*AMSE JOURNALS-AMSE IIETA publication-2017-Series: Advances A; Vol. 54; N°2; pp 278-289*
*Submitted Mar. 20，2017; Revised May 09, 2017; Accepted May 23, 2017*

# Alpha Normal Distribution and Its Application in China's Logistics Prosperity Index

*Xiaheng Zhang, **Ronggang Zhang, ***Lingling Li

*Business School, Northwest University of Political Science and Law, Xi'an 710000, China (zhangxiaheng@163.com)

**Business School, Northwest University of Political Science and Law, Xi'an 710000, China

***Business School, Northwest University of Political Science and Law, Xi'an 710000, China

## Abstract

Despite the extensive use in data processing, the symmetric distribution fails to yield satisfactory results. Based on the traditional normal distribution, the asymmetric distribution is often adopted to handle the features of asymmetric-tailed data. In view of the above, this paper proposes a new normal distribution, alpha normal distribution, to deal with skewed and heavy-tailed data. The validity of the proposed distribution was verified by fitting the data of China's logistics prosperity index (LPI).

## Key words

Skewed Data, Heavy-tailed data, Alpha normal distribution, Logistics Prosperity Index (LPI).

## 1. Introduction

To obtain the desired results, the data are by default normally distributed in many studies. Under real-world conditions, however, many data as too skewed or heavy-tailed to be interpreted by normal distribution, and should be depicted by more flexible statistical models. A viable option is the symmetrical distribution, a variation of normal distribution. Its density function is expressed as $f_{LN}(x;\mu,\sigma)=\frac{1}{\sqrt{2\pi}\sigma x}\exp\left[-\frac{(\ln x-\mu)^2}{2\sigma^2}\right]$, where x＞0. The symmetrical distribution does well in analysing skewed data of positive real numbers. It has been applied in the research of various fields, particularly the analysis of pollution concentration. Some representative studies are presented as follows.

Based on the data on air pollutant concentration in Beijing, Fan [1] looked for the general distribution pattern of air pollutant concentration in northern China during the winter, and concluded that the pollution concentration follows lognormal distribution, exponential distribution and other distribution patterns. Wayne [2] suggested that the pollutant concentration in different environments adheres to or approximates lognormal frequency distribution, that is, the logarithms of the measured concentrations are distributed in a near normal or Gaussian pattern in the plot of frequency distribution. Larry [3] held that symmetric distribution fits well in analysing the data on environmental pollution, and lognormal distribution is a good choice to characterize the intrinsically positive and often highly skewed environmental data. However, the scholar did not mention the implications of the use of the said distribution. Han et al. [4] collected the PM2.5 concentration from 12 monitoring stations, and verified that the concentration obeys the lognormal distribution. In 2012, Chen [5] discussed the distribution of air pollutant concentration in different seasons in Yulin, and revealed that the lognormal distribution is the most suitable model for SO2 concentration in winter. Through Maximum likelihood estimation, Han and Chen [6] verified the fitness of four theoretical distributions (lognormal, gamma, Pearson and generalized extreme value) to the daily average concentrations of PM10, SO2 and NO2 in Shanghai from June 1, 2000 to May 31, 2003, and evaluated the fitting results by the chi-square test. The results show that the most suitable distributions for PM10, SO2 and NO2 concentrations in Shanghai are lognormal, Pearson and generalized extreme value, respectively.

The symmetrical distribution has also been employed in decision analysis and quality evaluation. With attributes of unknown weights and valued as interval numbers, Xu and Lv [7] assumed that the interval numbers obey the normal distribution, presented a new attribute weighting strategy, and conducted a multi-attribute decision-making. Their research reveals that normal distribution is an effective tool for multi-attribute decision-making method with interval numbers. Zhao et al. [8] applied the normal distribution to assess the quality of regional geochemical exploration samples, and conducted a normal distribution test to convert the trace elements into normal distribution by log transformation.

In the above studies, the asymmetry and kurtosis of data deviate greatly from the lognormal distribution, making it necessary to develop even more flexible distribution to model asymmetric datasets.

