Journal of the Korean Statistical Society (2007), **36***: 4, pp 471–477*

ON THE LIMITING DISTRIBUTION FOR ESTIMATE OF PROCESS CAPABILITY INDEX

HYO-IL PARK[1] AND JOONG JAE CHO[2]

ABSTRACT

In this paper, we provide a new proof to correct the asymptotic normality for the estimate $\hat{C}_{pmk}$ of C_{pmk}, which is one of the well-known definitions of the process capability index. Also we comment briefly on the correction of the limiting distribution for $\hat{C}_{pmk}$ and on the use of re-sampling methods for the inference of C_{pmk}. Finally we discuss the concept of asymptotic unbiasedness.

AMS 2000 subject classifications. Primary 62G20; Secondary 62N05.
Keywords. Asymptotic normality, consistency.

1. INTRODUCTION

As a criterion on whether a production process is capable of producing items satisfying the quality requirements preset by the designer, the process capability index (PCI) has received substantial attention in the quality control. Also the definitions of PCI have been developed and evolved in accordance with the need to accommodate additional situations such as non-centering case of the process mean between two specifications of limits and/or inclusion of the target value into the process. Among them, C_p, C_{pk}, C_{pm} and C_{pmk} are well-known and widely used definitions of PCI and the corresponding estimates have been obtained by substituting empirical ones from data for the components in the expression of PCI's. Then considerable researches have been followed to show the asymptotic properties such as the consistency and asymptotic normality for each estimate by many statisticians. See, Chan *et al.* (1990), Chen and Hsu (1995) and Chen and

Received May 2006; accepted May 2007.
[1]Corresponding author. Department of Statistics, Chong-ju University, Chong-ju, Choongbook 360-761, Korea (e-mail: hipark@cju.ac.kr)
[2]Department of Information Statistics, Chungbuk National University, Chong-ju, Choongbook 361-769, Korea (e-mail: jjcho@chungbuk.ac.kr)

Pearn (1997), among others. However the first two papers, mentioned just before contain a common fault for dealing with the estimates of PCI's. This fault came from the efforts that both the works have been carried out by trying to deal with the estimates in a unified fashion for the definition of PCI's and the authors have used repeatedly and unconsciously the same mathematical methodology when they derived the asymptotic normality. As results, some of their conclusions are not reliable and even the limiting variances are not correct. Also in some cases, the limiting distribution is not normal. Therefore, in this paper, we consider to correct the results for the asymptotic normality for the estimates of PCI's. We consider only the correction of the paper by Chen and Hsu (1995), who dealt with C_{pmk} and mentioned briefly the correction for Chan *et al.* (1990). In order to clarify our arguments, first of all, we begin our discussion by stating the definition of C_{pmk}. we note that the definition of C_{pmk} allows the case of non-centering and accommodate the target value. In the following L and U are the lower and upper specification limits of the process, respectively. Then the definition of C_{pmk} is as follows:

$$C_{pmk} = \frac{\min\{U-\mu, \mu-L\}}{3\sqrt{E(X-T)^2}},$$

where X is a random variable and represents the quality characteristics of product and T is the target value. Also let

$$D = \frac{U-L}{2} \quad \text{and} \quad M = \frac{U+L}{2}.$$

Then C_{pmk} can be represented as follows:

$$C_{pmk} = \frac{D-|\mu-M|}{3\sqrt{\sigma^2+(\mu-T)^2}}, \tag{1.1}$$

where μ and σ^2 are the mean and variance of X, respectively. we note that from (1.1), when $\mu = M$, *i.e.*, the process mean and mid-point between the two specification limits coincide,

$$C_{pmk} = \frac{D}{3\sqrt{\sigma^2+(\mu-T)^2}} = \frac{U-L}{6\sqrt{\sigma^2+(\mu-T)^2}}, \tag{1.2}$$

which is also the definition of C_{pm}. C_{pm} only considers the target value of the process. Also if $\mu < M$,

$$C_{pmk} = \frac{\mu-L}{3\sqrt{\sigma^2+(\mu-T)^2}}, \tag{1.3}$$

and if $\mu > M$,

$$C_{pmk} = \frac{U-\mu}{3\sqrt{\sigma^2+(\mu-T)^2}}. \tag{1.4}$$

Chen and Hsu (1995) studied the asymptotic properties based on the definition (1.1). Therefore, they overlooked the simplified form (1.2) when $\mu = M$ and made incorrect conclusion for the asymptotic normality. Hence we consider the asymptotic properties based on (1.2)–(1.4) in a separated manner.

