BMC JOURNAL OF SCIENTIFIC RESEARCH
A Multidisciplinary Research Journal
ISSN: 2594-3421

A Study of Lognormal Model for Air Pollution Concentration

Prem Sagar Bhandari*

Abstract

The Lognormal distribution has been widely used to represent the type of air pollutant concentration distribution. There are different methods to estimate the distribution parameters such as the method of moments, percentiles, maximum likelihood estimation (MLE) and Bayesian method of estimation. In this study, the theoretical distribution Lognormal is used to fit the parent distribution of PM_{10} of Putalisadak of Kathmandu, Nepal. Two estimating methods namely method of maximum likelihood and method of moments are used to estimate the parameters of the theoretic distributions.

Keywords: *Method of moments, Maximum likelihood method, Probability distribution functions, PM_{10} Concentration*

Introduction

Air is essential for life itself, without it we could survive only a few minutes. However, most cities around the world, particularly in developing countries, are experiencing worsening air pollution due to a number of factors as uncontrolled urbanization, increasing number of fossils fuel burning vehicles and highly concentrated industries around these areas.

According to Daly and Zannetti (2007), air pollution is a emission of substances into the air from an anthropogenic, biogenic or geogenic source, that is either not part of the natural atmosphere or is present in higher concentrations than the natural atmosphere and may cause a short-term or long-term adverse effect.

Air pollution is a complex mixture of gases, dust, fumes and odours in amounts which could be harmful to human health or other ecosystems. Dust pollution from the massive demolition and reconstruction activities under road expansion and earthquake recovery in Kathmandu Valley has seriously affected the environment and public health. The levels of particulate matters (PM) in the ambient have risen to hazardous levels. The government has introduced several policies, legislation and standards related to air pollution in Nepal.

The impact of air pollution is broad, especially for human beings, where it can cause several significant effects including carcinogenic effects. There are very limited data and case studies on air pollution and almost no air pollution modelling using distribution functions in Nepal. Many types of probability distributions have been used to fit air pollutant concentrations including Weibull distribution [Wang and Mauzerall (2004)], lognormal distribution [Hadley and Toumi (2002)], gamma distribution [Singh (2004)] and Rayleigh distribution [Celik (2003)].

* Dr. Bhandari is Reader in Mathematics at Birendra Multiple Campus

From the study that has been done by Hadley and Toumi (2003), the two parameter Lognormal distribution can be a very good description of annual mean daily sulphur dioxide concentrations for a wide range of ambient levels, time periods and monitoring site types. The Lognormal distribution has a consistently better fit to the data than the normal distribution.

There are many techniques to estimate the distribution parameters, namely the method of moments, percentiles and maximum likelihood estimation (MLE) [Georgopoulos and Seinfeld (1982)] and Bayesian method of estimation [Sultan and Ahmad (2013)]. The method of moments was more widely used whereas the method of maximum likelihood provides the best estimate of the parameters [Mage and Ott (1984)].Like these, Bayesian method of estimation is also a better estimation procedure which provides more reliable estimates as it uses the prior information in terms of prior probability density function

In this study, the theoretical distribution Lognormal is used to fit the parent distribution of PM_{10} of oct.2013 to may 2014 for Putalisadak (one of the busiest streets) of Kathmandu, Nepal. Two estimating methods namely method of maximum likelihood and method of moments are used to estimate the parameters of the theoretic distributions. From the statistical properties of air pollutants, the probabilities of air pollutant concentration, the Ambient Quality Standards can be predicted.

2 PM_{10} Concentration Data and Lognormal Distribution

In this study, PM_{10} hourly concentration data in Putalisadak, Kathmandu, Nepal are used. The data [AQM (2014)] were prepared by Water Engineering and Training Center (P) Ltd., Dillibazar, Kathmandu and submitted to Department of Environment, Kupondol [Ukesh (2014)]. These observed data are used to estimate the parameters of the Lognormal distribution.