In 1985, Azzalini put forward the partial normal distribution by introducing the asymmetric parameter into the normal distribution. The asymmetric distribution has been widely used to model various asymmetric-tailed datasets [9]. The distribution is introduced as follows. If a random variable X obeys the normal distribution with the density function of $\varnothing(\mathrm{x};\alpha)=2\varphi(\mathrm{x})\Phi(\alpha\,\mathrm{x})$ (xϵR), where $\varphi(x)$ and $\Phi(x)$ are the density function and distribution function of the standard normal distribution, respectively; α is the shape parameter that controls skewness, then X follows the partial normal distribution with the shape parameter of α, denoted as X~SN(α). The density function is reduced to normal distribution when α=0, and equal to the maximum distribution of the two independent standard normal variables when α=1. The partial normal distribution is applicable to asymmetric data fitting, as it carries some properties of the original symmetric distribution.

Henze [10] further explored the probability expressions of the partial normal distribution and deduced the odd order moments of the distribution. Recently, Azzalini and Dalla Valle [11] investigated the multivariate skewed normal distribution, which is created by adding a shape parameter to normal distribution, and extended the generalized normal distribution to the multivariate case. Huang and Chen [12] introduced a skewed function based on a symmetric distribution and derived the general structure of partial normal distribution.

Unlike lognormal distribution, the asymmetric distribution has rarely been explored in an in-depth manner. The only exceptions are Gupta et al. [13] and Chen et al. [14]. The former examined numerous skewed data distributions, while the latter implemented partial distribution in environmental pollution analysis. Based on the traditional normal distribution, this paper proposes a new normal distribution, alpha normal distribution. After parameter estimation and statistical property analysis, the alpha normal distribution was applied in actual datasets, and was validated through comparison with normal distribution, partial normal distribution, and lognormal distribution.

## 2. Alpha Normal Distribution

Definition 1: If the density function of a random variable X is as follows:

$$f(\mathrm{x};\alpha)=\alpha\left[1-\Phi(\mathrm{x})\right]^{\alpha-1}\varphi(\mathrm{x}),\ \mathrm{x}\epsilon\mathrm{R},\ \alpha>0 \quad (1)$$

where $\varphi(x)$ and $\Phi(x)$ are the density and distribution functions of the standard normal distribution, respectively; α is the shape parameter that controls skewness; then X obeys the alpha normal distribution with the shape parameter of $\alpha$, denoted as X~AN(α).

The density function is reduced to standard normal distribution when α=1, tends towards the right when α>1, and tends towards the left when α<1.

The distribution function of the alpha normal distribution is expressed as:

$$F(x;\alpha)=1-[1-\Phi(x)]^{\alpha}, \ x\epsilon R, \ \alpha>0 \tag{2}$$

where $\Phi(x)$ is the distribution function of standard normal distribution; α is the shape parameter that controls skewness. The greater the value of α, the steeper the curve of the distribution function.

# 3. Properties of Alpha Normal Distribution

## 3.1 Expectation and Variance

If the random variable X~AN(α), its expectation is the following.

$$E(X)=\alpha\int_{-\infty}^{+\infty}x[1-\Phi(x)]^{\alpha-1}\varphi(x)\,dx$$

$$=-\alpha\int_{-\infty}^{+\infty}[1-\Phi(x)]^{\alpha-1}\,d\varphi(x)$$

$$=\alpha(1-\alpha)\int_{-\infty}^{+\infty}[\varphi(x)]^{2}[1-\Phi(x)]^{\alpha-2}\,dx$$

Let $C_m(\lambda)=\int_{-\infty}^{+\infty}[1-\Phi(\lambda x)]^{m}\varphi(x)\,dx$, m=0, 1, 2, …, λϵR

Thus, $\int_{-\infty}^{+\infty}\left[\frac{1}{2}-\Phi(\lambda x)\right]^{2m+1}\varphi(x)\,dx=0$.

whereas the integrand is a singular function [15], we have:

$$\int_{-\infty}^{+\infty}\left[\frac{1}{2}-\Phi(\lambda x)\right]^{2m+1}\varphi(x)\,dx$$

$$=\int_{-\infty}^{+\infty}\left[1-\Phi(\lambda x)-\frac{1}{2}\right]^{2m+1}\varphi(x)\,dx$$

$$= \int_{-\infty}^{+\infty}\left\{\sum_{i=0}^{2m+1}[1-\Phi(\lambda x)]^{2m+1-i}\binom{2m+1}{i}\left(-\frac{1}{2}\right)^{i}\right\}\varphi(x)\,dx$$

