Process capability estimation using fuzzy probability in unknown probability distribution functions

S.Majid Ebrahimi
Department of Industrial Engineering
Mazandaran University of Science and Technology, Babol, Iran

Amir Mastali
Department of Industrial Engineering
Mazandaran University of Science and Technology, Babol, Iran

Shahab Namazi
Department of Industrial Engineering
Shahid Beheshti University, Tehran, Iran.

Abstract

Different methods have been carried out in order to estimate process capability in the literature, which many of them such as process capability indices and probability plot are used from a probability distribution function (PDF). However, in real case problems, because of system constraints e.g. sampling cost and environmental limitations, obtaining a PDF may be difficult to investigate. In this paper, a fuzzy based approach to estimate process capability is presented. The approach conducted through gathering limited number of samples from the process without using any information regarding to PDF and its parameters. Statistical inference validated the proposed approach and significant levels in different sample sizes have been tested. Computational experiments indicated that the approach can be used in small and medium sample size efficiently.

Keywords
Process capability indices, Fuzzy probability, Fuzzy number, Expected value, Probability distribution functions

1. Introduction
In the case of statistical quality control and application of its tools such as Showhart control chart or new problem solving method like Six Sigma methodology, estimation of process capability index usually is considered. Especially methods that could quantify process specification in order to decision making is mentioned. Process capability analysis indicates measurement, and quantifying the process variation in order to reduce or fix it. Two kinds of variation are known that affect processes, first systematic variations result in process mean changes, second probability variation clarify in process variation. Totally, less variation in quality characteristics is expected that is equal to more stability and less defect rate.
In from one point of view process capability analysis is equal to understanding the probability distribution function and its parameters to determine the range of product specification that this process could satisfy. This approach considers just process specification and is not related to engineering specification. For instance it is possible to express the output of process with a Normal distribution with μ=1 and σ =0.001 centimeter. Therefore analysis of process capability leads to define $\mu \pm 3\sigma$ as natural tolerances limits as this process is capable to produce well with these characteristics [0.997, 1.003].
Other approaches are based on the comparison of tolerances and Process limits. On the other word, how many of product will be defect using this process? Or what is the defect rate of product considering engineering specification? If good belongs to specification limit with high quality the process will be capable.
Most application of process capability analysis is:

- Predicting the conformities of processes and tolerances
- Selection of production process for product designer or manufacturers
- Selecting the best sub contractor to produce a specified product
- Reduction of variation in a process
- Sample size and period of sampling estimation for process control aim

For analyzing of process capability some methods exist like using control charts, design of experiment and probability and histogram diagram.
In most cases we would like to have a decision criteria using just a number that clarifies the position and variation of processes comparing tolerances zone, therefore process capability indices are introduced. These indices are based on three different approaches:

- Ratio of tolerance zone to process zone
- Percent of non conforming product
- Using loss function.

Literature review survey contains many different methods that some of them will be reviewed in the next sections. As will be seen, all of the approaches that are known until know are based on specific probability distribution function specially normal PDF like as C_p, C_{pk} . In the next section we will introduce fuzzy process capability function and its estimation methods. Finally we can have process capability estimation without any information about kind of distribution functions and their parameters. Just a simple sample will be sufficient to this calculation, finally in numerical example section statistical inference shows validation of proposed method.

2. Litrature review

There are different indices that are given in literature review. In this case assume that quality characteristic x and corresponding random sample ($x_1,x_2,\ldots,x_n$) are normal, in fact $X \sim N(\mu,\sigma^2)$ .let LSL denoted the lower and USL the upper specification limit , $m = (USL + LSL)/2$ the mid point of tolerance interval (LSL,USL), t the target for μ ,which we assume that t=m. The simplest index is the process potential index is defined as Equation (1)

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

It was suggested by Kane .This index is ratio of tolerance region to process region. It is clear that this method considers the variation of process.6σ in denominator of above fraction is based on assumption of approximately normal data.
In order to reflect departures from the target value as well as change in the process variation, several order indexes have been suggested, like C_{pk} (Sullivan) and C_{pm} (Chan, Chen and Spring) as below:

$$C_{pk} == \frac{\min\{USL - \mu, \mu - LSL\}}{3\sigma} \tag{2}$$

$$C_{pm} = \frac{USL - LSL}{6\sqrt{\sigma^2 + (\mu - t)^2}} \tag{3}$$

6σ , 3σ in denominator of above fractions are based on assumption of approximately Normal data. In the other words main constrains in above indices its normality assumption.
The last method is more efficient because of target value and process mean consideration. The weak point of C_{pk} is that these methods are not mentioned as target value, and the main constraint for all of these methods are normal distribution probability .The other approach is to use non conforming ratio (NC) as an index for capability process that are identified by Carr(1991) for the first time:

$$NC = p = \Pr\{x \notin [L,U]\} = 1 - \left\{\varphi\left(\frac{U-\mu}{\sigma}\right) - \varphi\left(\frac{L-\mu}{\sigma}\right)\right\} \tag{4}$$

