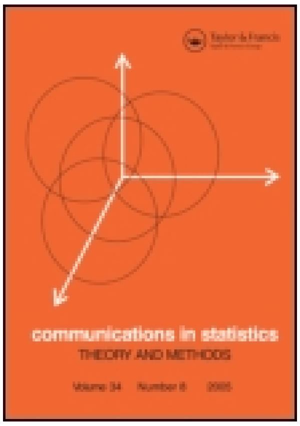

Communications in Statistics - Theory and Methods
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/lsta20

The asymptotic distribution of the estimated process capability index C_{pk}

Sy-Mien Chen [a] & W. L. Pearn [b]
[a] Department of Mathematics , Fu-Jen Catholic University , Taipei, Taiwan , ROC
[b] Department of Industrial Engineering & Management , National Chiao Tung University , Hsinchu, Taiwan, ROC
Published online: 27 Jun 2007.

To cite this article: Sy-Mien Chen & W. L. Pearn (1997) The asymptotic distribution of the estimated process capability index C_{pk} , Communications in Statistics - Theory and Methods, 26:10, 2489-2497, DOI: 10.1080/03610929708832061

To link to this article: http://dx.doi.org/10.1080/03610929708832061

COMMUN. STATIST.—THEORY METH., 26(10), 2489-2497 (1997)

THE ASYMPTOTIC DISTRIBUTION OF THE ESTIMATED PROCESS CAPABILITY INDEX $\tilde{C}_{pk}$

Sy-Mien Chen
Department of Mathematics
Fu-Jen Catholic University, Taipei, Taiwan, ROC

W. L. Pearn
Department of Industrial Engineering & Management
National Chiao Tung University, Hsinchu, Taiwan, ROC

Keywords and Phrases: process capability index; asymptotic distribution; process mean; process standard deviation.

ABSTRACT

Bissell (1990) proposed an estimator $\widehat{C'_{pk}}$ for the process capability index C_{pk} assuming that $P(\mu \geq m) = 0$, or 1, where μ is the process mean, and m is the midpoint between the upper and lower specification limits. Pearn and Chen (1996) considered a new estimator $\tilde{C}_{pk}$, which relaxes Bissell's assumption on the process mean. The evaluation of $\tilde{C}_{pk}$ only requires the knowledge of $P(\mu \geq m) = p$, where $0 \leq p \leq 1$. The new estimator $\tilde{C}_{pk}$ is unbiased, and the variance is smaller than that of Bissell's.

In this paper, we investigated the asymptotic properties of the estimator $\tilde{C}_{pk}$ under general conditions. We derived the limiting distribution of $\tilde{C}_{pk}$ for arbitrary population assuming the fourth moment exists. The asymptotic distribution provides some insight into the properties of $\tilde{C}_{pk}$ which may not be evident from its original definition.

1. INTRODUCTION

Process capability indices, whose purpose is to provide a numerical measure on whether a production process is capable of producing items satisfying the quality requirements preset by the designer, have received substantial attention in the quality control and statistical literature. The two most widely used capability indices are $C_p = \frac{USL-LSL}{6\sigma}$, and $C_{pk} = Min\{\frac{USL-\mu}{3\sigma}, \frac{\mu-LSL}{3\sigma}\}$, where USL is the upper specification limit, LSL is the lower specification limit, μ is the process mean, and σ is the process standard deviation. While the C_p index reflects only the magnitude of the process variation, the C_{pk} index takes into account the process variation as well as the location of the process mean relative to the specification limits.

For processes with two-sided specification limits, the process yield can be calculated as $F(USL) - F(LSL)$, where $F(\cdot)$ is the cumulative distribution function of the process characteristic. If the process is normal, then the process yield can be expressed as :

$$\Phi\{\frac{USL-\mu}{\sigma}\} - \Phi\{\frac{LSL-\mu}{\sigma}\},$$

where $\Phi(\cdot)$ is the cumulative function of the standard normal distribution. If the process is perfectly centered, then the process yield can be expressed alternatively as $2\Phi(3C_{pk}) - 1 = 2\Phi(3C_p) - 1$. For example, $C_{pk} = 1.00$ corresponds to process yield 99.73%. If the process is non-normal (symmetric or asymmetric), then these indices provide only approximate measures on the process performance.

