ICIC Express Letters
Volume **14**, Number **2**, February **2020**

ICIC International ©2020 ISSN 1881-803X
pp. 197–202

ANALYSIS OF THE ESTIMATION OF PROCESS CAPABILITY INDEX OF NON-NORMAL PROCESS DATA USING THE LOGNORMAL DISTRIBUTIONS

Chin-Chuan Wu and Shu-Fei Wu*

Department of Statistics
Tamkang University
No. 151, Yingzhuan Road, Tamsui District, New Taipei City 25137, Taiwan
*Corresponding author: 100665@mail.tku.edu.tw

Received August 2019; accepted November 2019

Abstract. *In recent years, process capability indices (PCIs) have been widely applied in quality control by most practitioners to assess whether the production process reaches a required level. However, the characteristic variable in many industrial production processes has non-normal distribution. This paper uses the Clements's method to estimate four non-normal process capability indices for lognormal distribution. A simulation study is done to analyze the influence of skewness and kurtosis on the precision of estimation of process capability indices for the given distribution.*

Keywords: Process capability index, Clements's method, Lognormal distribution

1. **Introduction.** In recent years, process capability analysis has been widely applied in the field of quality control to assess whether the production process is capable or not to reach the required specification limits. The most extensively used PCIs in industrial manufacturing are defined as follows:

$$C_p = \frac{USL - LSL}{6\sigma}, \tag{1}$$

$$C_{pu} = \frac{USL - \mu}{3\sigma}, \tag{2}$$

$$C_{pl} = \frac{\mu - LSL}{3\sigma}, \tag{3}$$

$$C_{pk} = \min\left(C_{pu}, C_{pl}\right), \tag{4}$$

where USL is the upper specification limit, LSL is the lower specification limit, μ is the process mean, and σ is the process standard deviation. The process capability indices C_p and C_{pk} are used in cases of bilateral specifications (target-the-better type); C_{pu} and C_{pl} are used in cases of unilateral specifications (larger-the-better type and smaller-the-better type quality characteristics). These four PCIs are generally defined based on three basic assumptions.

1) The collected data is under control.
2) The data are independent and identically distributed.
3) The collected process data is normally distributed.

However, some quality characteristics are not normally distributed. For the non-normal data, one way is to transform the data into normally distributed data using some mathematical functions. Johnson [5] utilized the moment method to build a system of distributions. Box and Cox [1] proposed a well-known power transformation. Somerville and Montgomery [9] used a square-root transformation to solve this problem. All these three methods are based on the technique of data transformations. Another simple way to deal

DOI: 10.24507/icicel.14.02.197

with non-normal data is to use the percentiles instead of means and variances to modify classical PCIs C_p, C_{pu}, C_{pl} and C_{pk} for non-normal data. The percentile based PCIs are defined as follows:

$$C_{p(q)} = \frac{USL - LSL}{x_{99.865} - x_{0.135}}, \tag{5}$$

$$C_{pu(q)} = \frac{USL - x_{50}}{x_{99.865} - x_{50}}, \tag{6}$$

$$C_{pl(q)} = \frac{x_{50} - LSL}{x_{50} - x_{0.135}}, \tag{7}$$

$$C_{pk(q)} = \min\left(C_{pu(q)}, C_{pl(q)}\right), \tag{8}$$

where $x_p = p * 100$th percentile value of non-normal data.

Clements [3] proposed the method of using non-normal percentiles to calculate the estimator of C_p and C_{pk} indices for non-normal Pearsonian distributions. Clements's method is to calculate the sample mean $\bar{X}$, the standard deviation S, the sample kurtosis and the sample skewness of the given data in advance. From Gruska et al. [4] and Kotz and Lovelace [7], we can use the sample kurtosis and skewness to find the corresponding standardized tails $Z_{0.135}$, $Z_{0.5}$ and $Z_{0.99865}$ of Pearson curves, where the skewness is ranging from -2 to 2 and kurtosis is ranging from -1.4 to 12.2. Once the standardized percentiles are obtained, the percentile value of non-normal data can be estimated as

$$\hat{x}_{0.135} = \bar{X} + Z_{0.135}S,\ \hat{x}_{50} = \bar{X} + Z_{0.5}S \text{ and } \hat{x}_{99.865} = \bar{X} + Z_{0.99865}S.$$

Replace the percentiles $x_{0.135}$, x_{50} and $x_{99.865}$ in the percentile based PCIs $C_{p(q)}$, $C_{pu(q)}$, $C_{pl(q)}$ and $C_{pk(q)}$ by the estimated percentile values $\hat{x}_{0.135}$, $\hat{x}_{50}$ and $\hat{x}_{99.865}$ of non-normal data, then we can obtain the related estimator $\hat{C}_{p(q)}$, $\hat{C}_{pu(q)}$, $\hat{C}_{pl(q)}$ and $\hat{C}_{pk(q)}$.

