The Power and Its Graph Simulations On Discrete and Continuous Distributions

by Budi Pratikno

Submission date: 16-Jun-2022 04:10PM (UTC+0700)
Submission ID: 1857841017
File name: MS-13427687-FINAL.pdf (547.18K)
Word count: 2897
Character count: 14224

The Power and Its Graph Simulations On Discrete and Continuous Distributions

Budi Pratikno[1*], Nailatul Azizah[2] and Avita Nur Azizah[3]

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

Abstract. We determined the power and its graph simulations on the discrete Poisson and Chi-square distributions. There are four important steps of the research methodology summarized as follow: (1) determine the sufficiently statistics (if possible), (2) create the rejection area (UMPT test is sometime used), (3) derive the formula of the power, and (4) determine the graphs using the data (in simulation). The formula of the power and their curves are then created using *R* code. The result showed that the power of the illustration of the discrete (Binomial distribution) depending on the number of trials *n* and bound of the rejection area. The curve of the power is sigmoid (*S*-curve) and tend to be zero when parameter shape (θ) greater 0.4. It decreases (started from theta = 0.2) as the parameter theta increases. In the Poisson context, the curves of the power of the Poisson distribution is not *S*-curve, and it is only depended on the parameter shape λ. We noted that the curve of the power of the Poisson is quickly to be one for *n* greater than 2 and λ less than 10. In this case, the size of the Poisson distribution is greater than 0.05, so it is not a reasonable thing even the power close to be one. In this context, we must have the maximum power and minimum size. In the context of Chi-square distribution, the graph of the power and size functions was depended on rejection region boundary (*k*). Here, we noted that *skewness* of the S-curve positive as the *k* increases. Similarly, the size are also depended of the *k* (and constant), and it decrease as the *k* increases. We here also noted that the power quickly to be one for large degree of freedom (*r*).

Keywords Poisson discrete distributions, Chi-square continuous distribution, parameter shape, and R-code.

1. Introduction

In the theory of statistics, there are three important concepts of the hypothesis testing in rejecting or accepting null hypothesis (H_0), namely (1) a probability error type I (α), (2) a probability error type II (β) and (3) power of the test $(\pi(\theta))$ (Wackerly, et al. [5]). Here, the power is a significant method to test the hypothesis on parameter shape. We then study more detail about the power of the hypothesis testing on some distributions. Furthermore, the power is defined as a probability to reject H_0 under H_1 on $H_0:\theta=\theta_0$ versus $H_1:\theta\neq\theta_0$, for parameter shape θ (Wackerly, et al. [5]).

Following to the previous research, many authors, such as Pratikno [2], Khan and Pratikno [22] and Khan [12], used 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. Moreover, Pratikno [2] and Khan et al. [11] used the power and size to compute the cdf of the bivariate noncentral *F (*BNCF*)* distribution in multivariate and multiple regression models. Here, many authors, such as Khan [12, 13, 14], Khan and Saleh [15,16,17, 20, 21], Khan and Hoque [19], Saleh [1], Yunus [6], and Yunus and Khan [7, 8, 9, 10], have contributed to the research of the power 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. [11] 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.[18]), so the *R* code is used.

Unlike previous research analyze focusing on continuous distribution, we only consider on two distributions (Poisson and Chi-square). To illustrate of the simple power, we present the power of the

Binomial distributions. Furthermore, the steps to compute the power of the Binomial, Poisson and Chi-square distributions are similar to the previous theory are: (1) we have to determine the sufficiently statistics (if possible), (2) we create the rejection area using *uniformly most powerful test* (UMPT, if needed), (3) we derive the formula of the power of the discrete and continuous distributions, and (4) finally, we graphically analysed the power. A simulation is then conducted using the generate data.

The concept of power and size (as initiate, Binomial distribution) of the testing hypothesis is presented in Section 2. The derived formula and the analysis of the power of the power and size of the Binomial, Poisson and Chi-square distributions are then given in Section 3. The conclusion is in Section 4.