The probability density function (pdf) for two parameters Lognormal distribution is given as

$$f(x;\mu,\sigma^2)=\frac{1}{x\sigma^2\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{\log x-\mu}{\sigma^2}\right)^2\right],\mu>0,\sigma^2 x>0,x>0 \qquad \text{... (1)}$$

The mean and variance of the Lognormal distribution are given by

$$E(X)=\alpha=\exp(\mu+\frac{\sigma^2}{2}) \qquad \text{... (2)}$$

$$V(X)=\beta=\exp(2\mu+\sigma^2)(\exp\sigma^2-1) \qquad \text{... (3)}$$

The commutative distribution function is given by

$$F(x;\mu,\sigma^2)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\frac{\log x-\mu}{\sigma^2}} e-\frac{x^2}{2}dx,\mu>0and\sigma^2>0,x>0 \qquad \text{... (4)}$$

Here, we use the following two well-known methods to estimate the parameters of the Lognormal distribution:

2.1 The Method of Maximum Likelihood

The method of maximum Likelihood is a popular estimation technique for most of distributions because it picks the values of the distribution parameters that make the data more likely than any other values of the parameters. This is accomplished by maximizing the likelihood function of the parameters with given data.

Let us consider the estimation of the parameters as α and β.

Also, let U*i* = log*x i*: *i* =1,2,..., *n*. Then, using the fact that (U_1, U_2, U_3, ..., U_n) is a random sample from normal distribution with parameters (μ, σ^2).

For Lognormal distribution equation (1), the likelihood function is defined as follows:

$$L = \prod_{i=1}^{n} f(x,\mu,\sigma^2) = \prod_{i=1}^{n} \frac{1}{x\sigma^2\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\log x-\mu}{\sigma^2}\right)^2\right]$$

or $$L = (\sigma^2)^{-n}(2\pi)^{-\frac{n}{2}} \prod_{i=1}^{n} x_i^{-1} \exp\left[-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{\log x-\mu}{\sigma^2}\right)^2\right] \qquad ..\ (5)$$

Taking log on both the sides of the likelihood function, we get

$$\log L = -n\log\sigma^2 - \frac{n}{2}\log(2\pi) - \log(\prod_{i=1}^{n} x_i) - \frac{1}{2}\sum_{i=1}^{n}\left(\frac{\log xi-\mu}{\sigma^2}\right)^2 \qquad ...\ (6)$$

Differentiating (6) with respect to μ, we get

$$\frac{\partial \log L}{\partial\mu} = -\frac{1}{2}\sum_{i=1}^{n} 2\left(\frac{\log xi-\alpha}{\beta}\right)\left[\frac{-\sigma^2-0}{(\sigma^2)^2}\right]$$

or $$\frac{\partial \log L}{\partial\mu} = \sum_{i=1}^{n}\left(\frac{\mu-\log x_i}{(\sigma^2)^2}\right) \qquad ...\ (7)$$

Again, differentiating (6) with respect to σ^2, we get

$$\frac{\partial logL}{\partial\sigma^2} = -\frac{n}{\sigma^2} - \frac{1}{2}\sum_{i=1}^{n}(logx_i-\mu)^2(-2)(\sigma^2)^{-3}$$

or $$\frac{\partial logL}{\partial\sigma^2} = -\frac{n}{\sigma^2} + \sum_{i=1}^{n}(logx_i-\alpha)^2(\sigma^2)^{-3} \qquad ...\ (8)$$

Now, taking $\frac{\partial logL}{\partial\mu} = 0$, we get

$\sum_{i=1}^{n}\left(\frac{\mu-logx_i}{(\sigma^2)^2}\right) = 0$ or $n\mu - \sum_{i=1}^{n} logx_i = 0$

Therefore, we get $$\hat{\mu}_{ML} = \frac{1}{n}\sum_{i=1}^{n} logx_i \qquad ...\ (9)$$

Again, taking $\frac{\partial logL}{\partial \sigma^2} = 0$, we get

$$-\frac{n}{\sigma^2} + \sum_{i=1}^{n}(logx_i - \mu)^2(\sigma^2)^{-3} = 0 \text{ or } \frac{n}{\sigma^2} = \sum_{i=1}^{n}(logx_i - \mu)^2(\sigma^2)^{-3}$$

or $$n(\sigma^2)^2 = \sum_{i=1}^{n}(logx_i - \mu)^2 \quad ... (10)$$

Therefore, we have $$\hat{\sigma}_{ML}^2 = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(logx_i - \hat{\mu}_{ML})^2} \quad ... (11)$$

Thus, the MLE of α and β are given by

$$\hat{\alpha}_{ML} = \exp\left[\hat{\mu}_{ML} + \frac{\hat{\sigma}_{ML}^2}{2}\right] \quad ... (12)$$

and $$\hat{\beta}_{ML} = \exp[2\hat{\mu}_{ML} + \hat{\sigma}^2{}_{ML}]\,[\exp(\hat{\sigma}_{ML}^2) - 1] \quad ... (13)$$