In this study, we consider the lognormal distribution data and use Clements's method to estimate the percentile based PCIs. One can see Singh et al. [8] for the Bayesian estimation and prediction for lognormal distribution. The structure of this research is organized as follows. In Section 2, the derivation of the first four central moments, skewness and kurtosis is developed. The steps to find the estimator of percentile based PCIs based on Clements's method are given in this section. In Section 3, a simulation study is done to analyze the influence of skewness and kurtosis on the precision of estimation of process capability indices for lognormal distribution and the analysis is given. At last, the conclusion is discussed in Section 4.

2. **The Lognormal Distribution.** Let random variable X have a two-parameter lognormal distribution with probability density function given by

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

From Johnson et al. [6], we can obtain the rth moment of X as $\mu'_r = E[X^r] = \exp\left(r\zeta + \frac{1}{2}r^2\sigma^2\right)$. From Wartmann [10], the rth central moment can be obtained as

$$\begin{aligned}
\mu_r &= E\left[(X - \mu'_1)^r\right] \\
&= \sum_{j=0}^{r}(-1)^j \binom{r}{j} \mu'_{r-j}\mu'^{j}_1 \\
&= \sum_{j=0}^{r}(-1)^j \binom{r}{j} \exp\left\{(r-j)\zeta + \frac{1}{2}(r-j)^2\sigma^2 + j\zeta + \frac{1}{2}j\sigma^2\right\} \\
&= \sum_{j=0}^{r}(-1)^j \binom{r}{j} \exp\left[r\zeta + \frac{1}{2}\left\{(r-j)^2 + j\right\}\sigma^2\right]
\end{aligned}$$

Let $\omega = \exp(\sigma^2)$, then we have

$$\begin{aligned} \mu_r &= e^{r\zeta} \sum_{j=0}^{r} (-1)^j \begin{pmatrix} r \\ j \end{pmatrix} \omega^{\{(r-j)^2+j\}/2} \\ &= \omega^{r/2} \left\{ \sum_{j=0}^{r} (-1)^j \begin{pmatrix} r \\ j \end{pmatrix} \omega^{(r-j)(r-j-1)/2} \right\} e^{r\zeta} \end{aligned} \tag{9}$$

Taking $r = 3$ and 4 in μ_r, we can obtain $\mu_3 = \omega^{3/2}(\omega - 1)^2(\omega + 2)e^{3\zeta}$ and $\mu_4 = \omega^2(\omega - 1)^2(\omega^4 + 2\omega^3 + 3\omega^2 - 3)e^{4\zeta}$. From Burr [2], we can obtain the skewness

$$\alpha_3 = \sqrt{\beta_1} = \frac{\mu_3}{\sigma^3} = \frac{\mu_3}{(\mu_2)^{3/2}} = \frac{\omega^{3/2}(\omega - 1)^2(\omega + 2)e^{3\zeta}}{\{e^{2\zeta}\omega(\omega - 1)\}^{3/2}} = (\omega - 1)^{1/2}(\omega + 2)$$

and kurtosis as

$$\alpha_4 = \beta_2 = \frac{\mu_4}{\sigma^4} = \frac{\mu_4}{(\mu_2)^2} = \frac{\omega^2(\omega - 1)^2(\omega^4 + 2\omega^3 + 3\omega^2 - 3)e^{4\zeta}}{\{e^{2\zeta}\omega(\omega - 1)\}^2} = \omega^4 + 2\omega^3 + 3\omega^2 - 3,$$

where $\alpha_3 > 0$ and $\alpha_4 > 3$.

Observe that both of skewness and kurtosis are independent of the location parameter ζ and only dependent on the shape parameter σ.

