PROCESS CAPABILITY ANALYSIS FOR NON NORMAL DATA

Ejaz Ahmed
College of Computer Science and
Information Systems, Karachi
Suboohi Safdar,
Department of Statistics
University of Karachi, Karachi

Abstract

Process capability analysis refers to the normal behavior of a process when operating in a state of statistical control. Drives to continuous improvement are usually associated with the process capability measures. Typically we assume that the processes follows normal probability distribution ensuring a high percentage of the process measurements falling between $\pm 3s$ of the process mean and the total spread amounts to about $6s$ variations. This article describes the estimation of C_p and C_{pk}, commonly used process capability indices (PCI), in case of non-normal data using the characteristics of Weibull distribution. Earlier work of Lovelace and Swain (2009) has been extended for this distributional assumption. Quantiles are estimated by probability plotting technique and then control limits are obtained to determine whether the process is in statistical control or not. Percentage points of the fitted distribution have been used to compute under the assumption of Weibull distribution. We have used Delta method (Stuart and Ord, 1987) to estimate parameters and their standard errors. These estimated parameters are then used to develop new PCIs. Average PCI values are given along with the standard errors.

Keywords: Process capability, Estimation, Non normal Data, Weibull Distribution

JEL Classification: C1160

1. Introduction

Process capability refers to the inherent ability of a process to produce homogeneous parts for a sustained time period under given conditions. Kane (1986) described six areas of applications of capability indices that include presentation of non conforming products, measuring continuous improvement, communication, prioritization, identifying direction for improvement and auditing a quality system. Deleryd, et. al. (1999) identified six critical factors for successful implementation of process capability studies. Tsim (1997) identified four key objectives of PCI including, among others the ability to compare different processes (unit less measure) and to identify the closeness to target (Taguchi Loss Function concept).

Process capability measures the variability of a process relative to its specification limit based on three assumptions, namely, (i) the process is itself in control, (ii) target value and specifications of a quality characteristics are specified, and (iii) the process measures quality characteristics that follow a normal distribution.

2. Process Capability Indices- Review

The most commonly used PCI, named C_p Index, measures the potential process performance (process consistency) which only reflects the consistency of the product quality characteristic. C_p measures process spread related to specification limits and hence the location of process mean is not considered. Assuming a quality characteristic to follow $N(m, s^2)$ with upper and lower specification limits USL and LSL, this measure is defined as $C_p = (USL - LSL)/6s$. The expected proportion of nonconforming products assuming $m = (LSL + USL)/2$ can be obtained as

$$
\begin{aligned}
E(p) &= \Pr(X < LSL) + p(X > USL) \\
&= \Pr\left(Z < \frac{LSL - m}{s}\right) + p\left(Z > \frac{USL - m}{s}\right) \\
&= 2\Pr(X < LSL) \\
&= 2\Phi\left(\frac{LSL - \frac{\text{USL} + \text{LSL}}{2}}{s}\right) = 2\Phi\left(\frac{LSL - USL}{2s}\right) \\
&= 2\Phi[-d/s] \qquad \text{..........1}
\end{aligned}
$$

Where $\Phi(.)$ denotes standard normal cumulative distribution function and $d = (USL - LSL)/2$. Constable and Hobbs (1992) defined 'capable' as percentage of output within specification while Montgomery (2001) recommended minimum C_p equals to 1.33, for an existing process, and 1.50 for a new process. Small values of C_p are bad sign but large values do not guarantee of acceptability in the absence of information about the values of the process mean.

As is obvious C_p depends on the true standard deviation (SD) of the process which is usually unknown. It therefore forces to use an estimator of the SD which then results in estimated PCI, based upon sample observations. The estimated PCI given as $\hat{C}_p = (USL - LSL)/6\hat{s}$ is evaluated using suitable unbiased estimator of s , such as $\overline{R}/d_2$ (Montgomery, 2001). An important ratio $\hat{C}_p/C_p$ is quite frequently used as it yields the statistics. $\hat{C}_p = \left[\sqrt{(n-1)/c^2_{n-1}}\right] C_p$ This relationship is useful in constructing confidence intervals or testing hypotheses. Kane (1986) introduced another PCI known as C_{pk} which depends on both mean and standard deviation of the process to deal with violation of centering assumptions and to measure the actual capability performance. Mathematically this index is described as $C_{pk} = Min\{(USL -.m)/3s, (m - LSL)/3s\}$ Obviously, $C_{pk} = \{(d - |m - (LSL + USL)/2|)/d\}C_p$. Under the assumption of normality exact confidence interval for C_{pk} involves the joint distribution of two random variables following non-central t-distribution. Nagata and Nagahata (1992) showed that $(1-a)100\%$ approximate confidence interval for C_{pk} is given by,

$$\hat{C}_{pk} \pm Z_{a/2}\sqrt{\frac{1}{9n} + \frac{\hat{C}_{pk}^2}{2(n-1)}}$$

These measures are unit less and permit comparison amongst hundreds of process emanating from a whole range of production processes and industries. However the methodology and inferences about the process capability indices do not remain too straight forward in the absence of normality assumption. Next sections of this article will discuss the PCIs when the underlying distribution is skewed.

