International Journal of Scientific Research in Science, Engineering and Technology (www.ijsrset.com)
© 2020 IJSRSET | Volume 7 | Issue 3 | Print ISSN: 2395-1990 | Online ISSN : 2394-4099
DOI : https://doi.org/10.32628/IJSRSET207358

Density Estimation of Neonatal Mortality Rate Using Empirical Bayes Deconvolution with Poisson Distributed Surrogate in Central Java Province Indonesia

Fevi Novkaniza[1,3], Khairil Anwar Notodiputro*[1], I Wayan Mangku[2], Kusman Sadik[1]
*[1]Department of Statistics, IPB University, Bogor, West Java, Indonesia
[2]Department of Mathematics, IPB University, Bogor, West Java, Indonesia
[3]Department of Mathematics, Universitas Indonesia, Depok, West Java, Indonesia

ABSTRACT

This article is concerned with the density estimation of Neonatal Mortality Rate (NMR) in Central Java Province, Indonesia. Neonatal deaths contribute to 73% of infant deaths in Central Java Province. The number of neonatal deaths for 35 districts/municipalities in Central Java Province is considered as Poisson distributed surrogate with NMR as the rate of Poisson distribution. It is assumed that each number of neonatal deaths by district/municipality in Central Java Province were realizations of unobserved NMR, which come from unknown prior density. We applied the Empirical Bayes Deconvolution (EBD) method for estimating the unknown prior density of NMR based on Poisson distributed surrogate. We used secondary data from the Health Profiles of Central Java Province, Indonesia, in 2018. The density estimation of NMR by the EBD method showed that the resulting prior estimate is relatively close to the Gamma distribution based on Poisson surrogate. This is implying that the suitability of the obtained prior density estimation as a conjugate prior for Poisson distribution.

Keywords: Deconvolution, Density, Empirical Bayes, Prior, The Mortality Rate

I. INTRODUCTION

The neonatal period is recognized as the most high-risk time in an infant's life and accepted as beginning at birth and ending at 28 completed days of life. Neonatal *death* has been defined by the World Health Organization (WHO) as "deaths among live births during the first 28 completed days of life" which can be further sub-divided into early neonatal deaths (deaths between 0 and 7 completed days of birth) and late neonatal deaths (deaths after seven days to 28 completed days of birth). Many countries concern with neonatal deaths, and WHO reported that 2.5 million children died in the first month of life in 2018 and approximately 7,000 neonatal deaths every day. It is estimated that four million neonatal deaths occur annually, and 98% of which occur in developing countries. The major direct causes of neonatal deaths vary from region to region and from country to country (Woldeamanuel, [14]). Maternal health before, during, and after pregnancy, conditions at the time of labor and delivery, and post-natal care of babies play a significant role in reducing neonatal mortality.

Neonatal Mortality Rate (NMR) is the number of neonates dying before reaching 28 days of age, per 1,000 live births in a specified year. The NMR is commonly accepted as a measure of the general health and wellbeing of a population because it is significant indexes of a country's state and trends of health or development and a component of the physical quality of life index. A decrease in NMR results in the improvement of infant mortality and survival, which can positively influence the national public state of health. NMR is a significant concern for the government for monitoring the current health programs and formulating policies for improving the current health situation. It can be a recommendation given to policymakers and health programs for improving the health childcare policies and development of focused and evidence-based health interventions to prevent neonatal deaths.

Some research studies are proposed to measure and to model NMR. Gonzales and Gilleskie [6] develop a method to adjust country-specific reported infant mortality figures that may misrepresent development within a country and some essential socioeconomic indicators. They proposed an "augmented" measure of mortality that includes both infant and late neonatal deaths that should be considered when assessing levels of social welfare in a country. Houweling et al. [7] developed regression models to predict the risk of neonatal death with characteristics known at the start of pregnancy, the start of delivery, and 5 minutes postpartum. They used data from rural and urban surveillance sites in South Asia (1,742 neonatal deaths from 49,632 live births). They assessed the model's discriminative ability using cross-validation between sites under the receiver operating characteristics curve.