We will use the kurtosis and skewness to find the corresponding standardized tails $Z_{0.135}$, $Z_{0.5}$ and $Z_{0.99865}$ of Pearson curves and then use sample mean and sample standard deviation to find the estimated percentiles $\hat{x}_{0.135} = \bar{X} + Z_{0.135}S$, $\hat{x}_{50} = \bar{X} + Z_{0.5}S$ and $\hat{x}_{99.865} = \bar{X} + Z_{0.99865}S$ for lognormal distribution. Replace the percentiles in the PCIs $C_{p(q)}$, $C_{pu(q)}$, $C_{pl(q)}$ and $C_{pk(q)}$ by the estimated percentile values $\hat{x}_{0.135}$, $\hat{x}_{50}$ and $\hat{x}_{99.865}$ of non-normal data, then we can obtain the corresponding estimators $\hat{C}_{p(q)}$, $\hat{C}_{pu(q)}$, $\hat{C}_{pl(q)}$ and $\hat{C}_{pk(q)}$.

3. **Simulation Study.**

The procedure of the process capability analysis using the lognormal is presented as follows.

Step 1. Consider a specific distribution for example Lognormal$(0, 0.1)$.

Step 2. Set up a target process capability index values (PCIs) for $C_{p(q)}$, $C_{pu(q)}$, $C_{pl(q)}$, $C_{pk(q)}$ as 1, 1.5 and 2 with nonconforming rates of 1350 ppm (parts per million), 3.4 ppm and 0.001 ppm.

Step 3. Find the related upper specification limit (USL) and lower specification limit (LSL) as follows: $USL = C_{pu(q)}(x_{99.865} - x_{50}) + x_{50}$ and $LSL = x_{50} - C_{pl(q)}(x_{50} - x_{0.135})$.

Step 4. Generate random sample of sizes $n = 50$, 100, 500 from the specific distribution chosen in Step 1.

Step 5. Calculate the sample mean $\left(\bar{X}\right)$, standard deviation (S), skewness $(\hat{\alpha}_3)$, and kurtosis $(\hat{\alpha}_4)$ from the generated data.

Step 6. Use the sample kurtosis and skewness to find the corresponding standardized tails $Z_{0.135}$, $Z_{0.5}$ and $Z_{0.99865}$ of Pearson curves and the percentile value of non-normal data can be estimated as

$$\hat{x}_{0.135} = \bar{X} + Z_{0.135}S,\ \hat{x}_{50} = \bar{X} + Z_{0.5}S \text{ and } \hat{x}_{99.865} = \bar{X} + Z_{0.99865}S.$$

Step 7. Replace $x_{0.135}$, x_{50} and $x_{99.865}$ in (5)-(8) by $\hat{x}_{0.135}$, $\hat{x}_{50}$ and $\hat{x}_{99.865}$, then we can obtain the estimators $\hat{C}_{p(q)}$, $\hat{C}_{pu(q)}$, $\hat{C}_{pl(q)}$ and $\hat{C}_{pk(q)}$.

Step 8. Repreat Steps 4-7 for 30 times and yield the mean and standard error of $\hat{C}_{p(q)}$, $\hat{C}_{pu(q)}$, $\hat{C}_{pl(q)}$ and $\hat{C}_{pk(q)}$.

For lognormal distribution, we consider the following five cases with increasing skewness and kurtosis for fixed $\xi = 0$.

Lognormal distribution			
ξ	σ	α_3	α_4
0	0.1	0.30	3.16
0	0.2	0.61	3.68
0	0.3	0.95	4.64
0	0.4	1.32	6.26
0	0.5	1.75	8.90

Table 1. The mean and SE of 30 estimated process capabilities for log-normal distribution

