International Journal of Multidisciplinary and Current Research ISSN: 2321-3124

Research Article Available at: http://ijmcr.com

Use of Lognormal distribution Model to Fit the Observed data for Air Pollutants Concentration

Prem Sagar Bhandari*

Tribhuvan University, Birendra Multiple Campus, Nepal

Received 21 Oct 2020, Accepted 18 Dec 2020, Available online 20 Dec 2020, **Vol.8 (Nov/Dec 2020 issue)**

Abstract

In this paper, the probability density function for two parameters lognormal distribution is used to fit the parent distribution of air pollutants PM_{10} for eight months from October 2013 to may 2014 of road Putalisadak of Kathmandu, Nepal. To estimate the parameters of the theoretic distributions, two estimating methods that is method of moments and Bayesian method are applied. From the study, it is found that the method of moments is best fit with the observed PM_{10} concentration of one of the busiest streets of Kathmandu of the year 2014.

Keywords: *Lognormal distribution, Air pollution PM_{10}, Moments method, Bayesian method, Estimation of parameters*

Introduction

Air pollution from different sources is a major problem in many countries. In most of the big cities in the world, particularly in developing countries like Nepal are facing bad air pollution because of uncontrolled unsystematic urbanization, increasing number of fossils fuel burning vehicles, an unmanaged industry and many more.

Air pollution is harmful to human health or other ecosystems. The massive demolition and reconstruction activities under road expansion and earthquake recovery in Kathmandu Valley have increased the air pollution. The levels of particulate matters in the ambient have gone to hazardous levels.

There are very limited data and case studies on air pollution and almost no air pollution modelling using distribution functions in Nepal. The different types of probability distributions can be used to fit air pollutant concentrations: lognormal distribution [6], Gamma distribution [9], Weibull distribution [12] and Rayleigh distribution [3].

Various authors have been used the Lognormal distribution to represent the type of air pollutant concentration distribution. Hadley and Toumi have found that 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 [7].

*Corresponding author's ORCDI ID: 0000-00000-0000-0000
DOI: https://doi.org/10.14741/ijmcr/v.8.6.6

Bhandari has used the lognormal distribution to fit the data applying the method of maximum likelihood and method of moments to estimate the parameters of theoretic distributions [1].

To estimate the distribution parameters, we have many methods namely the method of moments, percentiles and maximum likelihood estimation (MLE) [5] and Bayesian method of estimation [10]. The method of moments was more widely used whereas the method of maximum likelihood provides the best estimate of the parameters [8].Similarly, Bayesian method of estimation which provides more reliable estimates as it uses the prior information in terms of prior probability density function is also a better estimation procedure.

In this paper, the probability density function (pdf) for two parameters Lognormal distribution is used to fit the parent distribution of air pollutants PM_{10} for eight months from oct.2013 to may 2014 of road Putalisadak of Kathmandu, Nepal. To estimate the parameters of the theoretic distributions, two estimating methods, method of moments and Bayesian method are applied. From the study, we can predict the probabilities of air pollutant concentration and the Ambient Quality Standards.

2 Observed Concentration PM_{10} Data and Lognormal Distribution

For this study, we have taken hourly concentration PM_{10} data from a busy road Putalisadak, Kathmandu, Nepal. The data [AQM (2014)] were collected by Water Engineering and Training Center (P) Ltd., Dillibazar, Kathmandu and submitted to Department of Environment, Kupundol [11]. Here, we have used these observed data to estimate the parameters of the lognormal distribution.

For two parameters Lognormal distribution, the probability density function (pdf) is given by

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

Mean and variance of the lognormal distribution are given as

$$E(X) = \alpha = \exp\left(\mu + \frac{\sigma^2}{2}\right) \quad (2)$$

$$V(X) = \beta = \exp(2\mu + \sigma^2)\,(\exp\sigma^2 - 1) \quad (3)$$

And the commutative distribution function (cdf) is given as

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

Now, we have used two methods to estimate the parameters of the Lognormal distribution:

