 doi:10.20944/preprints202109.0526.v1

begindocument/before

 doi:10.20944/preprints202109.0526.v1

Article

Monitoring The Nonconforming Fraction With A Dynamic Scheme When Sample Sizes Are Time Varying

Javier Blanco [1,‡], Víctor H. Morales [1,‡*] and Carlos A. Panza [2,‡]

[1] Departamento de Matemáticas y Estadística, Universidad de Córdoba, Montería 230027, Colombia; jbm092@gmail.co
[2] Departamento de Estadística, Universidad Nacional de Colombia, Bogotá 111321, Colombia; capanzao@unal.edu.co
* Correspondence: vmorales@correo.unicordoba.edu.co
‡ These authors contributed equally to this work.

Abstract: In many practical applications, it is more convenient to characterize the quality of production processes or service operations throughout the count of nonconformities. In the context of SPC, noncorfomities are usually assumed to appear according to the binomial probability model. The conventional way for monitoring nonconformities invloves Shewhart-type control procedures based on both constant and time varying sample sizes. In this article, an EWMA scheme is proposed for monitoring the fraction of nonconforming items with time-varying sample sizes. The proposed control chart is referred to as the EWMAG-B and can be easily adapted to work with a constant sample size by fixing it at a needed value. By means of simulation, it was found out that the EWMAG-B chart outperforms the conventional p control chart in Phase II while detecting changes in the process level is wanted.

Keywords: Average run length; binomial distribution; control charts; EWMAG control chart; nonconforming fraction; p control chart; Phase II; SPC.

1. Introduction

SPC has been implemented for some time in many areas, including applications in industry, service operations, health surveillance and other activities where the monitoring of processes is required. Control charts have become the most widely used tool for process monitoring. The main goal of using control chart is to check processes for stability over time. In many practical applications, the quality of processes is often characterized by a continuous random variable, whose distributional parameters (or some function of them) need to be monitored in order to detect unwanted drops in them the soon as possible. The $\bar{X}$ and R charts are good enough illustrations for monitoring the mean level and the dispresion of a normal distributed quality characteristic. Montgomery [3] present some relevant insights on control chart designing for monitoring continuous processes.

However, it is not always easy or convenient in economic terms to record the measure of a quality characteristic. It is better sometimes to classify the producto of a process into one of two exhaustive and mutually exclusive categories: in compliant and non-compliant, for example. In these situations, process quality is formally expressed by a discrete counting random variable. Vargas [8], among many other authors, present the np, c and u control charts for monitoring the count of nonconformities of a process. The p control chart is the most commonly used procedure for monitoring the variations in the fraction of non-conforming items per sample. For instance, the p chart is used for monitoring the fraction of defective tires per sample in automotive industry.

 doi:10.20944/preprints202109.0526.v1

The p chart is a Shewhart-type control methodology and its design, as well as those of the aforementioned charts, strongly depends on assumptions that must necessarily be met. Although chart designing requires the size of available samples to be known and fixed, it is also based on asymptotic normal distributional properties. In most cases, as in healthcare surveillance applications, a fixed large enough sample size is almost impossible to guarantee. As pointed out in Alvarado & Retamal [1], the normal approximation works fairly acceptable just for $n > 30$, $np \geq 5$ and $np(1-p) \geq 5$. Quesenberry [6] states that when $n \leq 30$ and $p \geq 0.20$, the estimates of the normal parameters would inevitably be unreliable and could lead to an unacceptable performance of the traditional p control chart. Recall that the problem related to the independence of the available samples, which is not minor, has not even touched.

Recently, some methodologies have been proposed that work fairly well when the assumptions for the p chart design are not fully met. Aytaçoğlu & Woodall [2] propose a CUSUM methodology that overcomes aforementioned shortcomings. Nevertheless, the construction and interpretation of this proposal may represent some challenges for some untrained practitioners, given the relative complexity of chart designing and implementation. In SPC, a great effort has been focused on developing monitoring schemes whose control limits can be drawn not as rigidly as is done in traditional schemes. In this sense, Shen *et al.* [7] proposed the monitoring of Poisson-type counting data when sample sizes vary over time using an EWMA chart with probability control limits.

