EWMA *p* charts for detecting changes in mean or scale of normal variates

Sven Knoth

Sebastian Steinmetz

European University Viadrina Frankfurt (Oder)

Department of Business Administration and Economics

Discussion Paper No. 368

April 2015

ISSN 1860 0921

EWMA p charts for detecting changes in mean or scale of normal variates

Sven Knoth and Sebastian Steinmetz*

Department of Mathematics and Statistics, Helmut Schmidt University,

University of the Federal Armed Forces Hamburg

April 2015

Abstract

Methods of Statistical Process Control (SPC) are used for detecting deviations from regular processing. SPC is applied in manufacturing implementations where statistical tools are used to monitor the performance of production processes in order to identify and correct considerable changes in the process performance. Today, SPC methods are incorporated by organizations around the world as a suitable tool to improve product quality by reducing process variation. The current method of SPC is the application of control charts which are used to monitor process parameters (e. g., mean μ, standard deviation σ or percent defective p) over time. Well-established control chart schemes are, amongst others, exponentially weighted moving average charts (EWMA), cumulative sum charts (CUSUM) or, of course, the classical Shewhart charts. In this article, an EWMA control chart for variables calculating the percent defective $p = f(\mu, \sigma)$ will be presented where both process parameters are under risk to change. The scheme will be compared to several other control chart applications (EWMA $\bar{X}$, EWMA $\bar{X}$-S^2, and an alternative EWMA p chart). Numerical methods and Monte Carlo simulations are used for computing the average run length (ARL) as the measure of performance.

Keywords: SPC, ARL, control chart for variables, percent defective

*Corresponding author. E-Mail: steinmetz@hsu-hh.de.

1 Introduction

The objective of Statistical Process Control (SPC) is the utilization of statistical methods in order to monitor and improve processes. In addition to the academic discussion on methods of this part of applied statistics, it has a high level of relevancy for practice. Thus, these methods are applied for inspecting production or service processes in industry and economy alike.

Typically, the application of these methods is carried out by control charts. The concept of control charting dates back to the early 20th century when Walter A. Shewhart laid its basis. Shewhart developed a control chart for monitoring the percent defective of a manufacturing process called p chart. Essentially, the inspection of the percent defective is made under sampling by attributes. On the basis of a sample, an estimator of the current percent defective is plotted into a control chart which is kept under surveillance over time.

The general idea of operating a control chart consists of signaling an alarm if particularly specified thresholds (control limits) are exceeded. An alarm usually leads to stopping the process in order to eliminate the reasons for changes in the relevant process parameters. Characteristic of Shewhart type control charts is the feature of only taking the most recent sample into account. Therefore, Shewhart control charts perform very well when it comes to the detection of large changes in the process parameters.

The further development of the control chart fundamentals has progressed rapidly and a lot has been done since Shewhart introduced his original idea. Thus, numerous enhancements of the p chart exist to monitor other process parameters. Common schemes are charts inspecting the process mean μ (e. g., $\bar{X}$ chart) or the standard deviation σ (e. g., S chart). In these charts, the process parameters are usually monitored by their estimators $\bar{X}$ and S, respectively. In addition, new approaches were developed besides the Shewhart concept. In particular, control charts that take previous samples into account, have relevancy for both theory and practical application. This property leads to a better performance in case of small changes in the process parameters. For example, an exponentially weighted moving average scheme (EWMA: Roberts, 1959; Crowder, 1987; Lucas and Saccucci, 1990), or the cumulative sum chart (CUSUM: Page, 1954), were introduced. Other extensions deal with the type of parameter that is subject of inspection. While the popular p chart uses sampling by attributes, the $\bar{X}$ chart uses sampling by variables. By now, there are several combinations of control charts of, e. g., Shewhart, EWMA or CUSUM type, each for attributes or variables.

Krumbholz and Zöller (1995) introduced a newly designed p chart. In contrast to Shewhart's original idea, the percent defective is not sampled by attributes but by variables. Therefore, the percent defective is expressed as functional relationship of the mean and the standard deviation, $p = f(\mu, \sigma)$, which is inspected over time. The basic idea is to monitor both mean and standard deviation simultaneously by a single statistic and in this way by a single control chart. The typically one-sided p chart under sampling by variables uses only one design parameter (control limit) and signals an alarm as the estimator of p exceeds the specified control limit. Thus, the interpretation of the statistic and the control limit in terms of the percent defective can be intuitively and easily be done.

