

Linear combination of chi-squares for multinomial process monitoring

Combinação linear de qui-quadrados para monitoramento de processo multinomial

Ramzi Talmoudi¹ , Ali Achouri² , Hassen Taleb²

¹University of Carthage, LR18ES48 ENVIE Laboratory, Faculté des Sciences Economiques et de Gestion de Nabeul, Tunisia. E-mail: ramzi.telmoudi@fsegn.u-carthage.tn

²University of Carthage, ARBRE Laboratory, Tunisia

How to cite: Talmoudi, R., Achouri A., & Taleb H (2021). Linear Combination of Chi-Squares for Multinomial Process Monitoring. *Gestão & Produção*, 28(3), e41. <http://doi.org/10.1590/1806-9649-2020v28e41>

Abstract: Marcucci (1985) proposed a chi square goodness of fit statistic based generalized p-chart for multinomial process monitoring. A chi square distribution quantile was considered as a control chart limit. A weighted chi square goodness of fit statistic-based control chart is proposed for multinomial process monitoring in this paper, where more important weights are advocated to poor quality categories. The statistic distribution is approximated by a well-known linear combination of chi squares distribution. The approximation is assessed through a simulation, an extreme percentile of the approximated distribution is used as an upper control chart limit and a comparison is carried out with a chi square goodness of fit statistic-based control chart. The average run length is used as a benchmark and the comparison is performed using simulations considering two process shifts scenarios. Under some restrictions, the weighted statistic-based control chart allows an earlier detection of process shift in case of deterioration and postpones out of control signals in case of improvement. This benefit is clearer when the process is improved by a decrease in the poor quality probability category and an increase in the best quality category probability.

Keywords: Multinomial Process; Generalized p-Chart; Distribution Approximation; Simulation.

Resumo: Marcucci (1985) propôs um gráfico generalizado baseado na estatística de bondade de ajuste do qui-quadrado para monitoramento de processos multinominais. Um quantil de distribuição do qui quadrado foi considerado como limite do gráfico de controle. Um gráfico de controle baseado em estatística de qualidade de ajuste chi-quadrado ponderado é proposto para monitoramento de processo multinominal neste artigo, onde pesos mais importantes são defendidos para categorias de baixa qualidade. A distribuição estatística é aproximada por uma combinação linear bem conhecida de distribuição de qui-quadrados. A aproximação é avaliada por meio de uma simulação, um percentil extremo da distribuição aproximada é usado como um limite superior do gráfico de controle e uma comparação é realizada com um gráfico de controle baseado na estatística de qualidade do qui quadrado do ajuste. O comprimento médio de execução é usado como referência e a comparação é realizada por meio de simulações considerando dois cenários de deslocamento de processo. Sob algumas restrições, o gráfico de controle baseado em estatística ponderada permite uma detecção precoce de mudança de processo em caso de deterioração e adia sinais fora de controle em caso de melhoria. Esse benefício fica mais claro quando o processo é aprimorado

Received Oct. 23, 2020 - Accepted Jan. 25, 2021
Financial support: None.



This is an Open Access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

por uma diminuição na categoria de probabilidade de qualidade ruim e um aumento na probabilidade da categoria de melhor qualidade.

Palavras-chave: Processo Multinomial; Gráfico generalizado; Aproximação de distribuição; Simulação.

1 Introduction

Control charts for attributes are used when focus is on the classification of product units into categories rather than quality characteristic measurements. The inspected units are generally classified into conforming or nonconforming units; however, different criteria of nonconforming classification could be used following defect seriousness. Such processes are called multinomial processes.

Raz & Wang (1990) and Taleb & Limam (2002) have introduced and discussed the construction of univariate control charts for multinomial process monitoring using both probability and fuzzy theory. Moreover, (Taleb et al., 2006) discussed methods based on multivariate fuzzy multinomial control charts. A review of multinomial and multi attribute quality control charts is given by (Topalidou & Psarakis, 2009).

A probabilistic approach is considered in this paper as no classification ambiguity is considered. It is based on the work of (Duncan, 1950) who developed a chi square chart for controlling a set of percentages and (Marcucci, 1985) who introduced a one-sided generalized p-chart for multinomial process monitoring. (Cozzucoli, 2009) used an overall defectiveness index based two-sided multivariate p-chart to distinguish between process improvement and process deterioration by points less than LCL and greater than UCL, respectively. The overall defectiveness index is a weighted proportions of defects statistic where weights are between 0 and 1. It was noticed that the proposed control chart performance depends on weight values. (Yashchin, 2012) derived changepoint detection schemes based on generalized likelihood ratio tests. The considered schemes are based on predefined acceptable and unacceptable regions for the monitored parameters. (Li et al., 2014a) considered that the ordinal attribute levels of a categorical variable are formed by some thresholds of the latent continuous variable. An example where flash on the head of toothbrush is classified into four categories according to its length is considered. A simple ordinal categorical control chart to detect location shift in factor's latent continuous variable distribution was then proposed. (Li et al., 2014b) proposed a multivariate binomial and multinomial control chart. An EWMA-type control chart is implemented using log-linear models for characterizing the relationship among categorical factors. (Weiss, 2018) proposed a control chart for serially dependent categorical processes monitoring. (Perry, 2019) used an EWMA control chart for categorical processes where weighted category counts were used as a control chart statistic. The weight assigned to each category is the inverse of its expected count. (Marcucci, 1985) control chart statistic depends already on the inverse of the expected count. The aim of this work is to propose an improved (Marcucci, 1985) generalized p-chart rather than improvement detection. The proposed control chart is based on a weighted chi-square goodness of fit statistic and aims at earlier detection of process deterioration and delay improvement detection. A weighted goodness of fit statistic called Q statistic is presented in section 2, as the distribution of the Q statistic is unknown, the distribution approximation issue is discussed in section 3 and the approximated distribution is assessed in section 4. The Q statistic control chart is presented through an illustrative example in section 5, finally its performance is compared to Marcucci (1985) control chart in section 6.

2 The Weighted Goodness of Fit Statistic

Multinomial process output is classified into q categories. Let p_1, p_2, \dots, p_q be the probabilities of classification into the exhaustive and mutually exclusive categories with $\sum_{j=1}^q p_j = 1$ and p_j known $\forall j = 1, \dots, q$, where conform units are classified into the first category and worst quality units are classified into the q^{th} category. X_j is the observed number of output units classified into the j^{th} category. If m samples are selected randomly without replacement, then X_{ij} is the observed number of output units in the i^{th} sample and classified into the j^{th} category with $i = 1, \dots, m$ and let n_i be the sample sizes, it is common to suppose that all samples have the same sample size n .