In this article, it is proposed to adapt Shen *et al.*'s methodology to the special case of monitoring the nonconforming fraction in Phase II processes. The resulting control procedure is referred to as the EWMAG-B chart. This methodology avoid the need of using the asymptotic normal approximation for setting up the control limits and is especially developed to work with time-varying sample sizes. Through simulation, it was found out that the proposed EWMAG-B chart outperforms the traditional p chart in all the explored scenarios. As shown in the provided example, the EWMAG-B chart can be easily reformulated for dealing with a fixed known value of the sample size.

2. Theoretical framework

2.1. The EWMAG control chart

Shen *et al.* [7] first proposed an EWMA chart with probability control limits for monitoring Poisson count data with time varying sample sizes in Phase II. This chart is called "EWMAG" because exhibits the appealing property of having geometrically-disrtibuted run lengths. That is, the false alarm rate of the control scheme does not depend on the monitoring time, nor the size of the available samples. In addition, chart performance does not depend on the mechanism generating sample sizes, whatever its nature.

Let X_t, $t = 1, 2, \ldots$, be the count of an adverse event during the fixed time period $(t-1, t]$. Given the sample size n_t at time t, suppose that X_t independently follows the Poisson distribution with mean $\theta\, n_t$, where θ denotes the occurrence rate of the event of interest. To detect an abrupt change in the rate of occurrence from θ_0 to another unknown value $\theta_1 > \theta_0$, it is proposed the EWMA statistic

$$Z_j = (1-\lambda)Z_{j-1} + \lambda \frac{X_j}{n_j}, \quad j = 1, 2, 3, ..., t, \tag{1}$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 September 2021 doi:10.20944/preprints202109.0526.v1

where $Z_0 = \theta_0$ and $\lambda \in (0;1]$ is the smoothing parameter of the chart determining the weights of past observations. According to Shen *et al.* [7], the the upper control limit h_t of the EWMA statistic (1) a each monitoring moment t has to satisfy

$$\begin{aligned} P(Z_1 > h_{1,\alpha}|n_1) &= \alpha, \\ P(Z_t > h_{t,\alpha}|Z_{t-1} < h_{t-1,\alpha}, n_t) &= \alpha, \quad t > 1, \end{aligned} \tag{2}$$

where α is the previously established level of the false alarm rate.

It is worth mentioning that the probability control limit is determined just after the n_t value is observed at time t. Consequently, the EWMAG chart does not suffer from incorrect model assumptions on the mechanism generating the sizes of available samples. This property makes the proposed EWMAG chart significantly different from the previously proposed control charts.

The calculation of the conditional probability given in (2) can be a difficult task to solve analytically $h_t(\alpha)$. Computational procedures based on simulation via Monte Carlo and Markov chains are available. For more details, the reader is addressed to Shen *et al.* [7]. This methodology was also implemented by Morales & Vargas [5] for monitoring aggregated Poisson data in processes with time-varying sample sizes.

2.2. The EWMAG-B control chart

This section presents the theoretical and formal aspects of the EWMAG-B control chart, as a result of adapting the Shen *et al.*'s proposal to the special case of monitoring the nonconforming fraction.

Let X_t be the number of successful trials in a sequence of n Bernoulli independent trials at time t, $t = 1, 2, \ldots$. It is clear that X_t independently follows the binomial distribution with parameters $n_t \in \mathbb{N}$ and $0 < p < 1$, where n_t is the sample size at time t and p is the stable probability of having a successful trial, respectively. The main goal is to detect an abrupt change in the probability p from a known value p_0 to another unknown level $p_1 > p_0$ (it is also possible to make $p_1 < p_0$). The EWMAG-B chart uses the statistic given in (1) with $Z_0 = p_0$.