2.2 The Method of Moments

To compute the moments estimators μ and σ^2, we first find E (x) and E (x^2) of the Lognormal distribution. The moments of the Lognormal distribution are given by the following equation defined by Casella and Berger (2002):

$$E(x^t) = \exp(t\mu + \frac{t^2\sigma^2}{2}) \quad ... (14)$$

Using the above equation, we can write

$$E(x) = \exp(\mu + \frac{\sigma^2}{2}) \quad ... (15)$$

and $$E(x^2) = \exp(2\mu + 2\sigma^2) \quad ... (16)$$

Now, we set $E(x)$ equals to the first sample moment m_1 and $E(x^2)$ equals to the second sample moment m_2 . Thus, we have

$$m_1 = \frac{\sum_{i=1}^{n} x_i}{n} \quad \text{and} \quad m_2 = \frac{\sum_{i=1}^{n} x_i^2}{n} \quad ... (17)$$

Using the first part of equation (17) in equation (15), we get

$$e^{\mu + \frac{\sigma^2}{2}} = \frac{\sum_{i=1}^{n} x_i}{n} \quad ... (18)$$

Taking log on both sides of equation (18), we get

$$\mu + \frac{\sigma^2}{2} = log(\frac{\sum_{i=1}^{n} x_i}{n})$$

or $$\mu = log(\sum_{i=1}^{n} x_i) - log\, n - \frac{\sigma^2}{2} \quad ... (19)$$

Using the second part of equation (17) in equation (16), we get

$$e^{2\mu + 2\sigma^2} = \frac{\sum_{i=1}^{n} x_i^2}{n} \quad ... (20)$$

Taking log on both sides of equation (20), we get

$$2\mu + 2\sigma^2 = log(\frac{\sum_{i=1}^{n} x_i^2}{n})$$

or $$\mu = \frac{log(\sum_{i=1}^{n} x_i^2)}{2} - \frac{logn}{2} - \sigma^2 \quad ... (21)$$

Now, from equations (19) and (21), we get

$$log(\sum_{i=1}^{n} x_i) - log\, n - \frac{\sigma^2}{2} = \frac{log(\sum_{i=1}^{n} x_i^2)}{2} - \frac{logn}{2} - \sigma^2$$

or $$\sigma^2 = log(\sum_{i=1}^{n} x_i^2) - 2log(\sum_{i=1}^{n} x_i) + log(n) \quad ... (22)$$

Using the equation (22) in equation (21), we get

$$\mu = log(\sum_{i=1}^{n} x_i) - log\, n - \frac{log(\sum_{i=1}^{n} x_i^2) - 2log(\sum_{i=1}^{n} x_i) + log\, n}{2}$$

or $$\mu = 2log(\sum_{i=1}^{n} x_i) - \frac{3}{2} log\, n - \frac{log(\sum_{i=1}^{n} x_i^2)}{2} \quad ... (23)$$

Therefore, moments estimators are given by

$$\hat{\mu}_{Mo} = -\frac{log(\sum_{i=1}^{n} x_i^2)}{2} + 2log(\sum_{i=1}^{n} x_i) - \frac{3}{2} log\, n \quad ... (24)$$

and $$\hat{\sigma}_{Mo}^2 = log(\sum_{i=1}^{n} x_i^2) - 2log(\sum_{i=1}^{n} x_i) + log\, n \quad ... (25)$$

Thus, the moments estimates of α and β are given by

$$\hat{\alpha}_{MO} = \exp\left[\hat{\mu}_{MO} + \frac{\hat{\sigma}_{MO}^2}{2}\right] \quad ... (26)$$

and $$\hat{\beta}_{MO} = \exp[2\hat{\mu}_{MO} + \hat{\sigma}_{MO}^2]\,[\exp(\hat{\sigma}_{MO}^2) - 1] \quad ... (27)$$

3. Results and Discussion

The values of α and β derived from data for PM_{10} of Oct. to May, 2013-2014 are given in Table 1. From the Table, it can be seen that even though there were three different methods used for obtaining estimated values, the values obtained for α and β are quite similar.

