A Reliable Procedure on Performance Evaluation - A Large Sample Approach Based on the Estimated Taguchi Capability Index

Gu-Hong Lin
Professor, Department of Industrial Engineering and Management, National Kaohsiung University of Applied Sciences, Kaohsiung, Taiwan

Abstract—the Taguchi capability index $C_{pm}$ has been proposed to the manufacturing industry for measuring manufacturing capability. Contributions of the estimated Taguchi capability index based on subsamples have been proposed and arrested substantial research attention. In this paper, investigations based on the proposed estimator are considered under general conditions having fourth central moment exists. The limiting distribution of the considered estimator is derived. A reliable inferential procedure based on large samples is proposed. A demonstrate example is also provided to illustrate how the proposed approach may be applied for judging whether the process runs under the desirable quality requirement.

Key words: asymptotic, capability, Taguchi.

I. INTRODUCTION

Process capability indices, whose purpose is to provide numerical measures on whether or not a manufacturing process is capable of reproducing items satisfying the quality requirements preset by the customers or the product designers, have received substantial research attention in the quality control and statistical literature. The three basic capability indices $C_p$, $C_a$, and $C_{pk}$ have been defined as (e.g., [1]-[4])

$$C_p = \frac{USL - LSL}{6\sigma}, \quad C_a = 1 - \frac{|\mu - m|}{d},$$

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

where $USL$ and $LSL$ are the upper and lower specification limits preset by the customers, the product designers, $\mu$ is the process mean, $\sigma$ is the process standard deviation, $m = (USL + LSL)/2$ and $d = (USL - LSL)/2$ are the mid-point and half length of the specification interval, respectively. The index $C_p$ reflects only the magnitude of the process variation relative to the specification tolerance and, therefore, is used to measure process potential. The index $C_a$ measures the degree of process centering (the ability to cluster around the center) and is referred as the process accuracy index. The index $C_{pk}$ takes into account process variation as well as the location of the process mean. The natural estimators of $C_p$, $C_a$, and $C_{pk}$ can be obtained by substituting the sample mean $\bar{x} = \sum_{i=1}^{n} x_i/n$ for $\mu$ and the sample variance $s^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2 / (n - 1)$ for $\sigma^2$ in (1) and (2). The statistical properties and the sampling distributions of the natural estimators of $C_p$, $C_a$, and $C_{pk}$ have been widely investigated in literature (e.g., [2]-[8]).

The capability index $C_{pk}$ is a yield-based index [9]. However, the design of $C_{pk}$ is independent of the target value $T$, which can fail to account for process targeting (the ability to cluster around the target). For this reason, the index $C_{pm}$ was independently introduced (e.g., [10]-[11]) to take the process targeting issue into consideration. The index $C_{pm}$ is defined as

$$C_{pm} = \frac{USL - LSL}{6\sigma (\sigma^2 + (\mu - T)^2)}.$$

We note that the index $C_{pm}$ is not originally designed to provide an exact measure on the number of non-conforming items. But, $C_{pm}$ includes the process departure $(\mu - T)^2$ (rather than $6\sigma$ alone) in the denominator of the definition to reflect the degree of process targeting. Some $C_{pm}$ values commonly used as quality requirements in most industry applications are 1.00, 1.33, 1.50, 1.67, and 2.00. A process is called “inadequate” if $C_{pm} < 1.00$, called “capable” if $1.00 \leq C_{pm} < 1.33$, called “marginally capable” if $1.33 \leq C_{pm} < 1.50$, called “satisfactory” if $1.50 \leq C_{pm} < 1.67$, called “excellent” if $1.67 \leq C_{pm} < 2.00$, and is called “super” if $C_{pm} \geq 2.00$. The above six quality requirements and the corresponding $C_{pm}$ values are displayed in Table 1.

<table>
<thead>
<tr>
<th>Quality Condition</th>
<th>$C_{pm}$ Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inadequate</td>
<td>$C_{pm} &lt; 1.00$</td>
</tr>
<tr>
<td>Capable</td>
<td>$1.00 \leq C_{pm} &lt; 1.33$</td>
</tr>
<tr>
<td>Marginally Capable</td>
<td>$1.33 \leq C_{pm} &lt; 1.50$</td>
</tr>
<tr>
<td>Satisfactory</td>
<td>$1.50 \leq C_{pm} &lt; 1.67$</td>
</tr>
<tr>
<td>Excellent</td>
<td>$1.67 \leq C_{pm} &lt; 2.00$</td>
</tr>
<tr>
<td>Super</td>
<td>$C_{pm} \geq 2.00$</td>
</tr>
</tbody>
</table>

II. ESTIMATING $C_{pm}$ UNDER THE NORMALITY ASSUMPTION

A. The Estimated $C_{pm}$ Based on Single Sample

Assuming that the measurements of the characteristic

359
investigated, \((X_1, X_2, \ldots, X_n)\), are chosen randomly from a stable process which follows a normal distribution \(N(\mu, \sigma^2)\). Reference [9] proposed two approximate 100(1 - \alpha)% lower confidence bounds using the normal and chi-square distributions for \(C_{pm}\) from the distribution frequency point of view. Reference [2] investigated the statistical properties of the MLE of \(C_{pm}\). Reference [13] proposed a Bayesian procedure based on the MLE of \(C_{pm}\) without the restriction \(\mu = T\) on the process mean \(\mu\). Their results generalized those discussed in Reference [11].