Knoth and Steinmetz (2013) extended the idea of monitoring the functional relationship $p = f(\mu, \sigma)$ and transferred it to the EWMA control chart concept. As a first step, they applied an EWMA p chart only to monitor changes in the process parameter μ through the lens of its yield impact. Here, the standard deviation σ is assumed to be known and constant. Effects of a change in the location parameter on the performance of the control chart are analysed. Thereby, the average run length (ARL) is used as a measure of performance and several control chart schemes are compared to each other. Steinmetz (2014) studied the EWMA p chart, where σ is still assumed to be known for setting up the chart, changes in both process parameters μ and σ are now taken into account and the correspondingly changed ARL is deployed for comparing with other EWMA schemes.

Intention of the present paper is to apply the EWMA p idea to a more realistic scenario. It will be compared to well established EWMA control chart schemes and an alternative EWMA p approach. The previous EWMA p chart will be extended by a design using both estimators $\bar{X}$ and S. For the case that both parameters changed, this chart is expected to perform better than the simpler version. After a short introduction to performance measurement of control charts and the basic approach of the EWMA p chart, the paper will focus on the opposing schemes and the mutual performance comparisons. For illustration reasons, the differences in the performance will be presented using both ARL tables and graphical ARL profiles. For the ARL computations, numerical methods as well as Monte Carlo simulations are applied.

2 Performance measurement

Within control chart applications, the question of performance measurement is omnipresent. Nowadays, various approaches are available to compare different control charts in terms of their performance. Here, the ARL concept will be used. Within ARL considerations, the in-control and the out-of-control case is distinguished. Let ℓ be the random number of observations or samples until the control chart stops and therefore signals an alarm. While L_0 is denoted as the average number of observations or samples until the control chart signals the first false alarm (the process is actually still under control), L_1 expresses the average number of observations or samples until the first correct alarm occurs. For a process with given ideal (in-control) process parameters μ_0 and σ_0, the ARL is defined as

$$ARL = \begin{cases} L_0 = E_\infty(\ell), & \text{in-control case } \mu = \mu_0, \sigma = \sigma_0 \\ L_1 = E_1(\ell), & \text{out-of-control case } \mu \neq \mu_0 \vee \sigma \neq \sigma_0 \end{cases} .$$

Thereby, the more general $E_m(\ell)$ denotes the expected value of the run length ℓ for a particular change point m which is unknown but fixed and not random. The parameter m marks the point in time when either μ, σ or both parameters change. Solely utilizing the ARL is more specific, because only two extreme change point positions are considered, $m = 1$ (instant change) and $m = \infty$ (no change).

Because both process parameters can have an effect on the ARL, one can express the ARL as a function of μ and σ, $L_{\mu,\sigma}$. Within this paper, results are based on ARL computation using Monte Carlo simulations, the Markov chain approach (Brook and Evans, 1972; Lucas and Saccucci, 1990), and the collocation method (Knoth, 2005).

3 EWMA p chart design and properties

Based on the idea of Krumbholz and Zöller (1995) and merged with the general idea of EWMA control charts, an EWMA p chart under sampling by variables was introduced by Knoth and Steinmetz (2013) and more thoroughly studied by Steinmetz (2014). The chart keeps advantages of monitoring continuous data and observing p as a function of μ and σ. Thus, the EWMA p chart is also able to check μ and σ simultaneously by only one statistic which can be intuitively interpreted. So far, only the mean was estimated by $\bar{X}$ and the standard deviation σ was still assumed to be known. Within this paper, the model will be extended by utilizing both parameters estimates, now.

3.1 Setup

Continuing the explanations above, one can introduce a more general EWMA p chart. The model will be modified in order to analyze the influence of both changes in mean and standard deviation.

Let $X \sim \mathcal{N}(\mu, \sigma^2)$ be an *iid* characteristic with certain in-control values μ_0 and σ_0. Samples of size n are taken at time $i = 1, 2, \ldots$ The sample mean is calculated by $\bar{X}_i = 1/n \sum_{j=1}^{n} X_{ij}$ with variance $Var(\bar{X}_i) = \sigma^2/n$. The sample variance is given by $S_i^2 = 1/(n-1) \sum_{j=1}^{n} (X_{ij} - \bar{X}_i)^2$. Φ and Φ^{-1} denote the cumulative distribution function (cdf) of the standard normal distribution and its inverse. Lower (LSL) and upper (USL) specification limits in terms of X are given, $LSL < USL$. As already Krumbholz and Zöller (1995) mentioned, the percent defective p can be determined by:

$$p = h(\mu, \sigma) = \Phi\left(\frac{LSL - \mu}{\sigma}\right) + \Phi\left(\frac{\mu - USL}{\sigma}\right). \tag{1}$$

