ELSEVIER

European Journal of Operational Research 129 (2001) 683–688

EUROPEAN JOURNAL OF OPERATIONAL RESEARCH

www.elsevier.com/locate/dsw

Short Communication

C_{pk} index estimation using fuzzy numbers

Hong Tau Lee *

Department of Industrial Engineering and Management, National Chin-Yi Institute of Technology, Taiping, Taichung County 4111, Taiwan, ROC

Received 9 February 1999; accepted 3 November 1999

Abstract

Process capacity indices (PCIs) were developed and have been successfully used by companies to compete in and dominate the high-profit markets by improving the quality and the productivity since the past two decades. There is an essential assumption, in the conventional application, wherein the output process measurements are precise and distributed as normal random variables. Since the assumption of normal distribution is untenable, errors can occur if the C_{pk} index is computed using non-normal data. In the present study, we address the situation that the output of data from measurement of the quality of a product is insufficiently precise or scarce. This is possible when the quality measurement refers to the decision-maker's subjective determination. In such a situation, the linguistic variable that is easier to capture the decision-maker's subjective perception is applied to construct the PCI C_{pk}. The present approach can mitigate the effect when the normal assumption is inappropriate and extends the application of C_{pk} index. © 2001 Elsevier Science B.V. All rights reserved.

Keywords: Process; Capacity; Fuzzy; Index

1. Introduction

A common way to summarize the process performance is by using the process capacity indices (PCIs) which provide information with respect to engineering specification [10]. The measurement of performance of a process considering both the location and the dispersion information about it is referred as the PCI of interest in the present study, C_{pk}. It is defined as follows:

$$C_{pk} = (1 - k) \times \frac{T}{6s},$$

where $k = |\bar{x} - \mu|/(T/2)$, $\bar{x}$ and s are the mean and standard deviation calculated from a sample, μ is the central value of specifications and $T = \text{USL} - \text{LSL}$, the difference in the upper and lower bounds of specifications is usually explained as the toleration of specifications. This index can be explained as the multiplication of the capability of accuracy that concerns the location information of the process and the capability of precision that concerns the dispersion information of the process, respectively. Another definition for index C_{pk} is as follows:

* Tel.: +886-4-392-4505; fax: +886-4-393-4620.

PII: S 0 3 7 7 - 2 2 1 7 (9 9) 0 0 4 3 8 - 5

$$C_{pk} = \min\left(C_{pk}(U),\ C_{pk}(L)\right),$$

where

$$C_{pk}(U) = \frac{\mathrm{USL} - \bar{x}}{3s}, \quad C_{pk}(L) = \frac{\bar{x} - \mathrm{LSL}}{3s}.$$

Then index C_{pk} is the shorter standardized distance from the center of the process to either USL or LSL [1]. These two definitions represent the same thing and get the same result for a specific sample. Definition 1 is more concise and easy to operate. In this research, we refer the index C_{pk} to the first definition, while the procedure discussed below can be applied to another definition with no modification.