Indonesia's NMR between 1969 and 2018 was declining at a moderating rate to shrink from 49.1 deaths per 1,000 live births in 1969 to 12.7 deaths per 1,000 live births in 2018. Central Java is one of the provinces that contribute significantly to NMR in Indonesia. Triyanto *et al.* [13] proposed modified Multivariate Poisson Regression (MPR) to modeling the number of maternal, early neonatal, and postneonatal mortality in Central Java Province, Indonesia. They assumed that the covariance is a function of the covariate and the population size of each sample unit is different. From their studies, they found factors that significantly influence of early neonatal mortality is the percentage of pregnancy visit, the percentage of pregnant women get a tablet Fe3, the percentage of low birth weight, and the percentage of married women before age 17.

In this paper, we discuss a method to construct a density estimation of NMR because neonatal deaths contribute to 73% of infant deaths in Central Java Province, Indonesia. We applied the Empirical Bayes Deconvolution (EBD) method for estimating the density of NMR. The EBD method proposed by Efron [4] is a combined deconvolution method with an empirical Bayes strategy. One of the main problems in deconvolution is to estimate the density of the unknown random variable. In the Bayesian framework, an unknown prior density has realizations, which is unobservable. However, each of them produces observable value as surrogate according to a known probability mechanism. We assumed the number of neonatal deaths for 35 districts/municipalities in Central Java Province is considered as Poisson distributed surrogate with NMR as the rate of Poisson distribution, which comes from unknown prior density.

This paper is organized as follows: section 2 will present an overview of the concept of deconvolution and the EBD method. Section 3 and 4 presents the implementation of the EBD method for estimating the neonatal mortality rate based on a data set from the Health Profiles of Central Java Province Indonesia in 2018. The conclusion is given in the last section.

II. METHODS AND MATERIAL

A. Deconvolution

A deconvolution model can be written as:

$$W_i = X_i + e_i,\ i = 1,...,n \quad (1)$$

For $\{X_i\}$ are i.i.d. with unknown density $g(x)$ are independent real-valued sequences of random variables. The density function of e_i's is denoted by f_e, and we assume X and e are mutually independent. From probability theory, we know that the density of the sum of two independent random variables is equal to the convolution of the densities of both X and e. Under model (1), deconvolution is a method to estimate the density of $X, g(x)$ or the distribution function X based on the data W_i. Efron [4], Fan [5], Laird [9], Hall and Meister [8], and Meister [10] study the deconvolution problem on an additive model as in equation (1). Efron [4] proposed modeling strategies for deconvolution, which referred to as Empirical Bayes Deconvolution (EBD). EBD method estimates an unknown prior density of X which yielded unobservable $X_1,...,X_n$ but produces observable value or surrogate W_i according to a known probability density. Model (1) also can be considered as an additive measurement error model. The i.i.d random variables $e_1,...,e_n$ represent the measurement error or the contamination of the data. Ignoring measurement error can bring biased estimates and lead to erroneous conclusions to various degrees in data analysis. Carroll and Hall [1], Carroll and Hall [2] and Stefanski and Carroll [12] proposed the deconvolution kernel density estimator to recover the unknown density function from contaminated data, where the kernel idea and the Fourier inverse are used for constructing the estimator. Delaigle [3] have shown that the deconvolution kernel density estimator is consistent and derived asymptotic properties with only a little knowledge of the error density.

B. Empirical Bayes Deconvolution

Let unknown prior density $g(x)$ has an observed independent random sample of realizations $X_1, X_2,..., X_n$:

$$X_1, X_2,..., X_n :\ g(x).$$

Each X_i independently produces an observed random variable W_i with known probability densities for W_i given $X_i, W_i :\ p_{W_i}(w), i = 1,...,n$, and the marginal density of W_i

$$f(w) = \int p_W(w \mid x) g(x) dx.$$

According to Efron [4], for estimating the prior density $g(x)$ using sample observation $W_1,...,W_n$, we can use Empirical Bayes Deconvolution (EBD). The Empirical Bayes Deconvolution method is an estimation procedure $g(x)$ based on sample observations from $f(w)$. Efron [4] uses the likelihood approach to EBD problems with prior $g(x)$, which is modeled through exponential family density in Space- X, denote by T. T is assumed to be a finite discrete support set $T = (x_1,..., x_m)$ and by discretizing Space- X:

$$X \in T = \left\{x_{(1)}, x_{(2)},..., x_{(m)}\right\}$$

then prior distribution can be denoted as

$$g_j = \Pr\left\{X = x_{(j)}\right\}.$$

Prior density $g(x)$ is an m-vector $g = (g_1,..., g_m)$ that is which specifies the probability g_j on x_j:

$$g = g(\alpha) = \exp\left\{Q\alpha - \phi(\alpha)\right\}$$

with $\phi(\alpha) = \log \sum_{j=1}^{m} \exp\left\{Q_j^T \alpha\right\}$ where $\alpha =$ p-dimensional vector and $Q =$ known $m \times p$ structure matrix. The-j component of $g(\alpha)$:

$$g_j(\alpha) = \exp\left\{Q_j^T \alpha - \phi(\alpha)\right\},\ j = 1,..., m.$$

Define $p_{ij} = p_i\left(W_i \mid X_i = x_j\right)$ and denote P_i as m-vector $P_i = \left(p_{i1},..., p_{im}\right)^T$, then the marginal probability for W_i:

$$f_i(\alpha)=\sum_{j=1}^{m} p_{ij} g_j(\alpha)=P_i^T g(\alpha).$$

The loglikelihood function for the parameter vector $\alpha=(\alpha_1,...,\alpha_p)^T$ is:

$$l_i(\alpha)=\log f_i(\alpha)=\log P_i^T g(\alpha)$$

with p-dimensional first derivative vector and $p\times p-$ dimensional second derivative matrix

$$\dot{l}_i(\alpha)=\left(...,\frac{\partial l_i(\alpha)}{\partial \alpha_h},...\right)^T, \ddot{l}_i(\alpha)=\left(...,\frac{\partial^2 l_i(\alpha)}{\partial \alpha_h \partial \alpha_k},...\right)$$

for the maximum likelihood calculation. For W_i with n observation, the total loglikelihood $l(\alpha)=\sum_{i=1}^{n} l_i(\alpha)$ has first and the second derivative is:

$$\dot{l}(\alpha)=\sum_{i=1}^{n}\dot{l}_i(\alpha)=Q^T\sum_{i=1}^{n}B_i(\alpha)=Q^T B_+\alpha$$

where

$$B_i(\alpha)=\{b_{i1}(\alpha),...,b_{im}(\alpha)\}^T, b_{ij}(\alpha)=g_j(\alpha)\left\{\frac{p_{ij}}{f_i(\alpha)}-1\right\}$$

and

$$-\ddot{l}_i(\alpha)=Q^T\left[B_i(\alpha)B_i(\alpha)^T+B_i(\alpha)g(\alpha)^T+g(\alpha)B_i(\alpha)^T-diag\{W_i(\alpha)\}\right]Q.$$

Efron (2016) proposed the maximum likelihood estimate $\hat{\alpha}$ for α satisfies $Q^T B_+\hat{\alpha}=0$.

C. Neonatal Mortality Rate (NMR) in Central Java Province Indonesia

Children's health care program in Indonesia is intended to prepare healthy, intelligent, and quality future generations and to reduce child mortality. Some indicators of mortality rates related to children are Neonatal Mortality Rate (NMR), Infant Mortality Rate (IMR), and Child Mortality Rate (CMR). Mortality during the neonatal period is considered a useful indicator of both maternal and newborn health and care. WHO defines NMR as the number of deaths during the first 28 completed days of life per 1000 live births in a given year or period. Based on data from the Family Health Program in Central Java Province Indonesia, the trend of NMR, IMR, and CMR from 2014-2018 are shown in Fig.1:

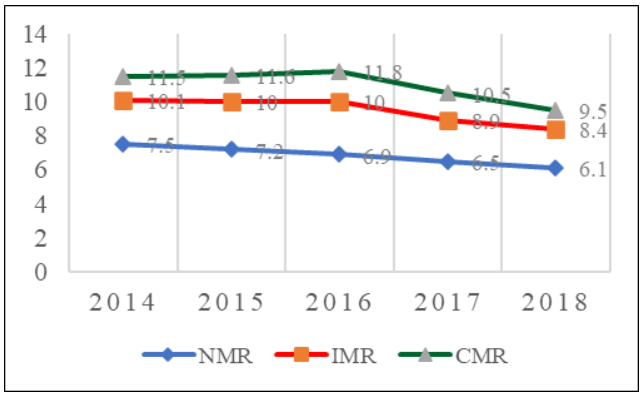

Figure 1 NMR, IMR and CMR Trend in Central Java Province

From Fig.1, the NMR has a downward trend from 2014 to 2018, but still 6.1 per 1,000 live birth. Kabupaten Rembang has the highest NMR in 2018 (11.7 per 1,000 live birth), followed by Kabupaten Grobogan (11.5 per 1,000 live births) and Kabupaten Banjarnegara (10.7 per 1,000 live births). Kota Surakarta had the lowest NMR (1.8 per 1,000 live birth) in 2018. *In this section,* we want to estimate NMR density in Central Java Province because neonatal mortality contributes to 73% of infant deaths. *We use secondary data set* from 35 districts/municipalities in Central Java Province from the Health Profiles of Central Java Province Indonesia in 2018. Using the EBD method, data analysis procedure as the following steps:

1. We denote the number of neonatal deaths for 35 districts/municipalities in Central Java Province Indonesia as $W_i, i=1,...,35$ and conducting the statistical descriptive of W_i
2. We assume W_i as the observed random variable which has Poisson distribution with rate X_i
3. X_i is the NMR which has an unknown probability density $g(x)$ and each X_i independently produces surrogate W_i according to Poisson density
4. Using surrogate data $W_1,...,W_{35}$, we choose finite discrete support set T for X_i based on the minimum and maximum value of surrogate data W_i and discretize the support set as many points as $W_i, T=\{X_{(1)},..,X_{(35)}\}$

5. We compute the prior density estimation $\hat{g}(x_i) = \Pr\{X_i = x_{(i)}\}, i = 1,...,35$ bias for $\hat{g}(x_i)$
6. We compare the CDF estimate of $X_i, \hat{G}(x_{(i)})$ and fitted Gamma distribution from surrogate W_i

III.RESULTS AND DISCUSSION

Neonatal deaths data in Central Java Province Indonesia during 2018 can be seen in Fig.2:

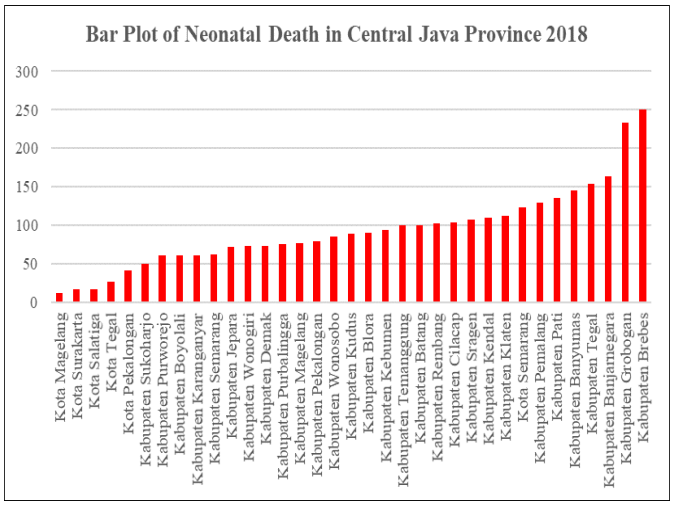

Figure 2 Number of Neonatal Deaths by districts/municipalities in Central Java Province

Kota Magelang has the minimum number of neonatal deaths (12), and Kab Brebes had the maximum number of neonatal deaths (250) in 2018. As we know, the distance between Kab Brebes to province capital is 174 km, but Kota Magelang is only 75.9 km. Data taken from Public Health Service of Central Java Province 2019, there are 10 hospitals, 1 maternity hospitals, 38 Public Health Center (Puskesmas) and 1771 Maternal & Child Health Center (Posyandu) in Kab Brebes. However, in Kota Magelang, there are 5 hospitals, 1 maternity hospitals, 5 Public Health Center (Puskesmas), and 197 Maternal & Child Health Center (Posyandu). Based on this fact, although Kab Brebes have more public health service than Kota Magelang, the number of neonatal deaths is still more significant. It means NMR for each regency/municipality can be different, so we cannot estimate the neonatal mortality rate for overall regencies/municipalities, but we estimate the probability density of NMR in Central Java Province Indonesia using a surrogate is the number of neonatal deaths in Central Java Province. We assumed surrogate has Poisson distribution with NMR the rate of Poisson. We use the EBD method for estimating NMR density and bias of NMR density in Central Java Province Indonesia by using the deconvolveR package [11] and they are shown in Fig.3 and Fig.4 as follows