2. ASYMPTOTIC NORMALITY

Suppose that we have a sample $X_1, \ldots, X_n$ from a production process with an unknown distribution function F having finite fourth moment. Let $\bar{X}_n$ and S_n^2 be the respective sample mean and variance based on $X_1, \ldots, X_n$. Then any reasonable estimate $\hat{C}_{pmk}$ of C_{pmk} may be as follows:

i) if $\mu = M$, then

$$\hat{C}_{pmk} = \frac{U-L}{6\sqrt{S_n^2+(\bar{X}_n-T)^2}},$$

ii) if $\mu < M$, then

$$\hat{C}_{pmk} = \frac{\bar{X}_n-L}{3\sqrt{S_n^2+(\bar{X}_n-T)^2}},$$

iii) if $\mu > M$, then

$$\hat{C}_{pmk} = \frac{U-\bar{X}_n}{3\sqrt{S_n^2+(\bar{X}_n-T)^2}}.$$

Upon the estimates just listed, we derive the asymptotic normality of

$$\sqrt{n}\left(\hat{C}_{pmk} - C_{pmk}\right)$$

for each case.

For the case i), first of all, we note the following expression with rationalizing the numerator,

$$\begin{aligned}
&\sqrt{n}\left(\hat{C}_{pmk} - C_{pmk}\right) \\
&= \frac{U-L}{6}\sqrt{n}\left(\frac{1}{\sqrt{S_n^2+(\bar{X}_n-T)^2}} - \frac{1}{\sqrt{\sigma^2+(\mu-T)^2}}\right) \\
&= \frac{U-L}{6}\sqrt{n}\frac{\sqrt{\sigma^2+(\mu-T)^2}-\sqrt{S_n^2+(\bar{X}_n-T)^2}}{\sqrt{S_n^2+(\bar{X}_n-T)^2}\sqrt{\sigma^2+(\mu-T)^2}}
\end{aligned}$$

$$= \frac{U-L}{6} \times \frac{-\sqrt{n}(S_n^2-\sigma^2)+\sqrt{n}(2T-\bar{X}_n-\mu)(\bar{X}_n-\mu)}{\sqrt{S_n^2+(\bar{X}_n-T)^2}\sqrt{\sigma^2+(\mu-T)^2}\{\sqrt{\sigma^2+(\mu-T)^2}+\sqrt{S_n^2+(\bar{X}_n-T)^2}\}}.$$

Before we proceed further, we state a result for the joint weak convergence of the first and second sample moments (Serfling, 1980). In the following, μ_k means the k^{th} central moment of X and $\xrightarrow{d}$ stands for the convergence in distribution. Also BN implies the bivariate normal distribution.

LEMMA 2.1. *If μ_4 exists, then*

$$\sqrt{n}\left(\bar{X}_n-\mu, S_n^2-\sigma^2\right) \xrightarrow{d} (Y,Q),$$

where $(Y,Q) \sim BN\left((0,0), \begin{pmatrix} \sigma^2 & \mu_3 \\ \mu_3 & (\mu_4-\sigma^4) \end{pmatrix}\right).$

Thus from Lemma 2.1, we see that with applying the Slutsky's theorem (Bickel and Duksum, 1977),

$$\sqrt{n}\left(\hat{C}_{pmk}-C_{pmk}\right) \xrightarrow{d} -\frac{(U-L)(\mu-T)Y}{6\{\sqrt{\sigma^2+(\mu-T)^2}\}^3} - \frac{(U-L)Q}{12\{\sqrt{\sigma^2+(\mu-T)^2}\}^3},$$

whose distribution is normal with mean 0 and variance σ_1^2, where

$$\sigma_1^2 = \frac{(U-L)^2}{36\{\sigma^2+(\mu-T)^2\}^3}\left\{(\mu-T)^2\sigma^2+\frac{1}{4}(\mu_4-\sigma^2)+(\mu-T)\mu_3\right\},$$

which is already obtained by Chan *et al.* (1990) but the symbol of square should be added in the numerator of their paper.