Marcucci (1985) proposed to use the following chi square (Equation 1) goodness of fit statistic for multinomial process monitoring:

$$Y_i^2 = \sum_{j=1}^q \frac{(X_{ij} - n_i p_j)^2}{n_i p_j} \quad (1)$$

where $n_i p_j$ is the expected count at the i^{th} sample and j^{th} category. The proposed statistic is a measure of overall deviations from expected counts and has a known distribution, hence, a quantile of the distribution is used as an upper control limit for the χ^2 control chart. Moreover, the goodness of fit statistic does not take into consideration whether poor quality categories expected count decrease or increase and all deviations have similar impact. The control chart statistic does not distinguish between quality improvement and quality deterioration.

For that purpose, different weights are attributed to the categories considering that deviations from expected counts in categories with poor process output quality have negative impact on the overall process output quality. Then, more important weights are advocated to poor quality categories. Let w_1, w_2, \dots, w_q be the categories constant weights. The overall deviations from expected counts assessment is given by the following Q statistic (Equation 2):

$$Q = \sum_{j=1}^q w_j \frac{(X_j - n p_j)^2}{n p_j} \quad (2)$$

if several samples are considered, then the Q statistic for the i^{th} sample is given by Equation 3:

$$Q_i = \sum_{j=1}^q w_j \frac{(X_{ij} - n_i p_j)^2}{n_i p_j} \quad (3)$$

3 Q Statistic distribution approximation

Q_i is considered as the control chart positive statistic and a one-sided control chart is the appropriate choice. In what follows, Q_i distribution is approximated by a linear combination of chi squares as the exact distribution is not known (Equation 4)

$$Q_i \approx \sum_{j=1}^q w_j \chi_{ij}^2(1) \tag{4}$$

Applications of linear combination of chi squares are presented in (Jensen & Solomon, 1972) and in (Mathai & Provost, 1992). (Kotz et al., 1967) used power series expansion to determine the approximated Q statistic distribution. (Davis, 1977) tried to solve the slow convergence of the power series for large values of the linear combination of chi squares using differential equations and assuming positive coefficients for a useful range of parameters. (Oman & Zacks, 1981) noticed that all available methods suffer of lengthy computations or are insufficiently accurate. However, the work of (Feiveson & Delaney, 1968) was not considered. (Moschopoulos & Canada, 1984) presented an easily programmed method for the distribution computation and better or at least as accurate as (Oman & Zacks, 1981) method, they noticed that their moment generating function-based method has equivalent results than (Imhof, 1961) method. Moreover, (Davies, 1980) presented an algorithm to compute the distribution function of linear combination of non-central chi square variables and (Castano-Martinez & Lopez-Blazquez, 2005) derived Laguerre expansions for the density and distribution functions for a sum of positive weighted noncentral chi square variables. (Bodenham & Adams, 2016) compared different approximations for weighted sum of chi squared random variables in the case of streaming data and noticed that Imhof's method is essentially exact, however, it is not suitable for a streaming data scenario. (Moschopoulos & Canada, 1984) point out that the linear combination of independent central chi-square random variables distribution function is given by Equation 5:

$$F(x) = \Pr(Q_i \leq x) = \prod_{i=2}^q b_i \sum_{j=0}^{\infty} a_j \int_0^x g_j(t) dt \tag{5}$$

where Q_i is expressed as in Equation 4, $g_j(t)$ is the p.d.f. of the Gamma distribution with a shape parameter $\lambda = \frac{q}{2} + j$ and a rate parameter $\beta = \frac{1}{2w_1}$. The coefficients $b_i = \left(\frac{w_1}{w_i}\right)^{\frac{1}{2}}$ and the coefficients $a_j = A_j^{(q)}$ where $A_j^{(i)}$ are determined recursively with

$$A_j^{(i)} = \sum_{k=0}^j A_k^{(i-1)} A(w_i, j-k), \quad A_r^{(2)} = A(w_2, r); A(w_i, r) = \left(\frac{1}{2}\right)_r \frac{\left(\frac{1-w_1}{w_i}\right)^r}{r!}; \left(\frac{1}{2}\right)_r = \left(\frac{1}{2}\right)\left(\frac{1}{2}+1\right)\dots\left(\frac{1}{2}+r-1\right);$$

$r = 0, 1, 2, \dots; i = 3, \dots, q; j = 0, 1, 2, \dots$ and $j-k = 0, 1, 2, \dots$

(Moschopoulos & Canada, 1984) provided some percentiles corresponding to different w_j and different degrees of freedom, they recommend using 40 terms in Equation 5 to determine numerically upper percentiles of the linear combination of chi squares. However, it is noticed that for extreme quantiles and large weight values, result

convergence needs more than 85 terms. Notice that if $w_j = 1, \forall j = 1, \dots, q$ then $Q_i \sim \chi^2(q)$ this property is used as a benchmark for program code validation. Figures 1,2 and 3 compare the approximated Q distribution to different chi square distributions and it is noticed that the approximated Q distribution depends on the category number and is different from a chi square distribution. Approximated Q quantiles are given in Table 1 for different values of q and different weights where α is the quantile order. (Feiveson & Delaney, 1968) approximated the linear combination of chi-squares distribution by a

gamma distribution with a rate parameter $\beta = \frac{1}{2} \frac{\sum_{i=1}^q w_i}{\sum_{i=1}^q w_i^2}$ and a shape parameter $\lambda = \frac{1}{2}$

$\frac{(\sum_{i=1}^q w_i)^2}{\sum_{i=1}^q w_i^2}$ and pointed out that large standard deviation of weights overestimate the true functional value in the right tail of the distribution.

Table 1. Linear Combination Percentiles for Different Weights and q Values.