For applying EWMA, an estimator of p has to be used, where process parameters μ and σ are replaced by their estimators $\bar{X}$ and S. The estimator $\hat{p}$ is finally given by

$$\hat{p}_i = h(\bar{X}_i, S_i) = \Phi\left(\frac{LSL - \bar{X}_i}{S_i}\right) + \Phi\left(\frac{\bar{X}_i - USL}{S_i}\right),$$

which is considerably biased. Given the simple $USL = 3 = -LSL$, one obtains, indeed, for the in-control target $h(\mu_0 = 0, \sigma_0 = 1) = 0.0027$ the expectation $E_\infty(\hat{p}_i) = 0.0133$.

The EWMA sequence is defined by

$$Z_i^p = (1 - \lambda) Z_{i-1}^p + \lambda \hat{p}_i \,,$$

where $0 < \lambda \leq 1$ specifies the usual EWMA weight parameter. It is initialized with a starting value of $Z_0^p = p_0 = h(\mu_0, \sigma_0)$ which represents the in-control case. Because only increasing p resembles deteriorated process quality, the EWMA p chart signals an alarm if $Z_i^p > c_{E,p}$. Thus, the run length $\ell_{E,p}$ is given by

$$\ell_{E,p} = \inf\left\{ i \in \mathbb{N} : Z_i^p > c_{E,p} \right\}.$$

The control limit $c_{E,p}$ is deployed for calibration at $\mu_0 = 0$, $\sigma_0 = 1$ (in-control), and the ARL is considered for situations of changed true process parameters μ and σ. The question is what impact a change would have on the performance of the control chart if certain changes in the process parameters occurred. The ARL results presented for the

EWMA p chart using $\bar{X}$ and S^2 are based on applying collocation to the ARL integral equation following Knoth and Steinmetz (2013). For details see the Appendix.

3.2 Distribution of $\hat{p}$

For ARL computation using the Markov chain approach, the cdf of $\hat{p}$ is required. Thus, the cdf of $\hat{p}$ in case of unknown (and hence estimated) mean and standard deviation is derived from the input given by Bruhn-Suhr and Krumbholz (1990) and Krumbholz and Zöller (1995). Let $0 < p < 1$. Set

$$\mu^* = \frac{LSL + USL}{2} \qquad \text{and} \qquad \sigma_p^* = \frac{LSL - USL}{2\Phi^{-1}(\frac{p}{2})} \quad .$$

The authors pointed out that for any σ with $0 < \sigma \leq \sigma_p^*$ there is exactly one $\mu = \mu_{p,\sigma} \geq \mu^*$ and one $\tilde{\mu} = \tilde{\mu}_{p,\sigma} \leq \mu^*$ that conform to $h(\mu_{p,\sigma}, \sigma) = h(\tilde{\mu}_{p,\sigma}, \sigma) = p$. By symmetry, it follows that

$$\tilde{\mu}_{p,\sigma} = \mu^* - (\mu_{p,\sigma} - \mu^*) = 2\mu^* - \mu_{p,\sigma} \qquad \text{and} \qquad \mu_{p,\sigma_p^*} = \mu^* \; .$$

Let the cdf of $\hat{p}$ in case of unknown μ and σ be denoted by

$$F_{\hat{p}}(p) = P\left(\hat{p} \leq p\right) = P\left(h(\bar{X}, S) \leq p\right) \; .$$

