# ESTIMATION OF PROCESS CAPABILITY INDEX FOR CONCENTRICITY

**Ashok Sarkar and Surajit Pal**
SQC & OR Unit
Indian Statistical Institute
110 Nelson Manickam Road
Aminjikarai, Madras 600 029, India

## Key Words

CPU; Concentricity; Extreme value distribution; Critical value.

## Introduction

Consider a hollow circular job where the centers of both inner and outer circles should be at a preassigned point. Due to the influence of several process parameters, the centers of both circles deviate from their ideal positions. Concentricity is the measure of the difference between the two centers. This is presented in Figure 1.

In the case of hollow right cylindrical jobs, where the axes of the inner circle and outer circles should be the same, concentricity is the measure of distance between the centers of two circles on a particular cross section. Concentricity plays an important role in the proper functioning of many components/jobs. Because a high value of concentricity results in malfunctioning of the component/job, it is necessary to control the concentricity within its specified maximum value.

The general practice of an organization is to control and improve the process by estimating the process capability index (PCI). Like other parameters of the component, it is essential that the manufacturing process be evaluated and controlled with respect to concentricity. The estimation of the PCI assumes that the underlying distribution is normal. But concentricity generally does not follow the normal distribution pattern. Hence, the problem is to estimate the process capability with respect to concentricity of the job.

In this article, the behavior of concentricity is explained through a distribution. The live data are then fitted to the distribution and, based on the fitted distribution, the PCI is estimated. Various methods of estimating the PCI when the underlying distribution is non-normal are also discussed. A simple method for the estimation of the PCI is suggested also.

The process capability indices are compared with the minimum recommended value to determine whether the process can be considered as capable. As the PCI depends on sample size, it should be compared with a critical value so that the process can be considered capable with a high probability. The critical values of PCI are calculated for various sample sizes.

## Distribution of Concentricity

Let us consider a cross section of a hollow circular component whose outer surface and inner surface are both

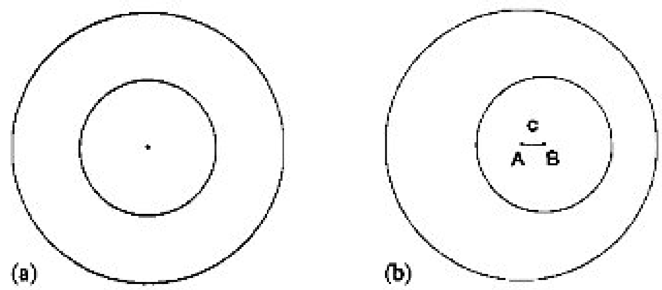


*Figure 1.* (a) Jobs without concentricity; (b) jobs having concentricity.

circular. Let A and B be the centers of the outer and inner circles, respectively, as given in Figure 2. The distance AB is a measure of concentricity.

If we draw a straight line through AB, then we can write

$$AD = AB + BC + CD$$

where AD is the radius of the outer circle ($R$), BC is the radius of the inner circle ($r$), AB is the measure for concentricity ($c$), and CD is the wall thickness ($y$). Therefore,

$$R = c + r + y \tag{1}$$

or

$$c = (R - r) - y. \tag{2}$$

Because $R - r$ is constant for a particular job, concentricity depends on the wall thickness $y$. The wall thickness $y$ is minimum in Eq. (2). If we assume that the wall thickness follows a normal distribution, then $c$ is an extreme value and, hence, the distribution of concentricity is an extreme value distribution.

### *Proof: In Eq. (2), the Wall Thickness Is Minimum*

Let us consider another wall thickness $y_1$ at any point. By joining the inner wall point with inner center, and outer wall point with outer center, we can form a triangle ABD, as shown in Figure 3, where AD is the outer circle radius ($R$), AB is the concentricity ($c$), BD is the sum of the wall thickness and the inner circle radius ($y_1 + r$). From the properties of a triangle we know

$$AD \le AB + BD \Rightarrow R \le c + y_1 + r.$$

Substituting the value of $R$ from the above equation, we get

$$c + r + y \le c + y_1 + r$$

or

$$y \le y_1.$$

Thus, $y$ is the minimum wall thickness. ■

Therefore, the distribution of concentricity can be explained by an extreme value distribution under appropriate assumptions. The type of distribution which can be fitted with the empirical data is