w_1, \dots, w_q	$\alpha = 0.99$	$\alpha = 0.975$	$\alpha = 0.95$	$\alpha = 0.9$	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.025$	$\alpha = 0.01$
0.10,0.40,0.50	0.0313	0.0593	0.0976	0.1643	2.1873	2.8175	3.4503	4.2908
1.00,1.00,1.00	0.1148	0.2157	0.3518	0.5843	6.2513	7.8147	9.3484	11.3448
1.00,4.00,5.00	0.3139	0.5936	0.9760	1.6438	21.8738	28.1754	34.5032	42.9080
0.40,0.50,5.00	0.1163	0.2212	0.3667	0.6273	14.1921	15.0605	15.2203	15.2930
0.20,0.30,10.0	0.1000	0.1936	0.3294	0.5922	7.4066	7.4389	7.4448	7.4516
0.10,0.20,0.40	0.0230	0.0434	0.0712	0.1191	1.5171	1.9716	2.4398	3.0766
0.20,0.30,0.50	0.0357	0.0672	0.1098	0.183	2.1214	2.7075	3.3021	4.1022
0.40,0.60,1.80	0.0871	0.1644	0.2696	0.4519	6.1697	8.182	10.2868	13.1681
0.05,0.10,0.40	0.0145	0.0276	0.0455	0.0769	1.262	1.7119	1.9708	2.0150
$e^{\left(\frac{j-q}{2}-1\right)}$	0.0702	0.1329	0.2189	0.3698	5.5635	7.4118	8.8541	9.1435
0.05,0.10,0.20,0.30	0.0396	0.0651	0.0964	0.1465	1.3374	1.6958	2.0582	2.5438
0.05,0.15,0.30,0.40	0.0524	0.0866	0.1289	0.1972	1.8617	2.3563	2.8533	3.5223
0.05,0.10,0.20,0.40	0.0427	0.0703	0.1044	0.1593	1.5717	2.0260	2.4940	3.1316
0.05,0.10,0.20,0.50	0.0452	0.0747	0.1112	0.1703	1.8141	2.3745	2.9585	3.7887
0.05,0.10,0.25,0.30	0.0419	0.0690	0.1024	0.1559	1.4461	1.8302	2.2152	2.7256
0.05,0.10,0.15,0.30	0.0367	0.0604	0.0893	0.1355	1.2335	1.5744	1.9249	2.4015
1.00,2.00,3.00,5.00	0.7018	1.1510	1.7007	2.5729	22.3576	28.2192	34.1631	42.1611
0.50,1.00,3.00,5.00	0.5030	0.8321	1.2422	1.9096	20.2029	26.0361	32.0016	40.8856
1.00,3.00,5.00,10.0	1.0618	1.7539	2.6133	4.0032	39.7681	51.1285	62.8500	80.0873
1.00,1.00,1.00,1.00	0.2971	0.4844	0.7107	1.0636	7.7794	9.4877	11.1432	13.2767
2.00,5.00,10.0,20.0	2.0285	3.3504	4.9922	7.6501	78.0008	100.7196	124.1711	158.3079
$e^{\left(\frac{j-q}{2}\right)}$	0.1869	0.3118	0.4702	0.7338	9.3129	12.3603	15.5398	19.9112
1.00,1.00,1.00,1.00,1.00	0.5542	0.8312	1.1454	1.6103	9.2363	11.0704	12.8325	15.0862
0.05,0.10,0.20,0.30,0.50	0.0981	0.1494	0.2093	0.3012	2.2890	2.8753	3.4764	4.4908
0.05,0.10,0.15,0.20,0.50	0.0849	0.1289	0.1801	0.2583	2.0067	2.5652	3.1473	3.9988
0.05,0.10,0.25,0.40,0.50	0.1094	0.1669	0.2346	0.3390	2.5866	3.2273	3.9065	4.7318
0.05,0.10,0.20,0.40,0.50	0.1044	0.1592	0.2235	0.3227	2.5004	3.1330	3.7664	4.6068
0.50,1.00,3.00,4.00,5.00	1.1371	1.7369	2.4427	3.5340	26.7899	33.2844	39.7262	48.2125
0.50,2.00,2.50,3.00,5.00	1.1809	1.7967	2.5147	3.6110	25.2738	31.2622	37.2654	45.2826
0.05,0.10,0.20,0.40,1.20	0.1267	0.1950	0.2769	0.4066	4.1783	5.5218	6.9284	9.6310

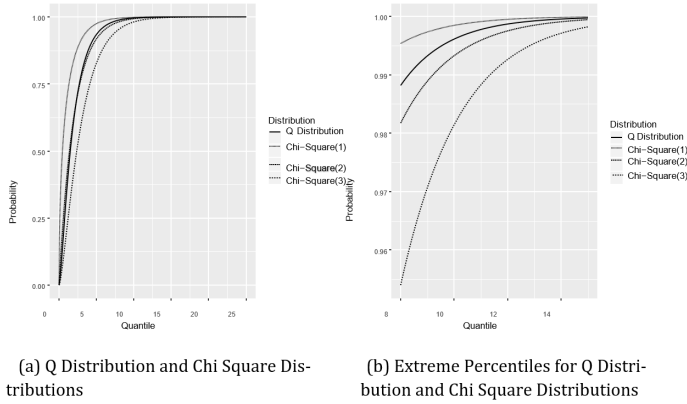


Figure 1. Approximated Three Categories Q Distribution with Weights $\frac{i}{q}$ Against Chi Square Distributions.

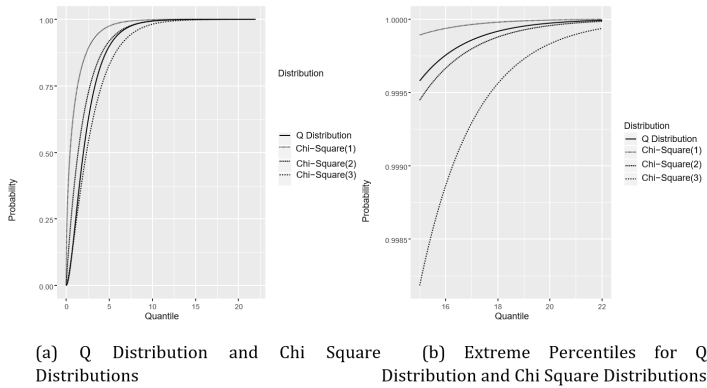


Figure 2. Approximated Four Categories Q Distribution with Weights $\frac{i}{q}$ Against Chi Square Distributions.

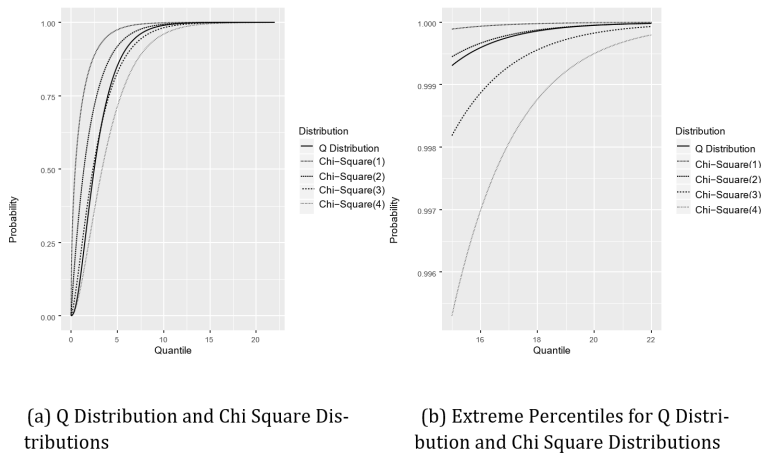


Figure 3. Approximated Five Categories Q Distribution with Weights $\frac{i}{q}$ Against Chi Square Distributions

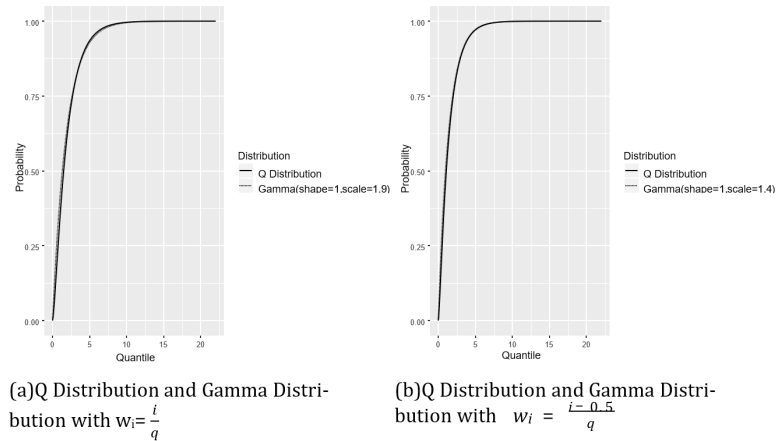


Figure 4. Approximated Three Categories Q Distribution with different weights against the gamma distribution.