Lognormal$(0, 0.1)$ with skewness = 0.30 and kurtosis = 3.16						
$C_{p(q)}$	1	Clements	1.5	Clements	2	Clements
mean	$n = 50$	1.298505	$n = 50$	1.947758	$n = 50$	2.597011
SE		0.266165		0.399248		0.532331
mean	$n = 100$	1.083829	$n = 100$	1.625744	$n = 100$	2.167659
SE		0.148619		0.222928		0.297238
mean	$n = 500$	1.063759	$n = 500$	1.595638	$n = 500$	2.127517
SE		0.095725		0.143588		0.191450
$C_{pu(q)}$	1	Clements	1.5	Clements	2	Clements
mean	$n = 50$	1.330011	$n = 50$	2.001350	$n = 50$	2.672690
SE		0.359901		0.548536		0.737702
mean	$n = 100$	1.131592	$n = 100$	1.698306	$n = 100$	2.265020
SE		0.226299		0.337342		0.448739
mean	$n = 500$	1.068982	$n = 500$	1.603543	$n = 500$	2.138103
SE		0.120981		0.183206		0.245728
$C_{pl(q)}$	1	Clements	1.5	Clements	2	Clements
mean	$n = 50$	1.322537	$n = 50$	1.983555	$n = 50$	2.644574
SE		0.276673		0.429860		0.584164
mean	$n = 100$	1.072723	$n = 100$	1.609632	$n = 100$	2.146541
SE		0.228885		0.348744		0.469202
mean	$n = 500$	1.100470	$n = 500$	1.653267	$n = 500$	2.206063
SE		0.221691		0.341943		0.462372
$C_{pk(q)}$	1	Clements	1.5	Clements	2	Clements
mean	$n = 50$	1.153978	$n = 50$	1.713188	$n = 50$	2.272399
SE		0.228893		0.337570		0.446759
mean	$n = 100$	0.952399	$n = 100$	1.426042	$n = 100$	1.898931
SE		0.108580		0.165198		0.223398
mean	$n = 500$	0.950331	$n = 500$	1.418651	$n = 500$	1.886971
SE		0.088941		0.133070		0.177440

Due to limited space, other tables for skewness = 0.61, 0.95, 1.32, 1.75 are available in authors' site.

From Table 1, we have the following findings.

1) When sample size increases, the estimated PCI is getting closer to the nominal one.
2) When sample size increases, the precision is getting better since the SE is getting small under fixed preassigned PCI values.
3) When the target value of PCI increases, SE is getting larger so that we can conclude that the performance of estimated PCI is getting worse.

For large sample $n = 500$, the box plots of 30 estimated PCIs are presented in Figure 1.

