

The Graphs of the CDF, Power and Its Interpretation on Several Types of Binomial Probability Distribution

Budi Pratikno^{1*}, Evita Luaria Wulandari¹, Jajang Jajang¹, Junita Sage Sianipar¹, and Mashuri Mashuri¹

¹Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Jenderal Soedirman, Purwokerto, Indonesia

*Corresponding author. Email: budi.pratikno@unsoed.ac.id

ABSTRACT

The research discussed the graphically analyzed of the cumulative distribution function (cdf), and the power function of hypothesis testing on the binomial distribution. In this research, we also showed (derived) the formula of the power function on special case of binomial such us Negative Binomial and the Geometric distribution. The result showed that the degree of freedom, bound of the rejection area, and parameter shape significantly affect to the curves of the power function. The curves of the power are sigmoid and they increase quickly to be one on the small parameter shape and large degree of freedom.

Keywords: Binomial distribution, the power function of hypothesis testing, R-code

1. INTRODUCTION

Here, we interested to study the use of the power and size in testing the hypothesis parameter and their graphically analyse for improving the inference population (Pratikno, [2]). Following, Wackerly, et. al. [5], we note that there are three definitions related to the hypothesis testing, namely a probability error type I (α), a probability error type II (β) and a **power function**. In addition, 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 , as a statistical technique to investigate the population inference. Moreover, the size is then defined as the probability to reject H_0 under H_0 . We then choose the maximum power and minimum sizes as the theoretical concept to compare the testing.

Many authors such as, Pratikno [2], Khan [12-14], Khan and Saleh [15,16,17, 20, 21], Khan and Hoque [19], Saleh [1], Yunus [6], and Yunus and Khan [7-10], already studied the power and size of the tests on the hypothesis testing. Furthermore, we noted from the previous research some authors studied the power in testing intercept with non-sample prior information (NSPI), such as Pratikno [2], Khan and Pratikno [20] and Khan [12]. They used the probability integral of the cumulative distribution function (cdf) to compute the power and size. Moreover, Pratikno [2] and Khan et al. [9] used the formula of the power to compute the cdf of the bivariate noncentral F (BNCF) distribution in

regression models. Others authors have also contributed to the research of the power in the context of the hypothesis testing, such as Khan [12-14], Khan and Saleh [15,16,17, 20, 21], Khan and Hoque [17], Saleh [1], Yunus [6], and Yunus and Khan [7-10]. Due to the complicated and hard computational, Pratikno [2] and Khan et al. [9] used the BNCF distribution to compute the power using *R-code* (see Pratikno [2] and Khan et.al. [18]).

The research methodology for investigating the power and size as follows: (1) we have to determine the sufficiently statistics, (2) we then create the rejection area using uniformly most powerful test (UMPT) to derive the formula of the power of the geometric distribution, and (3) we finally compute and figure the graphs using *R-code*.

In this paper, Section 1 presented the introduction. The graphically analyzed of the cdf of the Binomial, Negative Binomial and Geometric distributions, and the power function are given in Section 2. The conclusion is given in Section 3.

2. THE GRAPHICALLY ANALYZED OF THE PDF, CDF, AND THE POWER-SIZE OF THE BINOMIAL AND GEOMETRIC DISTRIBUTION

2.1. The cdf Graphs of the Binomial, Negative Binomial and Geometric Distributions

Following Pratikno et al. [4], the general probability mass function (pmf) of the **Binomial** distribution with X_i Bernoulli trials with parameter p , and number of trial n , is given as

$$p(x) = P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}; \tag{1}$$

with $\binom{n}{x} = \frac{n!}{n!(n-x)!}$

Similarly, 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 , $X \sim BN(r, p)$, is then presented as

$$p(x) = P(X = x) = \binom{x-1}{r-1} p^r (1-p)^{x-r}, \quad x = r, r+1, \tag{2}$$

For $r=1$, the distribution of the random variable X will be Geometric distribution, $X \sim BN(1, p) = \text{Geo}(r=1, p)$, with pmf is given as

$$p(x) = P(X = x) = p(1-p)^{x-1}, \quad x = 1, 2, 3, \tag{3}$$

Using the equation (1), (2) and (3), we presented the graphs (curves) of the cdf of the three distributions, $P(X \leq x)$ at Figures 1, 2, and 3, respectively.