2. The Power and Size of One-Side Hypothesis Testing

Following Pratikno [2], Khan [12,13,14], Khan and Saleh [15,16,17,20,21], Khan and Hoque [19], Saleh [1], Yunus [6], and Yunus and Khan [7, 8, 9, 10], we noted that the power and size of the provide tests a significant method to find the conclusion of the hypothesis testing parameter shape. Here, we must choose the maximum power and minimum size as a indicator. Here, the power and size are defined a probability to reject H_0 under H_1 in testing hypothesis, and probability to reject H_0 under H_0, respectively (Wackerly, et al. [5]). Following Pratikno [2], we then write the power and size in testing hypothesis, $H_0:\theta=\theta_0$ versus $H_1:\theta>\theta_0$ (or $H_1:\theta=\theta_1$) as, respectively,

$$\pi(\theta_1)=P(\text{reject } H_0 \mid \text{under } H_1)$$
$$=P(\text{reject } H_0 \mid \theta=\theta_1) \qquad (1)$$
$$\pi(\theta_0)=P(\text{reject } H_0 \mid \text{under } H_0)$$
$$=P(\text{reject } H_0 \mid \theta=\theta_0) \qquad (2)$$

where α is probability of type error I and β is probability of type error II. The details of the power and size in testing coefficient parameters on the regression models are found on Pratikno [2], and the power and size on several continuous distributions are also found Pratikno et al.[3,4].

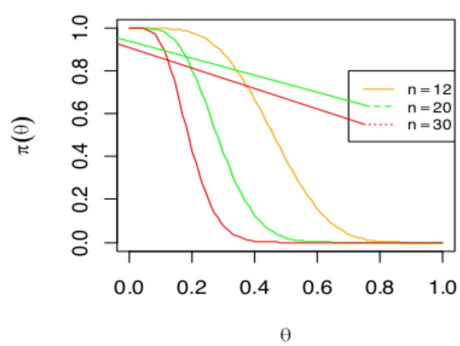
3. The Power and Size of Discrete and Continuous Distributions

3.1. The Power and Size of the Binomial Distribution

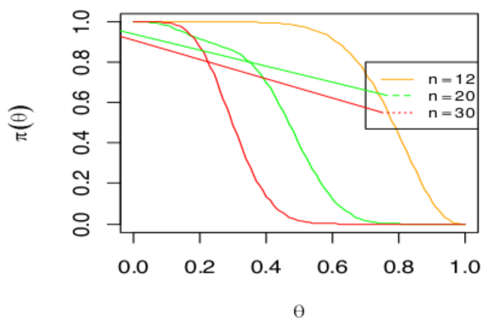
Following Pratikno [2], we firstly derived the formula of the power and size of the discrete Binomial distribution (as a review). The power and size of this distribution are computed in one-side hypothesis testing on several n and bound of the rejection areas. Let, X_i follows Bernoulli distribution with parameter θ. Take a trial $n=12$, then $Y=\sum_{j=1}^{n=12} X_j$ follows Binomial distribution with $n=12$ and $p=\theta$ and is written as $Y \sim B(n,\theta)$. Here, we decide (an example $\theta=0.7$) to test $H_0:\theta=0.7$ versus $H_1:\theta>0.7 (as\ \theta_1)$, with rejection area $\{(x_1,...,x_{12}):Y\leq 5\}$, therefore the power function on the binomial distribution is then given as

$$\pi(\theta)=P(\text{reject } H_0 \mid \text{under } H_1:\theta)$$
$$=\sum_{y=0}^{5}\binom{12}{y}\theta^{y}(1-\theta)^{12-y}$$
$$=\binom{12}{0}\theta^{0}(1-\theta)^{12}+\binom{12}{1}\theta(1-\theta)^{11}+\ldots+\binom{12}{5}\theta^{5}(1-\theta)^{7}$$
$$=\binom{12}{0}\theta^{0}(1-\theta)^{12}+\binom{12}{1}\theta(1-\theta)^{11}+\ldots+\binom{12}{5}\theta^{5}(1-\theta)^{7}$$
$$=(1-\theta)^{7}\left(1+7\theta+28\theta^{2}+84\theta^{3}+210\theta^{4}+462\theta^{5}\right) \qquad (3)$$

Using the equation (3) and *R-code*, we then produced the graphs (curves) of the power in Figure 1.



(a)



(b)

Figure 1. The power of the Binomial Distribution at several *n* and *Y=5* and *9*

Figure 1. showed that the curve of the power of the Binomial distribution (Figure 1 (a) and (b)) are sigmoid (*S* curve) and depends on the number of trials (*n*) and the bound of the rejection area (*Y*). They are going to be zero for $\theta > 0.4$. The curve decreases (started from $\theta = 0.2$) as the parameter increases. From Figure 1 (a) and (b), it is clear that both *n* and *Y* have significant affect on the shape of the curve (see Figure 1 (a) and (b), they move to the right). Here, the maximum power is one and the minimum power is zero. The size is then produced using the equation (3) under H_0, as

$$\begin{aligned}\alpha &= P(\text{reject } H_0 | \text{under } H_0)\\ &= P(Y \le 5 | \theta = 0{,}7)\\ &= 0.04.\end{aligned}$$