$=0$

$$\int_{-\infty}^{+\infty}[1-\Phi(\lambda x)]^{2m+1}\varphi(x)\,dx$$

$$= -\int_{-\infty}^{+\infty}\left\{\sum_{i=1}^{2m-1}[1-\Phi(\lambda x)]^{2m+1-i}\binom{2m+1}{i}\left(-\frac{1}{2}\right)^{i}\right\}\varphi(x)\,dx$$

$$= \sum_{i=1}^{2m+1}(-1)^{i+1}\binom{2m+1}{i}\left(\frac{1}{2}\right)^{i}\int_{-\infty}^{+\infty}[1-\Phi(\lambda x)]^{2m+1-i}\varphi(x)\,dx$$

Thus:

$$C_{2m+1}(\lambda) = \sum_{i=1}^{2m+1}(-1)^{i+1}\binom{2m+1}{i}\left(\frac{1}{2}\right)^{i}C_{2m+1-i}(\lambda)$$

Hence, we may reasonably draw the following:

$C_0(\lambda)=1$, $C_1(\lambda)=\frac{1}{2}C_0(\lambda)=\frac{1}{2}$, $C_2(\lambda)=\frac{1}{\pi}\arctan\sqrt{1+2\lambda^2}$, $C_3(\lambda)=\frac{3}{2\pi}\arctan\sqrt{1+2\lambda^2}-\frac{1}{4}$, etc.

Therefore, if α=1, 2, 3, 4, …, the E(X) is 0, $-\frac{1}{\sqrt{\pi}}$, $-\frac{3}{2\sqrt{\pi}}$, $-\frac{6\arctan\sqrt{2}}{\pi^{\frac{3}{2}}}$, …, respectively.

The moment generating function of $X^2$ is:

$$M_{x^2}(t) = E(e^{tX^2})$$

$$= \alpha\int_{-\infty}^{+\infty} e^{tx^2}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}[1-\Phi(x)]^{\alpha-1}\varphi(x)\,dx$$

$$= \frac{\alpha}{\sqrt{1-2t}}\int_{-\infty}^{+\infty}\varphi(u)\left[1-\Phi\left(\frac{u}{\sqrt{1-2t}}\right)\right]^{\alpha-1}du$$

$$= \frac{\alpha}{\sqrt{1-2t}}C_{\alpha-1}\left(\frac{1}{\sqrt{1-2t}}\right)$$

Thereby

$$E(X^2) = \alpha C_{\alpha-1}(1) + \alpha C'_{\alpha-1}(1)$$

$$= \alpha C_{\alpha-1}(1) + \frac{\alpha(\alpha-1)(\alpha-2)}{4\pi\sqrt{3}} C_{\alpha-3}\left(\frac{1}{\sqrt{3}}\right)$$

Therefore, if α=1, 2, 3, 4, …, the $E(X^2)$ is 1, 1, $1+\frac{3}{2\sqrt{\pi}}$, $1+\frac{\sqrt{3}}{\pi}$, …, respectively.

The variance can be obtained by $Var(X^2) = E(X^2) - [E(X)]^2$.

The central moment of order K is [16]:

$$E(X^k) = \int_{-\infty}^{+\infty} x^k \alpha [1-\Phi(x)]^{\alpha-1} \varphi(x)\, dx$$

$$= \alpha \int_{-\infty}^{+\infty} x^{k-1} [1-\Phi(x)]^{\alpha-1} x\varphi(x)\, dx$$

$$= \alpha \int_{-\infty}^{+\infty} (k-1)\, x^{k-2} [1-\Phi(x)]^{\alpha-1} \varphi(x)\, dx + \alpha \int_{-\infty}^{+\infty} (\alpha-1) x^{k-1} [1-\Phi(x)]^{\alpha-2} \varphi^2(x)\, dx$$

