European Journal of Operational Research 265 (2018) 589–597

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Stochastics and Statistics

A systematic look at the gamma process capability indices

Piao CHEN, Zhi-Sheng YE*

Department of Industrial Systems Engineering and Management, National University of Singapore, 117576, Singapore

ARTICLE INFO

Article history:
Received 23 May 2016
Accepted 11 August 2017
Available online 24 August 2017

Keywords:
Quality control
Ratio of maximized likelihoods
Probability of correct selection
Generalized confidence interval
Generalized pivotal quantity

ABSTRACT

Process capability indices (PCIs) are widely used to measure whether an in-control process conforms to manufacturing specifications. The normal distribution is assumed in most traditional applications of PCIs. Nevertheless, it is not uncommon that some quality characteristics have skewed distributions. In such cases, the gamma distribution is an appropriate model and percentile-based PCIs for the gamma process have been studied in the literature. In practical applications of PCIs, it is important to select an appropriate distribution between the normal and the gamma distributions based on historical data. In this study, we first construct a hypothesis test for model discrimination between the normal and the gamma distributions. Asymptotic distribution of the test statistic under the gamma process is derived. We then consider statistical inference for the percentile-based PCIs under the gamma process. The maximum likelihood method is used for point estimation and the method of generalized pivotal quantities is used for interval estimation. We demonstrate the proposed methods by a practical example.

1. Introduction

The process capability indices (PCIs) are widely used to assess whether an in-control process meets specification limits determined from engineering tolerances or customers' needs. They have become the common language for process quality between the customers and the suppliers (Müller & Haase, 2016). Numerous PCIs, such as the classical C_p, C_{pk}, C_{pm} and C_{pmk} (Hsu, Pearn, & Wu, 2008), have been proposed in the literature and industry. Among these PCIs, C_{pk} proposed by Boyles (1991) is probably the most frequently used due to its yield-based nature (Ryan, 2011, chap.7). To see this, assume that the process characteristic X follows a normal distribution $N(\mu, \sigma^2)$ with mean μ and standard deviation σ. Then C_{pk} is defined as

$$C_{pk} = \min\left\{\frac{USL-\mu}{3\sigma}, \frac{\mu-LSL}{3\sigma}\right\}, \tag{1}$$

where USL and LSL are the upper specification limit and the lower specification limit, respectively. For a given value of C_{pk}, the process yield is bounded between $2\Phi(3C_{pk})-1$ and $\Phi(3C_{pk})$, where $\Phi(\cdot)$ is the cumulative distribution function (CDF) of the standard normal distribution (Boyles, 1991). Generally, a lower value of C_{pk} implies a higher faction of defectives of the process. Therefore, a customer would usually specify a minimum value of C_{pk} in the purchasing contract (Wu, Aslam, & Jun, 2012).

It is necessary to emphasize that C_{pk} defined in (1) is meaningful only if the process characteristic X is normally distributed. Otherwise, the bounds of the process yield based on the value of C_{pk} could be misleading. In fact, the normality assumption is the basis for many commonly-used PCIs (e.g., Wu, 2012; Yeong, Khoo, Lee, & Rahim, 2013). Although the normal distribution seems appropriate for some manufacturing systems, it is very likely that a process characteristic has a non-normal distribution (Ryan, 2011, chap.7). For example, many quality characteristics such as diameter and roundness are often non-normal in manufacturing process (Ryan, 2011, chap.7). In the procurement process of oil and gas companies, the distribution of cycle time data is often skewed (Aldowaisan, Nourelfath, & Hassan, 2015). Also, the distribution of some chemical process such as zinc plating is found far from normal (Pyzdek, 1992). In addition, the normality assumption comes into question in many service and transaction systems (Hosseinifard, Abbasi, & Niaki, 2014).