For C_{pk}, if we assume $P(\mu \geq m) = 0$, or 1, where $m = (USL + LSL)/2$ is the mid-point between the upper and lower specification limits, then the Bissell (1990) estimator $\widehat{C_{pk}}$ is defined as follows: $\widehat{C_{pk}} = (USL - \overline{X}_n)/3S$ if $\mu \geq m$; otherwise, $\widehat{C_{pk}} = (\overline{X}_n - LSL)/3S$, where $\overline{X}_n = \Sigma_{i=1}^n X_i/n$ and $S = \{\Sigma_{i=1}^n(X_i -$

$\overline{X}_n)^2/(n-1)\}^{1/2}$ are conventional estimators of μ and σ, respectively, derived from a sample $X_1, X_2, \ldots, X_n$. Pearn and Chen (1996) considered a Bayesian-like estimator $\widehat{C''_{pk}}$ to relax Bissell's assumption on the process mean. The evaluation of the estimator $\widehat{C''_{pk}}$ only requires the knowledge of $P(\mu \geq m) = p$, $0 \leq p \leq 1$, which may be obtained from historical information of a stable process. If $P(\mu \geq m) = 0$, or 1, then the estimator $\widehat{C''_{pk}}$ reduces to Bissell's estimator. The estimator is defined as $\widehat{C''_{pk}} = \{d - (\overline{X}_n - m)I_A(\mu)\}/3S$, where $I_A(\mu) = 1$ if $\mu \in A, I_A(\mu) = -1$ if $\mu \notin A$, $A = \{\mu | \mu \geq m\}$ and $d = (USL - LSL)/2$.

Pearn and Chen (1996) showed that by multiplying the well-known correction factor b_f to the estimator $\widehat{C''_{pk}}$, where $b_f = \{2/(n-1)\}^{1/2}\Gamma((n-1)/2)\{\Gamma((n-2)/2)\}^{-1}$, an unbiased estimator $\tilde{C}_{pk} = b_f\widehat{C''_{pk}}$ can be obtained. Pearn and Chen (1996) also showed that on the assumption of normality, the distribution of the estimator $3(n)^{1/2}\widehat{C''_{pk}}$ is $t_{n-1}(\delta)$, a non-central t with $n-1$ degrees of freedom and non-centrality parameter $\delta = 3(n)^{1/2}C_{pk}$. In this paper, we investigate the asymptotic distribution of $\tilde{C}_{pk}$ under general conditions. We derived the limiting distribution of $\tilde{C}_{pk}$ for arbitrary population assuming the fourth central moment exists. The asymptotic distribution provides some insight into the properties of $\tilde{C}_{pk}$ which may not be apparent from its original definition. Consequently, some approximate statistical testing on whether a process is capable can be performed.

2. ASYMPTOTIC DISTRIBUTION OF $\tilde{C}_{pk}$

Let $X_1, \ldots, X_n$ be a random sample of measurements from a process which has distribution G with mean μ and positive variance σ^2. Let $\widehat{\mu} = \frac{\sum_{i=1}^n X_i}{n} = \overline{X}_n$, and $\widehat{\sigma^2} = \frac{\sum_{i=1}^n (X_i - \overline{X}_n)^2}{n-1} = S_{n-1}^2$, be the sample mean and sample variance, respectively.

Define $d = \frac{USL-LSL}{2}$, and $m = \frac{USL+LSL}{2}$. Then, the process capability index C_{pk} is defined as

$$C_{pk} = min\{\frac{USL-\mu}{3\sigma}, \frac{\mu-LSL}{3\sigma}\} = \begin{cases} [1 - \frac{\mu-m}{d}] \cdot \frac{d}{3\sigma} & \text{if } \mu \geq m, \\ [1 + \frac{\mu-m}{d}] \cdot \frac{d}{3\sigma} & \text{if } \mu < m. \end{cases}$$

When both the mean μ and the variance σ^2 of the measurements are unknown, an estimator proposed by Pearn and Chen (1996) is given by:

$$\tilde{C}_{pk} = b_f\widehat{C_{pk}} = b_f[\frac{d - (\overline{X}_n - m) \cdot I_A(\mu)}{3S_{n-1}}] = \begin{cases} [1 - \frac{\overline{X}_n - m}{d}] \cdot \frac{d}{3S_{n-1}/b_f} & \text{if } \mu \geq m, \\ [1 + \frac{\overline{X}_n - m}{d}] \cdot \frac{d}{3S_{n-1}/b_f} & \text{if } \mu < m. \end{cases}$$

Where $A = \{\mu | \mu \geq m\}$, and $b_f = \sqrt{\frac{2}{n-1}}\Gamma(\frac{n-1}{2})\Gamma(\frac{n-2}{2})^{-1}$.