The PCIs are used in industry. An underlying assumption is that the output process measurements are distributed as normal random variables. When normal distributions are assumed, however, different distributions are present such as skew, heavy-tailed, and short-tailed distributions and the percentages of non-conforming parts are significantly different from the computed PCIs indication. Experience shows that the normality assumption is often not met in real world application. Rivera et al. [11] have discussed these issues, using the data transformation technique to conform the data arising from non-normal distribution to the normality assumption, while the PCIs were computed for the transformed data. In the present study, we address a situation when the data obtained from the quality measurement of product are of linguistic value or are insufficiently precise, while the conventional assumption of any data distribution is no more appropriate. The situation arises since it is difficult to define the quality of the product using numerical values. For example, the buyer asks that the foaming of a polymer product must be suitable. Since the foaming process is influenced by the processing time, temperature and other uncontrollable factors it may cause over foaming or under foaming of the polymer product. However, due to over foaming, the density of the product is lower and the wear-resisting property of the product is poor. On the other hand, due to under foaming, the density of the product is higher and the elasticity of the product is hurt. Hence, the conventional quality measurement proves to be difficult for evaluating the performance of the process directly. In order to overcome such situations, fuzzy theory [13–16] is applied to construct the PCI. Fuzzy set theory, as developed by Zadeh [13–15], and the concept and arrangement of fuzzy numbers presented by Dubois and Prade [4–7] are applied to improve the presentation of the fuzzily defined system. Besides, Liang and Wang [9] have developed an algorithm for personnel selection, wherein, they first aggregate the decision-maker's linguistic assessments regarding subjective criteria weightings and ratings to obtain the fuzzy suitability index and its ranking value for each candidate. Lee [8] has discussed a manpower forecasting problem when the independent variables are fuzzy numbers in his regression model. In the present study, the linguistic variables are used to express the evaluation of the quality of a product. Consequently, C_{pk} is defined with fuzzy numbers. It can be applied to evaluate the capacity of a process where human judgments are involved in the evaluation of the process performance, since the index C_{pk} involves the location information and the dispersion information, simultaneously. These informations are presented by the mean and standard deviation of the producing process. These two statistics have been proven to be independent [12]. We treat them as non-interactive variables. In the next section, the membership functions of mean and standard deviation of fuzzy number are conducted. In Section 3, the C_{pk} index estimation presented by fuzzy numbers is constructed and its membership function is approximated. Besides, the defuzzy operation is also depicted in this section. Section 4 concludes the study.

2. The membership function for standard deviation of fuzzy number

The C_{pk} index is the multiplication of capability of accuracy and capability of precision of a process simultaneously. Besides, these two capacities are calculated from the mean and standard deviation of the data collected from the producing process. In order to obtain the membership function of C_{pk}, the membership functions of the two statistics must be identified beforehand. First of all, Proposition 1 is carried out for further use in the present study. Let us assume that $\tilde{x}_j,\ j=1,\dots,n$, $\tilde{y}_j,\ j=1,\dots,n$, be samples of the random variables of fuzzy number. $\tilde{x}$ and $\tilde{y}$ have triangular membership function as follows:

$$U_{\tilde{x}_j}(x_j)=\begin{cases}\frac{1}{p_j-o_j}(x_j-o_j), & o_j\leqslant x_j\leqslant p_j,\\ \frac{1}{p_j-q_j}(x_j-q_j), & p_j\leqslant x_j\leqslant q_j,\\ 0, & \text{otherwise,}\end{cases}$$

$$U_{\tilde{y}_j}(y_j)=\begin{cases}\frac{1}{m_j-l_j}(y_j-l_j), & l_j\leqslant y_j\leqslant m_j,\\ \frac{1}{m_j-n_j}(y_j-n_j), & m_j\leqslant y_j\leqslant n_j,\\ 0, & \text{otherwise,}\end{cases}$$

let

$$\bar{\bar{x}}=\frac{\sum_{j=1}^n\tilde{x}_j}{n},\quad \bar{\bar{y}}=\frac{\sum_{j=1}^n\tilde{y}_j}{n}$$

be the average operation for fuzzy numbers $\tilde{x}$ and $\tilde{y}$, respectively.

Proposition 1. *Define* COV $(\tilde{x},\tilde{y})=(1/n)\sum_{j=1}^n(\tilde{x}_j-\bar{\bar{x}})(\tilde{y}_j-\bar{\bar{y}})$ *as the covariance between the fuzzy numbers* $\tilde{x}$ *and* $\tilde{y}$. *Then the membership function for* COV$(\tilde{x},\tilde{y})$ *is obtained as follows*:

$$U_{\mathrm{COV}(\tilde{x}_j,\tilde{y}_j)}(r)=\begin{cases}\frac{-B_1}{2A_1}+\left[\left(\frac{B_1}{2A_1}\right)^2-\frac{C_1-r}{A_1}\right]^{1/2}, & C_1\leqslant r\leqslant C_3,\\ \frac{B_2}{2A_2}-\left[\left(\frac{B_2}{2A_2}\right)^2-\frac{C_2-r}{A_2}\right]^{1/2}, & C_3\leqslant r\leqslant C_2,\\ 0, & \textit{otherwise,}\end{cases}\tag{1}$$