4 Approximations assessment

Using (Moschopoulos & Canada, 1984) approximation the Q statistic approximated cumulative distribution is represented graphically against some chi square distributions. The considered weights are $w_j = \frac{j}{q}$ with $j = 1, \dots, q$. The Q statistic approximated cumulative distribution could be then compared graphically to $\chi^2(q-1)$ distribution. It becomes obvious that the upper control chart limit value based on linear combination of chi squares distribution approximation with $w_j = \frac{j}{q}$ is less than the upper limit determined from the chi square distribution q as $w_j \leq 1, \forall j = 1 \dots q$. Moreover, Figure 4 shows that the linear combination of chi squares distribution is more likely to be a gamma distribution rather than a chi square distribution and that it is sensitive to weight change. However, the Q statistic is a weighted chi square goodness of fit statistic, it would be interesting to compare graphically the linear combination of chi squares distribution and the weighted chi square goodness of fit statistic empirical distribution. In order to represent graphically the empirical distribution different multinomial distribution parameters were considered, 100000 samples were generated from each multinomial distribution and the Q statistic was computed for each sample of each multinomial distribution. Figure 5 illustrates both distributions where the empirical distribution is obtained for several category probabilities combinations and for huge sample size $n=20000$. It becomes obvious that the empirical distribution is sensitive to category probabilities changes and that the closest distribution to the approximation distribution is the empirical distribution with the lowest poor quality category classification probability.

A simulation is implemented to assess the Q statistic distribution approximation by linear combination of chi squares. As the approximated quantile with order 0.0027 would be used as the control chart upper control limit, then, first, the quantile $x_{0.0027}$ is determined for $q = 3, 4, 5$ using (5), then 5000 samples are generated from a multinomial distribution, different sample sizes will be used, finally the Q statistic is computed for each sample and the number of Q statistic values beyond the $x_{0.0027}$ are counted. These steps are repeated 5000 times, hence, the average count of Q statistic values beyond the $x_{0.0027}$ is determined and divided by 5000 in order to determine the proportion α of Q statistic values beyond the $x_{0.0027}$. The approximation is satisfactory if the determined proportions are close and

converge to the targeted value $\alpha = 0.0027$ when the sample size is increased. Simulation results are in Tables 2, 3 and 4.

The remarkable fact is that α does not converge exactly to the targeted value 0.0027 for all considered multinomial distributions, however, it is noticed computationally and graphically that the approximation is more efficient when $p_q \leq 0.01$. The quantile $X_{0.0027}$ could be used as an upper control limit for a Q control chart especially when $p_q \leq 0.01$. As in Figure 4 it was noticed that linear combination of chi squares distribution is more likely to be the gamma distribution, then, in what follows (Feiveson & Delaney, 1968) approximation is assessed for the Q distribution $X_{0.0027}$ quantile using the same steps explained previously. Results are in Tables 5, 6 and 7 where no approximation improvements are observed. Since Imhof's method the essentially exact and that (Moschopoulos & Canada, 1984) method is in fact the Imhof's method equivalent, then, only (Moschopoulos & Canada, 1984) approximation will be used for the control chart performance assessment where the A.R.L. is used as a benchmark.

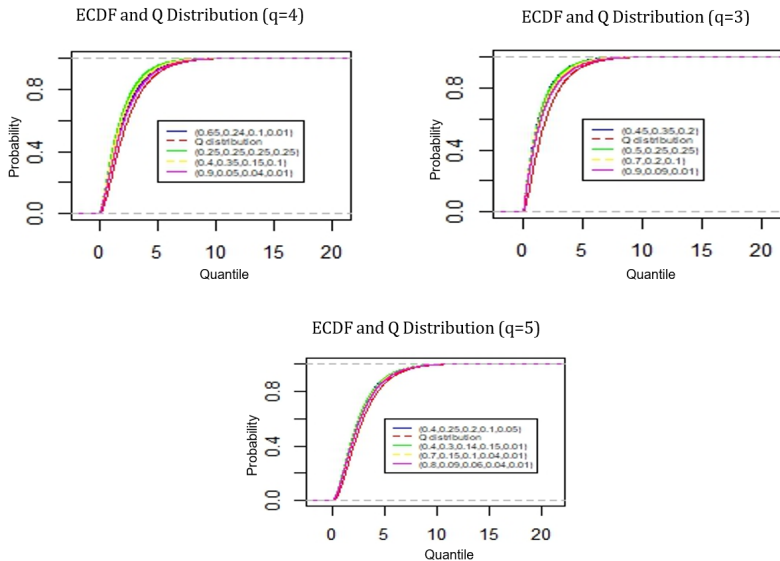


Figure 5. Q Statistic approximated cumulative distribution against empirical distributions.

5 Q Statistic Control Chart

Multinomial processes with q categories are considered. Quality levels are ordered and let the first category has the best quality level, the q^{th} category has the worst quality level. Let p_1, p_2, \dots, p_q be the probabilities to classify a given product unit into one of the q

categories with $\sum_{j=1}^q p_j = 1$. Different weights $w_1 \leq w_2 \leq \dots \leq w_q$ are assigned to the

considered categories. For a given taken randomly sample and including n product units, let X_j be the count of product units in the j^{th} category. The Q statistic is computed for the considered sample as Equation 2 and if several samples are considered, then the Q statistic for the i^{th} sample is given by Equation 3, where n_i is the size of the i^{th} sample and X_{ij} is the observed count of product units in the i^{th} sample and classified into the j^{th} category.