$$= (k-1)\, E(X^{k-2}) + \frac{\alpha(\alpha-1)}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^{k-1} [1-\Phi(x)]^{\alpha-2} \varphi(\sqrt{2}\, x) dx$$

$$= (k-1)\, E(X^{k-2}) + \frac{\alpha(\alpha-1)}{\sqrt{2\pi}\sqrt{2^k}} \int_{-\infty}^{+\infty} u^{k-1} \left[1-\Phi(\frac{u}{\sqrt{2}})\right]^{\alpha-2} \varphi(u)\, du$$

$$= (k-1)\, E(Y^{k-2}) + \frac{\alpha(\alpha-1) C_{\alpha-2}(\frac{1}{\sqrt{2}})}{\sqrt{2\pi}\sqrt{2^k}} E_{\alpha-2}(U_{k-1}),\ k=3,4,\ldots$$

where the density function of U is expressed as:

$$f(u) = \frac{\left[1-\Phi(\frac{u}{\sqrt{2}})\right]^{\alpha-2} \varphi(u)}{C_{\alpha-2}(\frac{1}{\sqrt{2}})},\quad C_{\alpha-2}(\frac{1}{\sqrt{2}}) = \int_{-\infty}^{+\infty} \left[1-\Phi(\frac{u}{\sqrt{2}})\right]^{\alpha-2} \varphi(u)\, du$$

### 3.2 Generation of Random Numbers

According to the definition of alpha normal distribution, the random numbers are generated by inverse function [17]. First, n evenly distributed random numbers were generated from the [0, 1] interval, and denoted as ri, i=1,2,…, n. Then, let $r_i = 1-[1-\Phi(x_i)]^{\alpha}$ to find the solution. In this research, the random numbers were generated with n=1,000 and α=2. Then, the author extracted some of the numbers through the comparison between density histogram and the distribution function map, and between the density function curve and density estimation curve.

## 4. Maximum Likelihood Estimation

To obtain the general form of alpha normal distribution, an unknown parameter μ and scale parameter σ were introduced to the density function. Assuming that Y is a random variable from alpha normal distribution AN(α) and that X=μ+σY (μ∈R; σ>0), then the density function of X is expressed below.

$$f(x;\mu,\sigma,\alpha)=\frac{\alpha}{\sigma}\left[1-\Phi(\frac{x-\mu}{\sigma})\right]^{\alpha-1}\varphi(\frac{x-\mu}{\sigma}) \tag{3}$$

Suppose $X_1$, $X_2$, …, $X_n$ are n random samples of formula (3), and let θ=(μ, σ, α), then there is the following log likelihood function [18]:

$$L(\theta)=\log\prod_{i=1}^{n}f(x_i;\mu,\sigma,\alpha)$$

$$=\log\left\{\alpha^n\sigma^{-n}\left[1-\Phi(\frac{x_1-\mu}{\sigma})\right]^{\alpha-1}\cdots\left[1-\Phi(\frac{x_n-\mu}{\sigma})\right]^{\alpha-1}\varphi(\frac{x_1-\mu}{\sigma})\cdots\varphi(\frac{x_n-\mu}{\sigma})\right\}$$

$$=n\log\alpha-n\log\sigma+(\alpha-1)\sum_{i=1}^{n}\log\left[1-\Phi(\frac{x_i-\mu}{\sigma})\right]-\frac{1}{2}\sum_{i=1}^{n}(\frac{x_i-\mu}{\sigma})^2 \tag{4}$$

Then, seek the partial derivative of the upper part of μ, σ and α, respectively. It is possible to obtain the following equations.

$$\frac{\partial L}{\partial\mu}=(\alpha-1)\sum_{i=1}^{n}\frac{\varphi(\frac{x_i-\mu}{\sigma})}{1-\Phi(\frac{x_i-\mu}{\sigma})}\frac{1}{\sigma}+\sum_{i=1}^{n}\frac{x_i-\mu}{\sigma^2}=0$$

$$\frac{\partial L}{\partial\sigma}=-\frac{n}{\sigma}+(\alpha-1)\sum_{i=1}^{n}\frac{\varphi(\frac{x_i-\mu}{\sigma})}{1-\Phi(\frac{x_i-\mu}{\sigma})}\frac{x_i-\mu}{\sigma^2}+\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{\sigma^3}=0$$