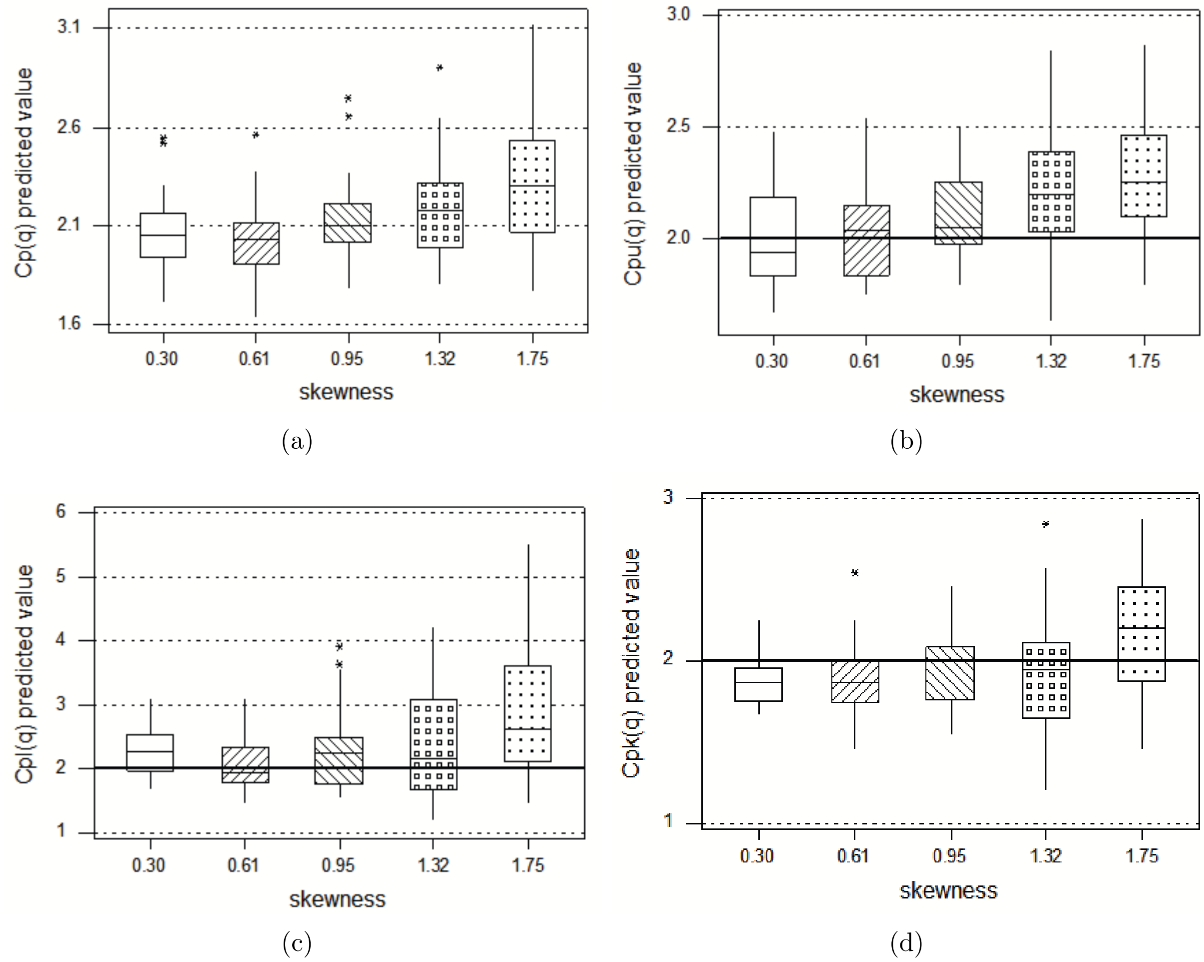

FIGURE 1. (a) Box plot for 30 $\hat{C}_{p(q)}$ under $C_{p(q)} = 2$, (b) box plot for 30 $\hat{C}_{pu(q)}$ under $C_{pu(q)} = 2$, (c) box plot for 30 $\hat{C}_{pl(q)}$ under $C_{pl(q)} = 2$, (d) box plot for 30 $\hat{C}_{pk(q)}$ under $C_{pk(q)} = 2$

From Figure 1, we have the following findings.

1) The accuracy is generally getting worse since the median is getting more deviated from the target value 2 when the skewness or kurtosis increases.
2) The precision is generally getting worse since the dispersion is getting larger when the skewness or kurtosis increases.
3) The estimation of $C_{pl(q)}$ performs the worse compared with the other three PCIs.

4. **Conclusions.** This paper utilizes the Clements's method to estimate four non-normal process capability indices for lognormal distribution. A simulation study is done to analyze the influence of skewness and kurtosis on the precision of estimation of process capability indices for the given two distributions. Generally speaking, the Clements's method employed to estimate four PCIs is effective for lognormal distribution. In the future, we can conduct the analysis on other distributions, for example, inverse Gaussian distribution.

REFERENCES

[1] G. E. P. Box and D. R. Cox, An analysis of transformation, *Journal of the Royal Statistical Society, Series B (Methodological)*, vol.26, no.2, pp.211-252, 1964.

[2] I. W. Burr, Cumulative frequency distribution, *The Annals of Mathematical Statistics*, vol.13, pp.215-232, 1942.

[3] J. A. Clements, Process capability calculations indices for non-normal distributions, *Quality Progress*, vol.22, pp.95-100, 1989.

[4] G. F. Gruska, K. Mirkhani and L. R. Lmberson, *Non-Normal Data Analysis*, Applied Computer Solutions, Inc., St. CLair Shores, Michigan, 1989.

[5] N. L. Johnson, System of frequency curves generated by methods of translation, *Biometrika*, vol.36, pp.149-176, 1949.

[6] N. L. Johnson, S. Kotz and N. Balakrishnan, *Continuous Univariate Distributions*, 2nd Edition, John Wiley, New York, 1994.

[7] S. Kotz and C. R. Lovelace, *Process Capability Indices in Theory and Practice*, Arnold, London, 1998.

[8] S. Singh, Y. M. Tripathi and S. J. Wu, Bayesian estimation and prediction based on lognormal record values, *Journal of Applied Statistics*, vol.44, no.5, pp.916-940, 2017.

[9] S. Somerville and D. Montgomery, Process capability indices and non-normal distributions, *Qual. Eng.*, vol.19, no.2, pp.305-316, 1996.

[10] R. Wartmann, Application of the lognormal distribution, *Newsletter for Mathematical Statistic*, vol.8, pp.83-91, 1956.