), Khan and Pratikno (2013), Khan (2003), Khan and Saleh (1995), Khan and Hoque (2003), Saleh (2006), Yunus (2010), and Yunus and Khan (2007). Besides that, Pratikno [2012], Khan and Pratikno [2013] and Khan [2003] studied also studied a testing intercept using non-sample prior information (NSPI). Moreover, Pratikno [2012] and Khan et al. [2016] used the power and size to compute the cdf of the bivariate noncentral *F (*BNCF*)* distribution of the pre-test test (PTT) with NSPI in some various regression models. Due to, the complicated computational of the probability integral of the probability distribution function (pdf) and cumulative distribution function (cdf) of the BNCF distribution, the R code is used. In the context of the power on the distribution, Pratikno et al. (2019, 2020), are already presented the formula and its value of the power on discrete and continuous distributions.

Furthermore, in this research, we then focused to analysis about the power and size on Geometric distribution. Here, we have several steps to compute the power of the distribution: (1) determine the sufficiently statistics, (2) compute the rejection area using *uniformly most powerful test* (UMPT), (3) derive the formula of the power of the distributions in testing hypothesis, and (4) plot the graphs of the power of their distributions.

In this paper, introduction is presented in Section 1. The methodology of the research is then given in Section 2. A simulation and graphically analysis of the power and size is obtained in Section 3, and Section 4 described the conclusion of the research.

2. Research Methodology

Step 1. We studied the basic concept of Negative Binomial and Geometric Distribution.
Step 2. We then refer to previous research, to derived the formula of the power and size of the Geometric Distribution using sufficiently statistics and uniformly *most powerful test* (UMPT) in obtaining the rejection area (RR).
Step 3. Based on step 2, we then produced and did graphically analysis of the graph of the power and size
Step 4. Finally, we draw the conclusion of the characteristics of the graph and choose the maximum power and minimum size.

3. Discussion Results

3.1 The Power on the Geometric Distribution

Following Pratikno et al. (2020), the probability mass function *(pmf)* of the Binomial distribution with X_i Bernoulli trials with parameter θ, $n = 20$, and $Y = \sum_{j=1}^{n=20} X_j$, $Y : B(n,\theta)$, is given as

$$f(x_i) = \binom{n}{x}\theta^x(1-\theta)^{n-x} \quad . \tag{1}$$

For testing the two-side hypothesis, $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$ on rejection area $\{(x_1,...,x_{12}) : Y \leq c\}$, we then derive the formula of the power and size are given as, respectively,

$$\begin{aligned}\pi(\theta) &= P(\text{Reject } H_0 \,|\, \text{under } H_1) = \sum_{y=0}^{c}\binom{20}{y}\theta^y(1-\theta)^{20-y} \\ &= \binom{20}{0}\theta^0(1-\theta)^{20-0} + \binom{20}{1}\theta^1(1-\theta)^{20-1} + \text{L} + \binom{20}{c}\theta^c(1-\theta)^{20-c}\end{aligned} \tag{2}$$

$$\begin{aligned}\alpha(\theta) &= P(\text{Reject } H_0 \,|\, \text{under } H_0) = \sum_{y=0}^{c}\binom{20}{y}\theta^y(1-\theta)^{20-y}\Big|_{\theta=\theta_0} \\ &= \binom{20}{0}\theta_0^0(1-\theta_0)^{20-0} + \binom{20}{1}\theta_0(1-\theta_0)^{20-1} + \text{L} + \binom{20}{c}\theta_0^c(1-\theta_0)^{20-c}\end{aligned} \tag{3}$$

From the equation (2), it is clear that the graph of the power is depended the parameter shape, and they will be sigmoid curve, and the curves tend to be one for large parameter shape . Similarly, from the equation (3), it is clear that the graph of the size is constant and they will then be straight line. Here we noted that the curve of the size is not depended on parameter shape.

In this research, we present the probability mass function of X as random variable of the negative Binomial distribution of success on n Bernoulli trials, with r success on x, $X : BN(r,\theta)$, as

$$p(x) = \binom{x-1}{r-1}\theta^{r}(1-\theta)^{x-r}, \quad x = r, r+1, ... \tag{4}$$

For $r = 1$, the distribution of the random variable X be Geometrik distribution, and it is then written as $X : BN(r,\theta) = Geo(r,\theta),\ r = 1$, with *pmf* is given as

