DESIGN OF CONTROL CHARTS FOR ATTRIBUTES MONITORING

Ângelo Márcio Oliveira Sant’Anna*, Carla Schwengber ten Caten†, Osiris Canciglieri Jr*

*Industrial and System Engineering Graduate Program, PUCPR,+55 41 3271 1332, Brazil, angelo.santanna@pucpr.br
†Industrial Engineering Graduate Program, UFRGS,+55 51 3308 3491, Brazil tencaten@producao.ufrgs.br

Abstract
Attributes charts are commonly used in monitoring quality characteristics of the proportion type and these charts assume that the monitored characteristics are binomially distributed. In this paper we propose a new control chart called Beta Charts, for monitoring proportion data (p). The Beta Chart presents the control limits based on the Beta probability distribution. It was applied for monitoring the quality characteristics in two real studies, and compared to the control limits present by Shewhart and Chen. 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 sensitivity to the average run length (ARL) in both in-control and out-of-control processes monitoring. The Beta Charts outperforms the control charts analyzed for monitoring this type data.

Keywords
Statistical quality control, control charts, proportion data, Beta distribution.

1 INTRODUCTION
Control Charts are commonly used in monitoring and detecting shifts in the production processes. Attribute control charts are important tools of statistical control to monitor processes with discrete data. p-Charts and np-Charts are more popular for monitoring nonconforming items, developed by W. Shewhart in 1924. Estimates of mean and variance are calculated assumption 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.

There are some rules that deal with the suppositions of symmetry and Normal distribution approximation. Is suggested [1] a Normal approximation to the Binomial distribution is satisfactory when np ≥ 10 and p is in the range (0.1 ≤ p ≤ 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; ...) and 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 proportion data is not new: [4] proposed a Binomial Q Chart to monitor nonconforming proportion using a nonlinear transformation for the control limits and demonstrated that the transformation to improve the Normal approximation to the Binomial distribution. [5] presented a modification to the p-Chart control limits for large sample sizes (n>10,000), noting that in this case the control limits are narrow, thus increasing the false alarm rate. [6] applied a power transformation (x^{0.2777}) for small nonconforming fraction (5ppm) as a
better Normal approximation to the Binomial distribution.

[7] proposed an adjustment to the p-Chart control limits and compared them with traditional p-Chart and Binomial Q Chart using the false alarm rate, while [8] compared the performance of four control charts in monitoring shifts of the nonconforming proportion in industrial processes and noted similarity in the performance of the Synthetic control chart and np-Chart over long time period of in-control process. [9] adapted the attribute control charts to monitor zero-inflated data and used the Blyth-Still interval with 3-sigma to calculate control limits, assumed that this data follows a Binomial and Poisson distribution.

When the data distribution in industrial process is asymmetric, the false alarm rate increases as the asymmetry because of the discrepancy between the shape of the data distribution and Normal distribution. In these cases to assume that the data distribution is known and to construct control charts with exact limits and which provide false alarms rate desired [10;11].

This paper proposes a Beta Control Chart aimed at monitoring proportion in industrial processes. This control chart assumption that the proportion data can be approximated by a Beta distribution and proposes new control limits based on this distribution.

The proposed Beta Chart was applied in two real studies, leading to slightly better results when compared with the CCs proposed by Shewhart, Chen [7] for monitoring proportion. Results show that our method performs well with asymmetrically distributed data commonly found in industrial scenarios. In addition, sensitivity analysis pointed our method as remarkably better than Shewhart and Chen CCs 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, while Section 4 brings the results in real studies and comparative analysis between the control charts, and Section 5 presents the sensitivity analysis. Section 6 concludes the paper.

## 2 PROBABILITY DISTRIBUTION

Let $Y$ be a random variable that measures the number of non-conforming items ($y$) in a sample size of $(n)$ independent items, $i = 1, 2, ..., m$. The probability of $Y$ [$P(Y_i = y)$] is defined by the binomial distribution,

$$P(Y = y) = \binom{n}{y} \pi^y (1 - \pi)^{n-y} \quad (1)$$

If the percentage of nonconforming products is measured in a data set ($\pi_i = y / n$) it is that, $Y_i \sim \text{Bin}(n_i ; \pi_i)$ and $\pi_i$ defines the nonconforming proportion. If the occurrence this nonconforming proportion is independent and identically distributed, and based on central limit theorem, it can be assumed that the nonconforming proportion follows the Normal probability distribution, for a sufficiently large $n$ [12].