wherein

$$A_1=\frac{1}{n}\sum_{j=1}^n(b_j-a_j)(e_j-d_j),$$

$$A_2=\frac{1}{n}\sum_{j=1}^n(c_j-b_j)(f_j-e_j),$$

$$B_1=\frac{1}{n}\sum_{j=1}^n(a_j(e_j-d_j)+d_j(b_j-a_j)),$$

$$B_2=\frac{1}{n}\sum_{j=1}^n(c_j(f_j-e_j)+f_j(c_j-b_j)),$$

$$C_1=\frac{1}{n}\sum_{j=1}^n a_jd_j,\quad C_2=\frac{1}{n}\sum_{j=1}^n c_jf_j,\quad C_3=\frac{1}{n}\sum_{j=1}^n b_je_j$$

and

$$a_j=o_j-\frac{\sum_{j=1}^n q_j}{n},\quad b_j=p_j-\frac{\sum_{j=1}^n p_j}{n},$$

$$c_j=q_j-\frac{\sum_{j=1}^n o_j}{n},\quad d_j=l_j-\frac{\sum_{j=1}^n n_j}{n},$$

$$e_j=m_j-\frac{\sum_{j=1}^n m_j}{n},\quad f_j=n_j-\frac{\sum_{j=1}^n l_j}{n}.$$

Proof. By the extension principle of fuzzy sets [11,13–15] and the definition of the triangular fuzzy number [4,5,7], the average and the subtracting operations of a set of triangular fuzzy numbers are still a triangular fuzzy number. For example, if $\tilde{A}=(a,b,c)$ and $\tilde{B}=(d,e,f)$ then

$$\tilde{C}=\frac{1}{2}(\tilde{A}+\tilde{B})=\left(\frac{a+d}{2},\frac{b+e}{2},\frac{c+f}{2}\right),$$

$$\tilde{D}=(\tilde{A}-\tilde{B})=(a-f,b-e,c-d).$$

By the above extension principles, the membership functions of the fuzzy number $\tilde{x}_j-\bar{\bar{x}}$ and $\tilde{y}_j-\bar{\bar{y}}$ are available.

From the membership functions of the fuzzy number $\tilde{x}_j-\bar{\bar{x}}$ and $\tilde{y}_j-\bar{\bar{y}}$, the α-cut interval sets for the two fuzzy numbers are

$$\left(\tilde{x}_j-\bar{\bar{x}}\right)^\alpha=[a_j+\alpha(b_j-a_j),c_j+\alpha(b_j-c_j)],$$

$$\left(\tilde{y}_j-\bar{\bar{y}}\right)^\alpha=[d_j+\alpha(e_j-d_j),f_j+\alpha(e_j-f_j)],$$

respectively. By the extension principle of the fuzzy sets, the multiplication of two α-cut interval sets results in a α-cut interval set as

$$\begin{aligned}Z_j^\alpha&=\left(\tilde{x}_j-\bar{\bar{y}}\right)^\alpha*\left(\tilde{y}_j-\bar{\bar{y}}\right)^\alpha\\&=\{[a_j+\alpha(b_ja_j)]*[d_j+\alpha(e_jd_j)],\\&\quad[c_j+\alpha(b_j-c_j)]*[f_j+\alpha(e_j-f_j)]\}.\end{aligned}$$

The membership function of the fuzzy number Z_j^α is facile.

Although the membership function of Z_j^α is obtained, it is not again a triangular fuzzy number. Hence, the summation of the n fuzzy number derived a fuzzy number with an approximated relationship function as Function 1 above. Fig. 1 describes the shape of the approximated function.

From Proposition 1, the membership function for $SS_{\tilde{x}} = (1/n)\sum_{j=1}^{n}(\tilde{x}_j - \bar{\bar{x}})(\tilde{x}_j - \bar{\bar{x}})$ as the sum of square error of fuzzy numbers $\tilde{x}$ is obtained as follows:

$$U_{SS_{\tilde{x}}}(s) = \begin{cases} \frac{-B_1}{2A_1} + \left[\left(\frac{B_1}{2A_1}\right)^2 - \frac{C_1 - s}{A_1}\right]^{1/2}, & C_1 \leqslant s \leqslant C_3, \\ \frac{B_2}{2A_2} - \left[\left(\frac{B_2}{2A_2}\right)^2 - \frac{C_2 - s}{A_2}\right]^{1/2}, & C_3 \leqslant s \leqslant C_2, \\ 0, & \text{otherwise,} \end{cases} \tag{2}$$