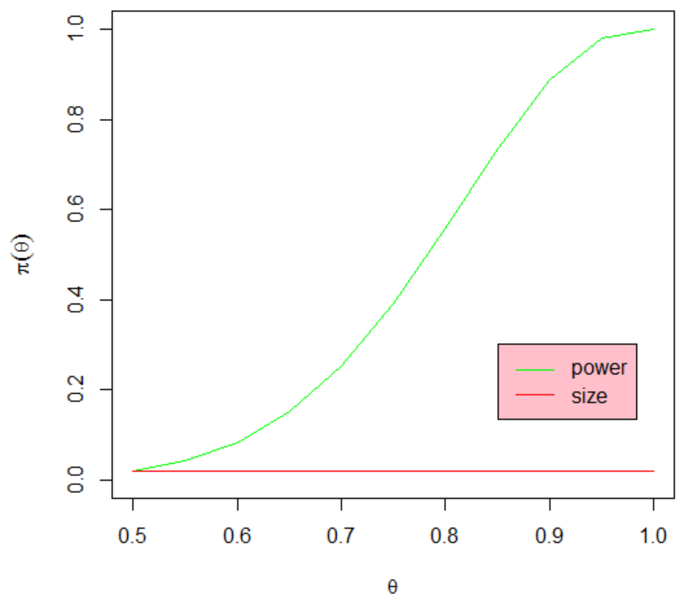
$$\begin{aligned} p(x) &= P(X = x) \\ &= \theta(1-\theta)^{x-1}, \quad x = 1, 2, 3, ... \end{aligned} \tag{5}$$

Similarly, following the equation (2), (3) and previous research, the power and size of the Geometric distribution are then derived in testing the two-side hypothesis, $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$, with $S = \sum_{i=1}^{n} X_i$ sufficiency statistics as a rejection area bound less than **12**, are

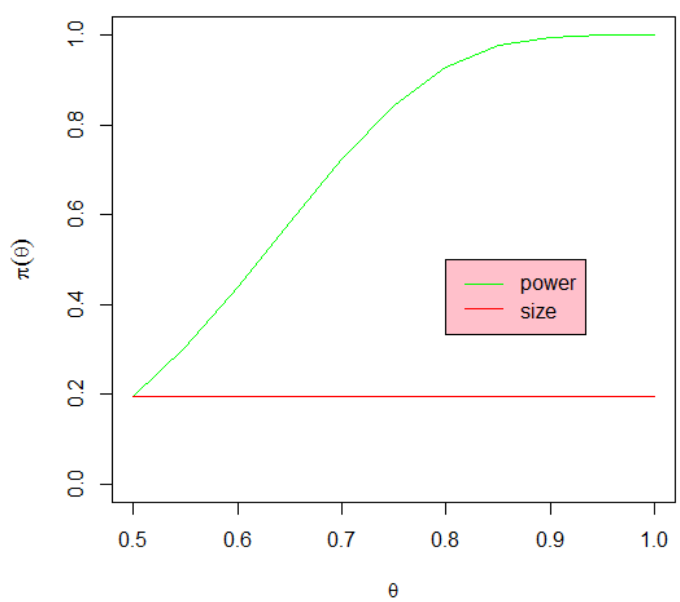
$$\begin{aligned} \pi(\theta) &= P(\text{Re}\,ject \quad H_0 \big| H_1 \ \ |_{\theta}) \\ &= P(s \leq 12 \big| s \sim BN(n,\theta)) \\ &= \sum_{s=10}^{12} \binom{s-1}{10-1} \theta^{10}(1-\theta)^{s-10} \end{aligned} \tag{6}$$

$$\begin{aligned} \alpha(\theta) &= P(\text{Re}\,ject \quad H_0 \big| H_0 \ \ |_{\theta_0}) \\ &= P(s \leq 12 \big| s \sim BN(n,\theta_0)) \\ &= \sum_{s=10}^{12} \binom{s-1}{10-1} (\theta_0)^{10} (\theta_0)^{s-10}, \ \theta_0 \text{ is determined} \end{aligned} \tag{7}$$

Following Pratikno et al. (2020) and using the equation (6) and (7), we then produced the graphs of the power of the Geometric distribution. Their graphs are then given in Figure 1 (a), (b) and (c).