Type 1:

$$F(x) = \begin{cases} 0, & x < 0 \\ \exp\left[-\exp\left(-\frac{x-\alpha}{\theta}\right)\right], & 0 \le x \le \infty; \end{cases}$$

Type 2:

$$F(x) = \begin{cases} 0, & x < \alpha \\ \exp\left[-\left(-\frac{x-\alpha}{\theta}\right)^{-k}\right], & x \ge \alpha, \end{cases}$$

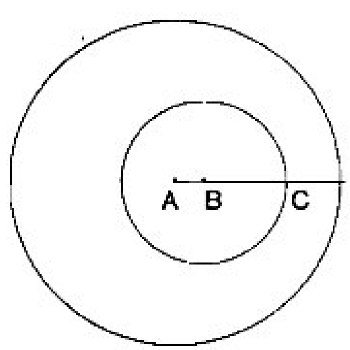


*Figure 2.* Jobs with concentricity.

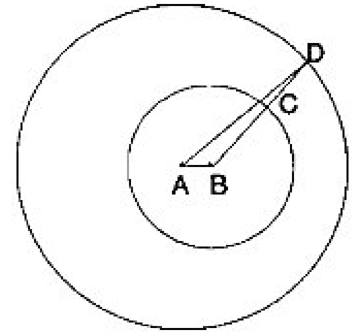


*Figure 3.* Comparison of wall thicknesses.

where $\alpha$ (>0), $\theta$ (>0), and $k$ (>0) are distribution parameters.

The Type 2 distribution can be transformed to Type 1 distribution by the simple transformation

$$Z = \log(X - \alpha).$$

Types 1 and 2 distributions are closely related to the Weibull distribution. In fact, if $X$ has a Type 1 extreme value distribution, then $e^X$ has a Weibull distribution.

## Fitting the Distribution

### *Estimation of Distribution Parameters*

The appropriate types of distribution to be fitted are decided considering the skewness of the empirical distribution based on data. The distribution parameters, $\alpha$ and $\theta$, of the Type 1 extreme value distribution can be estimated from the sample (1) by calculating

$$\hat{\theta} = 0.7797s \quad \text{and} \quad \hat{\alpha} = \bar{x} - 0.57722\hat{\theta}, \tag{3}$$

where $\bar{x}$ and $s$ are the sample average and the sample standard deviation, respectively. The distribution parameters of other types of extreme value distribution are estimated by equating the theoretical distribution function with the empirical distribution function at convenient points (2). A $\chi^2$ test of significance is to be done to test the fit of the extreme value distribution.

Alternatively, the empirical data after transformation can be plotted on the Weibull probability plotting paper to estimate the parameters of the Weibull distribution. From those estimated parameters of Weibull distribution, the parameters of the extreme value distribution can be estimated easily by using transformation techniques (3). The estimation of parameters through probability plotting may be tedious. However, this method can be implemented easily in electronic spreadsheets.

### *Case Example*

An organization is engaged in manufacturing of valve guides which are used in a diesel engine. The valve guide is a right hollow cylindrical bar made of cast iron. Besides other quality characteristics, concentricity plays a critical role in proper functioning of the valve.

The measurement of concentricity is obtained by rotating the job around a preassigned axis (ID axis) and then observing the deflection on a dial gauge. The maximum value of deflection is the measurement of concentricity.

Three sets of data on the concentricity of the valve guide from the process are collected and presented below. Extreme value distribution curves are fitted to the data and the fit is tested through a $\chi^2$ test for goodness of fit. The maximum value specified for concentricity is 0.40 unit.

*Data Set 1 (Table 1)*

The parameters of the extreme value (Type 1) distribution fitted are $\alpha = 0.154$ and $\theta = 0.0574$. The degrees of freedom in the $\chi^2$ test is $c - k - 1$, where $c$ and $k$ are the number of classes and number of parameters estimated, respectively. In this data set, $c = 7$ and $k = 2$. So the degrees of freedom in this test is $7 - 2 - 1 = 4$. The calculated $\chi^2 = 4.38$ is insignificant compared to the tabulated value of $\chi^2_{0.05,4} = 9.49$; hence, the fit is satisfactory.