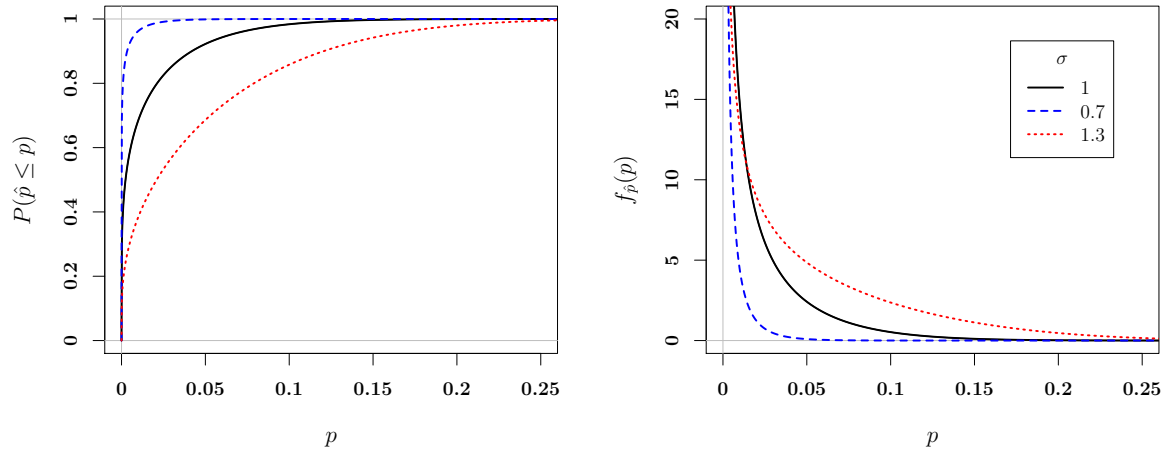
Bruhn-Suhr and Krumbholz (1990) derived

$$F_{\hat{p}}(p) = \int_0^{\alpha} \left\{ \Phi\left(\frac{\mu_{p,\tilde{\sigma}} - \mu}{\sigma} \sqrt{n} \right) - \Phi\left(\frac{\tilde{\mu}_{p,\tilde{\sigma}} - \mu}{\sigma} \sqrt{n} \right) \right\} \chi^2_{n-1}(x)\, dx \tag{2}$$

$$\text{with} \quad \alpha = \alpha(p) = \frac{n-1}{\sigma^2} \, \frac{(LSL - USL)^2}{4} \, \frac{1}{(\Phi^{-1}(\frac{p}{2}))^2} \quad \text{and} \quad \tilde{\sigma} = \tilde{\sigma}(x) = \sigma \sqrt{\frac{x}{n-1}} \, ,$$

where χ^2_{n-1} denotes the probability density function (pdf) of the χ^2 distribution with $n-1$ degrees of freedom. Figure 1 illustrates the shapes of the cdf and pdf of $\hat{p}$ for selected values of σ. The cdf $F_{\hat{p}}(p)$ is used for ARL computation of EWMA p charts applying the Markov chain approximation and, quite unusual, the collocation approach – see the Appendix for more details.

For applying numerical algorithms like the Nyström method or the common collocation approach to solve the ARL integral equation (Crowder, 1987, introduced it for the EWMA $\bar{X}$ chart), one usually deploys the pdf of $\bar{X}$, S^2 or $\hat{p}$. Differentiating $F_{\hat{p}}(p)$ in



(a) cdf (b) pdf

Figure 1: Distribution of $\hat{p} = h(\bar{X}, S)$: Cumulative distribution function (cdf) and density function (pdf) for $n = 5$, $\mu = 0$, $\sigma \in \{0.7, 1, 1.3\}$, and $USL = 3 = -LSL$.

(2) with respect to p and utilizing some arithmetics yields the following formula:

$$f_{\hat{p}}(p) = \int_0^{\alpha} \frac{\sqrt{n}/\sigma}{h'(\mu_{p,\tilde{\sigma}}, \tilde{\sigma})} \left[\varphi\left(\frac{\mu_{p,\tilde{\sigma}} - \mu}{\sigma} \sqrt{n} \right) + \varphi\left(\frac{-\mu_{p,\tilde{\sigma}} - \mu}{\sigma} \sqrt{n} \right) \right] \chi^2_{n-1}(x)\, dx\,. \qquad (3)$$

Thereby, $\varphi()$ denotes the pdf of the standard normal distribution and $h'(\mu, \sigma)$ stands for the derivative of $h(\mu, \sigma)$ with respect to μ. Both integrals, in (2) and (3), are calculated with Gauß-Legendre quadrature. Note the unpleasant behavior of the pdf, $f_{\hat{p}}(p)$, at $p = 0$ – it is unbounded and considerably steep, cf. Figure 1. The subsequent numerical problems while calculating the collocation definite integrals are treated by partial integration ending up again at the cdf (see Appendix).