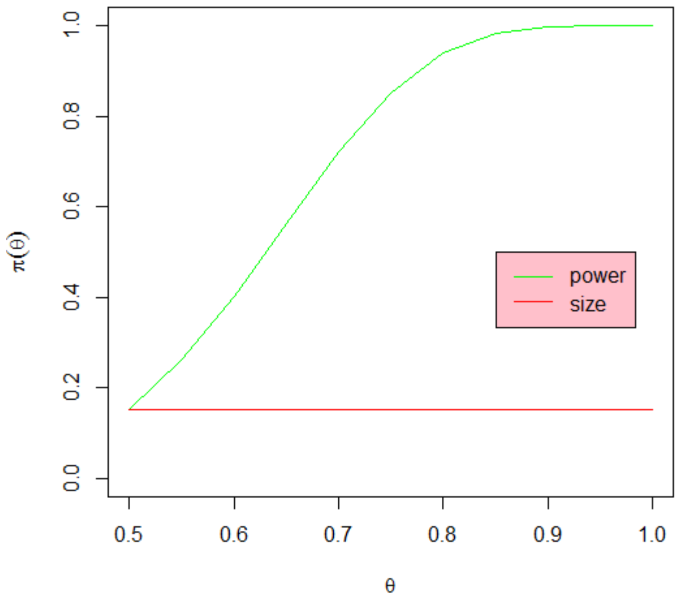


(a)



(b)

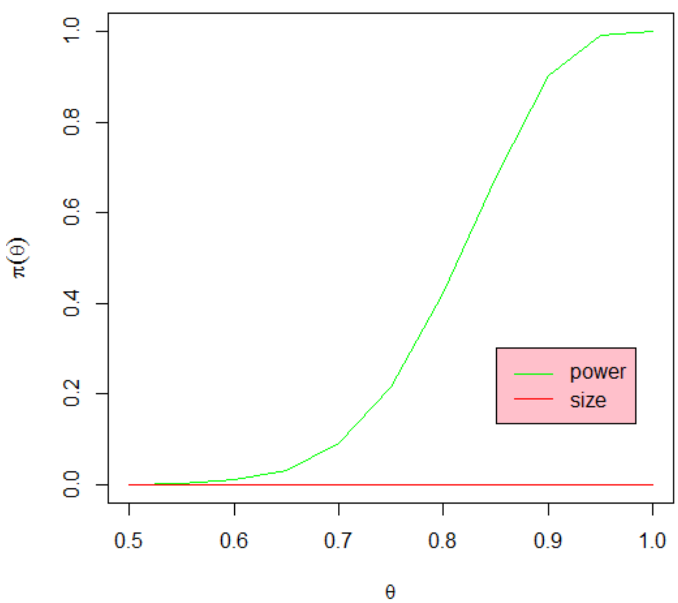
Figure 1 (a) and (b). The Power and Size of the Geometric distribution with $s \le 12$, $n = 10$ (a) and $n = 8$ (b)



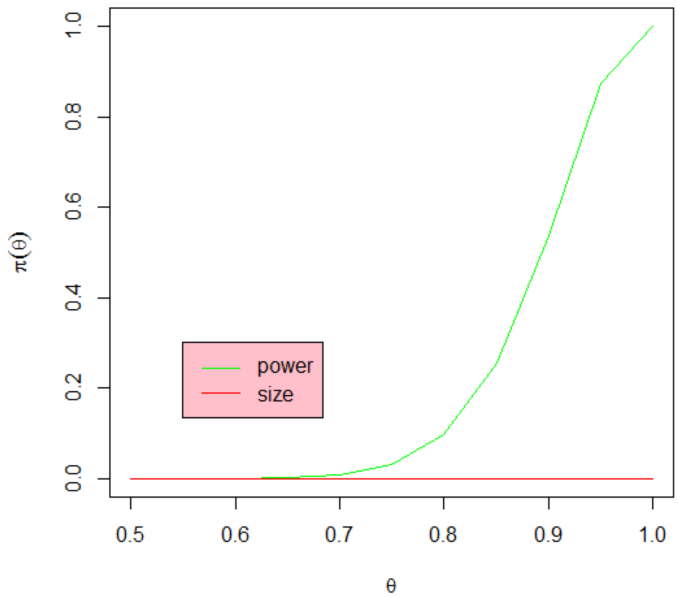
(c)

Figure 1 (c). The Power and Size of the Geometric distribution with $s \leq 15$, $n = 10$

Similarly, using the equation (6) and (7), we then also simulate for several graphs of the power and size on different $n = 21$ and 23 and $S=25$ and 27. The curves are then presented in Figure 2 (a), (b) and (c),

