On nonparametric Phase I analysis of individual observations from a change-point model perspective

C. Parpoula

Abstract Most of the univariate nonparametric control charts existing in the literature are designed for Phase II analysis, while little has been done in developing such Phase I control charts. Moreover, typical applications of control charts involve sub-grouped data, while recent advances have led to more and more instances where individual measurements are collected over time. Therefore, more research needs to be done regarding nonparametric Phase I analysis for small subgroup sizes and even for individual observations. More recent work dealt with this problem from a change-point model perspective utilizing rank or likelihood ratio-based statistics. Toward this end, in this paper, existing Phase I nonparametric control charts - directly applied to individual data - are compared in order to highlight their properties, pros and cons, as well as their efficacy under different distributions and shift sizes.
Key words and phrases: change-point model; control charts; individual data; nonparametric model; Phase I analysis.

1 Introduction

Despite the widely recognized importance of Phase I analysis in Statistical Process Control (SPC), most of the control charts existing in the literature are designed for Phase II analysis. For many decades, control charts could be found typically in manufacturing operation, while as nonmanufacturing applications continued to spread, new research challenges inevitably have arisen. For example, the underlying distribution in nonmanufacturing processes is usually unknown and not normal. Given these two concerns, nonparametric control charts are a useful and robust alternative to the prac-

C. Parpoula
Department of Psychology, Panteion University of Social and Political Sciences, Athens, Greece, e-mail: chparpoula@panteion.gr

titioner. Moreover, typical SPC applications involve sub-grouped data, while recent advances have led to more and more instances where individual measurements are collected over time. In this contribution, a simulation study is conducted comparing some of the most recent Phase I nonparametric control charts - developed from a change-point model perspective and directly applied to individual data - in order to highlight their properties, pros and cons, as well as their efficacy under different distributions and shift sizes. For a comprehensive review of the development of Phase I (and/or Phase II) nonparametric control charts for univariate (and/or multivariate) process monitoring, up until 2020, the interested reader may refer to Chakraborti and Graham [2]. The rest of the paper is organized as follows. In Section 2, we briefly discuss some of the existing univariate change-point model-based nonparametric control charts. Section 3 is devoted to simulation settings and results. Finally, in Section 4 some concluding remarks are made.

2 The existing methodologies

The most recently developed univariate nonparametric control charts utilize rank or likelihood ratio-based statistics, dealing with the problem of Phase I analysis from a change-point model perspective. Along these lines, Parpoula [5] investigated the in-control (IC) and out-of-control (OC) performance of such change-point model-based charts, considering that when the process is IC (stable), the individual observations are assumed to be independent and drawn from an unknown but common cumulative distribution function, whereas when the process is OC (unstable) the observations can be thought drawn by a general form of a nonparametric multiple change-point model describing processes subject to step, transient (and even isolated) shifts. Parpoula [5] examined the typical OC scenario of *a step change in the process mean*, assuming three different: a. underlying IC distributions and b. shift patterns regarding the random positions of the change times. Similar to the existing literature in which Phase I control charts are evaluated, we consider here an alternative OC scenario often encountered in practice, i.e., *a gradual shift in the process mean*. That is, for a given δ, the process mean $\mu(i)$, $i = 1, \ldots, m$, at the ith observation is defined by $\mu(i) = \mu_0 + \frac{(i-1)}{(m-1)} \times \delta\sigma_0$, where $\mu(1) = \mu_0$ and $\mu(m) = \mu_1 = \mu_0 + \delta\sigma_0$, and μ_0 and σ_0 are the IC mean and standard deviation, respectively, of a given distribution. The OC shift configurations considered here are of size $\delta = 0.25, 0.50, 1.00, 1.50, 2.00, 3.00$ (in units of standard deviations).