For the case ii) $\mu < M$, first of all, we note the following decomposition.

$$\sqrt{n}\left(\hat{C}_{pmk}-C_{pmk}\right) = \sqrt{n}\left(\frac{\bar{X}_n-L}{3\sqrt{S_n^2+(\bar{X}_n-T)^2}} - \frac{\mu-T}{3\sqrt{\sigma^2+(\mu-T)^2}}\right)$$

$$= \frac{\sqrt{n}}{3}\left(\frac{\bar{X}_n}{\sqrt{S_n^2+(\bar{X}_n-T)^2}} - \frac{\mu}{\sqrt{\sigma^2+(\mu-T)^2}}\right) \quad (2.1)$$

$$-\frac{L}{3}\sqrt{n}\left(\frac{1}{\sqrt{S_n^2+(\bar{X}_n-T)^2}} - \frac{1}{\sqrt{\sigma^2+(\mu-T)^2}}\right). \quad (2.2)$$

Then for (2.2), using the same arguments used for the case i), we obtain that

$$-\frac{L}{3}\sqrt{n}\left(\frac{1}{\sqrt{S_n^2+(\bar{X}_n-T)^2}}-\frac{1}{\sqrt{\sigma^2+(\mu-T)^2}}\right)$$
$$\xrightarrow{d} \frac{L(\mu-T)Y}{3\{\sqrt{\sigma^2+(\mu-T)^2}\}^3}+\frac{LQ}{6\{\sqrt{\sigma^2+(\mu-T)^2}\}^3}. \tag{2.3}$$

For (2.1), we note that

$$\frac{\sqrt{n}}{3}\left(\frac{\bar{X}_n}{\sqrt{S_n^2+(\bar{X}_n-T)^2}}-\frac{\mu}{\sqrt{\sigma^2+(\mu-T)^2}}\right)$$
$$=\frac{\sqrt{n}}{3}\left(\frac{\bar{X}_n-\mu}{\sqrt{S_n^2+(\bar{X}_n-T)^2}}+\frac{\mu}{\sqrt{S_n^2+(\bar{X}_n-T)^2}}-\frac{\mu}{\sqrt{\sigma^2+(\mu-T)^2}}\right)$$
$$=\frac{\sqrt{n}(\bar{X}_n-\mu)}{3\sqrt{S_n^2+(\bar{X}_n-T)^2}}+\frac{\mu}{3}\sqrt{n}\left(\frac{\mu}{\sqrt{S_n^2+(\bar{X}_n-T)^2}}-\frac{\mu}{\sqrt{\sigma^2+(\mu-T)^2}}\right).$$

Then again using the same arguments used for the case i), we have that

$$\frac{\sqrt{n}}{3}\left(\frac{\bar{X}_n}{\sqrt{S_n^2+(\bar{X}_n-T)^2}}-\frac{\mu}{\sqrt{\sigma^2+(\mu-T)^2}}\right)$$
$$\xrightarrow{d} -\frac{\{T(\mu-T)-\sigma^2\}Y}{3\{\sqrt{\sigma^2+(\mu-T)^2}\}^3}-\frac{\mu Q}{6\{\sqrt{\sigma^2+(\mu-T)^2}\}^3}. \tag{2.4}$$

Therefore, combining (2.3) and (2.4), we see that

$$\sqrt{n}\left(\hat{C}_{pmk}-C_{pmk}\right)\xrightarrow{d} -\frac{\{(T-L)(\mu-T)-\sigma^2\}Y}{3\{\sqrt{\sigma^2+(\mu-T)^2}\}^3}-\frac{(\mu-L)Q}{6\{\sqrt{\sigma^2+(\mu-T)^2}\}^3},$$

whose distribution is normal with mean 0 and variance σ_2^2, where

$$\sigma_2^2=\frac{\{(T-L)(\mu-T)-\sigma^2\}^2\sigma^2}{9\{\sigma^2+(\mu-T)^2\}^3}+\frac{(\mu-L)^2(\mu_4-\sigma^4)}{36\{\sigma^2+(\mu-T)^2\}^3}$$
$$+\frac{\{(T-L)(\mu-T)-\sigma^2\}(\mu-L)\mu_3}{9\{\sigma^2+(\mu-T)^2\}^3}.$$