Table 2. α for different sample sizes and $q = 3$.

p_1, p_2, p_3	$n=3000$	$n=1000$	$n=500$	$n=300$	$n=200$
0.8000,0.1600,0.0400	0.0018	0.0018	0.0019	0.0021	0.0024
0.6600,0.3000,0.0400	0.0016	0.0017	0.0019	0.0020	0.0023
0.8500,0.1000,0.0500	0.0018	0.0019	0.0019	0.0021	0.0022
0.7400,0.2000,0.0600	0.0016	0.0018	0.0018	0.0019	0.0020
0.5500,0.3800,0.0700	0.0014	0.0015	0.0016	0.0017	0.0019
0.8000,0.1300,0.0700	0.0016	0.0017	0.0018	0.0019	0.0021
0.9000,0.0700,0.0300	0.0020	0.0022	0.0023	0.0028	0.0037
0.7000,0.2700,0.0300	0.0017	0.0019	0.0021	0.0025	0.0034
0.9425,0.0500,0.0075	0.0025	0.0029	0.0043	0.0047	0.0053
0.5500,0.4425,0.0075	0.0022	0.0024	0.0036	0.0040	0.0047
0.9600,0.0300,0.0100	0.0024	0.0027	0.0033	0.0045	0.0056
0.6500,0.3400,0.0100	0.0021	0.0023	0.0028	0.0040	0.0047
0.9300,0.0500,0.0200	0.0021	0.0023	0.0026	0.0039	0.0038
0.5000,0.4800,0.0200	0.0017	0.0019	0.0021	0.0035	0.0029
0.9725,0.0200,0.0075	0.0026	0.0030	0.0045	0.0046	0.0060
0.9835,0.0100,0.0065	0.0027	0.0038	0.0050	0.0052	0.0095
0.9750,0.0200,0.0050	0.0026	0.0033	0.0049	0.0057	0.0054
0.9775,0.0175,0.0050	0.0027	0.0033	0.0050	0.0056	0.0059
0.9700,0.0200,0.0100	0.0024	0.0027	0.0033	0.0047	0.0059
0.9675,0.0200,0.0125	0.0022	0.0030	0.0034	0.0043	0.0058

Table 3. α for different sample sizes and $q = 4$.

p_1, p_2, p_3, p_4	$n=3000$	$n=1000$	$n=500$	$n=300$	$n=200$
0.8000,0.1000,0.0600,0.0400	0.0019	0.0021	0.0023	0.0025	0.0027
0.6000,0.2500,0.1100,0.0400	0.0018	0.0019	0.0021	0.0023	0.0024
0.5500,0.2500,0.1500,0.0500	0.0017	0.0017	0.0018	0.0021	0.0021
0.8500,0.0500,0.0500,0.0500	0.0019	0.0021	0.0022	0.0025	0.0027
0.7000,0.1500,0.0900,0.0600	0.0017	0.0018	0.0019	0.0020	0.0022
0.5000,0.2900,0.1500,0.0600	0.0016	0.0017	0.0017	0.0018	0.0020
0.6000,0.3400,0.0525,0.0075	0.0023	0.0030	0.0036	0.0044	0.0055
0.5400,0.2500,0.2025,0.0075	0.0020	0.0027	0.0032	0.0039	0.0051
0.6000,0.3700,0.0200,0.0100	0.0022	0.0027	0.0035	0.0051	0.0058
0.8000,0.1400,0.0500,0.0100	0.0023	0.0028	0.0034	0.0048	0.0054
0.7000,0.1700,0.1100,0.0200	0.0020	0.0022	0.0025	0.0029	0.0035
0.5500,0.3500,0.0800,0.0200	0.0020	0.0022	0.0025	0.0029	0.0036
0.9450,0.0300,0.0150,0.0100	0.0025	0.0031	0.0039	0.0054	0.0064
0.9700,0.0125,0.0100,0.0075	0.0027	0.0036	0.0047	0.0059	0.0070
0.9650,0.0150,0.0100,0.0100	0.0026	0.0033	0.0042	0.0060	0.0069
0.9675,0.0150,0.0100,0.0075	0.0027	0.0036	0.0046	0.0057	0.0069
0.9430,0.0300,0.0150,0.0120	0.0024	0.0030	0.0037	0.0051	0.0055
0.9425,0.0300,0.0150,0.0125	0.0024	0.0029	0.0038	0.0045	0.0061
0.9350,0.0300,0.0200,0.0150	0.0024	0.0028	0.0035	0.0041	0.0054

Table 4. α for different sample sizes and $q = 5$.

p_1, p_2, p_3, p_4, p_5	n=3000	n=1000	n=500	n=300	n=200
0.7400,0.10,0.070,0.050,0.0400	0.0020	0.0021	0.0023	0.0026	0.0029
0.6500,0.13,0.100,0.080,0.0400	0.0019	0.0019	0.0021	0.0023	0.0026
0.6400,0.12,0.100,0.080,0.0600	0.0017	0.0018	0.0019	0.0021	0.0022
0.5500,0.20,0.100,0.080,0.0700	0.0017	0.0017	0.0018	0.0020	0.0021
0.5500,0.20,0.120,0.080,0.0500	0.0018	0.0018	0.0020	0.0021	0.0024
0.8000,0.05,0.050,0.050,0.0500	0.0019	0.0020	0.0022	0.0024	0.0028
0.6000,0.25,0.070,0.050,0.0300	0.0020	0.0022	0.0024	0.0027	0.0031
0.4000,0.27,0.200,0.100,0.0300	0.0018	0.0019	0.0021	0.0023	0.0026
0.7000,0.20,0.050,0.030,0.0200	0.0022	0.0025	0.0029	0.0033	0.0043
0.5000,0.23,0.150,0.100,0.0200	0.0019	0.0021	0.0025	0.0028	0.0035
0.8000,0.10,0.050,0.040,0.0100	0.0023	0.0029	0.0036	0.0046	0.0059
0.6000,0.15,0.140,0.100,0.0100	0.0021	0.0026	0.0032	0.0041	0.0053
0.9100,0.05,0.020,0.010,0.0100	0.0026	0.0034	0.0044	0.0058	0.0073
0.9105,0.05,0.020,0.010,0.0095	0.0027	0.0034	0.0044	0.0056	0.0067
0.9425,0.03,0.010,0.010,0.0075	0.0027	0.0038	0.0048	0.0060	0.0078
0.9250,0.04,0.020,0.010,0.0050	0.0029	0.0041	0.0059	0.0071	0.0073
0.9200,0.05,0.015,0.010,0.0050	0.0029	0.0041	0.0059	0.0071	0.0075
0.9000,0.05,0.020,0.015,0.0150	0.0024	0.0029	0.0037	0.0047	0.0058
0.9080,0.05,0.020,0.010,0.0120	0.0026	0.0032	0.0040	0.0053	0.0065
0.9060,0.05,0.020,0.012,0.0120	0.0025	0.0032	0.0039	0.0051	0.0062
0.9030,0.05,0.020,0.010,0.0170	0.0025	0.0030	0.0038	0.0048	0.0057

Table 5. α for different sample sizes and $q = 3$ using Gamma distribution $x_{0,0027}$ quantile.