Since the PCIs based on the normality assumption could be meaningless (Ryan, 2011), several approaches have been suggested to handle the non-normal processes. The most straightforward way is to transform the non-normal data such that the transformed data will be approximately normal (e.g., Somerville & Montgomery, 1996). However, the transformation methods have drawbacks which are inherent in their utilization. First, transformation methods are computing-extensive (Tang & Than, 1999). For example, it is necessary to find the optimal parameter in the Box-Cox power transformation. Second, the approximate normality of the transformed data cannot be always guaranteed, and hence the PCIs may not be accurate. Finally, because of the problems associated

* Corresponding author.
E-mail address: yez@nus.edu.sg (Z.-S. YE).

http://dx.doi.org/10.1016/j.ejor.2017.08.024
0377-2217/

with translating the computed results with regard to the original scales, the use of transformed data is often not appealing to practitioners (Ryan, 2011). On the other hand, a more appropriate way is to fit a distribution to the process characteristic data, and use percentiles of this distribution to modify classical PCIs (Clements, 1989; Rodriguez, 1992). For example, a percentile-based C_{pk} is defined as

$$C^*_{pk} = \min\left\{\frac{USL - X_{0.5}}{X_{0.9987} - X_{0.5}}, \frac{X_{0.5} - LSL}{X_{0.5} - X_{0.0013}}\right\}, \tag{2}$$

where X_β is the 100βth percentile of the fitted distribution; USL and LSL are the upper specification limit and the lower specification limit, respectively. Notice that C^*_{pk} and C_{pk} coincide for a normal process.

Obviously, an appropriate distribution for the process is the premise of the use of the percentile-based PCIs. Among all the possible distributions, the gamma distribution is a popular choice and it has been extensively used to fit a variety of process characteristic data (e.g., Hsu et al., 2008; Aldowaisan et al., 2015). This is because the gamma distribution belongs to the Pearson family, which usually provides a reasonable curve flexibility for the process characteristic data (Rodriguez, 1992; Ryan, 2011). The gamma distribution itself is a very flexible distribution; it involves the exponential distribution and the χ^2 distribution as special cases. In fact, the gamma distribution governs a wide class of non-normal applications in numerous disciplines. For example, it is well known that the gamma distribution is a useful lifetime model (e.g., Chen & Ye, 2017c; 2017b). In addition, the gamma distribution plays an important role in some genetic research (e.g., Agarwala, Flannick, Sunyaev, Consortium, & Altshuler, 2013). It is also extensively used in environmental science (e.g., Villarini, Seo, Serinaldi, & Krajewski, 2014; Chen & Ye, 2017a).

Given the importance of the PCIs and the wide applications of the gamma distribution in non-normal processes, we focus on the percentile-based C^*_{pk} under the gamma processes assumption in this study. Evidently, the first important task is to select the appropriate distribution between the normal and gamma distributions. In the literature, a histogram of the process data is often plotted for qualitative model discrimination (Hsu et al., 2008). This method is simple and useful when the sample size is large and the histogram is very skewed or symmetric. Otherwise, it provides limited information in selecting the appropriate distribution. Therefore, quantitative methods for discrimination between the normal and the gamma distributions are needed. In this study, we treat this discrimination problem as a hypothesis test problem. Such a treatment can be found in Dumonceaux, Antle, 1973, Kundu and Manglick (2004), and Kim and Yum (2008), to name a few. In this study, the test statistic is constructed as the logarithm of the ratio of maximized likelihoods (RML) (Cox, 1961; 1962). Under the gamma process assumption, the asymptotic distribution of the test statistic is provided, which can be used to determine the probability of correct selection easily. A simulation is conducted to assess the Type I error and Type II error of the hypothesis under the proposed decision rule.