We then investigate the performance of existing competing Phase I nonparametric control charts, that is the Adjusted Generalized Likelihood Ratio (AdjGLR) test statistic-based chart, as in Parpoula and Karagrigoriou [6]; the RAdjGLR test statistic-based chart (identical to AdjGLR, except that a preliminary rank transformation of the original data is used), as in Parpoula and Karagrigoriou [6]; the Recursive Segmentation and Permutation (RS_P)-based chart (the "level" part of the RS_P procedure), as in Capizzi and Masarotto [1]; the Mann-Whitney (MW) test statistic-based chart, as in Hawkins and Deng [3]; the Cramer-von-Mises (CvM) and Kolmogorov-Smirnov (KS) test statistic-based charts, as in Ross and Adams [7]; the Empirical Likelihood Ratio (ELR) test statistic-based chart, as in Ning at al. [4]. Note that all considered charts are able to detect a single or multiple mean shifts in a sequence of individual observations, as well as both increases and decreases in the parameter being monitored.

3 Simulation settings & results

We follow similar simulation settings as in Parpoula [5]. Performance is evaluated using the probability of giving an alarm that has been estimated by simulations using 300,000 Monte Carlo replications. Note that 10,000 replications were used only for the ELR test statistic-based chart, as in Ning et al. [4], to sustain reasonable simulation execution times. We only consider one change-point location for the simplicity of the discussion. We set the nominal False Alarm Probability (FAP) to be 0.005. We consider two sample sizes $m = 50$ and $m = 100$. We consider a standard normal distribution, a negative exponential distribution of mean equal to 1, and a Student's t-distribution with 3 degrees of freedom, as in Parpoula [5]. The simulated IC signal probabilities for all considered charts can be found in Parpoula [5], thus are omitted here. Under the OC scenario "*the process mean undergoes a gradual shift*", the simulated OC signal probabilities are summarized in Table 1, for a given sample size, as a function of δ, which is a measure of the shift size.

Table 1 The OC signal probabilities for gradual mean shifts

Normal, m=50 (m=100)							
$\delta(\mu_0 = 0)$	AdjGLR	RAdjGLR	RS_P	MW	CvM	KS	ELR
0.25	0.011 (0.015)	0.011 (0.016)	0.009 (0.011)	**0.009 (0.014)**	0.009 (0.012)	0.009 (0.013)	0.006 (0.009)
0.50	0.028 (0.055)	0.030 (0.060)	0.022 (0.042)	**0.025 (0.056)**	0.025 (0.047)	0.025 (0.045)	0.016 (0.025)
1.00	0.138 (0.354)	0.150 (0.373)	0.117 (0.304)	**0.136 (0.363)**	0.131 (0.319)	0.116 (0.271)	0.083 (0.162)
1.50	0.406 (0.810)	0.427 (0.820)	0.364 (0.772)	**0.405 (0.817)**	0.387 (0.774)	0.332 (0.691)	0.238 (0.567)
2.00	0.730 (0.984)	0.746 (0.985)	0.692 (0.978)	**0.731 (0.985)**	0.709 (0.977)	0.628 (0.948)	0.507 (0.931)
3.00	0.989 (1.000)	0.990 (1.000)	0.986 (1.000)	**0.990 (1.000)**	0.987 (1.000)	0.967 (1.000)	0.953 (1.000)
Exponential, m=50 (m=100)							
$\delta(1/\lambda_0 = 1)$	AdjGLR	RAdjGLR	RS_P	MW	CvM	KS	ELR
0.25	0.008 (0.011)	0.009 (0.012)	0.008 (0.009)	0.008 (0.011)	0.008 (0.009)	0.008 (0.010)	**0.019 (0.027)**
0.50	0.011 (0.017)	0.017 (0.029)	0.013 (0.018)	0.014 (0.026)	0.014 (0.022)	0.015 (0.023)	**0.048 (0.065)**
1.00	0.022 (0.037)	0.043 (0.095)	0.029 (0.051)	0.038 (0.089)	0.037 (0.075)	0.036 (0.072)	**0.071 (0.126)**
1.50	0.034 (0.062)	0.079 (0.195)	0.047 (0.095)	0.070 (0.185)	0.069 (0.161)	0.064 (0.146)	**0.150 (0.266)**
2.00	0.045 (0.089)	0.121 (0.307)	0.066 (0.145)	0.109 (0.297)	0.106 (0.264)	0.097 (0.233)	**0.219 (0.400)**
3.00	0.065 (0.143)	0.208 (0.514)	0.101 (0.240)	0.193 (0.505)	0.188 (0.465)	0.167 (0.410)	**0.345 (0.662)**
Student, m=50 (m=100)							
$\delta(\mu_0 = 0)$	AdjGLR	RAdjGLR	RS_P	MW	CvM	KS	ELR
0.25	0.007 (0.011)	0.009 (0.013)	0.006 (0.007)	0.008 (0.011)	**0.008 (0.010)**	0.009 (0.012)	0.005 (0.009)
0.50	0.010 (0.015)	0.020 (0.038)	0.011 (0.013)	0.018 (0.035)	**0.019 (0.032)**	0.020 (0.035)	0.014 (0.026)
1.00	0.031 (0.051)	0.086 (0.212)	0.037 (0.062)	0.077 (0.204)	**0.082 (0.197)**	0.083 (0.195)	0.062 (0.140)
1.50	0.092 (0.165)	0.242 (0.563)	0.110 (0.215)	0.224 (0.555)	**0.238 (0.552)**	0.232 (0.534)	0.224 (0.437)
2.00	0.213 (0.361)	0.475 (0.861)	0.250 (0.468)	0.455 (0.860)	**0.476 (0.861)**	0.459 (0.842)	0.414 (0.724)
3.00	0.549 (0.706)	0.863 (0.997)	0.619 (0.856)	0.853 (0.997)	**0.872 (0.998)**	0.856 (0.997)	0.783 (0.958)