4 Competing control chart schemes

4.1 EWMA $\bar{X}$ chart

The first analyzed scheme besides EWMA p is the ordinary EWMA $\bar{X}$ chart. It was introduced by Roberts (1959) and represents an extension of the classical $\bar{X}$ chart for monitoring the process mean. It has been examined in several papers such as Crowder (1987) or Lucas and Saccucci (1990) amongst others. With weight parameter λ_X ($0 < \lambda_X \leq 1$), the EWMA $\bar{X}$ sequence at time i is given by $Z_i^X = (1 - \lambda_X) Z_{i-1}^X + \lambda_X \bar{X}_i$. For

$i \to \infty$, the variance of Z_i^X follows $\sigma^2_{Z^X_\infty} = \sigma^2/n \left(\frac{\lambda_X}{2-\lambda_X}\right)$. The run length of this EWMA control chart is then defined by its stopping time

$$\ell_{E,X} = \inf\left\{i \in \mathbb{N} \colon |Z_i^X - \mu_0| > c_{E,X}\sigma_{Z^X_\infty}\right\},$$

where $c_{E,X}$ is a constant design parameter. Moreover, $Z_0^X = \mu_0$ is the center line of the control chart and the initial value of the EWMA sequence. For the performance comparison, the above two-sided EWMA $\bar{X}$ control chart for various values of λ_X is utilized. Critical values $c_{E,X}$ are each calibrated for $\mu_0 = 0$ and $\sigma_0 = 1$ leading to $L_0 = 370.4$.

Even though the ordinary EWMA $\bar{X}$ is usually not set up for detecting changes in the standard deviation, it reacts on these changes in an appropriate way.

4.2 EWMA $\bar{X}$-S^2 charts

There are certain schemes monitoring mean and variance simultaneously – see Domangue and Patch (1991) for the so-called omnibus charts, Gan (1995) for combined EWMA $\bar{X}$-$\ln S^2$ control charts or Reynolds Jr. and Stoumbos (2001a,b, 2004) who presented some *ARL* studies. Here, the EWMA $\bar{X}$-S^2 scheme is applied – see Knoth (2007) and references therein for *ARL* calculation methods of adequate accuracy. The scheme is based on two simultaneous EWMA sequences. While the two-sided mean chart is implemented as in Section 4.1, the EWMA S^2 chart is applied with weight parameter λ_S ($0 < \lambda_S \leq 1$) and starting value $Z_0^S = \sigma_0^2$. Thus, the EWMA sequence of the variance chart at time i is given by $Z_i^S = (1-\lambda_S)Z_{i-1}^S + \lambda_S S_i^2$. The run length of the EWMA $\bar{X}$-S^2 scheme with a one-sided variance chart is finally defined by $\ell_{E,XS} = \min\left\{\ell_{E,X}, \ell_{E,S}\right\}$ following

$$\ell_{E,X} = \inf\left\{i \in \mathbb{N} \colon |Z_i^X - \mu_0| > c_{E,X}\sigma_{Z^X_\infty}\right\},$$
$$\ell_{E,S} = \inf\left\{i \in \mathbb{N} \colon Z_i^S - \sigma_0^2 > c_{E,S_{(u)}}\sigma_{Z^S_\infty}\right\},$$

where $c_{E,X}$ and $c_{E,S_{(u)}}$ are the design parameters of the EWMA $\bar{X}$ chart and the one-sided (upper) EWMA S^2 chart. Note that both designs use thresholds written as multiples of the asymptotic standard deviation of the corresponding EWMA sequence.

Another opponent is an EWMA $\bar{X}$-S^2 scheme with a two-sided variance chart which

is defined as follows:

$$\ell_{E,S} = \inf\left\{i \in \mathbb{N}\colon Z_i^S - \sigma_0^2 \notin [-c_{E,S_{(l)}}\sigma_{Z_\infty^S}, c_{E,S_{(u)}}\sigma_{Z_\infty^S}]\right\}.$$

Here, $c_{E,S_{(l)}}$ denotes the design parameter of the lower EWMA S^2 chart. Contrary to the EWMA $\bar{X}$ chart, the control limits design is not symmetric anymore. Therefore, the concept of ARL unbiased charts is used – see, for example, Knoth (2007) for more details. The numerical algorithms which are presented in the same paper and which are implemented in the R package spc, are utilized to obtain the ARL results for both EWMA $\bar{X}$-S^2 designs.

4.3 Alternative EWMA p chart

Merging the ideas of an EWMA p chart and simultaneously calculating the EWMA $\bar{X}$-S^2 sequences, an alternative EWMA p chart can be introduced. Referring to Section 4.2, the two EWMA statistics Z_i^X and Z_i^S are put into (1):

$$\hat{p}_i^{alt} = h(Z_i^X, Z_i^S)\,.$$

Again, if $\hat{p}_i^{alt}$ becomes large, the percentage defective level will be certainly increased. Thus, the run length $\ell_{E,p}^{alt}$ is defined as

$$\ell_{E,p}^{alt} = \inf\left\{i \in \mathbb{N} : \hat{p}_i^{alt} > c_{E,p}^{alt}\right\}.$$

Unfortunately, numerical calculation of the ARL is almost infeasible, so that its ARL analysis is performed exclusively with Monte Carlo simulations (10^8 repetitions).