Lemma 1: Define $\mu_k = E(X - \mu)^k$ as the k-th central moment of G. If μ_4 exists, then as $n \to \infty$,

$$\sqrt{n}(\overline{X}_n - \mu, S_{n-1}^2 - \sigma^2) \xrightarrow{L} N((0,0), \Sigma),$$

where

$$\Sigma = \left(\begin{bmatrix} \sigma^2 & \mu_3 \\ \mu_3 & \mu_4 - \sigma^4 \end{bmatrix} \right).$$

[Proof]: See page 72 of Serfling (1980).

Lemma 2: Let $\{\mathbf{u}(n)\}$ be a sequence of m-component random vectors and $\mathbf{b}$ a fixed vector such that $\sqrt{n}[\mathbf{u}(n) - \mathbf{b}]$ has a limiting distribution $\mathbf{N}(\mathbf{0}, \mathbf{T})$ as $n \to \infty$. Let $\mathbf{f}(\mathbf{u})$ be a vector-valued function of $\mathbf{u}$ such that each component $f_j(\mathbf{u})$ has a nonzero differential at $\mathbf{u} = \mathbf{b}$, and let $\frac{\partial f_j(\mathbf{u})}{\partial u_i}|_{\mathbf{u}=\mathbf{b}}$ be the (i, j)-th component of $\mathbf{\Phi_b}$. Then, $\sqrt{n}\{\mathbf{f}(\mathbf{u}(n)) - \mathbf{f}(\mathbf{b})\}$ has a limiting distribution $\mathbf{N}(\mathbf{0}, \mathbf{\Phi_b'}\mathbf{T}\mathbf{\Phi_b})$.

[Proof]: See Theorem 4.2.3 in T.W. Anderson (1984).

Theorem 1: If μ_4 exists, and $LSL \leq \mu \leq USL$, then
$\sqrt{n}[\tilde{C}_{pk} - b_f C_{pk})] \xrightarrow{L} N(0, \sigma^2_{pk1}) \cdot P(\mu \geq m) + N(0, \sigma^2_{pk2}) \cdot P(\mu < m)$.
Where
$\sigma^2_{pk1} = \frac{b_f^2}{9}\{1 + \frac{d}{2\sigma^4}(1 - \frac{\mu-m}{d})[2\mu_3 + \frac{d}{2}(\frac{\mu_3}{\sigma^2} - \sigma^2)(1 - \frac{\mu-m}{d})]\}$,
$\sigma^2_{pk2} = \frac{b_f^2}{9}\{1 - \frac{d}{2\sigma^4}(1 + \frac{\mu-m}{d})[2\mu_3 + \frac{d}{2}(\frac{\mu_3}{\sigma^2} - \sigma^2)(1 + \frac{\mu-m}{d})]\}$.
[Proof]:
For $u, v \in (LSL, USL)$, we define a function of u, v as follows:

$$g(u,v) = \begin{cases} [1 - \frac{u-m}{d}] \cdot \frac{d}{3\sqrt{v/b_f^2}} & \text{, if } \mu \geq m, \\ [1 + \frac{u-m}{d}] \cdot \frac{d}{3\sqrt{v/b_f^2}} & \text{, if } \mu < m. \end{cases}$$

Since
$$C_{pk} = min\{\frac{USL-\mu}{3\sigma}, \frac{\mu-LSL}{3\sigma}\} = \begin{cases} [1 - \frac{\mu-m}{d}] \cdot \frac{d}{3\sigma} & \text{if } \mu \geq m, \\ [1 + \frac{\mu-m}{d}] \cdot \frac{d}{3\sigma} & \text{if } \mu < m, \end{cases}$$
then $C_{pk} = g(\mu, \sigma^2)/b_f$.
$$\text{Also, } b_f\widehat{C_{pk}} = b_f[\frac{d-(\overline{X}_n - m)\cdot I_A(\mu)}{3S_{n-1}}] = \begin{cases} [1 - \frac{\overline{X}_n - m}{d}] \cdot \frac{d}{3S_{n-1}/b_f} & \text{if } \mu \geq m, \\ [1 + \frac{\overline{X}_n - m}{d}] \cdot \frac{d}{3S_{n-1}/b_f} & \text{if } \mu < m, \end{cases}$$

which implies that $b_f\widehat{C_{pk}} = g(\overline{X}_n, S^2_{n-1})$.
Hence, $\sqrt{n}[\tilde{C}_{pk} - b_f C_{pk})] = \sqrt{n}[b_f(\widehat{C_{pk}} - C_{pk})] = \sqrt{n}[g(\overline{X}_n, S^2_{n-1}) - g(\mu, \sigma^2)]$.