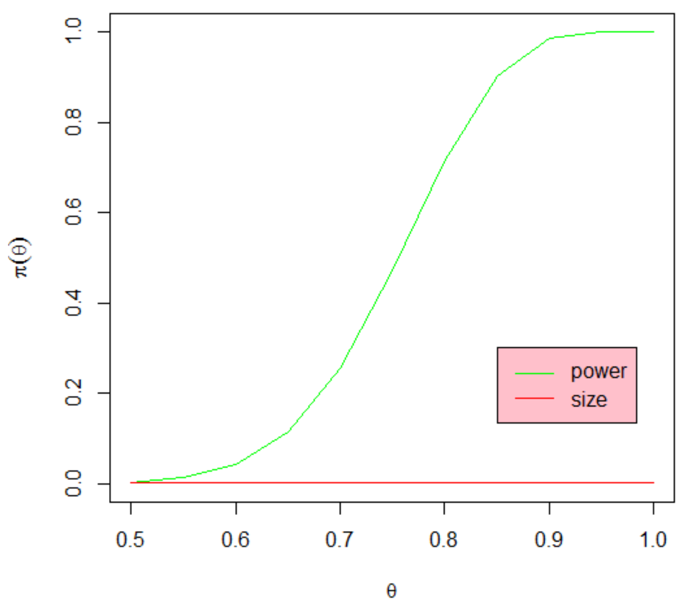


(a)



(b)

Figure 2 (a) and (b). The Power dnd Size of the Geometric Distribution with $s \le 25$, $n = 21$ (a) and $n = 23$ (b)



(c)

Figure 2 (c). The Power and Size of the Geometric distribution with $s \le 27$, $n = 21$

3.2 The Graphically Analysis

From Figure 1 (a) and (b), we see that the curves of the power are similar (sigmoid), the curve on the graph (b) more faster to be one than (a), it will be in short period on the parameter, that is $0 \leq \theta \leq 0.8$, but not in (a) $0 \leq \theta \leq 0.95$. They increase on $0.2 \leq \theta \leq 0.9$ (sigmoid), and then they tend to be flat (asymptote to one) on $\theta \geq 0.9$. This is due to the different value of *n*. Generally, we note that they (both of the graphs) will be maximum on one, and the minimum is zero. Following the previous research of Pratikno (2012), Khan and Pratikno (2013), Khan (2003), Khan and Saleh (1995), Khan and Hoque (2003), Saleh (2006), Yunus (2010), and Yunus and Khan (2007), we noted that the maximum power (one) and minimum size are chosen for the better indicator in testing. Note that both size are constant, and their values are *0.01 (n=10)* and *0.20 (n=8)*. We therefore conclude that the power on *n=8* is better than *n=10*. Furthermore, we did the power on $S \leq 15, n = 10$, Figure 1 (c) , the graph tend to be quick to be one in short period, $0 \leq \theta \leq 0.75$ and the size close to 0.20. It means that the *n* and *S* are really significant to change the curve of power and size.

We see from Figure 2 (a) and (b), the curves of the power are also sigmoid, the curve on the graph (a) more faster to be one than (b), their period on the parameter is around $0.95 \leq \theta \leq 1.0$. They increase on $0.65 \leq \theta \leq 1.0$ (sigmoid), and then they tend to be flat (asymptote to one) on $\theta \geq 0.95$. This is due to the different value of *n*. Generally, we also note that they will be maximum on one, and the minimum is zero. Following the previous research of Pratikno (2012), Khan and Pratikno (2013), we already noted that the maximum power (one) and minimum size are chosen for the better indicator in testing. Similarly, the both size are constant, and their values are close to *0.0* for both *n*. We therefore conclude that the power on *n=21* is better than *n=23.* Furthermore, we did the power on $S \leq 27, n = 21$, Figure 2 (c) , the graph tend to be quick to be one in short period, $0 \leq \theta \leq 0.85$ but the size is still close to 0.0. It means that the *n* and *S* are really significant to influence the graph of the power and size.

Noted that from the equation (6) and (7), the both formula of the power and size are depended *S* and *n*, I therefore conclude that the Figure 1 and 2 are really evidence that their graphs are really similar and suitable .

4 Conclusion

The research studied power and size on Geometric distribution. The rejection area is determined to find the formula of the power. Here, the *R-code* is also used to figure and graphically analysis of the curves (power and size). The result showed that the sample size *n* and rejection area of sufficiency statistics *S* are really significant to change the curves of the power and size.

References

Casella, G. dan Berger, R. L. (2002). *Statistical Inference, 2nd* ed. USA: Thomson Learning Inc.

Khan, S. and Pratikno, B. Testing Base Load with Non-Sample Prior Information on Process Load. *Statistical Papers*, 54 (3), 605-617, (2013).

Khan, S. Estimation of the Parameters of two Parallel Regression Lines Under Uncertain Prior Information. *Biometrical Journal, **44**, 73-90*, (2003).

Khan*, S.* and Saleh, A. K. Md. E. 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, 599-600*, (1995).