Once the gamma distribution is selected, the next important task is to make statistical inference for the percentile-based C^*_{pk}, as the true parameters are unknown in reality. A point estimator of C^*_{pk} is generally easy to obtain by the maximum likelihood (ML) method. Nevertheless, a lower confidence limit is of more interest in practice (Chang & Wu, 2008; Kotz & Lovelace, 1998; Ryan, 2011). Because C^*_{pk} is defined as the minimum of two functions, its lower confidence limit is quite difficult to obtain even under the normal process assumption. In this study, we use the idea of generalized confidence interval (GCI) to obtain the lower confidence limit of C^*_{pk} for a gamma process. Since it was introduced in Weerahandi (1995), the method of GCI has been successfully applied in many statistical inference problems. See Krishnamoorthy and Mathew (2004) and Hannig, Iyer, and Patterson (2006), among others. Generally, accurate coverage can be guaranteed based on the method of GCI. We first construct generalized pivotal quantities (GPQs) for the gamma parameters. The GPQ for C^*_{pk} can then be naturally constructed, based on which the GCI can be readily obtained. The performance of the constructed GCI is examined by a simulation study.

The rest of this paper is organized as follows. In Section 2, a hypothesis test for discrimination between the gamma and the normal distributions is constructed. We also derive the asymptotic distribution of the test statistic under the gamma process assumption. A simulation study is then conducted to assess the Type I error and Type II error under the proposed decision rule. Section 3 introduces the inference methods for the percentile-based C^*_{pk}. A simulation study verifies the performance of GCI in constructing the lower confidence limit of C^*_{pk}. In Section 4, a practical example is provided to show the usefulness of the proposed model discrimination method and the inference method. Section 5 concludes the paper.

2. Model discrimination

Assume the process characteristic X follows either a normal distribution or a gamma distribution. Let $X_1, \ldots, X_n$ be independent and identically distributed (iid) copies of X. We are interested in discriminating these two distributions based on the observed process characteristic data $x_1, \ldots, x_n$. The normal distribution $N(\mu, \sigma^2)$ with mean μ and standard deviation $\sigma > 0$ has a probability density function (PDF) as

$$f_N(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right], \quad -\infty < x < \infty, \tag{3}$$

and the gamma distribution $GA(k, \theta)$ has a PDF as

$$f_G(x; k, \theta) = \frac{\theta^k}{\Gamma(k)}x^{k-1}e^{-\theta x}, \quad x > 0, \tag{4}$$

where $k > 0$ is the shape parameter and $\theta > 0$ is the rate parameter. It is convenient to formulate this discrimination problem as a hypothesis test problem with null hypothesis H_0 and the alternative H_1 as Dumonceaux, Antle, 1973

$$H_0 : X \sim GA(k, \theta) \quad \text{and} \quad H_1 : X \sim N(\mu, \sigma^2). \tag{5}$$

The most commonly used test statistic for such a hypothesis is Cox's statistic (Cox, 1961; 1962), which is the logarithm of the ratio of the maximum likelihoods under both the null and alternative hypotheses. The widespread application of Cox's test can be found in Dumonceaux, Antle, 1973, Kundu and Manglick (2004), and Kim and Yum (2008), to name a few. For our problem, the RML is defined as

$$\text{RML} = \frac{\max\prod_{i=1}^n f_G(X_i; k, \theta)}{\max\prod_{i=1}^n f_N(X_i; \mu, \sigma)} = \frac{\prod_{i=1}^n f_G(X_i; \hat{k}, \hat{\theta})}{\prod_{i=1}^n f_N(X_i; \hat{\mu}, \hat{\sigma})}. \tag{6}$$

Here, $(\hat{k}, \hat{\theta})$ are the ML estimators of (k, θ), given as solution to the system of equations

$$\theta = \frac{k}{\bar{X}} \quad \text{and} \quad n\log(\theta) - n\psi(k) + \sum_{i=1}^n \log X_i = 0, \tag{7}$$

with $\bar{X} = \sum_i X_i/n$ and $\psi(x) = d\log\Gamma(x)/dx$. In addition, $(\hat{\mu}, \hat{\sigma})$ are the ML estimators of (μ, σ), given by

$$\hat{\mu} = \bar{X} \quad \text{and} \quad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2. \tag{8}$$