The upper control limit of the EWMAG-B chart must satisfy the expressions given in (2) and is determined just after the value of n_t is observed at time t. As mentioned above, due to the complexity of the conditional probability given in (2), needed $h_t(\alpha)$ values are found via simulation. The procedure for calculating the upper control limit is summarized in the following algorithm:

- **Step 1.** Set $t = 1$. Given $p = p_0$ and the first sample size n_1, generate a large enough number M of random values X_1 from the binomial distribution with parameters n_1 and p_0. Let $\hat{X}_{1,1}, \hat{X}_{1,2}, ..., \hat{X}_{1,M}$ be the generated values of X_1. Each of these values represent the number of successful trials in a sample of size n_1.
- **Step 2.** Transform each of the $\hat{X}_{1,i}$, $i = 1, \ldots, M$, into a "pseudo EWMA" $\hat{Z}_{1,i}$ value with the help of expression (1), such as indicated below

$$\hat{Z}_{1,i} = (1 - \lambda)p_0 + \lambda \frac{\hat{X}_{1,i}}{n_1}. \tag{3}$$

  There are obtained M "pseudo EWMA" values at the first monitoring moment.
- **Step 3.** Sort the M transformed $\hat{Z}_{1,i}$, $i = 1, \ldots, M$, values in ascending order and find the $(1 - \alpha) \times 100\%$ empirical quantile of the resulting ordered sequence of $\hat{Z}_{1,i}$ values. The found quantile is an estimate of the first control limit $h_1(\alpha)$, where α is the previously established false alarm rate.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 September 2021 doi:10.20944/preprints202109.0526.v1

- **Step 4.** Calculate the actual value of the EWMAG-B statistic Z_1 from the really observed X_1/n_1 value and compare it with the first control limit found in the preceding step. If $Z_1 \leq h_1(\alpha)$, assume that the process is in-control at the first monitoring moment and continue to the next monitoring moment. Otherwise, delete the actual observation from the analysis.
- **Step 5.** Proceed as follows for $t \geq 2$ if it is decided to continue:
  1. Retain only the first $M^* = \lfloor (1-\alpha) \times M \rfloor$ of the ordered "pseudo EWMA" values found in the preceding monitoring moment $t-1$.
  2. Once the sample size n_t is known, generate a large enough number M of random values X_t from the binomial distribution with parameters n_t and p_0. Let $\hat{X}_{t,i}$, $i = 1, \ldots, M$, be the generated values of X_t.
  3. Transform each of the $\hat{X}_{t,i}$, $i = 1, \ldots, M$, into a "pseudo EWMA" $\hat{Z}_{t,i}$ as indicated below
  $$\hat{Z}_{t,i} = (1-\lambda)\,\hat{Z}_{t-1} + \lambda\,\frac{\hat{X}_{t,i}}{n_t}, \tag{4}$$
  where $\hat{Z}_{t-1}$ is a value randomly picked from the set of M^* "pseudo EWMA" values obtained as said in Step 5-1. This is done to ensure that the process works under stable conditions.
  4. Sort the M transformed $\hat{Z}_{t,i}$, $i = 1, \ldots, M$, values in ascending order and repeat the whole algorithm from Step 3 for the given monitoring moment t.

If it is needed, both the lower and the upper control limits of the two-sided EWMAG-B control chart are determined in a similar way by finding the $\frac{\alpha}{2} \times 100\%$ and the $\left(1-\frac{\alpha}{2}\right) \times 100\%$ empirical quantiles, respectively, of the simulated marginal conditional distribution of the EWMA statistic at each monitoring moment. In this case, some minor adjustements should be introduced into the above described algorithm. This control scheme can also deal with fixed sample size by simply making $n_t = n$, where n is a constant value.