where

$$A_1 = \frac{1}{n}\sum_{j=1}^{n}(b_j - a_j)^2, \quad A_2 = \frac{1}{n}\sum_{j=1}^{n}(c_j - b_j)^2,$$

$$B_1 = \frac{1}{n}\sum_{j=1}^{n}2a_j(b_j - a_j), \quad B_2 = \frac{1}{n}\sum_{j=1}^{n}2c_j(c_j - b_j),$$

$$C_1 = \frac{1}{n}\sum_{j=1}^{n}a_j^2, \quad C_2 = \frac{1}{n}\sum_{j=1}^{n}c_j^2, \quad C_3 = \frac{1}{n}\sum_{j=1}^{n}b_j^2$$

and

$$a_j = o_j - \frac{\sum_{j=1}^{n} q_j}{n}, \quad b_j = p_j - \frac{\sum_{j=1}^{n} p_j}{n},$$

$$c_j = q_j - \frac{\sum_{j=1}^{n} o_j}{n}.$$

The membership function for the standard deviation

$$S_{\tilde{x}} = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(\tilde{x}_j - \bar{\bar{x}})(\tilde{x}_j - \bar{\bar{x}})}$$

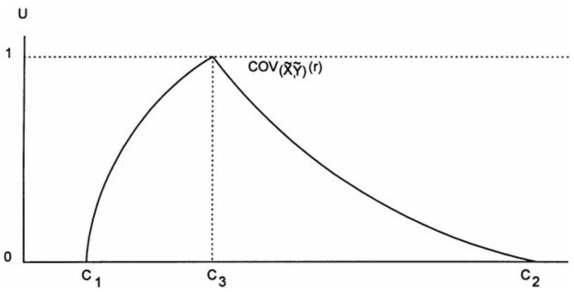

Fig. 1. Shape of the approximated membership Function 1.

is obtained immediately as follows:

$$U_{S_{\tilde{x}}}(s) = \begin{cases} \frac{-B_1}{2A_1} + \left[\left(\frac{B_1}{2A_1}\right)^2 - \frac{C_1 - s^2}{A_1}\right]^{1/2}, & C_1 \leqslant s \leqslant C_3, \\ \frac{B_2}{2A_2} - \left[\left(\frac{B_2}{2A_2}\right)^2 - \frac{C_2 - s^2}{A_2}\right]^{1/2}, & C_3 \leqslant s \leqslant C_2, \\ 0, & \text{otherwise,} \end{cases} \tag{3}$$

the meaning of each symbol in Eq. (3) is the same as Eq. (2). For simplicity, in practice, the approximation formula $U_{S_{\tilde{x}}}(s) \cong (C_1, C_3, C_2)$ that represents the triangular fuzzy number can be used [8].

3. Membership function of fuzzy index C_{pk}

The membership function of the index C_{pk} is obtained by first approximating the membership function to a triangular function for the capability of accuracy and the capability of precision, respectively. Finally, Proposition 1 is applied to conduct the membership function of the index.

Let the membership functions of the parameters LSL, USL and target value $\tilde{\mu}$ have the triangular form of (l, m, n), (o, p, q) and (w, y, z), respectively. Suppose the central value of membership function of $\bar{\bar{x}}$ is larger than y (if $\bar{\bar{x}}$ is not larger than y, then the conduction process is similar as the following and can be easily followed). By Proposition 1, the membership function of $\bar{\bar{x}} - \tilde{\mu}$ is a triangular form of

$$\left(\frac{\sum_{j=1}^{n} o_j}{n} - z, \frac{\sum_{j=1}^{n} p_j}{n} - y, \frac{\sum_{j=1}^{n} q_j}{n} - w\right).$$