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

According to the above equations, we have:

$$\hat{\alpha}=-\frac{n}{\sum_{i=1}^{n}\log\left[1-\Phi(\frac{x_i-\mu}{\sigma})\right]} \tag{5}$$

Substituting formula (5) to formula (4), and we have:

$$g(\mu,\sigma)=-\text{nlog}\left\{-\sum_{i=1}^{n}\left[1-\Phi(\frac{x_i-\mu}{\sigma})\right]\right\}-n\log\sigma-\sum_{i=1}^{n}\log\left[1-\Phi(\frac{x_i-\mu}{\sigma})\right]-\frac{1}{2}\sum_{i=1}^{n}(\frac{x_i-\mu}{\sigma})^2$$

Using the L-BFGS-B method in R software, the author obtained the estimated values of μ, σ and α.

## 5. Data Analysis

The research object is China's logistics prosperity index (LPI) from February 2014 to June 2017 (Table 1). First, the maximum likelihood and log likelihood were calculated for each parameter. Then, the model fitting was carried out with the AIC criterion (AIC=-2logL+2k) and the BIC criterion (BIC=-2logL+klogn). In the two criteria, k is the number of arguments, L is the maximum likelihood function, and n is the number of samples. The fitting results of the alpha normal function were contrasted with those of partial normal distribution, normal distribution and lognormal distribution. The density functions of the three contrast distributions are as follows:

$$f_{SN}(\text{x};\alpha,\mu,\sigma)=\frac{2}{\sigma}\Phi\left[\alpha(\frac{x-\mu}{\sigma})\right]\varphi(\frac{x-\mu}{\sigma}),\ \alpha,\ \mu\in R,\ \sigma>0$$

$$f_{N}(\text{x};\mu,\sigma)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(\text{x}-\mu)^2}{2\sigma^2}},\ \mu\in R,\ \sigma>0$$

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

Tab.1. China's LPI

| Data | Current index | Previous index | Difference | Rose% |
|---|---|---|---|---|
| 2017/6/25 | 0.558 | 0.577 | -0.019 | -3.29% |
| 2017/5/25 | 0.577 | 0.582 | -0.005 | -0.86% |
| 2017/4/25 | 0.582 | 0.554 | 0.028 | 5.05% |
| 2017/3/25 | 0.554 | 0.521 | 0.033 | 6.33% |
| 2017/2/25 | 0.521 | 0.525 | -0.004 | -0.76% |
| 2017/1/25 | 0.525 | 0.56 | -0.035 | -6.25% |
| 2016/12/25 | 0.56 | 0.593 | -0.033 | -5.56% |
| 2016/11/25 | 0.593 | 0.592 | 0.001 | 0.17% |
| 2016/10/25 | 0.592 | 0.59 | 0.002 | 0.34% |
| 2016/9/25 | 0.59 | 0.543 | 0.047 | 8.66% |
| 2016/8/25 | 0.543 | 0.548 | -0.005 | -0.91% |
| 2016/7/25 | 0.548 | 0.555 | -0.007 | -1.26% |
| 2016/6/25 | 0.555 | 0.542 | 0.013 | 2.40% |
| 2016/5/25 | 0.542 | 0.542 | 0 | 0.00% |
| 2016/4/25 | 0.542 | 0.529 | 0.013 | 2.46% |
| 2016/3/25 | 0.529 | 0.5 | 0.029 | 5.80% |
| 2016/2/25 | 0.5 | 0.533 | -0.033 | -6.19% |
| 2016/1/25 | 0.533 | 0.55 | -0.017 | -3.09% |
| 2015/12/25 | 0.55 | 0.542 | 0.008 | 1.48% |
| 2015/11/25 | 0.542 | 0.525 | 0.017 | 3.24% |
| 2015/10/25 | 0.525 | 0.522 | 0.003 | 0.57% |
| 2015/9/25 | 0.522 | 0.52 | 0.002 | 0.38% |
| 2015/8/25 | 0.52 | 0.522 | -0.002 | -0.38% |
| 2015/7/25 | 0.522 | 0.557 | -0.035 | -6.28% |
| 2015/6/25 | 0.557 | 0.58 | -0.023 | -3.97% |
| 2015/5/25 | 0.58 | 0.586 | -0.006 | -1.02% |
| 2015/4/25 | 0.586 | 0.58 | 0.006 | 1.03% |
| 2015/3/25 | 0.58 | 0.549 | 0.031 | 5.65% |
| 2015/2/25 | 0.549 | 0.563 | -0.014 | -2.49% |
| 2015/1/25 | 0.563 | 0.575 | -0.012 | -2.09% |
| 2014/12/25 | 0.575 | 0.565 | 0.01 | 1.77% |
| 2014/11/25 | 0.565 | 0.549 | 0.016 | 2.91% |
| 2014/10/25 | 0.549 | 0.564 | -0.015 | -2.66% |
| 2014/9/25 | 0.564 | 0.541 | 0.023 | 4.25% |
| 2014/8/25 | 0.541 | 0.568 | -0.027 | -4.75% |
| 2014/7/25 | 0.568 | 0.567 | 0.001 | 0.18% |
| 2014/6/25 | 0.567 | 0.552 | 0.015 | 2.72% |
| 2014/5/25 | 0.552 | 0.577 | -0.025 | -4.33% |
| 2014/4/25 | 0.577 | 0.53 | 0.047 | 8.87% |
| 2014/3/25 | 0.53 | 0.519 | 0.011 | 2.12% |
| 2014/2/25 | 0.519 | 0.515 | 0.004 | 0.78% |