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 proportion ($\pi_i$) obtained from the variable ($Y_i$) for each occurrence ($i = 1, 2, ..., m$) follows a Beta probability distribution, indexed by the parameters ($\theta_1, \theta_2$), where $\theta_1, \theta_2 > 0$ [13]. The Beta distribution can easily approximate other statistical distributions [13], while [14] describes a bound on the Beta distribution 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 $\theta_1$ and $\theta_2$, and its pdf can be written as,

$$f(y; \theta_1, \theta_2) = \frac{\Gamma(\theta_1 + \theta_2)}{\Gamma(\theta_1) \Gamma(\theta_2)} y^{\alpha-1} (1-y)^{\beta-1} \quad (2)$$

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

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

$$\theta_1 = \pi \left[ \frac{\pi(1-\pi)}{\sigma^2} - 1 \right] ; \quad \theta_2 = (1-\pi) \left[ \frac{\pi(1-\pi)}{\sigma^2} - 1 \right] \quad (3)$$
3 CONTROL CHARTS
The control limits of the traditional Shewhart chart for monitoring the nonconforming proportion 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.

\[
CL = \bar{p} \pm w \cdot \sqrt{\frac{p(1-p)}{n}}
\]  

(4)

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, is used the value \(w\) equals 3, due the approximation to the Normal distribution, corresponding a control region = 0.9973 and ARL\(_0\) = 370.

Chen extended the control limits proposed by [15] based on the Cornish-Fisher asymptotic correction of the first order for the control limits of the p Chart, in Eq. (5). The Cornish-Fisher correction has asymptotic properties to approximation F-Snedecor and Normal distributions [16].

\[
CL = \bar{p} + 3 \cdot \sqrt{\frac{p(1-p)}{n}} \pm \frac{4(1-2p)}{3n}
\]  

(5)

Note that the approximations proposed by Chen to the control limits are additive, without interfering in the shape of distribution.

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

\[
LCL = \bar{p} - w_1 \sqrt{s^2(\bar{p})}
\]  

\[
UCL = \bar{p} + w_2 \sqrt{s^2(\bar{p})}
\]  

(6)

where \(\bar{p}\) and \(s^2(\bar{p})\) represent the mean and variance of the proportion estimated through Eq. (2), \(w_1\) and \(w_2\) are constants that define the width of the control limits, corresponding a control region \((1 - \alpha)\) and a average run length (ARL\(_0\)) desired. The Beta Chart control limits are defined trough of cumulative distribution function (cdf) of the Beta distribution, give in \(F(y) = \Psi(\delta, \theta_1, \theta_2)\) is a continuous function for all value \(y\) and its accumulated when

\[
\Psi(\delta, \theta_1, \theta_2) = \int_{0}^{y} f(y; \theta_1, \theta_2) \, dy = 1. \quad \text{Thus, the constants values } w_1 \text{ and } w_2 \text{ of the Beta Chart control limits can be estimated using the Eq. [7].}
\]

\[
w_1 = \frac{\bar{p} - \Psi(\delta, \theta_1, \theta_2)}{\sqrt{\beta^2(\bar{p})}}
\]

\[
w_2 = \frac{\Psi(1 - \alpha, \theta_1, \theta_2) - \bar{p}}{\sqrt{\beta^2(\bar{p})}}
\]  

(7)

where \(\alpha\) e \([1 - \frac{\alpha}{2}]\) represent the percentiles of cdf of the random variable \(Y\) according to control region desired.

Although, the Beta chart proposed presents a rigorous theoretical fundament, this has the disadvantage in not being included in statistical software packages currently available. The Beta Chart proposed presents same practicality and operational simplicity as the p-Chart proposed by Shewhart, as: it has no restrictions to the sample size and frequency of sampling.

4 APPLIED STUDY
This section presents two examples of processes that proportions types QC are monitored. The examples illustrate real data sets published by the authors [3, p. 270] and [17, p.101], which will be analyzed using the control limits proposed by Shewhart, Chen and Beta Chart proposed. These examples were chosen to evaluate proportion data with small \((p = 0)\) and large \((p = 1)\) values.

Example 1 consists in data set of a manufacturing process of frozen orange juice concentrate in 30 batches of 50 packages. Example 2 consists in data set of the study of contaminated peanuts by toxic substances in 34 batches of 120 pounds (± 54 kg).