*Data Set 2 (Table 2)*

The parameters of the extreme value (Type 1) distribution fitted are $\alpha = 0.1022$ and $\theta = 0.0208$. The calculated

***Table 1.*** $\chi^2$ Test for Goodness of Fit for Data Set 1

| CLASS INTERVAL | OBSERVED FREQUENCY ($O$) | EXPECTED FREQUENCY ($E$) | $\chi^2$ $[(O - E)^2/E]$ |
|---|---|---|---|
| Below 0.10 | 12 | 11.20 | 0.06 |
| 0.10–0.14 | 39 | 29.26 | 3.24 |
| 0.14–0.18 | 33 | 36.29 | 0.30 |
| 0.18–0.22 | 26 | 28.85 | 0.28 |
| 0.22–0.26 | 16 | 18.20 | 0.27 |
| 0.26–0.30 | 9 | 10.21 | 0.14 |
| Above 0.30 | 10 | 10.98 | 0.09 |
| Total | 145 | 145 | 4.38 |

*Table 2.* $\chi^2$ Test for Goodness of Fit for Data Set 2

| CLASS INTERVAL | OBSERVED FREQUENCY ($O$) | EXPECTED FREQUENCY ($E$) | $\chi^2$ $[(O - E)^2/E]$ |
|---|---|---|---|
| Below 0.08 | 18 | 12.94 | 1.98 |
| 0.08–0.10 | 76 | 64.91 | 1.90 |
| 0.10–0.12 | 77 | 76.90 | 0.00 |
| 0.12–0.14 | 38 | 46.56 | 1.58 |
| 0.14–0.16 | 18 | 21.33 | 0.52 |
| 0.16–0.18 | 5 | 8.75 | 1.61 |
| Above 0.18 | 5 | 5.60 | 0.06 |
| Total | 237 | 237 | 7.65 |

$\chi^2 = 7.65$ is insignificant compared to the tabulated value of $\chi^2_{0.05,4} = 9.49$; hence, the fit is satisfactory.

*Data Set 3 (Table 3)*

The parameters of the extreme value (Type I) distribution fitted are $\alpha = 0.110$ and $\theta = 0.0535$. The calculated $\chi^2 = 13.50$ is insignificant compared to the tabulated value of $\chi^2_{0.05,8} = 15.51$; hence, the fit is satisfactory.

The summary statistics of the three data sets are presented in Table 4.

## Methods of Estimation of Capability Indices

From the above, it is clear that concentricity follows a non-normal distribution. Hence, the most commonly used methods of estimation for PCIs cannot be used, which assumes that the quality characteristic follows a normal distribution. Methods available for estimating the PCI where the quality characteristic follows a non-normal distribution are briefly discussed along with the demerits. Because only the upper limit is specified for concentricity, the process capability index CPU is considered.

### *Method 1*

Some of the statistical process control programs have calculated the process capability index CPU for non-normal distributions as

$$\text{CPU} = \frac{\text{USL} - \overline{X}}{4s}$$

where $\overline{X}$ is the sample mean and $s$ is the sample standard deviation. In the denominator of the calculation of the CPU, 4s has been chosen, instead of $3s$, because of positive skewness. The process is considered to be capable if $\text{CPU} \geq 1$.

*Table 3.* $\chi^2$ Test for Goodness of Fit for Data Set 3

| CLASS INTERVAL | OBSERVED FREQUENCY ($O$) | EXPECTED FREQUENCY ($E$) | $\chi^2$ $[(O - E)^2/E]$ |
|---|---|---|---|
| Below 0.06 | 28 | 23.67 | 0.79 |
| 0.06–0.08 | 37 | 28.90 | 2.27 |
| 0.08–0.10 | 49 | 38.47 | 2.88 |
| 0.10–0.12 | 48 | 41.81 | 0.92 |
| 0.12–0.14 | 34 | 39.47 | 0.76 |
| 0.14–0.16 | 26 | 33.78 | 1.79 |
| 0.16–0.18 | 18 | 27.03 | 3.01 |
| 0.18–0.20 | 17 | 20.63 | 0.64 |
| 0.20–0.22 | 13 | 15.25 | 0.33 |
| 0.22–0.24 | 10 | 11.03 | 0.10 |
| Above 0.24 | 26 | 25.97 | 0.00 |
| Total | 306 | 306 | 13.50 |

