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\textsubscript{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\textsubscript{10} concentration of one of the busiest streets of Kathmandu of the year 2014.

Keywords: Lognormal distribution, Air pollution PM\textsubscript{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].

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\textsubscript{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\textsubscript{10} Data and Lognormal Distribution

For this study, we have taken hourly concentration PM\textsubscript{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
obtained data to estimate the parameters of the lognormal distribution.

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

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

Mean and variance of the lognormal distribution are given as

\[ E(X) = \alpha = \exp(\mu + \frac{\sigma^2}{2}) \] (2)

\[ V(X) = \beta = \exp(2\mu + \sigma^2)(\exp \sigma^2 - 1) \] (3)

And the commutative distribution function (cdf) is given as

\[ (x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \frac{e^{-t^2}}{\sigma^2} \, dt, \mu > 0 \text{ and } \sigma^2 > 0, x > 0 \] (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 \( E(x^2) \) of the Lognormal distribution to obtain the moments estimators \( \mu \) and \( \sigma^2 \). According to Casella and Berger, the moments of the Lognormal distribution are given by [2]

\[ E(x^i) = \exp(t \mu + \frac{t^2 \sigma^2}{2}) \] (5)

Using the above equation, we have

\[ E(x) = \exp(\mu + \frac{\sigma^2}{2}) \] (6)

And

\[ E(x^2) = \exp(2\mu + 2\sigma^2) \] (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

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

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

\[ e^{\mu + \frac{\sigma^2}{2}} = \frac{m_2}{m_1} \] (9)

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

\[ \mu + \frac{\sigma^2}{2} = \log \left(\frac{m_2}{m_1}\right) \] (10)

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

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

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

\[ 2\mu + 2\sigma^2 = \log \left(\frac{\sum_{i=1}^{n} x_i^2}{n}\right) \] (12)

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

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

or

\[ \sigma^2 = \log \left(\frac{\sum_{i=1}^{n} x_i^2}{n}\right) - 2\log \left(\sum_{i=1}^{n} x_i\right) + \log(n) \] (13)

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

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

or

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

Thus, moments estimators are given as

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

Thus, moments estimators are given as

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

and

\[ \hat{\sigma}_{MO}^2 = \log \left(\frac{\sum_{i=1}^{n} x_i^2}{n}\right) - 2\log \left(\sum_{i=1}^{n} x_i\right) + \log(n) \] (16)

Therefore, the moments estimates of \( \alpha \) and \( \beta \) are as

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

and

\[ \hat{\beta}_{MO} = \exp \left[2\hat{\mu}_{MO} + \frac{\hat{\sigma}_{MO}^2}{2}\right][\exp(\frac{\hat{\sigma}_{MO}^2}{2}) - 1] \] (18)

2.2 Bayesian Estimation of Parameters

Consider the estimation of parameters \( \alpha \) and \( \beta \). At first, we obtain the posterior estimates of \( \mu \) and \( \sigma^2 \) and then simultaneously the posterior estimates for \( \alpha \) and \( \beta \). The joint prior pdf for \( \mu \) and \( \sigma^2 \) is considered as

\[ P(\mu, \sigma^2) \propto 1 \] (19)

From the Bayes theorem, joint posterior density of \( \mu \) and \( \sigma^2 \) is given by
The marginal density of $\sigma^2$ is obtained by integrating the joint posterior density of $\mu$ and $\sigma^2$ given in equation (25) over the range of $\mu$. It is given as

$$
\pi(\sigma^2 | x) = \int_{0}^{\infty} \pi(\mu, \sigma^2 | x) \, d\sigma^2
$$

where

$$
\pi(\mu, \sigma^2 | x) = \frac{1}{\beta} \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-\alpha}{2}\right)} \left(\frac{\beta}{\sigma^2}\right)^{\frac{n}{2}} \exp\left(-\frac{\beta}{2\sigma^2} \left(\frac{\mu - \sum_{i=1}^{n} \log x_i}{n}\right)^2\right)
$$

and $c$ is a normalizing constant.

Lindley (1961) explained if $P(\theta | x)$ is given by $P(\theta | x) = c P(\theta), P(x | \theta) , \text{ where c is the normalizing constant.}$

Then, the value of $c$ is obtained by using $c = \left[\int P(\theta) P(x | \theta) d\theta\right]^{-1}$. Therefore, we have $c = \left[\frac{n}{\beta} \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-\alpha}{2}\right)} \left(\frac{\beta}{\sigma^2}\right)^{\frac{n}{2}} \exp\left(-\frac{\beta}{2\sigma^2} \left(\frac{\mu - \sum_{i=1}^{n} \log x_i}{n}\right)^2\right)\right]^{\frac{1}{2}}$

Using the transformation $t = \frac{\sqrt{2n} \frac{\beta}{\sigma^2} \sum_{i=1}^{n} \log x_i}{\Gamma\left(\frac{n}{2}\right)} \sigma$, we get

$$
c^{-1} = \frac{\frac{\beta}{\sigma^2} \sum_{i=1}^{n} \log x_i}{\sqrt{2n} \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-\alpha}{2}\right)} \left(\frac{\beta}{\sigma^2}\right)^{\frac{n}{2}} \exp\left(-\frac{\beta}{2\sigma^2} \left(\frac{\mu - \sum_{i=1}^{n} \log x_i}{n}\right)^2\right)}
$$

2.2.1 Marginal Posterior Densities of $\mu$ and $\sigma^2$

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

$$
\pi(\mu | x) = \int_{0}^{\infty} \pi(\mu, \sigma^2 | x) \, d\sigma^2
$$

or

$$
\pi(\mu | x) = c \int_{0}^{\infty} \frac{1}{(\sigma^2)^{\frac{n}{2}}} \exp\left(-\frac{\beta}{2\sigma^2} \left(\frac{\mu - \sum_{i=1}^{n} \log x_i}{n}\right)^2\right) \, d\sigma^2
$$

or

$$
\pi(\mu | x) = c \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \beta^{\frac{n}{2}} \exp\left(-\frac{\beta}{2\sigma^2} \left(\frac{\mu - \sum_{i=1}^{n} \log x_i}{n}\right)^2\right)
$$

or

$$
\pi(\mu | x) = \frac{1}{\beta} \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \left(\frac{\beta}{\sigma^2}\right)^{\frac{n}{2}} \exp\left(-\frac{\beta}{2\sigma^2} \left(\frac{\mu - \sum_{i=1}^{n} \log x_i}{n}\right)^2\right)
$$

The marginal density of $\sigma^2$ is given in (26) is a student’s t pdf.
3 Results and Discussion

The obtained values of $\alpha$ and $\beta$ 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 $\alpha$ and $\beta$ 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 $\alpha$ and $\beta$ 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($\alpha$) and Beta($\beta$) values for PM$_{10}$ of the year 2014 of Putalisadak, Kathmandu (Nepal)

<table>
<thead>
<tr>
<th>Parameter Estimation</th>
<th>$\alpha$</th>
<th>$\beta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Method of moment</td>
<td>5.6769</td>
<td>0.0822</td>
</tr>
<tr>
<td>Bayesian method</td>
<td>5.6674</td>
<td>0.1151</td>
</tr>
</tbody>
</table>

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

[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.


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