The quality characteristics monitored in this paper are: percentage of nonconforming packages \((y_1)\) and proportion of ammonia unconverted \((y_2)\). The variable \(y_1\) is estimated by ratio between the number of nonconforming packages to the total number of packages in a lot, while the variable \(y_2\)
is the ratio between the volume of nonconforming raw material and the total volume produced. Figure 1 illustrates the histograms and density distribution of variables $y_1$ and $y_2$ with density Normal overlain, when the asymmetric distribution shape of the data can be seen. Summary statistics of the data investigated are showed in Table 1.

4.1 Comparative Analysis
The charts control limits was calculated for the probability false alarms ($\alpha = 0.0027$) in the monitoring process based on Normal distribution for Shewhart and Chen CCs and based on the Beta distribution for Beta Chart proposed. Table 2 shows the charts control limits calculated using sample estimates depicted in Table 2.

Next, the CCs $y_1$ and $y_2$ are depicted in Figure 2 e 3 to compare the three control limits (i.e. Shewhart; Chen and Beta Chart). The CCs detected two points outside control limits: sample 15 and sample 23 for the variable $y_1$ [see Figure 2]. [3] suggested that these points outside identified refer the introduction of raw materials new batches and new operator in the production line, respectively.

Figure 3 shows that the Shewhart and Chen CCs extrapolate the upper limit for monitoring $y_2$, which is restricted to the [0,1]-interval, while the Beta Chart do not extrapolate this region, considering in-control process. This result corroborates with the study conducted by [17].

Finally, the CCs with approximation to Normal distribution showed negative or above 1 values, while the Beta Chart proposed with control limits based on Beta distribution presented satisfactory estimates within the [0,1]-interval. Those points the Beta Chart proposed as a more robust for monitoring proportion type QCs.

### Table 1: Statistical summary

<table>
<thead>
<tr>
<th>Variables</th>
<th>Min.</th>
<th>Max.</th>
<th>Mean</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>$y_1$</td>
<td>0.080</td>
<td>0.480</td>
<td>0.2313</td>
<td>0.1778</td>
</tr>
<tr>
<td>$y_2$</td>
<td>0.99642</td>
<td>0.99987</td>
<td>0.99896</td>
<td>0.000001</td>
</tr>
</tbody>
</table>

### Table 2: Charts control limits for the variables $y_1$ and $y_2$

<table>
<thead>
<tr>
<th></th>
<th>Shewhart LCL</th>
<th>Shewhart UCL</th>
<th>Chen LCL</th>
<th>Chen UCL</th>
<th>Beta LCL</th>
<th>Beta UCL</th>
</tr>
</thead>
<tbody>
<tr>
<td>$y_1$</td>
<td>0.0525</td>
<td>0.4102</td>
<td>0.99596</td>
<td>1.0196</td>
<td>0.8882</td>
<td>0.4399</td>
</tr>
<tr>
<td>$y_2$</td>
<td>0.99642</td>
<td>0.99987</td>
<td>0.99896</td>
<td>0.000001</td>
<td>0.99419</td>
<td>0.9995</td>
</tr>
</tbody>
</table>

5 SENSITIVITY ANALYSIS
We now perform a sensitivity study to compare Shewhart, Chen CCs and Beta Chart in two scenarios: in-control and out-of-control; calculating the average run length (ARL$_0$) to different values of $p$. For the in-control process, we evaluate the ARL$_0$ until a false special cause is identified; the higher the ARL$_0$, the better [3].
suggested the ARL for comparison three schemes proposed in detecting shifts in the process. The ARL\(_0\) can be written as function of the type I error probability \(\alpha\), which is the probability of the control chart detecting a change in the in-control process monitored.

\[
ARL_0 = \frac{1}{\alpha}
\]  
(8)

where \(p_0\) is the average of the variable in-control process.

Similarly, for the out-of-control, we calculated the average run length until a true special cause is identified; the lowest the ARL\(_1\), the better. The ARL\(_1\) can be written as function of the type II error probability \(\beta\), which is the probability of the control chart not detecting a change in the out-of-control process.

\[
ARL_1 = \frac{1}{[1-\beta]}
\]  
(9)

where \(\lambda\) is the change induced in the process and, \(p_1\) the average of the variable out-of-process.