3. Simulation study

3.1. Performance evaluation

As usual, the performance of the proposed EWMAG-B control chart was evaluated in terms of the ARL. This is, the number of inspected samples until the scheme first signals. The performance of the EWMAG-B chart was compared to that of the traditional Phase II six-sigma p chart version for non-constant sample sizes. To compare the performance of both control procedures, a simulation-based study was conducted following the indications below.

3.2. Simulation settings

For all purposes, the target value of the nonconforming proportion is set to be $p_0 = 0.1$. As the work with the EWMAG-B chart can be a bit computationally demanding, both the EWMAG-B and the conventional p chart were calibrated to reach a nominal false alarm rate of $\alpha = 0.005$, leading to an in-control ARL value of 200 samples.

For the EWMAG-B scheme, the desired in-control ARL value is achieved by setting $\lambda = 0.1$ in the expression (1) and generating a set of $M = 50,000$ "pseudo EWMA" values at each monitoring moment t, just after the sample size n_t is known. Every "pseudo EWMA" value in the same set is based on the number of nonconformities found in a sample of size n_t with the nonconforming probability being equal to $p_0 = 0.1$. As the scheme is dealing with a nonconforming fraction, unwanted increases from $p_0 = 0.1$ are needed to be detected. That is why, the value of the upper control limit UCL_t is only needed to be calculated at each monitoring moment. The corresponding UCL_t value is approximated by the $(1-\alpha) \times 100\%$ quantile of the conditional marginal distribution of the EWMA statistic (1) obtained with the help of the $M = 50,000$ pseudo values. For

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 September 2021 doi:10.20944/preprints202109.0526.v1

the conventional p chart, the desired in-control ARL value is achieved by choosing the constant L_t determining the width of the in-control region, once the sample size n_t is known.

As stated before, theoretically the design of the EWMAG-B requires neither special assumptions on the sample sizes themselves nor on the mecanism generating them. However, the conventional p chart does. To ensure that both charts work under the same conditions, it is assumed that needed sample size n_t at moment t is uniformly distributed from 100 to 500. The uniform and other three scenarios for varying sample sizes were used in Shen *et al.* [7] to show that the performance of the EWMAG chart does not depend on the mechanism generating the sample sizes.

Out-of-control scenarios were considered by introducing increases of different magnitudes in the in-control value of the nonconforming proportion. Explored out-of-control scenarios included current 5%, 10% and 15% increase from $p_0 = 0.1$. Out-of-control ARL values for each combination of impacting factors were also estimated by computing the mean of 10,000 simulated run length values.

4. Results and discussion

Estimated ARL performance for both the EWMAG-B and the conventional p charts is reported in Table 1. As expected, all investigated control charts approximately exhibit the desired in-control performance. It is also noted that the greater the increase, the faster both charts detect it. However, the carried out simulation study suggests that for any of the explored out-of-control scenarios the EWMAG-B chart always takes considerably less time to detect planned current increases from the nominal value of the nonconforming proportion $p_0 = 0.1$ than the conventional chart. Other in-control p and ARL were also explored and similar results to those reported in Table 1 were obtained.

Table 1. Approximated ARL performance of the EWMAG-B and the conventional p charts.

Nonconforming Proportion	EWMAG-B Chart	p Chart
0.100	197.18	200.90
0.105	110.27	149.85
0.110	71.31	118.08
0.115	50.57	95.52

Furthermore, another simulation study was conducted to investigate the in-control ARL performance of both schemes when working with samples of the same size. Proposed control charts were calibrated to reach the same in-control ARL value for different fixed sample sizes. Obtained results are reported in Table 2 for $\alpha = 0.0027$ and $p_0 = 0.1$.

Table 2. Approximated in-control ARL performance of the EWMAG-B and the conventional p charts leading to $\alpha = 0.0027$.