Figures (1-3) shows an illustration of Cumulative density function F(x) plot for Lognormal distribution using α and β for PM_{10} of Oct. to May, 2013-2014 with different methods of parameter estimation. Two methods namely, maximum likelihood method and method of moments are used to estimate parameters for the Lognormal distribution. The values of estimated parameters are given in Table 1. For Lognormal distribution using Oct. to May 2013-2014, there is not much difference among MLE and method of moments for estimating the parameter.

Table 1: Estimation of Alpha α and Beta values for PM_{10} of the year 2014 of Putalisadak, Kathmandu (Nepal)

Parameter Estimation	Data	
	α	β
MLE	5.6641	0.1167
Method of moment	5.6769	0.0822

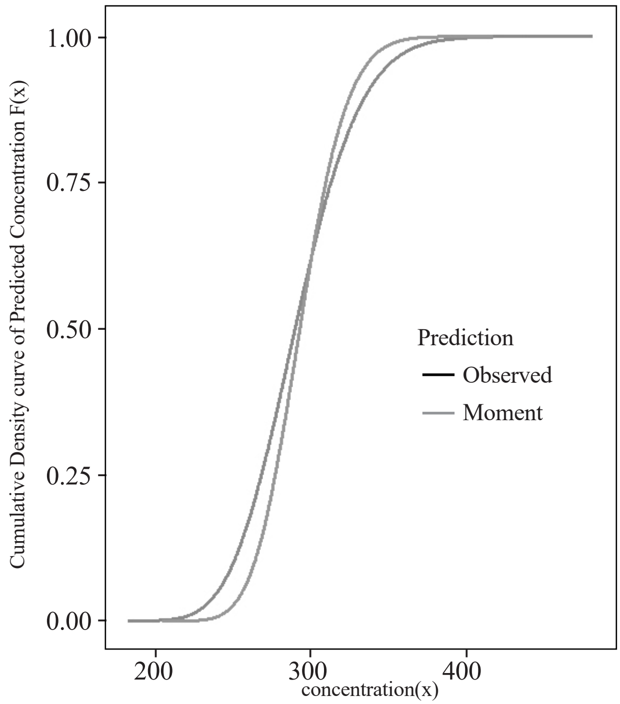

Fig.1 Cumulative Density Curve of Predicted Concentration for Observed and Moment Estimates of PM_{10}

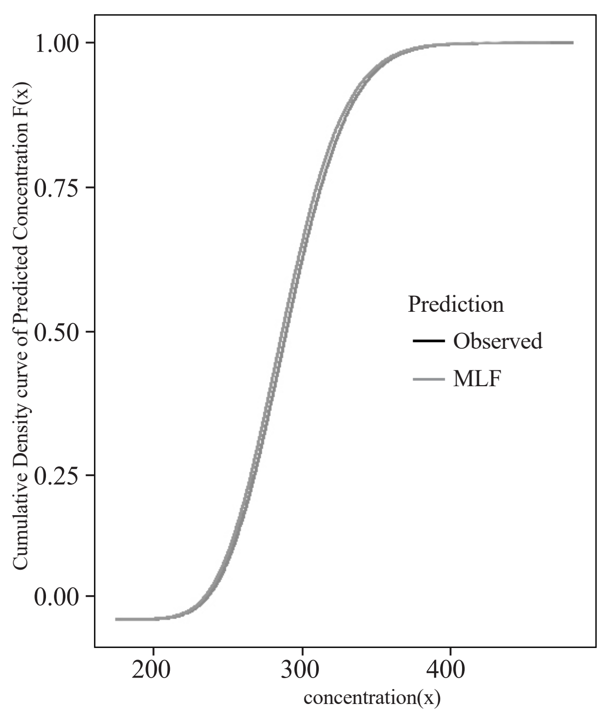

Fig. 2 Cumulative Density Curve of Predicted Concentration for Observed and MLE Estimates of PM_{10}