The sensitivity of the control charts investigated for \(p = 0.001\) and \(n = 1500\) shows similarity between the Shewhart and Chen charts for a process in-control (ARL\(_0\) = 226), although the Chart proposed by Chen provides faster in detection of changes in the parameter values \(p\) (\(p_1 = 0.002\), ..., \(p_4 = 0.01\)). The Beta Chart shows better sensitivity for the probability of false alarm and performance in detecting changes compared the charts investigated [see Table 3]. Similarly, it occur for the parameter value \(p\) (\(p_1 = 0.01\), ..., \(p_4 = 0.10\)) e \(n = 200\), when the Beta Chart present better performance in detecting changes [see Table 4].

This is due the modification proposed by authors having the correction with property positive asymmetry [16]. The control limits proposed to Betas Chart present the in-control average run length (ARL\(_0\) = 370) for 0.9973 of the probability (\(\alpha = 0.0027\)) and faster to detect changes for all the out-of-control scenarios, with lower values for ARL\(_1\). The control charts with approximation by Normal distribution present narrow limits, generating ARL\(_0\) less than 370 samples and false alarms more frequently. While that using the approximation by Beta distribution is possible to define an exact nominal value.

**Table 3: Comparison of the ARL for Shewhart, Chen CCs and Beta Chart for \(p_0 = 0.001\), \(n = 1500\), \(\lambda = 0.001\)**

<table>
<thead>
<tr>
<th>(p)</th>
<th>(LCL)</th>
<th>(UCL)</th>
<th>(\alpha)</th>
<th>(\beta)</th>
<th>ARL</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.001</td>
<td>0.00004</td>
<td>0.0004</td>
<td>0.004</td>
<td>0.004</td>
<td>226</td>
</tr>
<tr>
<td>0.002</td>
<td>0.0004</td>
<td>0.0008</td>
<td>0.008</td>
<td>0.008</td>
<td>12</td>
</tr>
<tr>
<td>0.003</td>
<td>0.0008</td>
<td>0.0012</td>
<td>0.012</td>
<td>0.012</td>
<td>3</td>
</tr>
<tr>
<td>0.005</td>
<td>0.0012</td>
<td>0.0016</td>
<td>0.016</td>
<td>0.016</td>
<td>1</td>
</tr>
<tr>
<td>0.01</td>
<td>0.0020</td>
<td>0.0024</td>
<td>0.024</td>
<td>0.024</td>
<td>1</td>
</tr>
</tbody>
</table>

**Table 4: Comparison of the ARL for Shewhart, Chen CCs and Beta Chart for \(p_0 = 0.01\), \(n = 200\), \(\lambda = 0.01\)**

<table>
<thead>
<tr>
<th>(p)</th>
<th>(LCL)</th>
<th>(UCL)</th>
<th>(\alpha)</th>
<th>(\beta)</th>
<th>ARL</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01</td>
<td>0.00006</td>
<td>0.0004</td>
<td>0.004</td>
<td>0.004</td>
<td>226</td>
</tr>
<tr>
<td>0.02</td>
<td>0.00014</td>
<td>0.0008</td>
<td>0.008</td>
<td>0.008</td>
<td>10</td>
</tr>
<tr>
<td>0.03</td>
<td>0.00021</td>
<td>0.0012</td>
<td>0.012</td>
<td>0.012</td>
<td>2</td>
</tr>
<tr>
<td>0.05</td>
<td>0.00038</td>
<td>0.0020</td>
<td>0.020</td>
<td>0.020</td>
<td>1</td>
</tr>
<tr>
<td>0.10</td>
<td>0.00064</td>
<td>0.0036</td>
<td>0.036</td>
<td>0.036</td>
<td>1</td>
</tr>
</tbody>
</table>

6 CONCLUSION

Although many CCs rely on the Binomial distribution for monitoring variables of the proportion type, there might be some noteworthy restrictions in that: (i) variables in industrial processes seldom have density symmetric; and (ii) Normal distribution becomes inappropriate approximation for small (larger) proportion measurements.

When applied to a two numerical examples, the Beta Chart led to more precise results when compared to Shewhart and Chen CCs used to proportion monitoring. Further, a sensitivity analysis with several values of proportion corroborated the superior performance of the Beta Chart compared to Shewhart and Chen CCs: the
proposed approximation yielded slightly larger ARL₀ and significantly smaller ARL₁.

The Beta Chart may be an adequate tool for monitoring the quality characteristics proportion type confined in the [0,1]-interval, allowing asymmetric control limits and false alarm rate desired.

7 ACKNOWLEDGEMENTS
The authors are grateful for the support by Pontifical University Catholic of Parana (PUCPR/Brazil).

8 REFERENCES