Case 1: When $\mu \geq m$, $u \geq m$.

Since
$$g(u,v) = [1 - \frac{u-m}{d}] \cdot \frac{d}{3\sqrt{v/b_f^2}} = [1 - \frac{u-m}{d}] \cdot \frac{d \cdot b_f}{3\sqrt{v}},$$
is a real valued function and is differentiable for all $u \in (LSL, USL)$, $v > 0$, with

$$\frac{\partial g}{\partial u}(u, v) = -\frac{b_f}{3\sqrt{v}},$$

and

$$\frac{\partial g}{\partial v}(u, v) = -\frac{d \cdot b_f}{6v^{\frac{3}{2}}}[1 - \frac{(u-m)}{d}].$$

Define

$$D_1 = (\frac{\partial g}{\partial u}|_{\mu,\sigma^2}, \frac{\partial g}{\partial v}|_{\mu,\sigma^2}),$$

then $D_1 \neq (0, 0)$.

Hence, by Lemma 1 and Lemma 2,

$$\sqrt{n}[\tilde{C}_{pk} - b_f C_{pk})] = \sqrt{n}[b_f(\widehat{C_{pk}} - C_{pk})] = \sqrt{n}[g(\overline{X}_n, S^2_{n-1}) - g(\mu, \sigma^2)]$$

$$\xrightarrow{L} N(0, \sigma^2_{pk1}),$$

where

$$\sigma^2_{pk1} = D_1 \Sigma D_1' = \frac{b_f^2}{9}\{1 + \frac{d}{2\sigma^4}(1 - \frac{\mu-m}{d})[2\mu_3 + \frac{d}{2}(\frac{\mu_3}{\sigma^2} - \sigma^2)(1 - \frac{\mu-m}{d})]\}$$

Case 2: When $\mu < m$, $u < m$.

Since

$$g(u, v) = [1 + \frac{u-m}{d}] \cdot \frac{d}{3\sqrt{v/b_f^2}} = [1 + \frac{u-m}{d}] \cdot \frac{d \cdot b_f}{3\sqrt{v}},$$

is a real valued function and is differentiable for all $u \in (LSL, USL)$, $v > 0$,

with

$$\frac{\partial g}{\partial u}(u, v) = \frac{b_f}{3\sqrt{v}},$$

and

$$\frac{\partial g}{\partial v}(u, v) = -\frac{d \cdot b_f}{6v^{\frac{3}{2}}}[1 + \frac{(u-m)}{d}].$$

Define

$$D_2 = (\frac{\partial g}{\partial u}|_{\mu,\sigma^2}, \frac{\partial g}{\partial v}|_{\mu,\sigma^2}),$$

then $D_2 \neq (0, 0)$.

Hence,

$$\sqrt{n}[\tilde{C}_{pk} - b_f C_{pk})] = \sqrt{n}[b_f(\widehat{C_{pk}^{r}} - C_{pk})] = \sqrt{n}[g(\overline{X}_n, S_{n-1}^2) - g(\mu, \sigma^2)]$$

$$\xrightarrow{L} N(0, \sigma_{pk2}^2),$$

where

$$\sigma_{pk2}^2 = D_2 \Sigma D_2' = \frac{b_f^2}{9}\{1 - \frac{d}{2\sigma^4}(1 + \frac{\mu - m}{d})[2\mu_3 + \frac{d}{2}(\frac{\mu_3}{\sigma^2} - \sigma^2)(1 + \frac{\mu - m}{d})]\}.$$

Since $P(\sqrt{n}[b_f(\widehat{C_{pk}^{r}} - C_{pk})] \leq t) =$

$$P(\sqrt{n}[b_f(\widehat{C_{pk}^{r}} - C_{pk})] \leq t | \mu \geq m) \cdot P(\mu \geq m) +$$
$$P(\sqrt{n}[b_f(\widehat{C_{pk}^{r}} - C_{pk})] \leq t | \mu < m) \cdot P(\mu < m),$$

then

$$\sqrt{n}[b_f(\widehat{C_{pk}^{r}} - C_{pk})] \xrightarrow{L} N(0, \sigma_{pk1}^2) \cdot P(\mu \geq m) + N(0, \sigma_{pk2}^2) \cdot P(\mu < m).$$