2.1 The Method of Moments

At first, we find E (x) and E (x^2) and of the Lognormal distribution to obtain the moments estimators μ and σ^2. According to Casella and Berger, the moments of the Lognormal distribution are given by [2]

$$E(x^t) = \exp\left(t\mu + \frac{t^2\sigma^2}{2}\right) \quad (5)$$

Using the above equation, we have

$$E(x) = \exp\left(\mu + \frac{\sigma^2}{2}\right) \quad (6)$$

And

$$E(x^2) = \exp(2\mu + 2\sigma^2) \quad (7)$$

Now, we let $E(x)$ as the first sample moment m_1 and $E(x^2)$ as the second sample moment m_2 . So, we have

$$m_1 = \frac{\sum_{i=1}^n x_i}{n} \quad \text{and} \quad m_2 = \frac{\sum_{i=1}^n x_i^2}{n} \quad (8)$$

Applying the equations (8) in equation (6), we have

$$e^{\mu+\frac{\sigma^2}{2}} = \frac{\sum_{i=1}^n x_i}{n} \quad (9)$$

After taking log on both sides of equation (9), we have

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

or $$\mu = \log\left(\sum_{i=1}^n x_i\right) - \log n - \frac{\sigma^2}{2} \quad (10)$$

Applying equation (8) in equation (7), we have

$$e^{2\mu+2\sigma^2} = \frac{\sum_{i=1}^n x_i^2}{n} \quad (11)$$

After taking log on both sides of equation (11), we have

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

or $$\mu = \frac{\log\left(\sum_{i=1}^n x_i^2\right)}{2} - \frac{\log n}{2} - \sigma^2 \quad (12)$$

From the equations (10) and (12), we have

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

or $$\sigma^2 = \log\left(\sum_{i=1}^n x_i^2\right) - 2\log\left(\sum_{i=1}^n x_i\right) + \log(n) \quad (13)$$

Applying the equation (13) in equation (12), we have

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

or $$\mu = 2\log\left(\sum_{i=1}^n x_i\right) - \frac{3}{2}\log n - \frac{\log\left(\sum_{i=1}^n x_i^2\right)}{2} \quad (14)$$

Thus, moments estimators are given as

or $$\mu = 2\log\left(\sum_{i=1}^n x_i\right) - \frac{3}{2}\log n - \frac{\log\left(\sum_{i=1}^n x_i^2\right)}{2} \quad (15)$$

Thus, moments estimators are given as

$$\hat{\mu}_{Mo} = -\frac{\log\left(\sum_{i=1}^n x_i^2\right)}{2} + 2\log\left(\sum_{i=1}^n x_i\right) - \frac{3}{2}\log n \quad (15)$$

and $$\hat{\sigma}^2_{Mo} = \log\left(\sum_{i=1}^n x_i^2\right) - 2\log\left(\sum_{i=1}^n x_i\right) + \log n \quad (16)$$

Therefore, the moments estimates of α and β are as

$$\hat{\alpha}_{MO} = \exp\left[\hat{\mu}_{MO} + \frac{\hat{\sigma}^2_{MO}}{2}\right] \quad (17)$$

and $$\hat{\beta}_{MO} = \exp[2\hat{\mu}_{MO} + \hat{\sigma}^2_{MO}]\,[\exp(\hat{\sigma}^2_{MO}) - 1] \quad (18)$$

2.2 Bayesian Estimation of Parameters

Consider the estimation of parameters α and β. At first, we obtain the posterior estimates of μ and σ^2 and then simultaneously the posterior estimates for α and β .

The joint prior pdf for μ and σ^2 is considered as

$$P(\mu,\sigma^2) \propto 1 \quad (19)$$

From the Bayes theorem, joint posterior density of μ and σ^2 is given by

$$\pi(\mu,\sigma^2|\underline{x}) \propto P(\mu,\sigma^2).P(\mu,\sigma^2|\underline{x}) \quad (20)$$

The likelihood function of random sample $(x_1, x_2, x_3, \dots \dots x_n)^T$ is given by

$$L(\mu,\sigma^2|x) = (\frac{1}{\sigma x\sqrt{2\pi}})^n \exp\,[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(logx_i-\mu)^2] \quad (21)$$