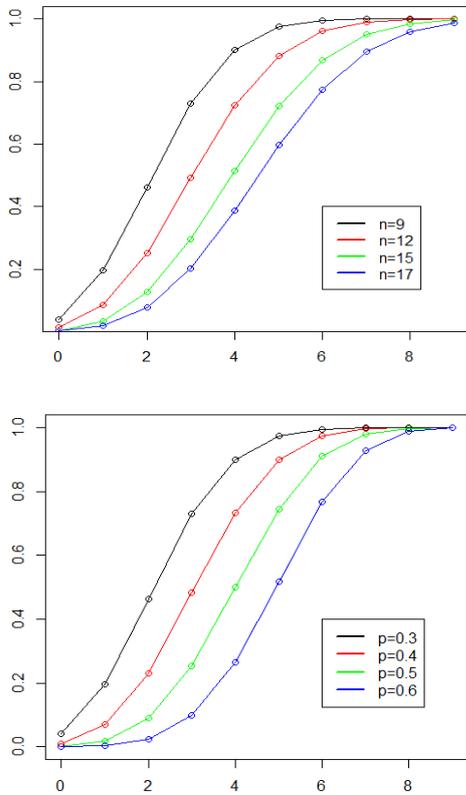


Figure 1. The graphs of the cdf of the binomial distribution on several n and p .

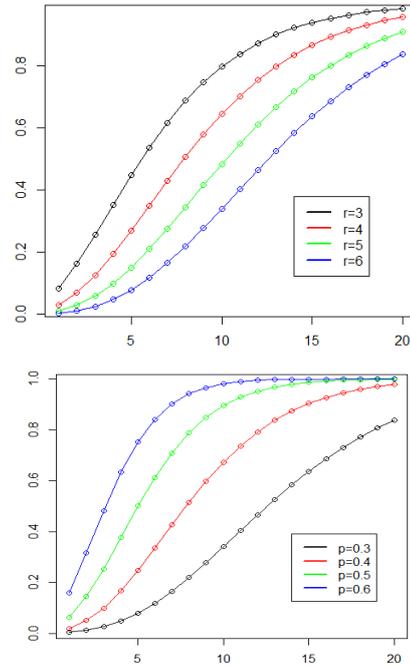


Figure 2. The graphs of the cdf of the Negative binomial distribution on several n and p .

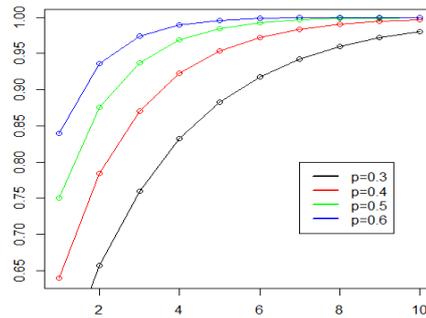


Figure 3. The graphs of the cdf of the Geometric distribution on several p .

From Figure 1, 2, and 3, we see that the all the curves are nonlinear (sigmoid). They increase as for small n and they increase as the p increase. They are going to be one (quickly) for large p (p increases). It means that both n and p really significant affect to the skew-ness of their curves.

2.2. The Graphs of the Power Functions

To derive the formula of the power function, we set the joint distribution of the random variable, X_1, \dots, X_n . Furthermore, we find sufficiency statistics and rejection area to define the power and size. Here, we then got the sufficiency statistics $S = \sum_i^m x_i$ that follows to specified distribution. In this case, we then find the rejection area using *most powerful (MP) test*. Due to the similarity of the distribution among Binomial, Negative Binomial and Geometric distribution, we then obtained the general

power function of the Binomial distribution to testing $H_0 : p = p_0$ versus $H_1 : p > p_0$ as

$$\begin{aligned} \pi(p) &= P(\text{reject } H_0 \mid \text{under } H_1 : p) \\ &= P\left(\sum_{i=1}^m X_i \geq k \mid p\right) = 1 - P(S \leq k-1 \mid p) \end{aligned} \quad (4)$$

Using the equation (4), we then produce the graphs of the power and size function on Binomial, Negative Binomial and Geometric distribution at the Picture of the bottom.