**B. The Estimated \(C_{pm}\) Based on \(m\) Subsamples**

In real-world practice, process information is often derived from sub-samples rather than from one single sample. A common practice of the process capability estimation in the manufacturing industry is to first implement a daily-based data collection program for monitoring the process stability, then to analyze the past “in control” data. Assuming that the measurements of the \(i\)-th production line investigated, \((X_{i1}, X_{i2}, \ldots, X_{in_i})\), are chosen randomly from a stable process which follows a normal distribution \(N(\mu_i, \sigma_i^2)\) for each \(i = 1, 2, \ldots, m\). The MLE of \(C_{pm}\) based on \(m\) subsamples of size \(n_i\) each:

\[
\hat{C}_{pm} = \frac{USL - LSL}{6\sqrt{s^2 + (\bar{X} - T)^2}} = \frac{d}{3\sqrt{\sum s^2_i}}
\]

where \(d = (USL - LSL)/2\) is half the length of the specification interval. As noted by Reference [12], the term \(s^2 + (\bar{X} - T)^2\) in the denominator of (4) is the UMVUE (uniformly minimum variance unbiased estimator) of the term \(\sigma^2 + (\mu - T)^2\) in the denominator of \(C_{pm}\), it is reasonable for reliability purpose. Under the assumption of normality, the exact cdf \(F_m(x)\) and pdf \(f_m(x)\) of \(\hat{C}_{pm}\) can be expressed in terms of the chi-square distribution and the normal standard distribution [12]:

\[
F_m(x) = 1 - \int_0^x U^{*}\{[(b^2/3n)/9x^2] - t^2\} \cdot [\phi(t + \xi \sqrt{N}) + \phi(t - \xi \sqrt{N})] dt
\]

\[
f_m(x) = \frac{b\sqrt{N}/(3x)}{u^{*}\{[(b^2/3n)/9x^2] - t^2\} \cdot \{[(2b^2/3n)/9x^3]\} \cdot [\phi(t + \xi \sqrt{N}) + \phi(t - \xi \sqrt{N})] dt
\]

for \(x > 0\), where \(b = d/\sigma, \xi = (\mu - T)/\sigma\), \(U^{*}\) and \(u^{*}\) are the cdf and pdf of the chi-square distribution \(\chi^2\) with \(n - 1\) degrees of freedom respectively, and \(\phi(\cdot)\) is the pdf of the standard normal distribution \(N(0, 1)\). These results generalize those discussed in Reference [11].

**III. THE LIMITING BEHAVIORS OF \(\hat{C}_{pm}\)**

**A. Asymptotic Distribution and Large Sample Properties of \(\hat{C}_{pm}\)**

In this subsection, asymptotic properties of \(\hat{C}_{pm}\) are investigated under general conditions. The limiting distribution of \(\hat{C}_{pm}\) is derived for arbitrary populations having fourth central moment \(\mu_4 = E(X - \mu)^4\) exists. Consequently, approximate manufacturing capability can be measured for processes under those described conditions, particularly, for those with near-normal distributions. Furthermore, consistent, asymptotically unbiased, and asymptotically efficient properties of \(\hat{C}_{pm}\) based under large samples are also presented. Let \(X_{ij}, X_{i2}, \ldots, X_{in_i}\) be a random sample of measurements from a process which has distribution \(G\) with mean \(\mu\) and variance \(\sigma^2\) for each \(i = 1, 2, \ldots, m\). Under the normality assumption, \((\bar{X}, s^2)\) is a MLE (maximum likelihood estimator) of \((\mu, \sigma^2)\) based on \(m\) sub-samples, where \(\bar{X} = \sum_{i=1}^m \sum_{j=1}^{n_i} x_{ij} / n\), and \(s^2 = \sum_{i=1}^m \sum_{j=1}^{n_i} (x_{ij} - \bar{X})^2 / N\). By substituting \(\bar{X}\) for \(\mu\) and \(s^2\)
Proof: From Theorem, we know that \( \sqrt{N} (C_{pm} - C_{pm}) \) {
\[ \text{d} \to N(0, \sigma_{pm}^2). \]
} Under the normality assumption, \( \mu_1 = 0 \) and \( \mu_2 = 3\sigma_1^2 \) implies that \( \sqrt{N} (C_{pm}^* - C_{pm}) \) {
\[ \text{d} \to N(0, \sigma_N^2). \]
} where \( \sigma_N^2 = (C_{pm})^2/2. \) The information matrix is \( I(\theta) = \)
\[
\begin{bmatrix}
1/\sigma_1^2 & 0 & 0 \\
0 & 1/(2\sigma_1^4) \\
0 & 1/(2\sigma_1^2) & \sigma_1^2/N
\end{bmatrix}
\]
Since the Cramer-Rao lower bound is achieved, the proof is complete.

VI. ACKNOWLEDGMENT
This research was partially supported by National Science Council of the Republic of China, under contract number NSC 95–2221–E–151–038-MY3.

REFERENCES

AUTHOR BIOGRAPHY
Gu-Hong Lin is a Professor at the Department of Industrial Engineering and Management, National Kaohsiung University of Applied Sciences, Kaohsiung, Taiwan. He received his Ph.D. degree in quality assurance from National Chiao-Tung University, Taiwan. His research interests include applied statistics and quality assurance.