*Table 4.* Summary Statistics of the Data Sets

| DATA SET | SAMPLE SIZE | AVERAGE | S.D. | SKEWNESS | KURTOSIS | % NONCONFORMANCE IN SAMPLE |
|---|---|---|---|---|---|---|
| 1 | 145 | 0.1872 | 0.0737 | 0.951 | 1.051 | 1.38 |
| 2 | 237 | 0.1142 | 0.0268 | 0.865 | 1.636 | 0.00 |
| 3 | 306 | 0.1412 | 0.0688 | 1.175 | 1.146 | 0.33 |

### *Method 2*

A non-normal curve is fitted to the data, and based on that, the proportion outside the tolerance limit is estimated. That proportion nonconforming is then converted to an equivalent Z-score for a normal distribution. Then the Z-score is divided by the appropriate factor to get the value of process capability index CPU. The process is considered to be capable if CPU ≥ 1 (see Ref. 4).

### *Method 3*

Clements (5) has proposed this method of calculating the CPU for any shape of distribution using the Pearson family of curves. From the sample data, the mean ($\bar{X}$), standard deviation ($s$), skewness (Sk), and kurtosis (Ku) are calculated. Then based on skewness and kurtosis values, standardized 99.865 percentile and standardized median values are read from the table to calculate the estimated 99.865 percentile ($U_p$) and estimated median ($M$). The process capability index CPU is then calculated as

$$\text{CPU} = \frac{\text{USL} - M}{U_p - M}.$$

The process is considered to be capable if CPU ≥ 1.

## Demerits of the Estimation Methods

Although Method 1 is used extensively by many industries because of its simplicity, there is no known statistical basis for the denominator to be chosen as four times this sample standard deviation.

Method 2 is too laborious to do manually. Fitting a non-normal curve to the empirical data may not be an easy task. Again, for a highly capable process, the PCI estimated using this method does not reflect the true capability (refer to the data set 2 of Table 5). This is primarily because of the properties of normal distributions.

Method 3 is also laborious to do manually. One may need a computer to use this method and have to refer to tables for getting the values of standardized median and standardized 99.865 percentile. Furthermore, concepts and calculations of skewness and kurtosis may be difficult to understand for shop-floor people.

As Methods 1–3 do not reflect the process behavior, we suggest a method in the next section which will truly reflect the process behavior. The method is simple enough to be followed by the shop-floor personnel.

## Proposed Method

An extreme value (Type 1) distribution curve whose parameters (α and θ) are estimated from the sample is fitted to the data. Then the process capability index CPU can be calculated as

$$\text{CPU} = \frac{\text{USL} - x_{0.5}}{x_{0.99865} - x_{0.5}}, \tag{4}$$

where $x_{0.5}$ and $x_{0.99865}$ are the 50th percentile and 99.865 percentile points, respectively. These two can be calculated easily by

$$x_{0.5} = \alpha - \theta \ln(-\ln 0.5) = \alpha + 0.3665\theta$$

and

$$x_{0.99865} = \alpha - \theta \ln(-\ln 0.99865) = \alpha + 6.607\theta.$$

Replacing these two values in Eq. (4), we find

$$\text{CPU} = \frac{\text{USL} - \alpha - 0.3665\theta}{6.2405\theta}. \tag{5}$$

There is a risk associated with this method. If the empirical data do not fit the extreme value distribution (Type I), this method of estimating CPU can be misleading.

The above method [Eq. (4)] can be further simplified. If we substitute the value of α and θ from Eq. (3) to Eq. (5) by $\bar{x}$ and $s$, respectively, then we get

$$\text{CPU} = \frac{\text{USL} - \bar{x} + 0.1643s}{4.8657s}.$$

Generally, a process is called capable if the true value of CPU ≥ 1. Then, for a capable process

*Table 5.* CPU Using Different Methods

| DATA SET | SAMPLE SIZE | METHOD 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| 1 | 145 | 0.72 | 0.636 | 0.716 | 0.628 | 0.613 |
| 2 | 237 | 2.67 | 1.323 | 2.410 | 2.232 | 2.268 |
| 3 | 306 | 0.94 | 0.783 | 1.004 | 0.808 | 0.800 |

$$\frac{USL - \bar{x} + 0.1643s}{4.8657s} \geq 1$$

or

$$USL - \bar{x} \geq 4.7014s$$

or

$$\frac{USL - \bar{x}}{4.7014s} \geq 1.$$

Thus, if we formulate

$$CPU = \frac{USL - \bar{x}}{4.7s},$$

that will serve the purpose of estimating the PCI. Thus, in this method [Eq. (5)], one has to calculate $\bar{x}$ and $s$ from the sample and then CPU can be easily estimated. There will be some amount of error in estimating the true value of CPU.