The membership function of $T/2$ is also a triangular form of $((1/2)(o - n), (1/2)(p - m), (1/2)(q - l))$. Consequently, the membership function of $1 - k$ is approximated as a triangular form of

$$\left(1 - \left(\frac{\sum_{j=1}^{n} o_j}{n} - z\right)\left(\frac{2}{q - l}\right), 1 - \left(\frac{\sum_{j=1}^{n} p_j}{n} - y\right)\right.$$

$$\left.\left(\frac{2}{p - m}\right), 1. - \left(\frac{\sum_{j=1}^{n} q_j}{n} - w\right)\left(\frac{2}{o - n}\right)\right).$$

Besides, the membership function of $T/6s$ is also approximated to the triangular form of

$$\left((o-n)\left(\frac{1}{6C_2}\right),\ (p-m)\left(\frac{1}{6C_3}\right),\ (q.-l)\left(\frac{1}{6C_1}\right)\right).$$

Let

$$a = 1-\left(\frac{\sum_{j=1}^{n} o_j}{n}-z\right)\left(\frac{2}{q-l}\right),$$
$$b = 1-\left(\frac{\sum_{j=1}^{n} p_j}{n}-y\right)\left(\frac{2}{p-m}\right),$$
$$c = 1-\left(\frac{\sum_{j=1}^{n} q_j}{n}-w\right)\left(\frac{2}{o-n}\right),$$
$$d=(o-n)\left(\frac{1}{6C_2}\right),\quad e=(p-m)\left(\frac{1}{6C_3}\right),$$
$$f=(q-l)\left(\frac{1}{6C_1}\right).$$

Finally, Proposition 1 is again applied to obtain the membership function for index C_{pk} as follows:

$$U_{C_{pk}}(I)\cong\begin{cases}\frac{-B_1}{2A_1}+\left[\left(\frac{B_1}{2A_1}\right)^2-\frac{C_1-I}{A_1}\right]^{1/2}, & C_1\leqslant I\leqslant C_3,\\ \frac{B_2}{2A_2}-\left[\left(\frac{B_2}{2A_2}\right)^2-\frac{C_2-I}{A_2}\right]^{1/2}, & C_3\leqslant I\leqslant C_2,\\ 0, & \text{otherwise},\end{cases}\tag{4}$$

where

$$A_1=(b-a)(e-d),\quad A_2=(c-b)(f-e),$$
$$B_1=a(e-d)+d(b-a),$$
$$B_2=c(f-e)+f(c-b),$$
$$C_1=ad,\quad C_2=cf,\quad C_3=be.$$

After the membership function of C_{pk} is obtained, the next step is to calculate the ranking values for the membership function of C_{pk} for all quality evaluations. It converts the fuzzy number C_{pk} to a crisp score for the purpose of comparison. Many fuzzy ranking methods have been proposed [2,3]. For simplicity of calculation and problem solving capacity, the method proposed by Chen [3] is applied to rank the fuzzy suitability index. Let $G(I)$ be the suitability index for fuzzy index C_{pk}, then define the two other sets which are the maximizing set $M=\{(I,f_M(I))\,|\,I\in R\}$ and the minimizing set $m=\{(I,f_m(I))\,|\,I\in R\}$ with two related functions as

$$f_M=\begin{cases}\frac{(I-I_1)}{(I_2-I_1)}, & I_1\leqslant I\leqslant I_2,\\ 0, & \text{otherwise},\end{cases}$$
$$f_m=\begin{cases}\frac{(I_2-I)}{(I_2-I_1)}, & I_1\leqslant I\leqslant I_2,\\ 0, & \text{otherwise},\end{cases}$$

where

$$I_1=\inf D,\quad I_2=\sup D,\qquad D=\bigcup_{i=1}^{p} D_i,$$
$$D_i=\left\{I\,\middle|\,u_{G_i(I)}>0\right\},\quad i=1,2,\ldots,p,$$

and p is the number of processes being considered to compare. If the value of p is 1 that means we just want to know the capacity of a specific process, then we can set $I_1=0$ and $I_2=1$. Thereafter, define the optimistic utility $U_{M(G)}=\sup_I\left(u_{G(I)}\Lambda u_{M(I)}\right)$ and the pessimistic utility $U_{m(G)}=1-\sup_I\left(u_{G(I)}\Lambda u_{m(I)}\right)$. After certain necessary manipulations, the values of $U_{M(G)}$ and $U_{m(G)}$ for the problem in the present study are obtained as follows:

$$U_{M(G)}=\frac{-(B_2+1)}{2A_2}-\left[\left(\frac{-(B_2+1)}{2A_2}\right)^2-\frac{C_2}{A_2}\right]^{1/2},$$
$$U_{m(G)}=\frac{-(B_1+1+2A_1)}{2A_1}-\left[\left(\frac{-(B_1+1+2A_1)}{2A_1}\right)^2-\left(\frac{B_1+C_1}{A_1}+1\right)\right]^{1/2}.$$