From (19) and (21), the joint posterior density of μ and σ^2 is given by

$$\pi(\mu,\sigma^2|x) \propto (\frac{1}{\sigma x\sqrt{2\pi}})^n \exp\,[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(logx_i-\mu)^2] \quad (22)$$

$$\pi(\mu,\sigma^2|\underline{x}) = \frac{c}{\sigma^2}\exp\left(-\frac{\beta}{2\sigma^2}\right)\exp\,[\frac{-n}{2\sigma^2}\{\mu-\frac{\sum_{i=1}^{n}logx_i}{n}\}^2 \quad (23)$$

where $\beta = \sum_{i=1}^{n}(logx_i)^2 - \frac{\left(\sum_{i=1}^{n}logx_i\right)^2}{n}$ and c is a normalizing constant.

Lindley (1961) explained if $P(\theta|\underline{x})$ is given by $P(\theta|\underline{x}) = cP(\theta).P(\underline{x}|\theta)$, where c is the normalizing constant. Then, the value of c is obtained by using $c = \left[\int P(\theta).P(\underline{x}|\theta)d\theta\right]^{-1}$.

or $c^{-1} = \int_0^\infty\int_{-\infty}^{\infty}\pi(\mu,\sigma^2|\underline{x})d\mu\, d\sigma^2$ $\quad c^{-1} = \int_0^\infty\int_{-\infty}^{\infty}\frac{1}{(\sigma^2)^{n/2}}\exp\left(-\frac{\beta}{2\sigma^2}\right)\exp\{\frac{-n}{2\sigma^2}\left(\mu-\frac{\sum_{i=1}^{n}logx_i}{n}\right)^2\}d\mu\, d\sigma^2$

Using the transformation $t = \frac{\sqrt{n}(\mu-\frac{\sum_{i=1}^{n}logx_i}{n})}{\sigma}$, we get

$$c^{-1} = \sqrt{\frac{2\pi}{n}}\int_0^\infty\frac{\exp(-\frac{\beta}{2\sigma^2})}{(\sigma^2)^{n-1/2}}d\sigma^2 \quad \text{or} \quad c^{-1} = \sqrt{\frac{2\pi}{n}}\frac{\Gamma(\frac{n-3}{2})}{(\frac{\beta}{2})^{\frac{n-3}{2}}}$$

Therefore, we have

$$c = \sqrt{\frac{n}{2\pi}}\frac{(\frac{\beta}{2})^{\frac{n-3}{2}}}{\Gamma(\frac{n-3}{2})} \quad (24)$$

From equation (23), we get

$$\pi(\mu,\sigma^2|\underline{x}) = \sqrt{\frac{n}{2\pi}}\frac{(\frac{\beta}{2})^{\frac{n-3}{2}}}{\Gamma(\frac{n-3}{2})}\frac{1}{(\sigma^2)^{n/2}}\exp\left(-\frac{\beta}{2\sigma^2}\right)\exp\left[\frac{-n}{2\sigma^2}\left\{\mu-\frac{\sum_{i=1}^{n}logx_i}{n}\right\}^2\right] \quad (25)$$

2.2.1 Marginal Posterior Densities of μ and σ^2

The marginal density of μ is obtained by integrating out $\sigma^2$from (25) and is given as

$$\pi(\mu|\underline{x}) = \int_0^\infty\pi(\mu,\sigma^2|\underline{x})\,d\sigma^2$$

or $\pi(\mu|\underline{x}) = c\int_0^\infty\frac{1}{(\sigma^2)^{n/2}}\exp[\left(-\frac{1}{2\sigma^2}\right)\{\beta+n\left(\mu-\frac{\sum_{i=1}^{n}logx_i}{n}\right)^2\}]d\sigma^2$

or $\pi(\mu|\underline{x}) = c\,\frac{\Gamma(\frac{n}{2}-1)2^{1/2-1}}{\{\beta+n\left(\mu-\frac{\sum_{i=1}^{n}logx_i}{n}\right)^2\}^{\frac{n}{2}-1}}$

or $\pi(\mu|\underline{x}) = \sqrt{\frac{n}{\beta}}\frac{1}{B(\frac{1}{2},\frac{n-3}{2})[1+\frac{n}{\beta}(\mu-\frac{\sum_{i=1}^{n}logx_i}{n})^2]^{\frac{n-2}{2}}}$ (26)