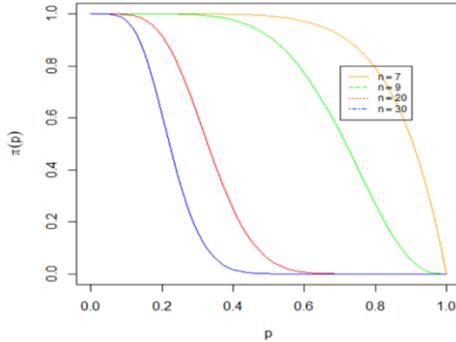


Figure 4. The power of the Binomial distribution with the rejection area 4 on several n

We see from Figure 4., the curves are sigmoid and tend to be quickly to be one for large n and small p .

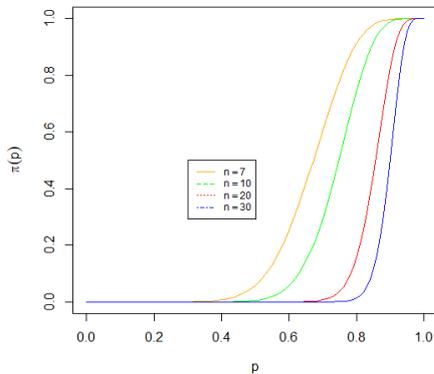


Figure 5. The power of the Negative Binomial distribution with the rejection area 6 on several n

From Figure 5., we see that the curves of the power are also sigmoid, and they are quickly to be zero for small n and p .

Furthermore, we noted that the graphs on Figure 6 are produced using the formula of the power

$$\pi(p) = \sum_{s=3}^6 \binom{s-1}{3-1} p^3 (1-p)^{s-3}$$

in testing $H_0 : p = 0.3$ versus $H_1 : p > 0.3$ with the m values are 3 and 4.

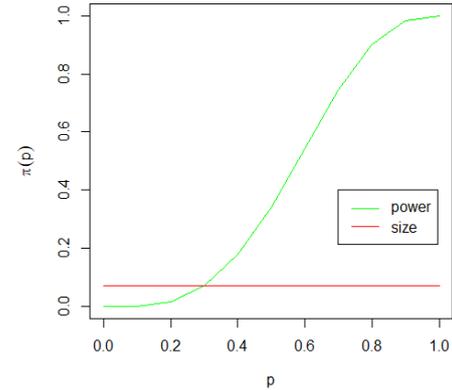
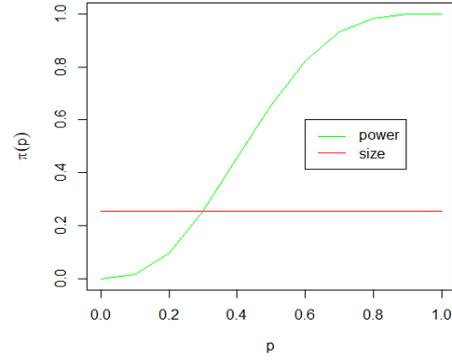


Figure 6. The power and size of the Geometric distribution with $RR = \{(x_1, x_2, \dots, x_m) \mid s \geq 6\}$

We see from Figure 6., it is clear that the power and size increase for small (lower) m and they decrease as the m increase. Note that the curve of the power function is sigmoid, but the size is constant. To evaluate the values of the size, we presented the manually computation form the power function

$$\pi(p) = 1 - \left[\sum_{s=0}^5 \binom{9}{s} p^s (1-p)^{9-s} \right]$$

under H_0 when $p_0=0.3$ and $m=3$,

$$\alpha = \pi(0.3) = 1 - \left[(0.7)^9 + \dots + 126(0.3)^5 (0.7)^4 \right] = 0.026$$

$p_0 = 0.3$ and $m = 4$, $\alpha = \pi(0.3) = 0.07$, respectively.