It is clear that the size is constant and is less than 0.05, as expected

3.2. The Power and Size of the Poisson Distribution

Let, $X_1, \ldots, X_n$ follow Poisson distribution, the probability distribution function (pdf) of random variable *X* is then given by

$$f(x,\lambda) = \frac{e^{-\lambda}\lambda^{x_i}}{x_i!} \qquad (4)$$

with $x = 0,1,2,\ldots,\text{and } \lambda > 0.$ The pdf curve of the Poisson distribution (positive skew) tends to be normal for large values λ, where the center of the pdf curve always moves to the right when λ increases.

To find the power, we then derive sufficient statistics and a rejection area using factorization theorem and UMP test. In other words, let *S* be sufficient statistics, The join distribution of the Poisson distribution is then expressed as

$$f(x_1,\ldots,x_n;\lambda) = g(s,\lambda)\,h(x_1,\ldots,x_n) \qquad (5)$$

where the pdf of the join distribution of the Poisson distribution is

$$\begin{aligned}f(x_1,\ldots,x_n;\lambda) &= \prod_{i=1}^{n} f(x_i,\lambda)\\ &= \prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{x_i}}{x_i!} = \frac{e^{-n\lambda}\lambda^{\sum x_i}}{\prod_{i=1}^{n} x_i!}. \qquad (6)\end{aligned}$$

We therefore conclude that $S = \sum_{i=1}^{n} x_i$ sufficient statistics, this is due to the equation (6) can be expressed as

$$\begin{aligned}f(x_1,\ldots,x_n;\lambda) &= \frac{e^{-n\lambda}\lambda^{\sum x_i}}{\prod_{i=1}^{n} x_i!}\\ &= \left(\frac{e^{-n\lambda}}{\prod_{i=1}^{n} x_i!}\right)\left(\lambda^{s}\right)\\ &= h(x_1,\ldots,x_n)\,g(s,\lambda) \qquad (7)\end{aligned}$$

The rejection area is then derived using *uniformly most powerful* (UMP) test as follow. Using the properties of *maximum likelihood ratio* (MLR) of the $f(x_1,\ldots,x_n;\lambda)$ on $S = \sum_{i=1}^{n} X_i$ $\left(\sum_{i=1}^{n} X_i > k\right)$ and UMP-*test*, we then get the probability to reject H_0 under H_0 (the size or α) and the probability to reject H_0 under H_1 (the power) in testing $H_0 : \lambda = \lambda_0$ versus $H_1 : \lambda > \lambda_0$, are, respectively,

$$\begin{aligned}\alpha = \alpha(\lambda) = \alpha^{*} &= P\left(\sum_{i=1}^{n} X_i > k | \lambda_0\right)\\ &= P\left(\sum_{i=1}^{n} X_i > Pois(n\lambda)\right)\\ &= \sum_{x=0}^{n} \frac{e^{-(n\lambda_0)}(n\lambda_0)^{x}}{x!} \qquad (8)\end{aligned}$$

$$\pi(\lambda) = (\text{Probability reject } H_0 \text{ under } H_1)$$
$$= P\left(\sum_{i=1}^{n} X_i > Poi(n\lambda) | \lambda\right)$$
$$= 1 - P\left(\sum_{i=1}^{n} X_i \le Pois(n\lambda) | \lambda\right)$$
$$= 1 - \left(\sum_{x=0}^{n} \frac{e^{-n\lambda}(n\lambda)^x}{x!} | \lambda\right) \qquad (9)$$

Using the equation (8) and (9), we presented the graph of the power of the Poisson distribution (Figure 2.), and the value of the size and power for *n=3* in testing $H_0 : \lambda = 1$ versus $H_1 : \lambda > 3$, respectively, as