Fixed sample size	EWMAG-B Chart	p Chart
50	351.32	312.13
100	358.91	326.87
200	367.98	295.31
300	371.03	345.01

To reach the established nominal FAR value, it is needed the in-control ARL value to be 372 approximately, whatever the scheme with geometrically distributed run lengths. As can be seen in Table 2, a more stable performance of the EWMAG-B chart with respect

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 September 2021 doi:10.20944/preprints202109.0526.v1

to the nominal in-control ARL value of 372 is observed. Conversely, the in-control ARL performance of the six-sigma p chart seems to depend strongly on the sample size used and only approaches the preset nominal 372 value as n increases. This can perhaps be explained by the fact that the p chart design is based on the asymptotic normal approximation of the maximum likelihood estimator of the proportion p. Other in-control p and ARL values were also explored and similar results to those reported in Table 2 were obtained.

5. Example

A local dairy company with a well-known experience in the market provided a dataset on tetrapack boxes that are used for packing yogurt in 120 g presentation. The use of the dataset is allowed for academic and illustrative purposes only. In the following, the design of both the traditional six-sigma p and the EWMAG-B chart based on the available data is shown. Some interesting findings will be addressed as well.

Nonconforming tetrapack boxes may result from a forming process due to improper sealing allowing packed yogurt to spill out. A retrospective phase I analysis was conducted based on 50 initial batches of $n = 250$ boxes each selected during a period of continuous operation of the production process.

The two-sided traditional six-sigma p chart that was constructed for the analysis of the initial batches showed that a 7% fraction of boxes per batch is poorly sealed. Then, the lower and upper control limits for checking process stability over time were set to be $LCL = 0.0216$ and $UCL = 0.1184$, respectively. For further analysis, only the last 20 batches of the dataset will be used.

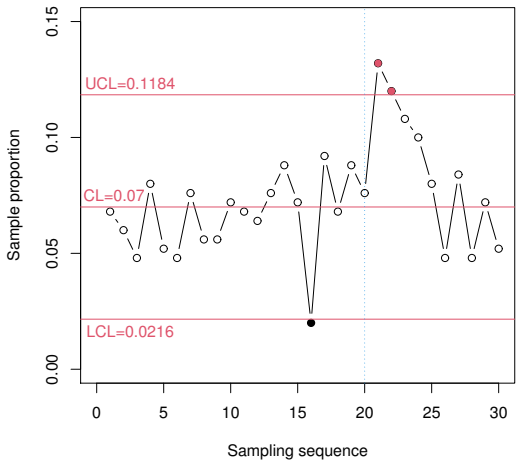

Figure 1. Two-sided traditional six-sigma p control chart for the nonconforming fraction in the tetrapack boxes example.

To check the stability of the process, 10 additional batches of $n = 250$ boxes each were collected. Nonconforming fractions corresponding to the new additional batches were plotted in the p chart obtained from the conducted Phase I analysis. The chart is presented in Figure (1). The vertical dashed blue line in the plot indicates the monitoring moment corresponding to the last batch inspected before collecting the new batches. The traditional p chart detects an increase in the target noncoconforming fraction as soon as the online monitoring is restarted. Samples 21 up to 25 are known to be out-of-control because the boxes in those batches were formed from a lower quality material. The chart

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 September 2021 doi:10.20944/preprints202109.0526.v1

just correctly identifies the non-conformity of the 21st and 22nd batches. In addition, the chart shows a signal at the 16th sample that is known to be a false alarm.

The two-sided EWMAG-B chart with smoothing parameter $\lambda = 0.1$ and nominal false alarm rate $\alpha = 0.0027$ was also built for the same 30 batches by assuming the value $p_0 = 0.07$ as the in-control nonconforming fraction. The respective plot is shown in Figure (2). The proposed methodology detects the change to the new level at the 23rd monitoring moment. Two samples later than the traditional p chart does. However, contrary to the traditional chart, the EWMAG-B chart shows stronger evidence of increasing in the nominal nonconforming fraction. To be more exact, the process seems to work at a better level from the beginning of the monitoring.

