Robust Multivariate Process Capability Indices

Herrera Acosta Roberto José
Universidad del Atlántico,
Gestión de la Calidad Research Group, Colombia

Mendoza Mendoza Adel Alfonso
3i+d Research Group,
Universidad del Atlántico, Colombia

Cabarcas Reyes Juan Carlos
3i+d Research Group,
Universidad del Atlántico, Colombia

Copyright © 2018 Herrera Acosta Roberto José, Mendoza Mendoza Adel and Cabarcas Reyes Juan Carlos. This article is distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

The evaluation of performance with respect to the specifications in production processes is generally measured by means of univariate capacity indicators, but this is inadequate when there are quality characteristics, adding that these characteristics have extreme values by their nature. It is suitable for this type of robust process indicators that use distances such as Canberra's that more accurately express the actual process conditions. The results obtained with the Canberra distance made the values given by the traditional indicators that used the Euclidean distance more flexible, showing that this new indicator is adjusted to processes with extreme values.

Keywords: Performance, Capability, Distances

1 Introduction

In processes where more than one quality characteristic is involved and these in turn are correlated, it is necessary to apply a multivariate indicator that allows us
to know the quality of the process and simultaneously measure the existing correlations. This research proposes a modification of K. S. Chen's proposal (2003); the Euclidean distance as a traditional element is replaced by the Canberra distance as a new expression that allows, under the assumption of multivariate normality, to be robust to the extreme data presented in some quality characteristics, as is the case, for example, of solid bulk products [5].

There are other measures of multivariate capacity, under the assumption of multivariate normality, which, despite having a sound theoretical basis, do not have the expected reception; this is the case of the proposal proposed by Hubele N. (2000), such as the one proposed by Bothe (1991) [8][1]. The first is based on tolerance zones, where a dimensional abstraction of the phenomenon is sometimes necessary; the second presents an evaluation of the joint probability measure of compliance with the specifications required in the design, for each of the quality characteristics.

2 Statistical distances

The concept of statistical distance allows the geometrical interpretation of many classical techniques, from univariate to multivalent analysis, equivalent to representing these objects as points within a metric space adequate to the conditions of the evaluated process. This interpretation of statistical distance is possible not only when quantitative variables are available, but also when the quality variables or characteristics evaluated are of a general nature, provided that it is meaningful to obtain a measure of proximity between the elements or objects. [4].

The notion of these statistical distances together with their properties constitutes an important tool, both in mathematical statistics and in data analysis; it allows to construct hypothesis contrasts, study asymptotic properties of estimators, and compare parameters. In data analysis, distance is a very intuitive concept that allows us to obtain geometric representations that are easy to understand. In the case of research work, it is a useful tool for the interpretation of the data structure.

3 Theoretical foundation and methodology

3.1 Euclidean distances

The concept of Euclidean distance is based on the following expression: either
two objects or individuals defined as \( x_i = (x_{i1}, \ldots, x_{ip}) \) \( x_j = (x_{j1}, \ldots, x_{jp}) \) \( i, j \)

results of measuring \( p \) variables \( X_1, \ldots, X_p \) on them. The Euclidean distance is defined in the following expression,

\[
\varphi_E^2(i, j) = \sum_{k=1}^{p} (x_{ik} - x_{jk})^2 = (x_i - x_j)^T (x_i - x_j)
\]

(1)

The Euclidean distance, although invariant due to orthogonal transformations, has the disadvantage of being sensitive to changes in the scale of the variables. It is therefore recommended to apply it in phenomena characterized by their homogeneity in the variables and the lack of knowledge of the matrix of covariance variances.

### 3.2 Canberra distances

It should be noted that the Euclidean type distances, has a disadvantage because the attributes with high values, extreme values, have a great weight in the value of affinity, while the low values, practically have no importance. On the other hand, the distance of Canberra, which is a variant of the Minkowski distance subsequently applied by Lance and Williams (1996), allows overcoming the previous disadvantages, defining its metric by the following expression. [1]

\[
\varphi_C^2(i, j) = \sum_{k=1}^{p} \frac{|x_{ik} - x_{jk}|}{|x_{ik}| + |x_{jk}|}
\]

(2)

The classical sample deviation is transformed using the Canberra metric to obtain the mathematical expression,

\[
\varphi_G = \frac{\sum_{k=1}^{p} |x_{ik} - x_{jk}|}{n} \frac{|x_{ik}| + |x_{jk}|}{n-1}
\]

(3)

### 4 Multivariate Capability Indices

The calculation of the capability index for processes with \( p \) variables, Bothe, D. (1991) presents a simple way: the minimum percentage of nonconformities for each of the characteristics \( P_1, P_2, \ldots, P_p \). is selected. So in the whole process the minimum size is \( P = \min\{P_1, P_2, \ldots, P_p\} \) [4].

Krishnamoorthi (1990) proposed the \( PC_p \) and \( PC_{pk} \) indices, extensions of the \( C_p \)
and \( C_{pk} \) indice, where it is assumed that the process average \( \mu \) is equal to the target value \( T \) [1]. On the other hand Davis, Kaminsky and Saboo (1992) proposed an index \( R = U / r \), where \( U \) is the radius of the specification region, \( r \) is the standard deviation of the quality characteristic [9]. Karl, Morisette and Taam (1994) later expanded the concept of multivariate capacity initially proposed by the latter author [2]. Bothe (2006) considered the radial distance as the difference between the target or target value \( T \) and the real value, as a quality criterion to evaluate the capacity of a process [3].