3. Case under Non-Normality

Numerous authors have discussed the construction and interpretation of PCIs under non-normal process behavior. Chen at al. (1988) proposed PCIs with distribution free tolerance intervals to estimate s while Clement (1989) and McCormack et. al. (2000) proposed empirical non-normal percentiles to evaluate both C_p and C_{pk}. Lovelace and Swain (2009) discussed the construction of C_p and C_{pk} assuming process behavior following a Log-Normal distribution. They used 99.875^{th} and 0.135^{th} quantiles to estimate both PCIs. Their proposed capability indices are given below,

$$C_p = \frac{USL - LSL}{X_{0.99865} - X_{0.00135}}$$

$$C_{pk} = Min\{C_{pu}.C_{pl}\}$$

$$where$$

$$C_{pu} = \frac{USL - Median}{X_{0.99865} - Median}$$

$$C_{pl} = \frac{Median - LSL}{Median - X_{0.00135}}$$

Chen and Chen (2004) compared these two process capability measures using four approximating methods, including bootstrap estimators. Since skewed distributions are not too common in production or service industries comparatively smaller proportion of literature is devoted to address PCIs in non-normal behaviors. Lovelace and Swain (2009) discussed both capability indices assuming process data following a Log-Normal distribution. Pal (2005) assumed process distribution to follow Generalized Lambda distribution and evaluated PCIs. We intend to discuss the construction of PCIs when the process distribution follows a Weibull distribution. Interested readers are recommended to refer to Munechika (1992), Pyzdek (1992), Kotz and Johnson (1993), Somerville and Montgomery (1996), and Pal (2005) for further details.

4. Process Capability Index- Weibull Distribution Case

Weibull distribution was originally proposed to describe data from life testing commonly used in reliability, fatigue and survivor analysis. Weibull distribution is a very important distribution and has been widely used in studies related to, for example, earthquakes, flood, breaking strengths, and reliability under censored or truncated situations. However parameter estimation is not easy especially in case of three parameter Weibull distribution. Quantiles are recommended to be used while estimating parameters. In this article we used delta method (Stuart and Ord, 1987) to find the standard error of the estimates. Later these estimated standard errors are used to construct our proposed process capability index.

The three-parameter Weibull distribution probability density function is given by

$$f(x)=\frac{b}{a}\left(\frac{x-g}{a}\right)^{b-1}\exp\left[-\left(\frac{x-g}{a}\right)^{b}\right]$$

where $x > g,\ g > 0, a > 0, b > 0$. We may derive the two-parameter Weibull distribution by assuming the location parameter "equal to zero". Many distributions are special cases of Weibull distribution, for example exponential distribution is a transformed form of Weibull distribution with shape parameter $b = 1$

We consider the two-parameter Weibull distribution with probability function given as $F(x) = 1 - \exp\left[-(x/a)^{b}\right]$, where x>0. A linear regression model was developed using the cumulative distribution function, as described below.

$$F(x) = 1 - \exp\left[-(x/a)^{b}\right]$$

$$\Rightarrow \ln(x) - \ln(a) = \left[\ln\{-\ln(1-F(x))\}\right]/b$$

$$\Rightarrow \ln(x) = \ln(a) + \left[\ln\{-\ln(1-F(x))\}\right]/b \quad \text{......2}$$

The last expression (equation 2) is equivalent to simple linear regression model, $y = b_0 + b_1 z + e$, where $y = \ln(x), b_0 = \ln(a), b_1 = 1/b$, and $z = \left[\ln\{-\ln(1-F(x))\}\right]$

Least square estimators of both parameters were determined and standard errors were obtained using the Delta method (Stuart and Ord, 1987) as given below:

$$V(\hat{a}) = \left[\frac{\partial}{\partial a}\exp(b_0)\right]^2 V(b_0) = [\exp(b_0)]^2 V(b_0) = [\exp(b_0)]^2\left[\frac{\sum y_i^2}{n\sum (y_i - \bar{y})^2}\right]s^2 \text{ -------------- } (3)$$

$$V(\hat{b}) = \left[\frac{\partial}{\partial b}\left(\frac{1}{b_1}\right)\right]^2 V(b_1) = \left[\left(-\frac{1}{b_1^2}\right)\right]^2 V(b_1) = \left[\left(\frac{1}{b_1}\right)\right]^4 V(b_1) = \left[\left(\frac{1}{b_1}\right)\right]^4\left[\frac{1}{\sum (y_i - \bar{y})^2}\right]s^2 \text{ ------ } (4)$$

5. Simulation Results and Newly Proposed PCI

The results of this section are based on a simulation study that we conducted to determine the effectiveness of our proposed measure of process capability. Details of the simulation procedure is listed below.