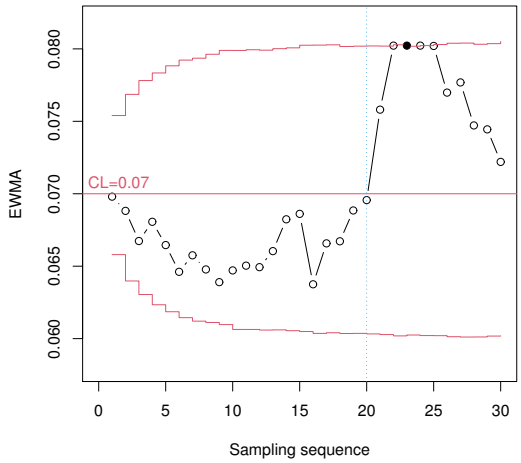

Figure 2. Two-sided EWMAG-B control chart with $\lambda = 0.1$ and $\alpha = 0.0027$ for the nonconforming fraction in the tetrapack boxes example.

6. Conclusions and recommendations

An EWMA control chart with varying-in-time control limits is proposed for monitoring the nonconforming fraction in samples of different sizes. The scheme is an adapted version of the EWMAG scheme first proposed by Shen *et al.* [7] and neither makes assumptions on the sizes of available samples nor on the mechanism generating them prior the monitoring.

By means of simulation, it was possible to establish that the new control chart outperforms the conventional p chart when the main objective is the detection of unwanted increases in the nominal proportion of non-conformity in Phase II. The provided example shows that the proposed control methodology can be easily redisigned to work with a fixed value of the sample sizes in two-sided schemes as well.

Funding: This research received no external funding.

Data Availability Statement: Dataset supporting reported results was kindly provided by the quality control laboratory of a local dairy company that prefers anonymity and allowed the use of the data for academic and illustrative purposes only.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Abbreviations

 doi:10.20944/preprints202109.0526.v1

The following abbreviations are used in this manuscript:

ARL	Average run length
CUSUM	Cumulative sums
EWMA	Exponentially weighted moving averages
EWMAG	EWMA chart with geometric run length distribution
EWMAG-B	EWMAG for binomial-distributed nonconformities
FAR	False alarma rate
SPC	Statistical Process Control

References

1. Alvarado, H.; Retamal, M.L. La aproximaación binomial por la normal: una experiencia de reflexión sobre la práctica. *Revista Paradigma* **2010**, *31(2)*, 89–108.
2. Aytaçoğlu, B.; Woodall, W.H. Dynamic probability control limits for CUSUM charts for monitoring proportions with time-varying sample sizes. *Quality and Reliability Engineering International* **2019**, *36(2)*, 592–603.
3. Montgomery, D. *Introduction to statistical quality control*, 7th ed.; Publisher: John Wiley & Sons, Inc., USA, 2013.
4. Montgomery, D. The control chart for fraction nonconforming. In *Introduction to statistical quality control*; John Wiley & Sons, Inc., USA, 2013; pp. 301–304.
5. Morales, V.H.; Vargas, J.A. Monitoring aggregated Poisson data for processes with time-varying sample sizes *Revista Colombiana de Estadística* **2017**, *40(2)*, 243–262.
6. Quesenberry, C.P. (Department of Statistics, North Carolina State University, USA). SPC binomial q-charts for short or long runs, Thechnical Report, 1990.
7. Shen, X.; Zou, C.; Jiang, W.; Tsung, F. Monitoring Poisson count data with probability control limits when sample sizes are time varying. *Naval Research Logistics*, **2013**, *60(8)*, 625–636.
8. Vargas, J.A. *Control estadístico de calidad*, 2nd ed.; Publisher: Universidad Nacional de Colombia, Colombia, 2015.
9. Zhou, Q.; Zou, C.; Wang, Z.; Jiang, W. Likelihood-based EWMA charts for monitoring Poisson count data with time-varying sample sizes. *Journal of the American Statistical Association*, **2012**, *107(499)*, 1049–1062.