K. S. Chen, et al. (2003) present multivariate capability index to assess \( p \) characteristics assuming normality and independence [6]. Based on the characteristics of the process, the capability index for multiple characteristics was selected from \( MC_{pk}^T \), which is obtained as shown below:

\[
MC_{pk}^T = \frac{1}{3} \Phi^{-1} \left\{ \prod_{j=1}^{p} (\Phi(3S_{pkj})-1)+1 \right\}
\]

(4)

where \( S_{pkj} \) denotes the value of the \( j \)-th characteristic for \( j: 1,2,\ldots,N \) and \( p \) is the number of characteristics, \( S_{pkj} = \min\{S_{puj};S_{pj}\} \) is the location capability indices calculated as shown by the equations

\[
S_{pkj} = \min\left\{S_{pu} = \frac{leu-\mu}{3\sigma}; S_{pl} = \frac{\mu-lei}{3\sigma}\right\}
\]

(5)


<table>
<thead>
<tr>
<th>( MC_{pk}^T )</th>
<th>Process yield</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00</td>
<td>0.997 300 204</td>
</tr>
<tr>
<td>1.24</td>
<td>0.999 800 777</td>
</tr>
<tr>
<td>1.33</td>
<td>0.999 933 927</td>
</tr>
<tr>
<td>1.50</td>
<td>0.999 993 205</td>
</tr>
<tr>
<td>1.67</td>
<td>0.999 999 456</td>
</tr>
<tr>
<td>2.00</td>
<td>0.999 999 998</td>
</tr>
</tbody>
</table>

The proposal indicates that the overall indices is \( C_{p}^T = \frac{1}{3} \Phi^{-1}\{\prod_{j=1}^{p}(P)\} \) where \( P \) is the measure of probability of compliance with the specifications, considering...
that process observations have a normal distribution $N \sim (\mu, \sigma^2)$. This probability measure is defined as $P = \{[\prod_{j=1}^{p} p(lei \leq x \leq les)_j] = \{[\prod_{j=1}^{p} p(z_l \leq Z \leq z_u)_j]\}$.

Including the distance of Canberra in the multivariate capability index implies that equation (5) is modified as follows,

$$S_{\theta_{Gj}} = \min \left\{ S_{\theta_{Gu}} = \frac{\text{les} - \mu}{3\theta_G}; S_{\theta_{Gl}} = \frac{\mu - \text{lei}}{3\theta_G} \right\}$$

(6)

### 5 Simulation and Results

A simulation is performed generating random numbers of three quality characteristics, for each of the two variables for the first variable the conditions are: $\text{les} = 100$, $\text{lei} = 88$ with parameters $\sigma_1 = 10$ and $\mu_1 = 92$, the second variable $\text{les} = 30$, $\text{lei} = 22.5$ with parameters $\sigma_2 = 1$ and $\mu_2 = 25$, and third variable $\text{les} = 60$, $\text{lei} = 50$ with parameters $\sigma_2 = 0.8$ and $\mu_2 = 55$; testing the assumption of multinormality, using equations (1) through (6). To perform this test, a complementary package of software R called MVN was used and the Mardia test was used to verify normality.

Table 2. Results of the multivariate indicators for a run of 35000 $MC_{p_k}^T$ for three variables.

| $\delta$ | Capability index | | |
|----------|------------------|-----------------|
|          | Multivariate modified with Canberra distance | Multivariate capability index of K. S. Chen | |
| 0.00     | 1.0339           | 0.1322          | |
| 0.05     | 0.8762           | 0.1248          | |
| 0.10     | 0.7023           | 0.1171          | |
| 0.15     | 0.1110           | 0.5836          | |
| 0.20     | 0.1076           | 0.4879          | |
| 0.25     | 0.1068           | 0.4103          | |
| 0.30     | 0.1003           | 0.3236          | |
| 0.35     | 0.0859           | 0.2524          | |
| 0.40     | 0.0871           | 0.2127          | |
| 0.45     | 0.0842           | 0.1743          | |
| 0.50     | 0.0805           | 0.1470          | |
| 0.55     | 0.0792           | 0.1191          | |

Source: Authors
Table 2, shows an appreciable difference between the indicator with the Canberra metric proposal and the multivariate indicator with the Euclidean distance proposal; for example, for a 5\% increase in variability $\delta$, the proposed indicator gave a result of 0.8762, while the classic K.S. Chen (2003) proposal indicates that it is extremely low, of course this shows that this indicator is very sensitive to extreme data, see Figure 1.

![Figure 1. Comparison of multivariate capacity indices](image)

When performing a hypothesis contrast, using ANOVA analysis of variance: the P-value is 0.007, indicating that there is a statistically significant difference between the means of the two capability indices with a 95.0\% confidence level.

### 6 Conclusions

The modified Canberra distance proved to be a better statistic for measuring the multivariate capability index, in contrast to the Euclidean distance used to calculate the variability of the data when the information is multivariate, with homogeneous values for each of the quality characteristics. This is because the Euclidean distance does not consider, among others, the association of variables; this shows an advantage by modifying the measure of variability in the present study, achieving a better indicator in processes that contain extreme values. In addition, the Canberra capacity index was unbiased with respect to information from the processes that have extreme data, so it gives robustness to the indicator. In contrast to the classic multivariate capability index of K.S. Chen (2003), with a
high degree of sensitivity to extreme values, thus giving values of capability indicators far removed from the actual process conditions.

Acknowledgements. The authors thank the Universidad del Atlántico, for its support in the preparation of this article.

References


Received: August 27, 2018; Published: September 25, 2018