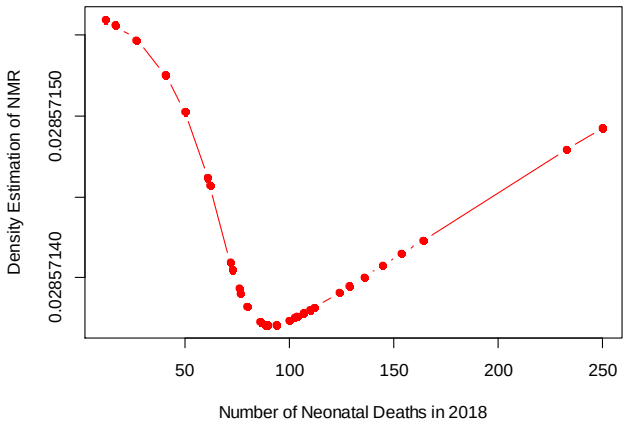

Figure 3 Density Estimation of NMR in Central Java Province

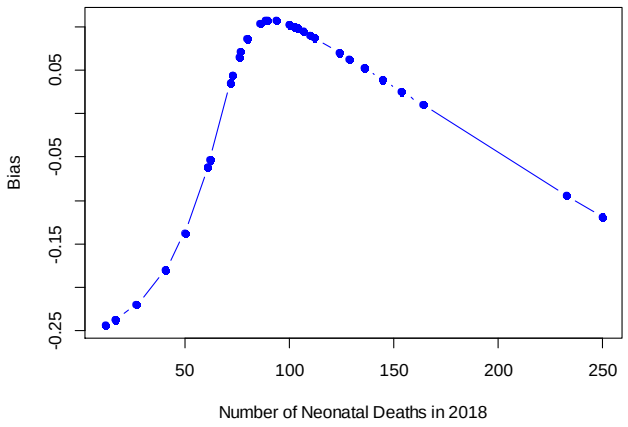

Figure 4 The bias of NMR Density Estimation in Central Java Province

From Fig.3, Kab Kebumen has a lower NMR probability estimation. However, Kota Surakarta dan Kota Salatiga has the highest one. Kab Kebumen has 94 neonatal deaths with NMR probability estimate 0.02857137. Kota Surakarta dan Kota Salatiga which

has 17 neonatal deaths with NMR probability estimate 0.028571. Kab Brebes, which has the highest number of neonatal deaths, i.e., 250, but NMR probability estimate, is only 0.02857149. As we know that Gamma distribution is conjugate prior for Poisson distribution, we fitted the number of neonatal deaths into Gamma distribution, and we got the parameter estimate (shape=2.924, scale 32.15). We also estimate cumulative density function (CDF) of NMR, and we compared it with fitted CDF Gamma distribution in Fig.5 as follows:

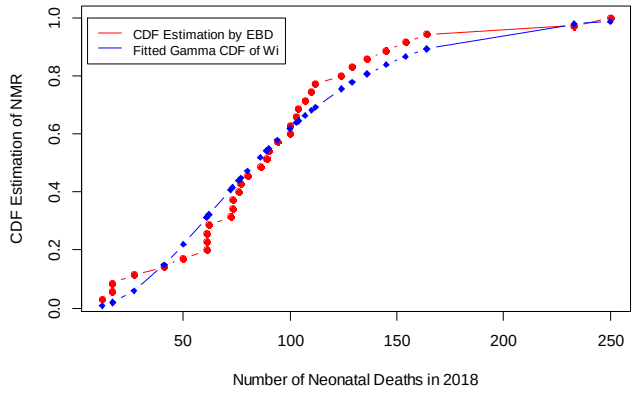

Figure 5 CDF Estimation of NMR in Central Java Province

Figure 5 showed that the resulting prior CDF estimate of NMR based on the EBD method relatively close to the fitted Gamma CDF based on Poisson surrogate, which is the number of neonatal deaths. This is implying that the suitability of the obtained prior density estimation.