1. Fifteen samples each of size 30 were drawn randomly from a Weibull distribution with different scale and constant location parameters.

2. Regression model was fitted to equation (2) and estimators and standard errors for $\ln(a)$ and $(1/b)$ were obtained.

3. Variance of $\ln(a)$ and $(1/b)$ were obtained while fitting least square regression model.

4. Delta method was used to determine variances of a and b. Delta methods says that if $f = g(g)$ then $Var(g) = [g'(g)]^2 [Var(g)]$.

5. Process capabilities $\hat{C}_p = (USL - LSL)/6\hat{s}$ were estimated, replacing $\hat{s}$ with standard errors obtained in step 4.

6. Averages and standard errors of process capability indices are reported in Table 1.

Table 1: Table showing PCIs based on Weibull Parameters

a	b		Alpha	Beta	se(Alpha)	se(Beta)	Cp(Alpha)	Cp(Beta)
2	3	**Mean**	1.678	81.970	0.140	0.001	1.225	279.682
		SE	0.141	214.749	0.024	0.001	0.219	125.525
3	3	**Mean**	2.417	144.614	0.327	0.000	0.761	69807.179
		SE	0.239	248.027	0.458	0.000	0.208	254108.036
4	3	**Mean**	3.287	250.767	0.749	0.000	0.551	104157.087
		SE	0.356	753.433	0.990	0.000	0.374	394228.414
5	3	**Mean**	3.375	104.995	0.875	0.000	0.451	14053.322
		SE	0.315	305.475	1.021	0.001	0.251	53594.197

Control charts for a with $6s$ variations can be easily plotted constructing $UCL = \hat{a} + 3[se(\hat{a})]$ and $LCL = \hat{a} - 3[se(\hat{a})]$. Table 1 reveals that the estimated values for b are highly deviated from the original parametric values. The proposed method is recommended when the process behavior follows a Weibull distribution and the characteristic of interest is a .

6. References

1. Chan, Lk., Cheng, S. W., and Spiring, F. A. (1988). A New Measure of Process Quality: C_{pm} . Journal of Quality Technology, 20, 162-175.

2. Chen, J. P. and Chen, K. S. (2004). Comparing the capability of two processes using Cpm. Journal of Quality Technology, 36, 329-335.

3. Clements JA (1989) Process Capability Calculations for non normal distributions. *Quality Progress,* 95-100

4. Constable, G. K and Hobbs, J. R, (1992), Small Samples and Non Normal Capability, Trans, ASQC *Quality Congress* 1-7.

5. Deleryd, M., Deltin, J. and Klefsjo, B. (1999). Critical factors for successful implementation of process capability studies. Quality Management Journal, 6, 40-59.

6. Kane, V. E. (1986), "Process Capability Indices," *Journal of Quality Technology*, 18, 41 -52.

7. Kotz, S. and Johnson, N. L. (1993), *Process Capability Indices*, London: Chapman & Hall.

8. Kotz, S. and Johnson, N. L. (2002). Process Capability Indices—A Review, 1992–2000. Journal of Quality technology, 34, 2-19.

9. Lovelace, C. R. and Swain, J. J. (2009). Process capability analysis methodology for zero bound, non-normal process data. Quality Engineering, 21, 190-202.

10. McCormark, D. W. Jr., Harris, I. R., Hurwitz, A. M. and Spagon, P. D. (2000). Capability indices for non normal Data. *Quality Engineering,* **12**, 489-495

11. Montgomery, D. C. (2001). *Introduction to Statistical Quality Control, 4th ed.*, John Wiley. New York.

12. Munechika, M. (1992). Studies on Process capability in machining processes. *Rep. Stat. Appl. Res JUSE*,**39**, 14-29.

13. Nagata, Y. & Nagahata, H. (1992). Approximate formulas for the confidence intervals of process capability indices. *Okayama Econ. Rev.*, **25,** 301-314.

14. Pal, S. (2005). Evaluation of non normal process capability indices using generalized lambda distribution. *Quality Engineering*, **17, 77**-85.

15. Pyzdek, T. (1992). Process Capability Analysis Using Personal Computers, *Quality Engineering, 4,* 419-440

16. Somerville, S. and Montgomery D. C. (1996). Process Capability Indices and non-normal distributions. *Quality Engineering,* 19, 305-316

17. Stuart, A. and Ord, J. K. (1987). Kendall's advanced theory of statistics, Volume 1. Charles Griffin & Company Limited.