Expert Systems with Applications 39 (2012) 10236–10243

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

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

Beta control charts for monitoring fraction data

Ângelo Márcio Oliveira Sant'Anna [a,*], Carla Schwengber ten Caten [b,1]

[a] Industrial & Systems Engineering Graduate Program, Pontifical Catholic University of Parana, Parana, Brazil
[b] Department of Industrial Engineering, Federal University of Rio Grande do Sul, Osvaldo Aranha Street, 99, Porto Alegre, Brazil

ARTICLE INFO

Keywords:
Beta distribution
Fraction data
Control charts
Statistical quality control

ABSTRACT

p-Charts and *np*-Charts are commonly used in monitoring variables of the fraction type and these charts assume that the monitored variables are binomially distributed. In this paper we propose a new control chart called Beta Charts, for monitoring fraction data (p). The Beta Chart presents the control limits based on the Beta probability distribution. It was applied for monitoring the variables in three real studies, and it was compared to the control limits with three schemes. The comparative analysis showed that: (i) Beta approximation to the Binomial distribution was more appropriate with values confined in the [0,1]-interval; and (ii) the charts proposed were more sensitive to the average run length (ARL), in both in-control and out-of-control processes monitoring. The Beta Charts outperform the control charts analyzed for monitoring fraction data.

1. Introduction

Statistical Process Control (SPC) is commonly used in monitoring and detecting shifts in the production processes. Attribute control charts are important tools found in SPC to monitor processes with discrete data. *p*-Charts and *np*-Charts are more popular for monitoring nonconforming items, developed by Shewhart in 1924. Estimates of mean and variance are calculated assuming a Binomial probability distribution with n and p parameters to the number of nonconforming items, and the control limits are calculated based on the Normal distribution approximation.

The variables of the fraction type are observations expressed in the [0,1]-interval. In this paper, the fraction lays within the [0,1]-interval representing: (i) a percentage - the ratio between two discrete numbers (e.g. the number of defective parts in a production lot); and (ii) a proportion – the ratio between two continuous numbers (e.g. the volume of alcohol in wine).

There are some rules that deal with the suppositions of symmetry and Normal distribution approximation. Schader and Schmid (1989) suggested that a Normal approximation to the Binomial distribution is satisfactory if two rules are satisfied: (i) $np\,(1-p) \geqslant 9$ and (ii) $np \geqslant 5$ when $0 < p \leqslant 0.5 \leqslant (1-p)$. Similarly, Fleiss, Levin, and Paik (2003) described an approximation as satisfactory, when the size of p is in the range ($0.3 \leqslant p \leqslant 0.7$) and n is extremely large for $np \geqslant 5$ and $n(1-p) \geqslant n$, the variance $p\ (1-p)$ remains constant, while Montgomery (2005) indicated that a Normal approximation to the Binomial distribution is satisfactory when $np \geqslant 10$ and p is in the range ($0.1 \leqslant p \leqslant 0.9$).

In many studies the *p*-Charts are used in situations where the parameter p is considered small (i.e. $p = 0.001$; 0.01; 0.05; 0.1; etc.). In these cases a Binomial distribution is quite skewed and the approximation by a Normal distribution is not satisfactory, as it allows values: negative or greater than one.

The use of alternative control charts for monitoring fraction data is not new: Quesenberry (1991) proposed a Binomial Q Chart to monitor nonconforming fraction using a nonlinear transformation for the control limits and demonstrated that it approximates the Normal distribution closer to the Binomial. Heimann (1996) presented a modification of the *p*-Chart control limits for large sample sizes ($n > 10{,}000$), noting that in this case the control limits are narrow, thus the false alarm rate increases. Schwertman and Ryan (1997) suggested modifications of the *np*-Chart control limits to fit on the Normal approximation when $p < 0.03$, while, Chen (1998) proposed an adjustment to the *p*-Chart control limits and compared them with the traditional *p*-Chart and the Binomial Q Chart using the false alarm rate. Sim and Lim (2008) adapted the attribute control charts to monitor zero-inflated data and used the Blyth-Still interval with 3-*sigma* to calculate control limits assuming that this data follows a Binomial and Poisson distribution.