Finally for the case iii), since

$$\frac{U}{3}\sqrt{n}\left(\frac{1}{\sqrt{S_n^2+(\bar{X}_n-T)^2}}-\frac{1}{\sqrt{\sigma^2+(\mu-T)^2}}\right)$$
$$\xrightarrow{d} -\frac{U(\mu-T)Y}{3\{\sqrt{\sigma^2+(\mu-T)^2}\}^3}-\frac{UQ}{6\{\sqrt{\sigma^2+(\mu-T)^2}\}^3},$$

we have,

$$\sqrt{n}\left(\hat{C}_{pmk} - C_{pmk}\right) \xrightarrow{d} -\frac{\{(U-T)(\mu-T)-\sigma^2\}Y}{3\{\sqrt{\sigma^2+(\mu-T)^2}\}^3} - \frac{(U-\mu)Q}{6\{\sqrt{\sigma^2+(\mu-T)^2}\}^3},$$

whose distribution is normal with mean 0 and variance σ_3^2, where

$$\sigma_3^2 = \frac{\{(U-T)(\mu-T)-\sigma^2\}^2\sigma^2}{9\{\sigma^2+(\mu-T)^2\}^3} + \frac{(U-\mu)^2(\mu_4-\sigma^4)}{36\{\sigma^2+(\mu-T)^2\}^3} + \frac{\{(U-T)(\mu-T)-\sigma^2\}(U-\mu)\mu_3}{9\{\sigma^2+(\mu-T)^2\}^3}.$$

3. Concluding Remarks

Chan *et al.* (1990) considered the following form for the definition of C_{pk} when the target value T is the mid-point of the interval $[L, U]$:

$$C_{pk} = (1-K)C_p = (1-K)\frac{U-L}{6\sigma},$$

where $K = (2|T-\mu|)/(U-L)$. Also we note that when $\mu = T$, we have

$$C_{pk} = \frac{U-L}{6\sigma},$$

since $K = 0$ for $\mu = T$ which is just the definition of C_p. In spite of that fact, Chan *et al.* (1990) considered the limiting distribution for $\sqrt{n}(\hat{C}_{pk} - C_{pk})$ with the estimate, $\hat{C}_{pk}$ containing the expression, $|T - \bar{X}_n|$. As a result, in case that $\mu = T$, they came to a fallacious conclusion as in Chan and Hsu (1995). Therefore, one should use the result of C_p when $\mu = T$.

Chan and Hsu (1995) also considered the asymptotic unbiasedness property for $\hat{C}_{pmk}$. The asymptotic unbiasedness may be a little more stronger concept than the consistency in the sense that only the existence of the second moment is required for the asymptotic unbiasedness whereas the consistency requires the existence of the fourth moment in this specific case. However since we have used the finite fourth moment in proving the asymptotic normality, it is enough to deal with the consistency for the actual statistical consideration.

For the statistical inference such as confidence interval and hypothesis testing for C_{pmk} using $\hat{C}_{pmk}$, one has to obtain the null distribution for $\hat{C}_{pmk}$. However the assumption that the underlying distribution F is unknown prohibits one from obtaining the null distribution of $\hat{C}_{pmk}$. Or even if one knew the distribution F, it would be difficult to find exact null distribution. Therefore, it would be natural that one has to take asymptotic approach and so use the asymptotic normality for each case for the inferences. This means that the limiting variances should be consistently estimated. Then this can be achieved by substituting empirical ones for each component in the expression of the limiting variances as Chan and Hsu (1995) did. However this approach may induce the instability the estimates of the limiting variances since too many components should be estimated. In order to alleviate this unpleasant situation, one may consider applying the re-sampling methods such as bootstrap method (Efron, 1982) and/or the permutation principle (Good, 2000). This will be one of our future research topics.

REFERENCES

BICKEL, P. J. AND DOKSUM, K. A. (1977). *Mathematical Statisics: Basic Ideas and Selected Topics, Vol I*, Holden-Day, San Francisco.

CHAN, L. K., XIONG, Z. AND ZHANG, D. (1990). "On the asymptotic distribution of some process capability indices", *Communications in Statistics-Theory and Methods*, **19**, 11–18.

CHEN, S. M. AND HSU, N. F. (1995). "The asymptotic distribution of the process capability index C_{pmk}", *Communications in Statistics-Theory and Methods*, **24**, 1279–1291.

CHEN, S.-M. AND PEARN, W. L. (1997). "The asymptotic distribution of the estimated process capability index $\widetilde{C}_{pk}$", *Communications in Statistics-Theory and Methods*, **26**, 2489–2497.

EFRON, B. (1982). *The Jackknife, the Bootstrap and Other Resampling Plans*, CBMS-NSF, Regional Conference Series in Applied mathematics, SIAM, Philadelphia.

GOOD, P. (2000). *Permutation Test: A Practical Guide to Resampling Methods for Testing Hypotheses*, 2nd ed., Springer, New York.

SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics*, John Wiley & Sons, New York.