This approach based on the stepwise loss function as below:

$$Loss = \begin{cases} 0 & L \le x \le U \\ 1 & otherwise \end{cases} \tag{5}$$

Using fraction conforming was suggested by Fleig (2000) defined as 1-NC (nonconforming).Although these methods are not restrict to normal PDF but determination for distribution for its calculations are needed. As mentioned above, the “6” in Equation c_p has been associated with the idea that a normal distribution for X provides a satisfactory approximation At a relatively early date, Clements (1989), in an influential paper, suggested that “6σ” be replaced by the length of the interval between the upper and lower 0.135 percentage points of the distribution of X (this is 6σ for a normal N(μ, σ2) distribution). The new PCI is

$$C'_p = \frac{U - L}{(\varepsilon_{1-a} - \varepsilon_a)} = \frac{2d}{\varepsilon_{1-a} - \varepsilon_a} \tag{6}$$

Where ξa is defined by $Pr[X \le \xi a] = a$, taking $a = 0.00135$, so that $\xi 1-a$, ξa are the upper and lower 0.135 percentiles of the distribution of X. [For
a $N(\mu, \sigma 2)$ distribution $\xi 1-a = \mu+3\sigma$, $\xi a = \mu-3\sigma$.]The corresponding definition for C_{pk} is

$$C'_{pk} = \frac{d - |\varepsilon_{0/5} - M|}{\frac{1}{2}(\varepsilon_{1-a} - \varepsilon_a)} \tag{7}$$

The replacement of the expected value (e.g. μ) by the median(ξ0.5) is a natural, though not essential choice. Among other methods we note that Kotz and Johnson (1993b) suggest replacing "6σ" by "5.15σ,"because the approximate relation Pr { μ - 2/575σ < X< μ + 2/575σ} = 0/99. Varies only a little among gamma distributions (see Merrington and Pearson (1958)). This might be use-full if there are insufficient data to provide sufficiently accurate fitting. On the other hand, there are two important drawbacks: the most important being the limitation to about 1 percent for the value of p when the value of the modified Cp is 1, which is too high in most cases; and the other, somewhat less formidable, drawback is the required limitation to gamma-type distributions, though this includes a wide range of distributional shapes, from exponential to normal (as a limiting case).
The preceding discussion relates to properties of PCIs which depend on knowledge of the values of parameters (μ, σ, etc.) in their formulation. Generally, values of these parameters are not known precisely, but have to be estimated from data obtained from samples of produced items. It is quite possible that the estimated values may differ substantially from the actual values, leading to serious discrepancies between estimated and actual properties (e.g. value of p). So far as we are aware, the most prevalent methods of estimating the basic PCIs are to replace μ by σ by, where $X_1, X_2, \ldots, X_n$ are n independent values of X, or by an appropriate multiple of the range (greatest Xi – least Xi). Very often, estimation is based on values from a series of samples, combining the individual sample estimates into a single estimate.
Recent investigations of asymptotic properties of estimators (spearheaded by Hallin and Seoh (1999)) indicate the importance of determining the sample sizes n for which asymptotic results are adequate. In many instances the appropriate values of n far exceed The "folklore values" of 25 or 30 suitable in the case of asymptotic results associated with the t-distribution.
Unusually, Little and Harrelson (1993) suggest using 2tn−1,0.9986S, in place of 6S, in the denominator to estimate Cp. Use of a multiplier of the average range in a number of subsamples of small size (< 10, usually) as an estimator for σ has been prominent in quality assurance for a long time. It is clear that all estimations methods based on determination of PDF of data. this constrain result in large sample gathering in other to have enough confidence to validation of data .for PDF determination some restriction like as high cost sampling or other limitation interact us to proposed a method that have validity with out any attention to kind of PDF. Therefore we will introduce fuzzy process capability index in next section.