The calculation results are listed in Table 2.

Tab.2. Fitting Results of Alpha Normal Function, Partial Normal Distribution, Normal Distribution and Lognormal Distribution

| Model | Log likelihood | AIC | BIC |
|---|---|---|---|
| Normal distribution | -496.271 | 997.068 | 1005.738 |
| Partial normal distribution | -477.173 | 956.297 | 969.755 |
| Lognormal distribution | -469.052 | 942.571 | 949.317 |
| Alpha normal distribution | -461.278 | 931.192 | 946.025 |

As shown in Table 1, the alpha normal distribution has the smallest AIC and BIC values on the dataset of China's LPI, and the highest value of likelihood. The lognormal distribution and partial normal distribution share similar results. The normal distribution, which is symmetrical in nature, fails to capture the skewed and heavy-tailed properties of the data. The results show that the alpha normal distribution can depict the dataset features in an accurate manner.

Furthermore, the alpha normal distribution boasts the highest goodness of fit, whether in skewness or kurtosis, and stands out as the best distribution of the LPI. The second best distribution is the popular lognormal distribution, which fits the skewness well. The partial normal distribution performs poorer than the lognormal distribution on capturing the right tail but outperforms the latter on fitting the thick left tail. Therefore, the partial normal distribution and the alpha normal distribution are more suitable for LPI analysis. The two distributions provide better models for asymmetric heavy-tailed data in real life.

Next, the author tested if there is any significant difference between normal distribution and alpha normal distribution in a given dataset. Two assumptions were made as follows.

H0: If α=1, the sample comes from a normal distribution.

H1: If α≠1, the sample comes from an alpha normal distribution.

The likelihood ratio was introduced to the test: $-2\log(\mathrm{LRT}) = -2\log \frac{l_N(\hat{\mu}, \hat{\sigma})}{l_{AN}(\hat{\alpha}, \hat{\mu}\ \hat{\sigma})} = 67.518$. At the significance level of 0.05, $x_1^2 = 3.86$, which rejects the original hypothesis. Hence, the alpha normal distribution has a higher goodness of fit than the normal distribution. The conclusion echoes the fitted density curves of these distributions. Thus, the alpha normal distribution is an important choice considering that no normal distribution applies to the analysis of asymmetric datasets.