Finally, the ranking value is obtained by a linear combination of the value of $U_{M(G)}$ and $U_{m(G)}$ as $U_{T(G)}=\alpha(U_{M(G)})+(1-\alpha)(U_{m(G)}),\ 0\leqslant\alpha\leqslant 1.$ The value of α can be manipulated to reflect the scheduler's risk taking attitude. Although, the calculation is a little tedious, however, with a formula it can be easily calculated by a computer. The larger value of $U_{T(G)}$ indicates that the mean of quality measurement of the sample is closer to the target of specification and the quality of the sample is more coincident and stable.

4. Conclusion

The objective of the present study is to propose a method to calculate the C_{pk} index when precise quality cannot be identified and must be referred to the decision-maker's subjective evaluation of the quality of product. The fuzzy set theory and the linguistic variables are applied to tackle these problems. While applying this approach, the practitioner must first define the membership function of each linguistic variable according to the property of the quality to be measured. The membership function for statistic mean and standard deviation can be approximated according to the membership function of the linguistic variables. Finally, the membership function of the C_{pk} index is conducted and the ranking value is calculated. Although, in the present research, we confine that all the fuzzy numbers have triangular membership function. This can be released for other types of membership functions but the complexity of applying the extension principle increased. However, the exposition approach proposed in our paper is generic enough to be applicable to the fuzzy number with other types of membership functions.

References

[1] F. Alsup, R.M. Watson, Practical Statistical Process Control, Van Nostrand Reinhold, New York, 1993.
[2] L.M. Campos, A. Gonzalez, A subjective approach for ranking the fuzzy numbers, Fuzzy Sets and Systems 29 (1989) 145–153.
[3] S.H. Chen, Ranking fuzzy numbers with maximizing and minimizing set, Fuzzy Sets and Systems 17 (1985) 113–129.
[4] D. Dubois, H. Prade, Operations on fuzzy numbers, International Journal of Systems Science 9 (1978) 613–626.
[5] D. Dubois, H. Prade, Fuzzy Sets and Systems: Theory and Applications, Academic Press, New York, 1980.
[6] D. Dubois, H. Prade, Addition of interactive fuzzy numbers, IEEE Transactions on Automatic Control 26 (1981) 926–936.
[7] D. Dubois, H. Prade, Possibility Theory: An Approach to Computerized Processing of Uncertainty, Plenum Press, New York, 1986.
[8] H.T. Lee, Fuzzy regression model with fuzzy input and output data for manpower forecasting, Fuzzy Sets and Systems 119 (2) (2001) 19–27.
[9] G.-S. Liang, M.-J.J. Wang, Personnel selection using fuzzy MCDM algorithm, European Journal of Operational Research 78 (1994) 22–33.
[10] P.E. McCoy, Using performance indexes to monitor production processes, in: Quality Progress, February 1991, pp. 49–55.
[11] L.A.R. Rivera, N.F. Hubele, F.P. Lawrence, C_{pk} index estimation using data transformation, Computers Industrial Engineering 29 (1–4) (1995) 55–58.
[12] J. Shuster, A simple method of teaching the independence of $\bar{x}$ and s^2, The American Statistician 27 (1) (1973).
[13] L.A. Zadeh, Fuzzy sets, Information and Control 8 (1965) 338–353.
[14] L.A. Zadeh, The concept of a linguistic variable and its application to approximate reasoning, parts 1 and 2, Information Science 8 (1975) 199–249, 301–357.
[15] L.A. Zadeh, The concept of a linguistic variable and its application to approximate reasoning, part 3, Information Science 9 (1976) 43–80.
[16] H.J. Zimmermann, Fuzzy Set Theory and Its Application, Kluwer Academic Publishers, Reading, 1991.