The marginal density of σ^2 is obtained by integrating the joint posterior density of μ and σ^2 given in equation (25) over the range of μ . It is given as

$\pi(\sigma^2|\underline{x}) = c\int_{-\infty}^{\infty}\frac{1}{(\sigma^2)^{n/2}}\exp\left(-\frac{\beta}{2\sigma^2}\right)\exp\{-\frac{n}{2\sigma^2}\left(\mu-\frac{\sum_{i=1}^{n}logx_i}{n}\right)^2\}d\mu$

or $\pi(\sigma^2|\underline{x}) = c\frac{\exp(-\frac{\beta}{2\sigma^2})}{(\sigma^2)^{n/2}}\int_{-\infty}^{\infty}\exp\{-\frac{n}{2\sigma^2}\left(\mu-\frac{\sum_{i=1}^{n}logx_i}{n}\right)^2\}\, d\mu$

or $\pi(\sigma^2|\underline{x}) = c\frac{\exp(-\frac{\beta}{2\sigma^2})\sqrt{2\pi}}{(\sigma^2)^{n/2}\sqrt{n}}$

or $\pi(\sigma^2|\underline{x}) = c\frac{\exp(-\frac{\beta}{2\sigma^2})\beta^{n-3/2}}{(\sigma^2)^{n/2}2^{n-3/2}\Gamma(\frac{n-3}{2})}$ (27)

2.2.2 Posterior estimate of μ and σ^2

The marginal density is given in (26) is a student's t pdf. Thus, the posterior estimates of μ is given as

The marginal density is given in (26) is a student's t pdf. Thus, the posterior estimates of μ is given as

$$\hat{\mu}_{BA} = E(\mu|\underline{x}) = \sqrt{\frac{n}{\beta}}\frac{1}{B(\frac{1}{2},\frac{n-3}{2})}\int_{-\infty}^{\infty}\frac{\mu d\mu}{[1+\frac{n}{\beta}(\mu-\frac{\sum_{i=1}^{n}logx_i}{n})^2]^{\frac{n-2}{2}}} \quad (28)$$

Using the transformation $t = \sqrt{\frac{n}{\beta}}(\mu-\frac{\sum_{i=1}^{n}logx_i}{n})\sqrt{n-3}$, we have

$$\hat{\mu}_{BA} = \frac{\sum_{i=1}^{n}logx_i}{n\sqrt{n-3}\,B(\frac{1}{2},\frac{n-3}{2})}\int_{-\infty}^{\infty}\frac{dt}{[1+\frac{t^2}{n-3}]^{n-2/2}}$$

or $\hat{\mu}_{BA} = \frac{\sum_{i=1}^{n}logx_i}{n}$ (29)

which is the posterior estimate for μ under uniform prior. Now, the posterior estimate of σ^2 can be obtained from equation (27) as

$$\hat{\sigma}^2_{BA} = \int_0^\infty\frac{\sigma^2\exp(-\frac{\beta}{2\sigma^2})\beta^{n-3/2}}{(\sigma^2)^{n-1/2}2^{n-3/2}\Gamma(\frac{n-3}{2})}d\sigma^2$$

which can be written as $\hat{\sigma}^2_{BA} = \frac{\beta}{n-3}$ (30)

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

$$\hat{\alpha}_{BA} = \exp\left[\hat{\mu}_{BA}+\frac{\hat{\sigma}^2_{BA}}{2}\right] = \exp\,[\frac{\sum_{i=1}^{n}logx_i}{n}+\frac{\beta}{2(n-3)}] \quad (31)$$

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

or $\hat{\beta}_{BA} = \exp\left[2\,\frac{\sum_{i=1}^{n}logx_i}{n}+\frac{\beta}{n-3}\right][\exp\left(\frac{\beta}{n-3}\right)-1]$ (32)

3 Results and Discussion

The obtained values of α and β from the observed data for PM_{10} of Putalisadak are given in Table 6.1. From the table, we find that the values obtained for α and β are almost same even though methods are different to obtain estimated values,

Figure (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 method of moments and Bayesian method of parameter estimation. The values of estimated parameters are given in Table 1. For Lognormal distribution using Oct. to May 2013-2014, there is not much difference between method of moments and Bayesian method for estimating the parameter.