Khan, S. and Hoque, Z. Preliminary test estimators for the multivariate normal mean based on the modified W, LR and LM tests. *Journal of Statistical Research, Vol 37, 43-55, (*2003).

Khan, S., Pratikno, B. Ibrahim, A.I.N., and Yunus, R.M. The correlated bivariate noncentral *F* distribution and Its application. *Communications in Statistics—Simulation and Computation*, 45 3491–3507 (2016).

Pratikno, B, Parwati, M.S.R and Bon, A.T. (2020). The Decision Case on Gaussian Binary Data*Proceedings of the 5th NA International Conference on* Industrial Engineering and Operations Management Detroit, Michigan, USA, August 10-14, 2020, IEOM Detroit Conference, Link proceedings papers or abstracts here : http://ieomsociety.org/detroit2020/proceedings/

Pratikno, B ., Layyinah, S. Y., Pratidina, G. M., dan Suryaningtiyas, Y. D. (2019). *The Power of Weibull and Exponential Distribution On Testing Parameters Shape*. Journal of IOP Conference Series Vol 255(1), pp 012029.

Pratikno, B.. Test of Hypothesis for Linear Models with Non-Sample Prior Information. Unpublished PhD Thesis, University of Southern Queensland, Australia, (2012).

Saleh, A. K. Md. E. Theory of preliminary test and Stein-type estimation with applications. *John Wiley and Sons, Inc., New Jersey,* (2006).

Sirait, H., Sukono, Sundari, S., and Kalfin. (2020a). Ratio Estimator of Population Mean using Quartile and Skewness Coefficient. *International Journal of Advanced Science and Technology,* 29(6), pp. 3289-3295
Proceedings of the

Sirait, H., Karolin, S., Sukono, Kalfin, Bon, A.T. (2020b). Ratio estimator for population variations using additional information on simple random sampling. *International Conference on Industrial Engineering and Operations Management* (Detroit Michigan, USA)

Wackerly, D. D., Mendenhall, W., dan Scheaffer, R. L. (2008). *Mathematical Statistics with Applications*, *7th ed.* USA: Thomson Learning, Inc.

Yunus, R. M. Increasing power of M-test through pre-testing. *Unpublished PhD Thesis, University of Southern Queensland, Australia*, (2010).

Yunus ,R. M. and Khan, S. 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).

Acknowledgement
I thankfully to the colleague for providing me the ideas and data.

Biographies

Budi Pratikno, currently works as a lecturer in Department of Mathematics, Universitas Jenderal Soedirman, author of several international journals related to testing intercepts with non-sample prior information (NSPI), which are widely published in international journals such as Statistical Papers, Far East Journal of Mathematics Science, Springer, JSTA, ISSOS, IJET, IJAST, IOP series, IEOM and others. In the study of NSPI as the main research the author, we try to improve population inference using non-sample. The author completed his undergraduate education at the Mathematics Department (UGM), then continued his Masters in Statistical Science at La Trobe University, Melbourne, Australia, and his S3 in Statistical Science (Ph.D) at the University of Southern Quensland (USQ), Toowoomba, Brisbane, Australia.

Mashuri, currently works as a lecturer in Department of Mathematics, Universitas Jenderal Soedirman, author of several international journals related to mathematical dynamic, which are widely published in international journals such as Far East Journal of Mathematics Science, Advances and Applications in Fluid Mechanics, Far East Journal of Applied Mathematics and others. The author completed his undergraduate education (S.Si) at the Mathematics Department (UNPAD), then continued his Masters (M.Si.) and doctorate program (Dr.) in Mathematics at ITB Bandung, Indonesia.

Kurniasih, A. and Azizah, N. are students in Department of Mathematics, Universitas Jenderal Soedirman.

Abdul Talib Bon is a professor of Production and Operations Management in the Faculty of Technology Management and Business at the Universiti Tun Hussein Onn Malaysia since 1999. He has a PhD in Computer Science, which he obtained from the Universite de La Rochelle, France in the year 2008. His doctoral thesis was on topic Process Quality Improvement on Beltline Moulding Manufacturing. He studied Business Administration in the Universiti Kebangsaan Malaysia for which he was awarded the MBA in the year 1998. He's bachelor degree and diploma in Mechanical Engineering which his obtained from the Universiti Teknologi Malaysia. He received his postgraduate certificate in Mechatronics and Robotics from Carlisle, United Kingdom in 1997. He had published more 150 International Proceedings and International Journals and 8 books. He is a member of MSORSM, IIF, IEOM, IIE, INFORMS, TAM and MIM.