3. Fuzzy Process Capability Index

As up to standard goods, process capability is equal to below probability value:

$$p = \int_L^U f(x)dx \tag{8}$$

That f(x) is the probability distribution of quality characteristic. In fact this relation means, discrete probability for [LSL, USL] interval. If we use indicator function to this definition we could have below equation:

$$p = \int_{-\infty}^{\infty} \mu(x) f(x)dx \tag{9}$$

That

$$\mu(x) = \begin{cases} 1 & x \in [L, U] \\ 0 & x \notin [L, U] \end{cases} \tag{10}$$

This concept is using of stepwise loss function that is defined up to tolerance region. fig 3.1 provide fuzzy quality ,when X=t ,μ(t)=1, which shows products are completely up-to-standard ,when X=LSL or USL,μ(LSL)=μ(USL)=0,it shown product are totally below standards with both sides symmetrical. The curve shown in Fig 3.1 is worked out through the fuzzy statistical, processing of users suitability appraisals. However, in actual calculation, it can sometimes be simplified: using a linear line segment rather than a curve. The process capability index can be defined as the probability of fuzzy up-to-standard products turn out in the process of production it is labeled as $\tilde{C}_p$. When the quality index of process products is a continues random variable,

$$\tilde{C}_p = \int_{-\infty}^{\infty} \mu(x) f(x) dx \tag{11}$$

In the formula, f(x) is the probability density function of x ,μ(x) is the membership function of fuzzy up-to-standard products. When the quality characteristic of process is a discrete random variable X and the value of X is xi (i=1,..,n).

$$\tilde{C}_p = \sum_{i=1}^{n} \mu(x_i) f(x_i) \tag{12}$$

In this formula $p_x(x_i)$ is the probability when $X=x_i$, $\mu(x_i)$is the membership function of fuzzy up-to-standard product. Obviously we have to know the value of the probability density of quality characteristic and the membership function of fuzzy up to standard products, before we can work out $\tilde{C}_p$.

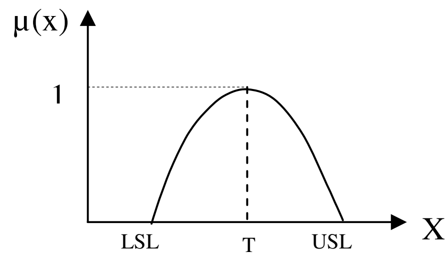
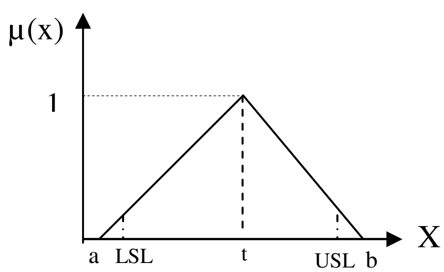


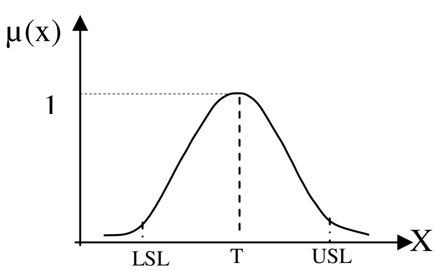
Figure 1: Fuzzy tolerance interval for up-to-standard product

The membership function that we define for up to standard product, in fact, is a fuzzy interval .Since then we use LR fuzzy number especially normal and triangular type in order to have simple and reasonable calculation.figure2.2 and figure 2.3 shows triangular and normal membership function type, respectively.



$$\mu(x) = \begin{cases} \dfrac{X-a}{t-a} & a \le X \le t \\ \dfrac{b-X}{b-t} & t \le X \le b \\ 0 & otherwise \end{cases}$$

Figure 2: Triangular membership function and its function



$$\mu(x) = e^{-1/2\left(\frac{m-x}{\lambda}\right)^2}$$

Figure 3: Normal membership function and its function

As an example if we have normal membership function and normal distribution, we could calculate fuzzy index as below integral:

$$\tilde{C}_p = \int_{-\infty}^{\infty} e^{-\frac{(x-t)^2}{2\lambda^2}} \frac{1}{\sqrt{2\pi}\sigma} e^{-1/2(\frac{x-\mu}{\sigma})^2} dx \tag{13}$$

If we solve it we have below answer:

$$\tilde{C}_p = (\frac{\lambda}{\sqrt{\sigma^2 + \lambda^2}}) e^{-(\mu - t)^2 / 2(\sigma^2 + \lambda^2)} \tag{14}$$

For index estimation two ways is applicable:

***a. definition bases method*:** in this method we put parameter estimated from sample to estimation of fuzzy index. For example if our sample has normal PDF we use $\overline{X}$ for μ and S^2 for σ^2.after this work it is enough to solve its integral or sigma. Below equation is sample of this way to normal data that solved analytically:

$$\hat{C}_{\tilde{p}} = (\frac{\lambda}{\sqrt{S^2 + \lambda^2}}) e^{-(\bar{x} - T)^2 / 2(s^2 + \lambda^2)} \tag{15}$$

For using of this way we have to find PDF of data and its parameter and it make many statistical calculations that we faced it in literature review.

b. expected value based method: regarding to definition of fuzzy process capability index and comparison to expected value definition we could have:

$$C_{\tilde{p}} = \int_{-\infty}^{\infty} \mu(x) f(x) dx = E(\mu(x)) \tag{16}$$

That it means mean of membership function of data is a good estimation to fuzzy index as below:

$$\hat{C}_{\tilde{p}} = \hat{E}(\mu(x)) = \overline{\mu(x)} = \frac{1}{N}\sum_{i=1}^{N} \mu(x_i) \tag{17}$$

For using this method, it is sufficient to gather a sample from population and without any inference about distribution of data and its parameter, just with calculation of membership value of sample data and mean of them we have output of this method.