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

Parameter Estimation	Data	
	α	β
Method of moment	5.6769	0.0822
Bayesian method	5.6674	0.1151

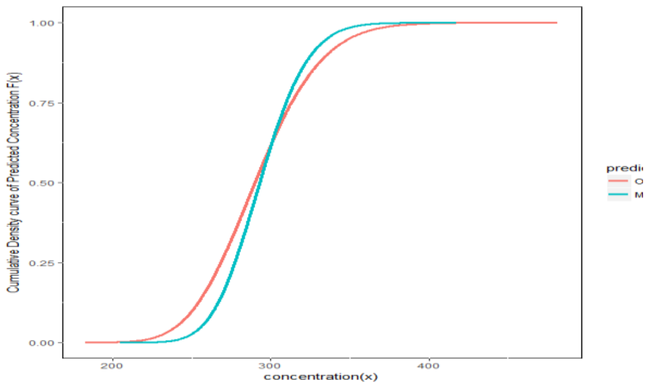

Fig.1 Cumulative Density Curve of Predicted Concentration for Observed and Moment Estimates of PM_{10}

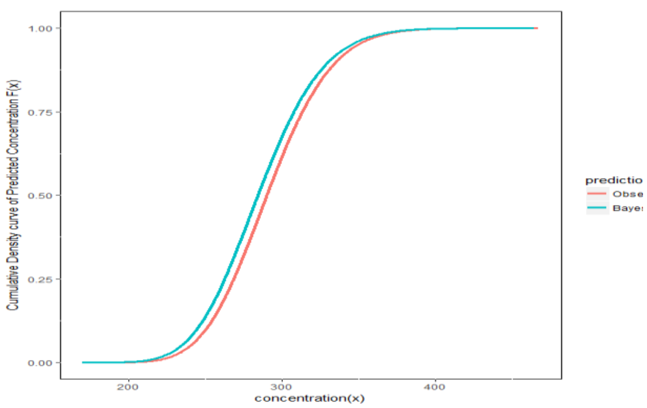

Fig.2 Cumulative Density Curve of Predicted Concentration for Observed and Bayes Estimates of PM_{10}

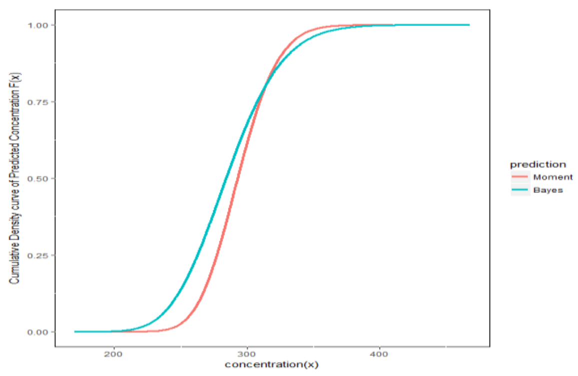

Fig. 3 Cumulative Density Curve of Predicted Concentration for Moment and Bayes Estimates of PM_{10}

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. It is seen that 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 Bayesian method for estimating the parameter.

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. The two parametric estimating methods are almost same not much difference between them. It is found that the method of moments is best fit with the observed PM_{10} concentration of one of the busiest streets of Kathmandu of the year 2014. So it is good to use moments method to fit the parent data between the method of moments and Bayesian method for estimating the parameter.

References

[1]. Bhandari,P.S.(2018). A Study of Lognormal Model for Air Pollution Concentration.BMC Journal of Scientific Research, 2,107-114.
[2]. Casella, G. and Berger, R. L. (2002): Statistical Inference (Second Edition), World Press.
[3]. 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.
[4]. 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.
[5]. Georgopoulos, P.G. and Seinfeld,J.H. (1982). Statistical distribution of air quality concentrations. Environmental Science and Technology, 16, 401A-416A.
[6]. 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.

[7]. 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.

[8]. 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.

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

[10]. 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.

[11]. Ukesh.com/Annual_Report_kathmandu-2013-2014.pdf

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

[13]. 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.