IV.CONCLUSION

In this article, we presented the density estimation of Neonatal Mortality Rate (NMR) using the empirical Bayes deconvolution method (EBD) in Central Java Province, Indonesia. We used secondary data from the Health Profiles of Central Java Province Indonesia in 2018, and we assumed that the number of neonatal deaths in every district/municipality as Poisson distributed surrogate with NMR as the rate of Poisson distribution. We estimated NMR density as prior density, and we also computed the bias. The estimated prior density and bias pf NMR have a quadratic pattern, which are smaller values for the number of neonatal deaths more than 100. However, we do not consider or include any covariate that can use as a possible predictive or explanatory variable of the number of neonatal deaths. In the next research, we will investigate the prior density estimation using some covariates that related to neonatal deaths and density estimation for any distributed surrogate in the EBD method, not only for conjugate prior for Poisson distribution.

V. ACKNOWLEDGMENTS

The first authors thank the Indonesian Endowment Fund for Education (LPDP) Scholarship under Beasiswa Unggulan Dosen Indonesia-Dalam Negeri (BUDI-DN) batch 2016 (Number: 20161141021499) which financially supported the doctoral study.

VI.REFERENCES

[1] Carroll, R.J., Hall, P. (1988). Optimal rates of convergence for deconvolving a density, *Journal of the American Statistical Association,* 83(404):1184-1186.

[2] Carroll, R.J., Hall, P. 2004. Low order approximations in deconvolution and regression with errors in variables, *Journal of the Royal Statistical Society: Series B (Statistical Methodology),* 66(1):31-46.

[3] Delaigle, A. 2008. An alternative view of the deconvolution problem. Statistica Sinica, 18(3): 1025-1045.

[4] Efron, B. 2016. Empirical Bayes Deconvolution Estimates. *Biometrika*, 103,1, 1-20.

[5] Fan, J. 1991. On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist*. 19, 1257–72.

[6] Gonzales, R.M & Donna, G. 2017. Infant Mortality Rate as a Measure of a Country's Health: A Robust Method to Improve Reliability

and Comparability. Demography,54(2): 701–720. DOI: 10.1007/s13524-017-0553-7.

[7] Houweling, T. A. J. et al. 2019. A prediction model for neonatal mortality in low- and middle-income countries: an analysis of data from population surveillance sites in India, Nepal, and Bangladesh. *International Journal of Epidemiology*, Volume 48, Issue 1, Pages 186–198, https://doi.org/10.1093/ije/dyy194.

[8] Hall, P., Meister, A. 2007. A ridge-parameter approach to deconvolution. *Ann. Statist.* 35, 1535–58.

[9] Laird, N. 1978. Nonparametric maximum likelihood estimation of a mixed distribution. *J. Am. Statist. Assoc.* 73, 805–11.1st Edition, Chapman & Hall/CRC Interdisciplinary Statistics.

[10] Meister, A. 2009. Deconvolution problems in nonparametric statistics. Lectures Note in Statistics, Springer, 5-105. DOI: 10.1007/978-3-540-87557-42.

[11] Repository CRAN. Efron B and Narasimhan B. Package 'deconvolveR'. 2019. Empirical Bayes Estimation Strategies. Version 1.1

[12] Stefanski, L.A., Carroll, R.J. 1990. Deconvolving kernel density estimators. *Statistics*, 21(2):169-184.

[13] Triyanto, Yeni K, Henny E.C. 2018. Modified Multivariate Poisson Regression to Modeling the Number of Mortality of Maternal, Early Neonatal, and Post Neonatal in Central Java Province, Indonesia. *Applied Mathematical Sciences*, Vol. 12, 2018, no. 28, 1373 – 1382.

[14] Woldeamanuel, B.T, 2018. Statistical Analysis of Neonatal Mortality: A Case Study of Ethiopia. *Journal of Pregnancy and Child Health*., Vol 5(2).

Cite this Article

Fevi Novkaniza, Khairil Anwar Notodiputro, I Wayan Mangku, Kusman Sadik, "Density Estimation of Neonatal Mortality Rate Using Empirical Bayes Deconvolution with Poisson Distributed Surrogate in Central Java Province Indonesia", International Journal of Scientific Research in Science, Engineering and Technology (IJSRSET), Online ISSN: 2394-4099, Print ISSN : 2395-1990, Volume 7 Issue 3, pp. 206-212, May-June 2020. Available at doi : https://doi.org/10.32628/IJSRSET207358
Journal URL : http://ijsrset.com/IJSRSET207358