For example if tolerance zone is [11.5, 14] and target value equal to 12.75 and define this membership function on the tolerance zone:

$$\mu_{nor}(x) = e^{-1/2(\frac{x - 12/75}{0/633})^2} \tag{18}$$

In addition we take 200 sample from a normal population with μ=12.8 and σ2=0.24.

Using first estimation method result in:

$$\hat{C}_{\tilde{p}} = (\frac{\lambda}{\sqrt{S^2 + \lambda^2}}) e^{-(\bar{x} - t)^2 / 2(s^2 + \lambda^2)} = 0/93847 \tag{19}$$

Also, second method has approximately same value:

$$\hat{C}_{\tilde{p}} = \frac{1}{N}\sum_{i=1}^{N} \mu_{nor}(x_i) = 0/9382 \tag{20}$$

4. Computational result

In order to compare results of purpose methods and validation, we calculate fuzzy index for several condition. Three kind of PDF is selected as shown in column 1, and then with attention to target value, USL and LSL, parameters of membership function is presented. "Mean A(x)" column shows result of method "b" that is no need for distribution information. In order to calculate it, we choose sample size equal 30. SCWP columns is process capability index of samples, using probability function (method a).

Table 1: Numerical results for validation of two methods

Distribution Function	target	USL	LSL	landa 1	landa 2	Mean A(x)	SCWP
N(12.8,0.24)	12.75	14	11.5	0.5106	0.5106	0.9926228	0.99257
N(12.8,0.24)	12	14	11.5	0.20427	0.817078	0.61216247	0.61216
N(12.8,0.24)	13.25	14	11.5	0.714943	0.306404	0.872535748	0.87265
$X^2(5)$	5	10	0	2.042695	2.042695	0.977527081	0.9771
$X^2(5)$	3	10	0	1.225617	2.859773	0.73215132	0.73119
$X^2(5)$	8	10	0	3.268312	0.817078	0.681189393	0.70488
U(5,15)	10	17	3	2.859773	2.859773	0.990109297	0.98979
U(5,15)	7	17	3	1.634156	4.08539	0.727241313	0.74613
U(5,15)	14	17	3	4.493929	1.225617	0.054808339	0.07529

In order to validate the two methods, we use T-test with significant level 0.05.finally, hypothesis test shows that the result of two proposed methods are statistically the same (t value=0.960861).

5. Conclusion

In this paper, two methods based on fuzzy concepts are suggested to process capability estimation. Results show that suggested methods have the same response. Strong point of this method is that it does not deal with PDF kind of sample and uses just sample data. In addition in small sample size, proposed methods are valid. Simplifying the calculations leads to practical usage of it.

References

1. Boyles, R. A., 1991, "The Taguchi Capability Index," Journal of Quality Technology, 23, pp.17–26.
2. Carr, W. E., 1991, "A New Process Capability Index: Parts per Million," Quality Progress, 24, pp. 152.
3. Chan,L.K.,Chen ,S.W. and Spring .F.A. ,1988, "Anew measure of process capability : C_{pm}," Journal of Quality Technology, 20, pp. 132-175.
4. Kane, V. E., 1986, "Process Capability Indices," Journal of Quality Technology, 18, pp. 41–52.
5. Klir,G.J., BO,Y. ,1995, "Fuzzy set and Fuzzy Logic, Theory and application," Prentice-hall Inc, New Jersey.
6. Kotz S., Johnson N. L., 2002, "Process Capability Indices—A Review 1992–2000," Journal of Quality Technology, 34, No. 1.
7. Montgomery, D. C., 1996, "Introduction to Statistical Quality Control," 3rd ed. John Wiley & Sons, New York, NY.
8. Pearn, W., Kotz, L. S., Johnson, N. L., 1992, "Distributional and Inferential Properties of Process Capability Indices," Journal of Quality Technology, 24, pp. 216–231.
9. Sullivan, L. P., 1985, Letters, Quality Progress, 18, pp. 7-8.
10. Yongting, C., 1996, "Fuzzy quality and analysis on fuzzy probability," Fuzzy set and system, 83, pp. 283-290.
11. Zadeh, L.A., 1967, "Fuzzy Set. Inform. Control," 8, pp. 338-340.