Table 1 shows that for any given control chart the signal probability is higher as the shift size δ increases and if the sample size is larger. The AdjGLR is generally dominated by the RAdjGLR test statistic-based chart which generally detects the shift with a high probability in almost all considered cases. However, Parpoula [5] pointed out that its attained IC signal probabilities are generally larger than the nominal FAP, and therefore may not be preferred in practice. The MW, ELR and CvM test statistic-based charts are the best performing charts under a normal, exponential and Student distribution, respectively. The RS_P and KS-test statistic based charts also provide reasonably good detecting power under the various distributions considered.

4 Concluding remarks

In this paper, we examined the performance of competing Phase I nonparametric control charting techniques (based on a change-point model formulation) for monitoring the process mean with individual observations. If the prevalent concern in Phase I analysis is that the process may incur a gradual mean shift, the derived results indicate that the MW, ELR and CvM test statistic-based charts are effective nonparametric Phase I control charts, with a satisfactory and robust performance under normal and nonnormal processes. It would also be worthwhile to study the effectiveness of these change-point model-based approaches for monitoring the process variability.

References

1. Capizzi, G., G. Masarotto, "Phase I distribution-free analysis of univariate data", *J. Qual. Technol.* 45:273-284, 2013.
2. Chakraborti, S., M.A. Graham, "Nonparametric (distribution-free) control charts: an updated overview and some results",*Qual. Eng.* 31:523-544, 2019.
3. Hawkins, D.M, Q. Deng, "A nonparametric change-point control chart", *J. Qual. Technol.* 42:165-173, 2010.
4. Ning, W., A.B. Yeh, X. Wu, B. Wang, "A nonparametric Phase I control chart for individual observations based on empirical likelihood ratio", *Qual. Reliab. Eng. Int.* 31:37-55, 2015.
5. Parpoula, C., "Phase I Nonparametric Control Charts for Individual Observations: A Selective Review and Some Results", In Y. Dimotikalis et al. eds. *Applied Modeling Techniques and Data Analysis 1: Computational Data Analysis Methods and Tools*, Vol. 7, Part 2, Chapter 13, pp. 233-248, iSTE Wiley, London, 2021.
6. Parpoula, C., A. Karagrigoriou, "On Change-Point Analysis-Based Distribution-Free Control Charts with Phase I Applications", In Koutras M., Triantafyllou I., eds. *Distribution-free Methods for Statistical Process Monitoring and Control*, Springer International Publishing, 157-182, 2020.
7. Ross, G.J, N.M Adams, "Two nonparametric control charts for detecting arbitrary distribution changes", *J. Qual. Technol.* 44:102-116, 2012.