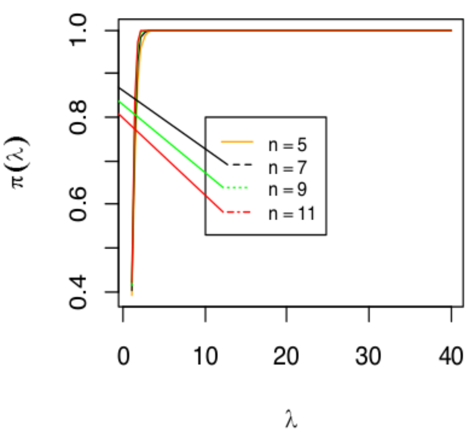


Figure 2. The power of the Poisson Distribution at several *n*

$$\alpha = 1 - \left(\sum_{x=0}^{n} \frac{e^{-n\lambda_0}(n\lambda_0)^x}{x!}\right) = 0.40779 > 0.05,$$
$$\pi(\lambda) = 1 - \left(\sum_{x=0}^{n} \frac{e^{-n\lambda}(n\lambda)^x}{x!}\right) = 0.99985 \simeq 1.$$

From Figure 2, we see that the power of the Poisson distribution tends to be 1 when $\lambda < 10$. Here, the simulation of the *n* has not influenced to the curve of the power yet. Thus, we conclude value of *n* whether small or large does not affect change the shape of the curve of the power. Similarly, for large λ, the shape of the curve of the power does not change. Here, the size 0.408 is greater than 0.05 (too high), and it is not as expected.

3.3. The Power and Size of the Chi-square Distribution

Let, X be a random variable that follows Chi-square distribution, The probability distribution function (pdf) of random variable X is then given by

$$f(x) = \frac{1}{2^{\frac{r}{2}} \tau\left(\frac{r}{2}\right)} x^{\frac{r}{2}-1} e^{-x/2}, x \ge 0 \qquad (10)$$

with r is the degree of freedom (as parameter). The cdf of this distribution is the written as

$$F_x(x) = \int_0^x f(x) = \int_0^x \frac{1}{2^{\frac{r}{2}} \tau\left(\frac{r}{2}\right)} x^{\frac{r}{2}-1} e^{-x/2}$$

The power of this distribution in testing parameter shape $H_0 : r = r_0$ versus $H_0 : r > r_0$ (r_0 is determined as 1), is then obtained as

$$\pi(r) = P(\text{reject } H_0 \mid \text{under } H_1, r_0 = r)$$
$$= P(S > k | r)$$
$$= 1 - P(S \le k | r)$$
$$= 1 - \int_0^k \frac{1}{2^{\frac{r}{2}} \tau\left(\frac{r}{2}\right)} s^{\frac{r}{2}-1} e^{-s/2} ds$$
$$= 1 - \left(-\frac{1}{\tau\left(\frac{r}{2}\right)}\right) \left[\tau\left(\frac{r}{2}, \upsilon^{2/r}\right)\right]_0^k$$
$$= 1 + \frac{\left[\tau\left(\frac{r}{2}, \frac{s}{2}\right)\right]_0^k}{\tau\left(\frac{r}{2}\right)} \qquad (11)$$

Here, $s = \sum_{i=1}^{n} x_i$ is a sufficiency statistics and $\upsilon = \left(\frac{s}{2}\right)^{r/2}$. Using the equation (11), we then produced the graphs of the power and size as presented in Figure 3 and Figure 4.