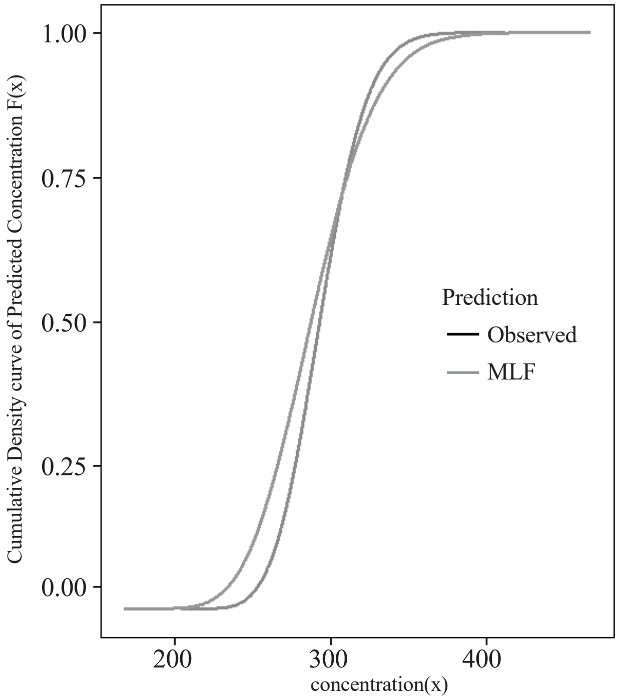

Fig. 3 Cumulative Density Curve of Predicted Concentration for Moment and MLE Estimates of PM_{10}

Prediction of the Probability Exceeding the Kathmandu Ambient Quality Standards Figures (1-3) show the CDF's for PM_{10} concentration in Kathmandu for the year of 2014. The values of estimated parameters are $\hat{\alpha}_{MO} = 5.6769$ and $\hat{\beta}_{MO} = 0.0822$ PM_{10} of oct. 2013 to may 2014 using method of moments. Even though the CDF plot for PM_{10} of the year 2014 showed approximately similar distribution, the distribution using method of moment gives better fit compared to the maximum likelihood estimate (MLE) especially at higher PM_{10} concentrations.

4. Conclusion

The results in table 1 and Figures from (1-3) of this work show that the PM_{10} concentration distribution can be represented by Lognormal model. Among the two parametric estimating methods, there is not much different between them. For that method of moments with , $\hat{\alpha}_{MO} = 5.6769$ and $\hat{\beta}_{MO} = 0.0822$ PM_{10} of the year 2014 is robust than other. The probabilities exceeding Kathmandu Ambient Quality Standards can be predicted by the lognormal model for PM_{10} concentration.

References

Casella, G. and Berger, R. L. (2002): Statistical Inference (Second Edition), World Press.

Celik, A. N. (2003): A Statistical Analysis of Wind Power Density Based on the Weibull and Rayleigh Models at the Southern Region of Turkey. Journal of Renewable Energy, 29, pp. 593-604.

Daly, A. and Zannitti, P. (2007): **An** introduction to air pollution-Definitions, Classification and History, Chapter 1 of ambient air pollution. The Arab School of Science and Technology (ASST) and The EnviroComp Institute.

Georgopoulos, P.G. and Seinfeld,J.H. (1982): Statistical distribution of air quality concentrations. Environmental Science and Technology, 16, 401A-416A.

Hadley, A. and Toumi, R. (2002): Assessing Changes to the Probability Distribution of Sulphur Dioxide in the UK Using Lognormal Model. Journal of Atmospheric Environment, 37 (24), pp. 455-467.

Hadley, A. and Toumi, R. (2003): Assessing changes to the probability distribution of sulphur dioxide in the UK using a lognormal model. Atmospheric Environment, 37, 1461-1474.

Mage, D.T. and Ott, W.R. (1984): An evaluation of the method of fractals, moments and maximum likelihood for estimating parameters when sampling air quality data from a stationary lognormal distribution. Atmospheric Environment, 18, 163-171.

Singh, P. (2004): Simultaneous Confidence Intervals for the Successive Ratios of Scale Parameters. Journal of Statistical Planning and Inference, 36 (3), pp. 1007-1019.

Sultan,R. and Ahmad,S.P.(2013): Comparison of Parameters of Lognormal Distribution Based On the Classical and Posterior Estimates. Journal of Modern Applied Statistical Methods, Vol. 12, No. 2, 304-313.

Wang, X. and Mauzerall, D. L. (2004): Characterizing Distributions of Surface Ozone and its Impact on Grain Production in China, Japan and South Korea 1990 and 2020. Journal of Atmospheric Environment, 38 (74), pp. 4383-4402.

Ukesh.com/Annual_Report_kathmandu-2013-2014.pdf

[AQM Data (2014), Water Engineering and Training Center (p) Ltd Dillibazar, Kathmandu,Submitted to Department of Environment Kupondol, Lalitpur.