Bourke (2008) compared the performance of four control charts by monitoring shifts of the nonconforming fraction in industrial processes. Thereby he noted similarities in the performance of the *Synthetic* control chart and *np*-Chart over a long time period of in-control process. With similar purposes, Aebtarm and Bouguila (2011) compared the performance with eleven control charts for monitoring defects with Poisson distribution. Hsieh, Tong, and

* Corresponding author. Address: Imaculada Conceicao Street, 1155, Curitiba 80215-901, Brazil. Tel.: +41 3271 1332.
E-mail addresses: angelo.santanna@pucpr.br (Â.M.O. Sant'Anna), tencaten@producao.ufrgs.br (C.S. ten Caten).
[1] Tel.: +55 51 3308 4423.

0957-4174/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2012.02.146

Wang (2007) applied fuzzy theory to monitor wafer defects Poisson distributed in manufacture process. while Croux, Gelper, and Mahieu (2011) applied Holt-Winters method for presence of autocorrelation in the dependent process and Lin, Chou, Wang, and Liu (2012) proposed fitting this data type using autoregressive moving average.

When the data distribution in industrial processes is asymmetric, the false alarm rate increases as the asymmetry because of the discrepancy between the shape of the data distribution and Normal distribution. Ferrell (1958) and Nelson (1979) suggested in these cases to assume that the data distribution is known and to construct control charts with exact limits, which provide desired false alarm rates.

This paper proposes a Beta control chart for monitoring variables of fraction type in industrial processes. This control chart assumes that the fraction data can be approximated by a Beta distribution and proposes new control limits based on this distribution.

The proposed Beta Chart was applied in three real studies leading to better outcomes when compared with the Shewhart, Ryan (1989) and Chen (1998) control charts for monitoring variables of fraction type. Results show that our method monitors well asymmetrically distributed data commonly found in industrial scenarios. In addition, sensitivity analyses demonstrate that our method is remarkably better than Shewhart, Ryan's (1989) and Chen's (1998) charts in both, in-control and out-of-control process monitoring.

This paper is organized as follows. Section 2 briefly presents a review on Binomial and Beta probability distributions, while Section 3 introduces the control charts surveyed, and Section 4 depicts the Beta Chart. Section 5 brings the outcomes of Beta Charts in real studies and the comparative analyses between the control charts, and Section 6 presents the sensitivity analyses. In Section 7 concluding remakes are provided.

2. Probability distribution

Let Y be a random variable that measures the number of nonconforming items (y_i) in a sample size of (n_i) independent items, i = 1, 2, ..., m. The probability of Y ($P\{Y_i = y\}$) is defined by the binomial distribution,

$$P(Y_i = y) = \binom{n}{y}\pi^y(1-\pi)^{n-y} \tag{1}$$

If the percentage of nonconforming products is measured in a data set ($\pi_i = y_i/n_i$) which defines $Y_i \sim \text{Bin}(n_i; \pi_i)$ and π_i as the nonconforming fraction. If the occurrence of this nonconforming fraction is independent and identically distributed, and based on central limit theorem, it can be assumed that the nonconforming fraction follows the Normal probability distribution for a sufficiently large n (Montgomery, 2005).

Johnson, Kotz, and Balakrishnan (1995) describes that a random variable Y Binomially distributed has the mean and variance given by, respectively,

$$E(Y) = n\pi \quad \text{and} \quad Var(Y) = n\pi(1-\pi) \tag{2}$$

The approximation to the Binomial distribution by a Beta distribution may be more appropriate because the Beta density function can present a variety of form. Thus, it is assumed that the random variable (Y_i) follows the Binomial distribution and the fraction (π_i) obtained from the variable (Y_i) for each occurrence (i = 1,2,...,m) follows a Beta probability distribution, indexed by the parameters (θ_1,θ_2), where θ_1, θ_2 > 0.