p_1, p_2, p_3	n=3000	n=1000	n=500	n=300	n=200
0.8000,0.1600,0.0400	0.0022	0.0023	0.0024	0.0029	0.0036
0.6600,0.3000,0.0400	0.0021	0.0021	0.0022	0.0028	0.0034
0.8500,0.1000,0.0500	0.0022	0.0023	0.0026	0.0026	0.0029
0.7400,0.2000,0.0600	0.0020	0.0020	0.0021	0.0022	0.0028
0.5500,0.3800,0.0700	0.0018	0.0018	0.0019	0.0019	0.0023
0.8000,0.1300,0.0700	0.0020	0.0020	0.0022	0.0022	0.0025
0.9000,0.0700,0.0300	0.0024	0.0025	0.0028	0.0032	0.0039
0.7000,0.2700,0.0300	0.0021	0.0022	0.0024	0.0027	0.0036
0.9425,0.0500,0.0075	0.0029	0.0034	0.0058	0.0058	0.0056
0.5500,0.4425,0.0075	0.0024	0.0028	0.0053	0.0056	0.0049
0.9600,0.0300,0.0100	0.0028	0.0032	0.0041	0.0050	0.0056
0.6500,0.3400,0.0100	0.0023	0.0028	0.0034	0.0041	0.0049
0.9300,0.0500,0.0200	0.0026	0.0027	0.0032	0.0043	0.0039
0.5000,0.4800,0.0200	0.0021	0.0022	0.0026	0.0036	0.0033
0.9725,0.0200,0.0075	0.0030	0.0035	0.0060	0.0064	0.0062
0.9835,0.0100,0.0065	0.0030	0.0042	0.0067	0.0065	0.0120
0.9750,0.0200,0.0050	0.0030	0.0042	0.0056	0.0059	0.0074
0.9775,0.0175,0.0050	0.0030	0.0042	0.0056	0.0060	0.0059
0.9700,0.0200,0.0100	0.0027	0.0034	0.0042	0.0049	0.0062
0.9675,0.0200,0.0125	0.0027	0.0034	0.0038	0.0060	0.0057

Table 6. α for different sample sizes and $q = 4$ using Gamma distribution $X_{0,0027}$ quantile.

p_1, p_2, p_3, p_4	n=3000	n=1000	n=500	n=300	n=200
0.8000,0.1000,0.0600,0.0400	0.0024	0.0026	0.0028	0.0030	0.0033
0.6000,0.2500,0.1100,0.0400	0.0022	0.0024	0.0025	0.0026	0.0029
0.5500,0.2500,0.1500,0.0500	0.0024	0.0025	0.0027	0.0030	0.0032
0.8500,0.0500,0.0500,0.0500	0.0021	0.0022	0.0022	0.0024	0.0025
0.7000,0.1500,0.0900,0.0600	0.0022	0.0022	0.0023	0.0026	0.0026
0.5000,0.2900,0.1500,0.0600	0.0020	0.0020	0.0022	0.0023	0.0024
0.6000,0.3400,0.0525,0.0075	0.0028	0.0034	0.0044	0.0051	0.0060
0.5400,0.2500,0.2025,0.0075	0.0025	0.0031	0.0040	0.0047	0.0054
0.6000,0.3700,0.0200,0.0100	0.0028	0.0033	0.0040	0.0054	0.0065
0.8000,0.1400,0.0500,0.0100	0.0025	0.0027	0.0030	0.0036	0.0039
0.7000,0.1700,0.1100,0.0200	0.0024	0.0027	0.0030	0.0036	0.0040
0.5500,0.3500,0.0800,0.0200	0.0024	0.0027	0.0030	0.0036	0.0040
0.9450,0.0300,0.0150,0.0100	0.0030	0.0037	0.0046	0.0058	0.0068
0.9700,0.0125,0.0100,0.0075	0.0033	0.0042	0.0055	0.0067	0.0076
0.9650,0.0150,0.0100,0.0100	0.0032	0.0040	0.0050	0.0063	0.0076
0.9675,0.0150,0.0100,0.0075	0.0033	0.0042	0.0056	0.0065	0.0077
0.9430,0.0300,0.0150,0.0120	0.0031	0.0036	0.0046	0.0060	0.0063
0.9425,0.0300,0.0150,0.0125	0.0030	0.0036	0.0043	0.0053	0.0067
0.9350,0.0300,0.0200,0.0150	0.0029	0.0034	0.0040	0.0047	0.0059

Table 7. α for different sample sizes and $q = 5$ using Gamma distribution $X_{0,0027}$ quantile

p_1, p_2, p_3, p_4, p_5	n=3000	n=1000	n=500	n=300	n=200
0.7400,0.10,0.070,0.050,0.0400	0.0025	0.0026	0.0028	0.0031	0.0034
0.6500,0.13,0.100,0.080,0.0400	0.0024	0.0025	0.0026	0.0029	0.0031
0.6400,0.12,0.100,0.080,0.0600	0.0022	0.0023	0.0024	0.0025	0.0027
0.5500,0.20,0.100,0.080,0.0700	0.0021	0.0022	0.0023	0.0024	0.0026
0.5500,0.20,0.120,0.080,0.0500	0.0023	0.0023	0.0024	0.0026	0.0029
0.8000,0.05,0.050,0.050,0.0500	0.0025	0.0026	0.0028	0.0030	0.0033
0.6000,0.25,0.070,0.050,0.0300	0.0025	0.0027	0.0030	0.0033	0.0037
0.4000,0.27,0.200,0.100,0.0300	0.0022	0.0024	0.0026	0.0029	0.0032
0.7000,0.20,0.050,0.030,0.0200	0.0028	0.0030	0.0035	0.0040	0.0049
0.5000,0.23,0.150,0.100,0.0200	0.0024	0.0026	0.0030	0.0033	0.0040
0.8000,0.10,0.050,0.040,0.0100	0.0029	0.0035	0.0041	0.0055	0.0064
0.6000,0.15,0.140,0.100,0.0100	0.0026	0.0031	0.0037	0.0049	0.0057
0.9100,0.05,0.020,0.010,0.0100	0.0033	0.0041	0.0050	0.0068	0.0081
0.9105,0.05,0.020,0.010,0.0095	0.0033	0.0041	0.0052	0.0063	0.0075
0.9425,0.03,0.010,0.010,0.0075	0.0034	0.0045	0.0056	0.0070	0.0086
0.9250,0.04,0.020,0.010,0.0050	0.0035	0.0047	0.0068	0.0077	0.0082
0.9200,0.05,0.015,0.010,0.0050	0.0035	0.0047	0.0068	0.0078	0.0083
0.9000,0.05,0.020,0.015,0.0150	0.0031	0.0036	0.0045	0.0053	0.0067
0.9080,0.05,0.020,0.010,0.0120	0.0032	0.0039	0.0048	0.0062	0.0071
0.9060,0.05,0.020,0.012,0.0120	0.0031	0.0038	0.0047	0.0061	0.0069
0.9030,0.05,0.020,0.010,0.0170	0.0031	0.0037	0.0044	0.0056	0.0065

Higher weight values are assigned to categories with poor quality in order to detect earlier process deterioration and delay process improvement detection. It is proposed that $w_j = \frac{j}{q}$.