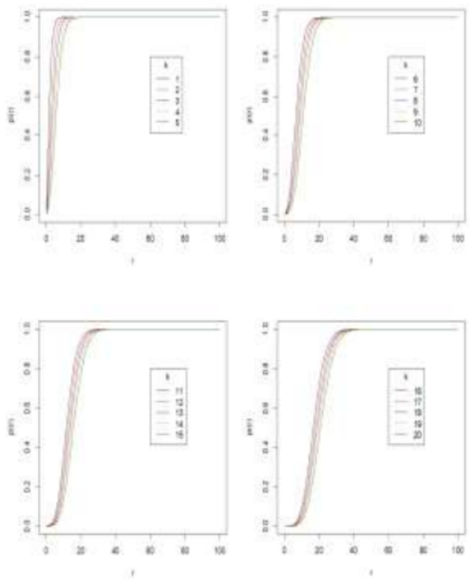


Figure 3. The power of the Chi-square Distribution at several k

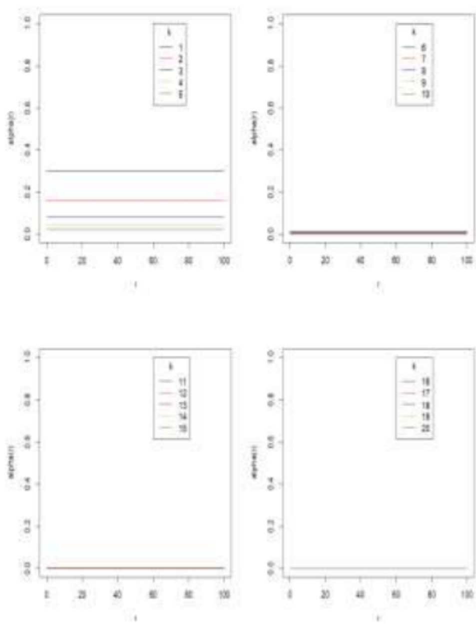


Figure 4. The Size of the Chi-square Distribution at several k

From Figure 4., we see that the curves of the power depend on of the values k . They skew to the right (S-curve positive) as k increases. Similarly, we see from Figure 4 that the size are constant and depended of the k, and they decrease as the k increases. To illustrate the values of the size of the Chi-square distribution, we present a simulation *k=5* and *k=10*, on *r =1*, as below

For *k=5*,

$$\alpha = \pi(1) = 1 + \frac{\left[\tau\left(\frac{r}{2}, \frac{s}{2}\right)\right]_0^5}{\tau\left(\frac{r}{2}\right)}$$

$$= 1 + \frac{\left[\tau\left(\frac{r}{2}, \frac{5}{2}\right) - \tau\left(\frac{r}{2}, 0\right)\right]}{\tau\left(\frac{r}{2}\right)}$$

$$\approx 0.025.$$

For *k=10*,

$$\alpha = \pi(1) = 1 + \frac{\left[\tau\left(\frac{r}{2}, \frac{s}{2}\right)\right]_0^{10}}{\tau\left(\frac{r}{2}\right)}$$

$$= 1 + \frac{\left[\tau\left(\frac{r}{2}, 5\right) - \tau\left(\frac{r}{2}, 0\right)\right]}{\tau\left(\frac{r}{2}\right)}$$

$$\approx 0.002.$$

4. **Conclusion**

To find the power of the Poisson distribution, we consider sufficient statistics and *UMP test* for getting the rejection area. In the Binomial distribution context, the curve of the power depends on the number of trials n and the bound of the rejection area. The curves tend to zero when $\theta > 0.4$, and it decreases (started from $\theta = 0.2$) as the parameter increases. We also note that, the curve is sigmoid (*S* curve). In the Poisson distribution context, the result showed that the power of the Poisson (not sigmoid, *S* curve) tends to be 1 on several simulation $n(n \geq 2)$ and $\lambda(\lambda < 10)$. In the context of chi-square distribution, we note that the curves of the power depend on of the k and the skewness of the S-curve is positive as k increases. On the size context, we note that the size is constant. The size also depends on of the k and decrease as k increases.

Acknowledgment

I thankfully to the LPPM UNSOED for providing me granting of research

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)*, 106 (2), 463-474, 2018.

[4] B. Pratikno, Power of hypothesis testing parameters shape of the distributions. *Far East Journal of Mathematical Science (FJMS)*, 110 (1), 15-22, 2019

[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 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.

[8] R. M. Yunus and S. Khan. Increasing power of the test through pre-test a robust method. Communications in Statistics-Theory and Methods, **40**, 581-597, 2011a.

[9] R. M. Yunus and S. Khan. M-tests for multivariate regression model. Journal of Nonparamatric Statistics, **23**, 201-218, 2011b.

[10] R. M. Yunus and S. Khan. The bivariate noncentral chi-square distribution - Acompound distribution approach. Applied Mathematics and Computation, **217**, 6237-6247, 2011c.

[11] 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, 45 3491–3507, 2016.

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

[13] S. Khan. Estimation of parameters of the multivariate regression model with uncertain prior information and Student-t errors. Journal of Statistical Research, **39**(2) 2005, 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, **37**, 247-260, 2008.

[15] 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, 599-600, 1995.

[16] 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, **39**, 1-17, 1997.

[17] 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, **42**(4), 451-473, 2001.

[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, **31** (1), 51-72, 2002.

[19] 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, Vol 37, 43-55, 2003.

[20] S. Khan. and A. K. Md. E. Saleh. Estimation of intercept parameter for linear regression with uncertain non-sample prior information. Statistical Papers. **46**, 379-394, 2005.

[21] S. Khan. and 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, **37**(16), 2564-258, 2008.

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

The Power and Its Graph Simulations On Discrete and Continuous Distributions

ORIGINALITY REPORT

8%	%	%	8%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Universitas Jenderal Soedirman Student Paper	7%
2	Submitted to Hellenic Open University Student Paper	1%
3	Submitted to Higher Education Commission Pakistan Student Paper	<1%

Exclude quotes	On	Exclude matches	Off
Exclude bibliography	On		