3. CONCLUSION

There several steps to derive the power function on distributions. The important step is finding the rejection area using UMPT test. The result showed that the power function is depended n and p , and they tend to be sigmoid and quickly to be one for large p (except on Binomial distribution)

AUTHORS' CONTRIBUTIONS

All authors, BP, ELW, JSS, JJ, and MM, have contributions about CONCEPT, METHOD, EDITING, and ANALYSIS. The first author BP provided feedback, discussed result and contributed to the final manuscript.

ACKNOWLEDGMENTS

We thankfully to the LPPM UNSOED for providing us granting of research of the BLU UNSOED, 2021.

REFERENCES

- [1] A. K. Md. E. Saleh. Theory of preliminary test and Stein-type estimation with applications. John Wiley and Sons, 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 of Mathematical Science (FJMS)*, vol. 106, no. 2, 2018, pp. 463-474.
- [4] B. Pratikno, Power of hypothesis testing parameters shape of the distributions. *Far East Journal of Mathematical Science (FJMS)*, 2019, vol. 110, no. 1, pp.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 Southern Queensland, Australia, 2010.
- [7] R. M. Yunus, 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.
- [8] R. M. Yunus, S. Khan. Increasing power of the test through pre-test a robust method. *Communications in Statistics-Theory and Methods*, vol. 40, 2011a, pp. 581-597.
- [9] R.M.Yunus, S. Khan. M-tests for multivariate regression model. *Journal of Nonparametric Statistics*, vol. 23, 2011b, pp. 201-218.
- [10] R. M. Yunus, S. Khan. The bivariate noncentral chi-square distribution - Acompound distribution approach. *Applied Mathematics and Computation*, vol. 217, 2011c, pp. 6237-6247.
- [11] S. Khan, B. Pratikno, A.I.N. Ibrahim, R.M Yunus, The correlated bivariate noncentral F distribution and Its application. *Communications in Statistics—Simulation and Computation*, vol. 45, 2016, pp. 3491–3507.
- [12] S. Khan. Estimation of the Parameters of two Parallel Regression Lines Under Uncertain Prior Information. *Biometrical Journal*, vol. 44, 2003, pp. 73-90.
- [13] S. Khan. Estimation of parameters of the multivariate regression model with uncertain prior information and Student-t errors. *Journal of Statistical Research*, vol. 39, no. 2, 2005, pp. 79-94.
- [14] S. Khan. Shrinkage estimators of intercept parameters of two simple regression models with suspected equal slopes. *Communications in Statistics - Theory and Methods*, vol. 37, 2008, pp. 247-260.
- [15] S. Khan, 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*, pp. 599-600.
- [16] S. Khan, A. K. Md. E. Saleh. Shrinkage pre-test estimator of the intercept parameter for a regression model with multivariate Student-t errors. *Biometrical Journal*, vol. 39, 1997, pp. 1-17.
- [17] S. Khan, A. K. Md. E. Saleh. On the comparison of the pre-test and shrinkage estimators for the univariate normal mean. *Statistical Papers*, vol. 42, no. 42001, 2020 , pp. 451-473.
- [18] 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*, vol. 31, no. 1, 2022, pp. 51-72.
- [19] S. Khan, Z. Hoque. Preliminary test estimators for the multivariate normal mean based on the modified W, LR and LM tests. *Journal of Statistical Research*, vol. 37, 2003, pp. 43-55.
- [20] S. Khan, A. K. Md. E. Saleh. Estimation of intercept parameter for linear regression with uncertain non-sample prior information. *Statistical Papers*. Vol. 46, 2005, pp. 379-394.
- [21] S. Khan, A. K. Md. E. Saleh. Estimation of slope for linear regression model with uncertain prior information and Student- t error. *Communications in Statistics - Theory and Methods*, vol. 37, no. 16, 2008, pp. 2564-258.
- [22] S. Khan, B. Pratikno, Testing Base Load with Non-Sample Prior Information on Process Load. *Statistical Papers*, vol. 54, no. 3, 2013, pp. 605-617.