The Q statistic is a weighted distance between the observed counts and theoretical counts computed under the hypothesis that count occurrence in different categories are independent. In order to control the Q statistic values, a one-sided control chart is considered where the Q statistic distribution quantile of order α is the upper control limit, where α is the targeted false alarm rate. Since the Q statistic distribution is unknown, the α quantile is approximated by a quantile of the distribution in Equation 5 as explained previously. Q statistic control chart is illustrated through the following example. The example of (Marcucci, 1985) is reconsidered, where in brick manufacturing, a brick could be classified as suitable for all uses (standard), or (chipped-face) if sound but not suitable for all uses or unacceptable for use (cull). Classification probabilities are 95%, 3%, 2% respectively and the 95th percentile of the $\chi^2(2)$ distribution is used as the upper control limit for the Y_1^2 statistic defined in Equation 1. In the same way the 95th percentile of the approximated distribution of the Q statistic is used as the upper control limit for Q statistic control chart where $w_j = \frac{j}{J}$ with $j=1,2,3$ in this case. Hence, 5.99 and 5.47 are the control limits for the χ^2 chart and the Q chart respectively. Table 8 gives the control charts statistics computation. Both charts behave identically and show that samples 5, 10, 11 and 14 are out of control.

Table 8. Control Charts Statistics Computation.

Brick Classification						
Chipped						
Time	Standard	Face	Cull	Total	Y_i^2	Q_i
1	242	8	4	254	0.25	0.24
2	199	5	3	207	0.58	0.48
3	228	10	5	243	1.05	0.69
4	193	5	3	201	1.78	0.38
5	214	15	3	232	10.05	6.83
6	132	4	2	138	0.22	0.21
7	206	7	5	218	0.13	0.12
8	146	5	4	155	0.30	0.28
9	207	7	7	221	1.57	1.53
10	174	24	8	206	57.44	38.71
11	223	12	10	245	8.66	7.40
12	204	12	5	221	4.59	3.03
13	196	8	8	212	3.76	3.66
14	225	10	10	245	6.52	6.03
15	225	7	5	237	0.16	0.01
16	141	2	5	148	2.75	2.30

6 Control Charts Comparison

A simulation is implemented in order to compare between the (Marcucci, 1985) χ^2 control chart and the Q statistic-based control chart with a UCL approximated by a linear combination of independent chi squares distribution quantile. The upper control limits for the Q chart are 10.6475, 11.5732 and 12.4866 for $q = 3$, $q = 4$ and $q = 5$ respectively. A multinomial distribution is used as an in-control situation, large number of samples are generated and the order of the first out of control statistic value is determined. These steps are repeated 5000 times and the inverse of the average order is considered as the A.R.L. when the process is in control. Two scenarios are then considered to represent out of control situations. The A.R.L. is computed in the case of process improvement and in the case of process deterioration.

Table 9. Average Run Length for $q = 3$.

In Control Process	n=3000		n=1000		n=500		n=300		n=200	
	χ^2	Q	χ^2	Q	χ^2	Q	χ^2	Q	χ^2	Q
0.9725,0.02,0.0075	316.87	370.13	246.70	332.84	210.47	222.92	188.27	216.04	127.96	167.82
Process Improvement										
0.9835,0.0100,0.0065	1.23	1.68	14.13	61.17	172.15	405.29	366.79	350.37	302.48	302.79
0.9750,0.0200,0.0050	52.42	41.43	425.85	1179.14	407.01	1505.66	439.29	957.015	292.88	628.05
0.9775,0.0175,0.0050	23.30	26.26	485.46	1602.78	1297.15	3544.37	1166.94	1784.83	620.21	1050.71
Process Deterioration										
0.9425,0.05,0.0075	1.00	1.00	1.01	1.03	1.19	1.39	1.69	2.15	2.34	3.24
0.9700,0.02,0.0100	12.68	10.44	30.79	31.11	48.10	38.92	56.80	53.74	46.44	51.44
0.9675,0.02,0.0125	2.16	1.96	6.88	6.79	14.15	11.52	21.15	19.30	20.64	21.50
0.6500,0.34,0.0100	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Simulation results are in Tables 9, 10 and 11. It is noticed that Q control chart produces less false alarms than the χ^2 control chart and that both are appropriate for large samples $n \geq 300$. The Q chart has a higher A.R.L. than the goodness of fit control chart in case of process improvement and a smaller A.R.L. in case of process deterioration. The Q control chart shows less points in average beyond the control chart limit in case of process improvement and allows earlier detection in case of process deterioration, however, these main benefits are observed specifically when the process improvement and process deterioration concern only extreme categories, for $q = 4$ and $q = 5$ and only when the UCL for the Q chart is efficiently approximated.

Table 10. Average Run Length for $q = 4$.

In Control Process	n=3000		n=1000		n=500		n=300		n=200	
	χ^2	Q	χ^2	Q	χ^2	Q	χ^2	Q	χ^2	Q
0.9450,0.030,0.015,0.0100	325.38	396.67	276.96	313.40	222.44	257.66	184.64	185.12	144.78	154.62
Process Improvement										
0.9700,0.0125,0.01,0.0075	1.00	1.00	1.83	6.13	15.2	160.15	299.03	852.26	733.43	738.50
0.9650,0.0150,0.01,0.0100	1.01	1.11	3.44	15.84	29.86	116.25	161.11	176.65	220.72	184.93
0.9675,0.0150,0.01,0.0075	1.01	1.07	3.00	11.77	32.64	300.08	490.02	993.93	870.21	806.19
Process Deterioration										
0.9430,0.030,0.015,0.0120	40.21	30.46	79.43	62.24	90.47	77.39	95.81	72.61	87.45	72.47
0.9125,0.030,0.015,0.0125	23.81	17.03	59.08	45.64	72.56	60.15	78.59	58.14	74.97	59.59
0.9350,0.030,0.020,0.0150	1.84	1.77	6.93	6.39	13.21	12.25	19.22	16.77	23.28	20.28

Table 11. Average Run Length for $q = 5$.

In Control Process	n=3000		n=1000		n=500		n=300		n=200	
	χ^2	Q	χ^2	Q	χ^2	Q	χ^2	Q	χ^2	Q
0.91,0.05,0.02,0.01,0.01	324.3	380.0	276.5	297.1	214.2	231.8	173.9	170.7	134.8	147.2
Process Improvement										
0.9425,0.03,0.010,0.01,0.0075	1.0	1.0	1.8	6.7	10.6	92.7	59.8	227.9	180.9	243.9
0.9250,0.04,0.020,0.01,0.0050	1.8	2.3	25.6	62.3	123.9	301.7	224.1	375.2	232.6	345.4
0.9105,0.05,0.02,0.01,0.0095	350.6	478.9	309.5	403.5	247.33	297.7	192.6	211.3	164.6	161.6
Process Deterioration										
0.900,0.05,0.02,0.015,0.015	1.6	1.5	5.9	5.1	11.9	10.0	17.2	14.0	22.1	17.6
0.908,0.05,0.02,0.010,0.010	49.6	32.9	91.9	65.1	98.9	75.3	95.8	73.0	94.3	67.9
0.906,0.05,0.02,0.012,0.012	19.7	16.3	48.2	40.1	60.1	52.1	62.1	53.0	64.1	51.6
0.903,0.05,0.02,0.010,0.017	1.8	1.5	6.7	4.7	13.2	9.4	19.9	13.4	24.9	15.7

7 Conclusion



A weighted goodness of fit statistic is proposed to improve the (Marcucci, 1985) one-sided control chart performance in the case of multinomial process monitoring. As the proposed statistic distribution is not known, the control limit is approximated by a linear combination of independent chi squares distribution quantile. Performance assessment through simulation was performed for the cases where the UCL is well approximated and considering process improvement and process deterioration scenarios. Main benefits of earlier deterioration detection and postponing improvement detection are observed specifically for processes with less than 1% of product units into the worst quality category. Moreover, control chart performance improvement is clearer for $q = 4$ and $q = 5$. As the linear combination of chi squares offers poor approximation for the weighted chi squares statistic distribution, better performance is expected with the exact distribution of the control chart statistic.