## Conclusion

This paper introduces the alpha normal distribution, a novel distribution pattern with normal distribution as a special case. Specifically, the basic properties were analysed, the random numbers were generated, and the parameters were predicted by maximum likelihood estimation. Based on the density function curve, it is found that the alpha normal distribution is left skewed and right skewed. Then, this new asymmetric distribution was compared to the normal distribution and the popular lognormal distribution. The results show that the alpha normal distribution is more flexible in handling skewed or heavy-tailed data, and is thus suitable for fitting asymmetric datasets. Then, the dataset of China's LPI was introduced for fitting, and statistical judgement criteria were adopted to compare the alpha normal distribution with partial normal distribution, normal distribution and lognormal distribution. It is concluded that the proposed distribution outshines the contrast distributions in data fitting. All in all, this research provides a new statistical for analysing asymmetric datasets in real life.

## Acknowledgements

We acknowledge the financial support from the "Young Academic Innovation Team of Northwest University of Political Science and Law", the Special research program of Shaanxi Provincial Department of Education of the "Operational Mechanism and Implementation Path of E-commerce in the Precision Poverty Reduction Strategy" (17JK0795).

## References

1. S.J. Fan, Study on concentration distribution of urban air pollutants, 1993, Studies in Logic, no. 1, pp. 163-167.
2. W.R. Ott, A physical explanation of the lognormality of pollutant concentrations, 1990, Journal of the Air & Waste Management Association, vol. 40, no. 10, pp. 1378.
3. L.G. Blackwood, The lognormal distribution, environmental data, and radiological monitoring, 1992, Environmental Monitoring & Assessment, vol. 21, no. 3, pp. 193.
4. J. Han, Z.G. Dai, W.T. Li, Analysis of PM2.5 concentration and meteorological conditions in haze weather of Xi'an city, 2014, Environmental Pollution & Control, vol. 36, no. 2, pp. 52-56.
5. J. Chen, Statistical characteristics of air pollutant concentration and its meteorological impact in Yulin, 2015, Chang'an University Press, Xi'an.
6. H.D. Kan, B.H. Chen, Statistical distributions of ambient air pollutants in Shanghai, China, 2004, Biomedical and Environmental Sciences, vol. 17, no. 3, pp. 366-372.
7. G.L. Xu, Y.J. Lv, Approach of multi-attribute decision-making based on normal distribution internal number, 2011, Systems Engineering, vol. 29, no. 9, pp. 120-123.

8. Z.F. Zhao, H. Yan, L. Yao, Application of normal distribution to estimate the quality of regional geochemical survey samples, 2013, Rock and Mineral Analysis, vol. 32, no. 1, pp. 96-100.
9. A. Azzalini, A class of distributions which includes the normal ones, 1985, Scandinavian Journal of Statistics, vol. 12, no. 2, pp. 171-178.
10. H. Norbert, A probabilistic representation of the “skew-normal” distribution, 1986, Scandinavian Journal of Statistics, vol. 13, no. 4, pp. 271-275.
11. A. Azzalini, A. Capitanio, Statistical applications of the multivariate skew normal distribution, 1999, Journal of the Royal Statistical Society, vol. 61, no. 3, pp. 579-602.
12. W.J. Huang, Y.H. Chen, Generalized skew-Cauchy distribution, 2007, Statistics & Probability Letters, vol. 77, no. 11, pp. 1137-1147.
13. R.D. Gupta, R.C. Gupta, Analyzing skewed data by power normal model, 2008, Test, vol. 17, no. 1, pp. 197-210.
14. M.M. Chen, J.H. Ma, N.N. Ji, The alpha normal distribution and its application to environment pollution, 2016, Statistics & Information Forum, vol. 31, no. 6, pp. 22-27.
15. B.C. Arnold, R.J. Beaver, A. Azzalini, Skewed multivariate models related to hidden truncation and/or selective reporting, 2002, Test, vol. 11, no. 1, pp. 7-54.
16. R.C. Gupta, R.D. Gupta, Generalized skew normal model, 2004, Test An Official Journal of the Spanish Society of Statistics & Operations Research, vol. 13, no. 2, pp. 501-524.
17. Y. Xue, L.P. Chen, R statistical modeling and R software, 2007, Tsinghua University Press， Peking, pp. 485-490.
18. S.S. Mao, Y.M. Cheng, X.L. Pu, Probability and Mathematical Statistics Tutorial, 2004, Higher Education Press, Peking, pp. 199-202.