The process is considered to be capable if CPU is greater than or equal to the critical value. The critical values for different sample sizes are discussed in the next section.

## Estimation of the CPU Index

The CPU indices are estimated using those five methods for the three data sets given earlier and are presented in Table 5. From the data sets, it can be observed that the CPU values estimated using method 4 and method 5 are more or less equal. For data set 2, the CPU value estimated using method 2 is clearly underestimated. For data set 3, the CPU value estimated using method 3 indicates that the process can be considered as capable. This may not be true, as there is nonconformance in the sample.

## Critical Values

In this section, we shall consider the critical values which are to be compared with the estimated CPU index in order that the process be considered capable with a high probability.

It is known that when the characteristic ($x$) follows a normal distribution, the capability index CPU follows a noncentral $t$-distribution. But, when the characteristic ($x$) follows an extreme value distribution, it is difficult to find the distribution of CPU. By using transformation techniques, the distribution can be found. Here, we try to find the critical values using the computer simulation programs. For finding the critical values of capability index CPU, only method 4 and method 5 are considered.

For a sample of size $n$, 1000 sample are generated, and for each sample, the CPU value is calculated. From the empirical distribution of those 1000 CPU values, 95 percentile and 99 percentile points are calculated. These are the 95% and 99% critical values, respectively, with which the estimated CPU values are to be compared to decide whether a process can be considered capable or not. For $n$ = 25(25)300, the critical values with 95% and 99% confidence levels are calculated and are shown in Table 6.

*Table 6.* Critical Values for Different Sample Sizes

| SAMPLE SIZE | CRITICAL VALUES OF CPU: METHOD 4, 95% | METHOD 4, 99% | METHOD 5, 95% | METHOD 5, 99% |
|---|---|---|---|---|
| 25 | 1.519 | 1.787 | 1.501 | 1.760 |
| 50 | 1.325 | 1.472 | 1.314 | 1.455 |
| 75 | 1.258 | 1.365 | 1.249 | 1.352 |
| 100 | 1.214 | 1.319 | 1.206 | 1.307 |
| 125 | 1.198 | 1.278 | 1.191 | 1.269 |
| 150 | 1.173 | 1.248 | 1.167 | 1.240 |
| 175 | 1.160 | 1.228 | 1.154 | 1.220 |
| 200 | 1.153 | 1.216 | 1.147 | 1.208 |
| 225 | 1.140 | 1.200 | 1.135 | 1.194 |
| 250 | 1.132 | 1.189 | 1.128 | 1.181 |
| 275 | 1.123 | 1.183 | 1.118 | 1.176 |
| 300 | 1.121 | 1.177 | 1.117 | 1.171 |

## Acknowledgments

The authors are grateful to Mr. D. K. Manna for his suggestions which led to substantial improvement in the contents and presentation. We are indebted to Mr. Bryan Dodson for his valuable comments and suggestions which also led to substantial improvement in the contents.

## References

1. Johnson, N. L. and Kotz, S., *Distribution of Statistics*, John Wiley and Sons, New York, 1970.
2. Dasgupta, R., On the Distribution of Eccentricity, *Sankhya, Series A*, 55, 226–232 (1993).
3. Dodson, B., *Weibull Analysis: with Software*, ASQC Quality Press, Milwaukee, WI, 1994.
4. Owen, D. B., *Handbook of Statistical Tables*, Addison-Wesley, Reading, MA, 1962.
5. Clements, J. A., Process Capability Calculations for Nonnormal Distribution, *Qual. Prog.* (September 1989).

*About the Authors:* Shri Ashok Sarkar has a bachelor's degree in textile technology and a postgraduate diploma in SQC and OR. He is working as a specialist in the SQC & OR unit of the Indian Statistical Institute, Madras and is engaged in consultancy services in quality assurance areas. He is also a faculty member of various courses offered by the institute.

Shri Surajit Pal has an M.Sc. in applied mathematics and has a postgraduate diploma in SQC and OR. He is working as a specialist in the SQC & OR unit of the Indian Statistical Institute, Madras, and is engaged in consultancy services in quality assurance areas. He is also a faculty member of various courses offered by the institute.