Moitra (1990) states that the two parameter Beta distribution can model a large variety of variable since its probability density function (pdf) can assume several shapes. Johnson et al. (1995) corroborates that the Beta distribution can easily approximate other statistical distributions, while Teerapabolarn (2008) describes that the Beta distribution is bounded for modeling more specific cases.

The Beta distributions family comprises all probability distributions which present a random variable Y, the pdf depends of the parameters θ_1 and θ_2, and its pdf can be written as,

$$f(y;\,\theta_1,\,\theta_2) = \frac{(\theta_1+\theta_2)}{(\theta_1)(\theta_2)}y^{\theta_1-1}(1-y)^{\theta_2-1} \tag{3}$$

$\Gamma(\theta)$ is a Gamma function assessed at point θ, i.e. with $(\theta) = \int_0^\infty y^{\theta-1}e^{-y}$, $\theta > 0$.

Johnson et al. (1995) describes that a random variable Y with Beta distribution of two parameters has a mean and variance given by, respectively,

$$E(Y) = \frac{\theta_1}{\theta_1+\theta_2} \quad \text{and} \quad Var(Y) = \frac{\theta_1\theta_2}{(\theta_1+\theta_2)^2\cdot(\theta_1+\theta_2+1)} \tag{4}$$

For the Beta to the approximation Binomial distribution, the parameters can be written as Eq. (5) below

$$\theta_1 = \pi\left[\frac{\pi(1-\pi)}{\sigma^2}-1\right] \quad \text{and} \quad \theta_2 = (1-\pi)\left[\frac{\pi(1-\pi)}{\sigma^2}-1\right] \tag{5}$$

3. Control charts

The control limits of the traditional Shewhart chart for monitoring the nonconforming fraction are determined by Eq. (6), assuming that the sample size (n) is too large that a Binomial distribution is approximately symmetrical on the mean ($\bar{p}$). This implies that this distribution can be approximated to a Normal distribution.

$$CL = p \pm w\cdot\sqrt{\frac{\bar{p}(1-\bar{p})}{n}} \tag{6}$$

where w is a constant that sets the width of the control limits corresponding to a control region $(1-\alpha)$ and a desired average run length until a false alarm (ARL_0). Usually, the value w equals 3 is used due to the Normal distribution approximation, corresponding a control region = 0.9973 and ARL_0 = 370.

Ryan (1989) proposed control limits for nonconforming fraction suggesting an improvement in the approximation to Normal distribution, which is given in Eq. (7).

$$\begin{aligned} LCL &= \bar{p} - 3\sqrt{\tfrac{\bar{p}(1-\bar{p})}{n}} + \tfrac{1{,}25}{n} \\ UCL &= \bar{p} + 3\sqrt{\tfrac{\bar{p}(1-\bar{p})}{n}} + \tfrac{1{,}15}{n} \end{aligned} \tag{7}$$

Chen (1998) extended the control limits proposed by Winterbottom (1993) based on the Cornish-Fisher asymptotic correction of the first order for the control limits of the p Chart, in Eq. (8). Hill and Davis (1968) describes that the Cornish-Fisher correction has asymptotic properties to approximation F-Snedecor and Normal distributions.

$$CL = \bar{p} \pm 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}} + \frac{4(1-2\bar{p})}{3n} \tag{8}$$

Note that the approximations proposed by Ryan (1989) and Chen (1998) to the control limits are additive, without interfering in the shape of distribution.

4. Beta control chart

The Beta control chart (Beta Chart) proposed in this paper is for monitoring variables measured in fractions (percentage or proportion), which usually follow non-Normal and asymmetric distributions. This control chart uses the Beta probability distribution to calculate control limits.