References

- Bodenham, D. A., & Adams, N. M. (2016). A comparison of efficient approximations for a weighted sum of chi-squared random variables. *Statistics and Computing*, 26(4), 917-928. <http://dx.doi.org/10.1007/s11222-015-9583-4>.
- Castano-Martinez, A., & Lopez-Blazquez, F. (2005). Distribution of a sum of weighted noncentral chi-square variables. *Test*, 14(2), 397-415. <http://dx.doi.org/10.1007/BF02595410>.
- Cozzucoli, P. C. (2009). Process monitoring with multivariate p-Control chart. *International Journal of Quality, Statistics, and Reliability*, 2009, 1-11. <http://dx.doi.org/10.1155/2009/707583>.
- Davies, R. B. (1980). Algorithm AS 155: The distribution of a linear combination of chi-square random variables. *Journal of the Royal Statistical Society. Series C, Applied Statistics*, 29(3), 323-333.
- Davis, A. W. (1977). A differential equation approach to linear combination of independent chi-squares. *Journal of the American Statistical Association*, 72(357), 212-214. <http://dx.doi.org/10.1080/01621459.1977.10479941>.
- Duncan, A. J. (1950). A Chi-Square chart for controlling a set of percentages. *Industrial Quality Control*, 7, 11-15.
- Feiveson, A. H., & Delaney, F. C. (1968). *The distribution and properties of a weighted sum of chi squares* (NASA Technical Note, D-4575). Houston, Texas: Manned Spacecraft Center.
- Imhof, J. P. (1961). Computing the distribution of quadratic forms in normal variables. *Biometrika*, 48(3-4), 419-426. <http://dx.doi.org/10.1093/biomet/48.3-4.419>.
- Jensen, D. R., & Solomon, H. (1972). A Gaussian approximation to the distribution of a definite quadratic form. *Journal of the American Statistical Association*, 67(340), 898-902. <http://dx.doi.org/10.1080/01621459.1972.10481313>.
- Kotz, S., Johnson, N. L., & Boyd, D. W. (1967). Series representations of distributions of quadratic forms in normal variables. I. Central case. *Annals of Mathematical Statistics*, 38(3), 823-837. <http://dx.doi.org/10.1214/aoms/1177698877>.
- Li, J., Tsung, F., & Zou, C. (2014a). A simple categorical chart for detecting location shifts with ordinal information. *International Journal of Production Research*, 52(2), 550-562. <http://dx.doi.org/10.1080/00207543.2013.838329>.
- Li, J., Tsung, F., & Zou, C. (2014b). Multivariate binomial/multinomial control chart. *IIE Transactions*, 46(5), 526-542. <http://dx.doi.org/10.1080/0740817X.2013.849830>.
- Marcucci, M. (1985). Monitoring multinomial processes. *Journal of Quality Technology*, 17(2), 86-91. <http://dx.doi.org/10.1080/00224065.1985.11978941>.
- Mathai, A. M., & Provost, S. B. (1992). *Quadratic forms in random variables: Theory and Applications* (vol. 126). New York: Marcel Dekker Inc.

- Moschopoulos, P. G., & Canada, W. B. (1984). The distribution function of linear combination of chi-squares. *Computers & Mathematics with Applications (Oxford, England)*, 10(4/5), 383-386. [http://dx.doi.org/10.1016/0898-1221\(84\)90066-X](http://dx.doi.org/10.1016/0898-1221(84)90066-X).
- Oman, S. D., & Zacks, S. (1981). A mixture approximation to the distribution of a weighted sum of chi-squared variables. *Journal of Statistical Computation and Simulation*, 13(3-4), 215-224. <http://dx.doi.org/10.1080/00949658108810498>.
- Perry, M. B. (2019). An EWMA control chart for categorical processes with applications to social network monitoring. *Journal of Quality Technology*. <http://dx.doi.org/10.1080/00224065.2019.1571343>.
- Raz, T., & Wang, J. (1990). On the construction of control charts using linguistic variables. *International Journal of Production Research*, 28(3), 477-487. <http://dx.doi.org/10.1080/00207549008942731>.
- Taleb, H., & Limam, M. (2002). On fuzzy and probabilistic control charts. *International Journal of Production Research*, 40(12), 2849-2863. <http://dx.doi.org/10.1080/00207540210137602>.
- Taleb, H., Limam, M., & Hirota, H. (2006). Multivariate fuzzy multinomial control charts. *Quality Technology & Quantitative Management*, 3(4), 437-453. <http://dx.doi.org/10.1080/16843703.2006.11673125>.
- Topalidou, E., & Psarakis, S. (2009). Review of multinomial and multi-attribute quality control charts. *Quality and Reliability Engineering International*, 25(7), 773-804. <http://dx.doi.org/10.1002/qre.999>.
- Weiss, C. H. (2018). Control Charts for Time-Dependent Categorical Processes. In: S. Knoth & W. Schmid (Eds.) *Frontiers in statistical quality control 12* (pp. 211-231). Springer International Publishing.
- Yashchin, E. (2012). On detection of changes in categorical data. *Quality Technology & Quantitative Management*, 9(1), 79-96. <http://dx.doi.org/10.1080/16843703.2012.11673279>.

ERRATUM: Linear combination of chi-squares for multinomial process monitoring

Ramzi Talmoudi¹ , Ali Achouri² , Hassen Taleb²

¹University of Carthage, LR18ES48 ENVIE Laboratory, Faculté des Sciences Economiques et de Gestion de Nabeul, Tunisia. E-mail: ramzi.telmoudi@fsegn.u-carthage.tn

²University of Carthage, ARBRE Laboratory, Tunisia

Due to desktop publishing error the article “Linear combination of chi-squares for multinomial process monitoring” (DOI <https://doi.org/10.1590/1806-9649-2021v28e41>), published in *Gestão & Produção*, 28(3), e41 was published with an error.

On page 1, where the text reads:

ORIGINAL ARTICLE

It should read:

THEMATIC SECTION: STATISTICAL PROCESS MONITORING AND CONTROL

The publisher apologizes for the errors.