Where

$$\sigma_{pk1}^2 = \frac{b_f^2}{9}\{1 + \frac{d}{2\sigma^4}(1 - \frac{\mu - m}{d})[2\mu_3 + \frac{d}{2}(\frac{\mu_3}{\sigma^2} - \sigma^2)(1 - \frac{\mu - m}{d})]\},$$
$$\sigma_{pk2}^2 = \frac{b_f^2}{9}\{1 - \frac{d}{2\sigma^4}(1 + \frac{\mu - m}{d})[2\mu_3 + \frac{d}{2}(\frac{\mu_3}{\sigma^2} - \sigma^2)(1 + \frac{\mu - m}{d})]\}.$$

Thus, the asymptotic distribution of $\sqrt{n}[\tilde{C}_{pk} - b_f C_{pk})]$ is a mixture of normal distributions provided that the weights $P(\mu \geq m)$ or $P(\mu < m)$ is known. ◫

<u>**Theorem 2:**</u> If μ_4 exists, and $\mu \in (LSL, USL)$, then $\sqrt{n}[\tilde{C}_{pk} - C_{pk}] \xrightarrow{L} N(0, \sigma_{pk1}^2) \cdot P(\mu \geq m) + N(0, \sigma_{pk2}^2) \cdot P(\mu < m)$.

Where σ_{pk1}^2 and σ_{pk1}^2 are defined as in Theorem 1.

[Proof]: Since $\sqrt{n}[\tilde{C}_{pk} - C_{pk}] = \sqrt{n}[b_f \widehat{C_{pk}^{r}} - C_{pk}] = \sqrt{n}[b_f(\widehat{C_{pk}^{r}} - C_{pk})] + \sqrt{n}[b_f \cdot C_{pk} - C_{pk})] = \sqrt{n}[b_f(\widehat{C_{pk}^{r}} - C_{pk})] + \sqrt{n}(b_f - 1) \cdot [C_{pk} - C_{pk}]$, and $b_f \overset{n \to \infty}{\to} 1$,

then the result follows from Theorem 1 and Slutsky's Theorem (Loève (1978)).

3. CONCLUSIONS

Bissell (1990) proposed an estimator $\widehat{C_{pk}}$ to calculate C_{pk} value assuming that $P(\mu \geq m) = 0$ or 1. Pearn and Chen (1996) considered a new estimator $\tilde{C}_{pk}$, which relaxes Bissell's assumption on the process mean. The evalution of $\tilde{C}_{pk}$ only requires the knowledge of $P(\mu \geq m) = p$, where $0 \leq p \leq 1$. The new estimator $\tilde{C}_{pk}$ is unbiased, and the variance is smaller than that of Bissell's estimator. In this paper, we investigated the asymptotic properties of the estimator $\tilde{C}_{pk}$. We showed that the limiting distribution of $\tilde{C}_{pk}$ for arbitrary population is a contamination of normal distributions provided that the fourth moment exists.

Thus, if the knowledge of $P(\mu \geq m) = p$, and $P(\mu < m) = 1 - p$ is given, then the asymptotic distribution of the estimator $\tilde{C}_{pk}$ is a contamination of two normal distributions. The estimator treats the process as a mixture of two manufacturing processes . Such situation occurs when the raw materials or components come from two different suppliers, or the machines, equipments have two different conditions, or there are two different groups of workmanship involved in the process.

ACKNOWLEDGMENTS

The authors are grateful to the editor's and referees' suggestions that improve this presentation.

BIBLIOGRAPHY

Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*, second edition. John Wiley, New York.

Bissell, A. F. (1990). "How Reliable is Your Capability Index?," *Applied Statistics*, 39, 331-340.

Chan, L., Xiong, Z. and Zhang, D. (1990). "On the Asymptotic Distributions of Some Process Capability Indices," *Communication in Statistics- Theory and Methods*, 19(1), 11-18.

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

Loève, M. (1978). *Probability Theory II*, 4th edition, Springer-Verlag, Berlin and New York.

Pearn, W. L., Kotz, S. and Johnson, N. L. (1992). "Distribution and Inferential Properties of Process Capability Indices," *Journal of Quality Technology*, 24, 216-231.

Pearn, W. L. and Chen, K. S. (1996). "A Bayesian-Like Estimator Of C_{pk}," *Communication in Statistics - Simulations and Computation*, 25(2), 321-329.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*, John Wiley and Sons, New York, 1-125.

Received October, 1996; Revised May, 1997.