5 Comparison

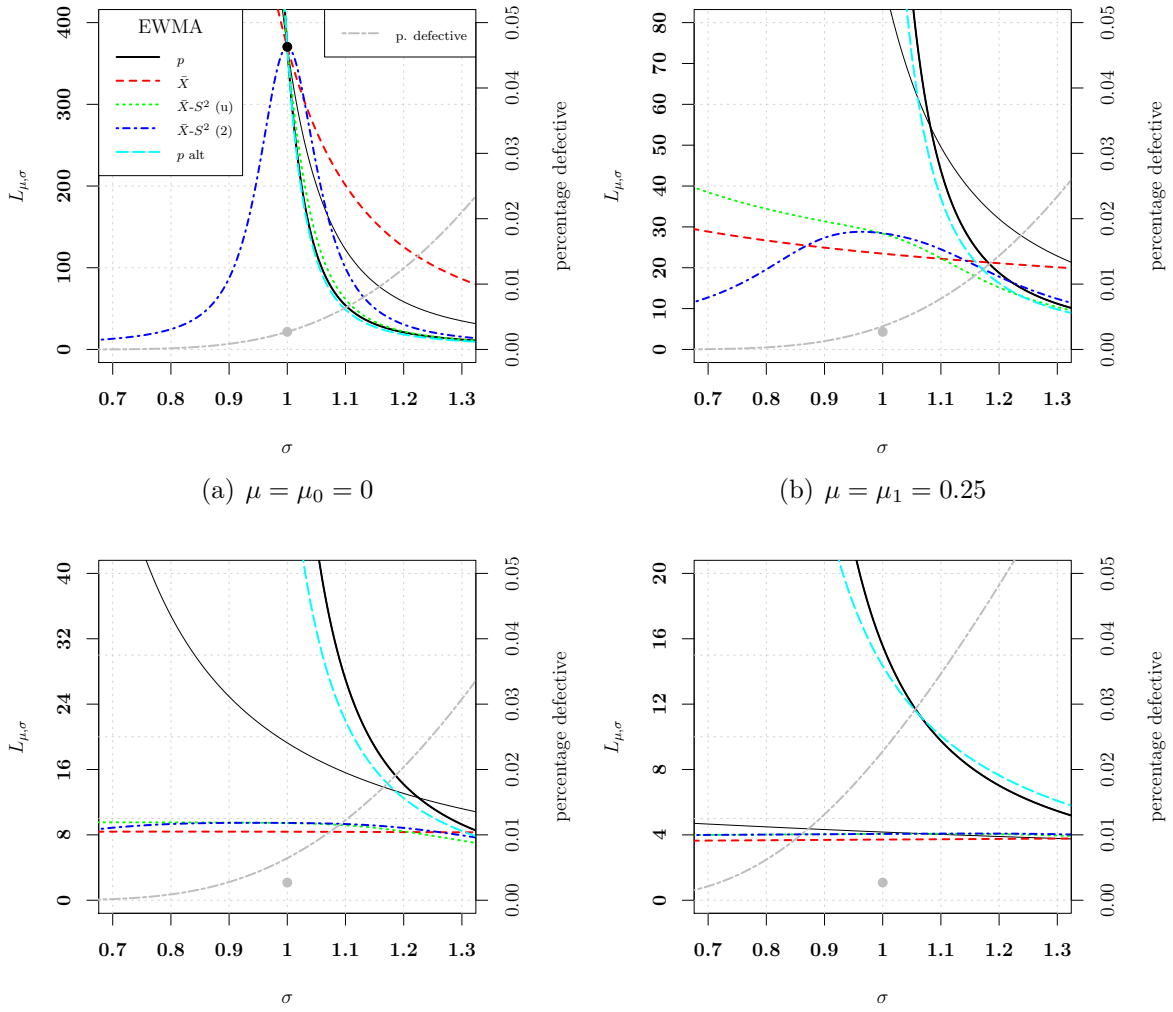
In this section, the different EWMA schemes will be compared to each other in case of changing mean and standard deviation. Thus, Table 1 shows the ARL results for various values of λ and changes in μ and σ. Each scheme is set up with a critical value calibrated for the in-control situation ($L_0 = 370.4$, $\mu_0 = 0$, $\sigma_0 = 1$), $n = 5$ and $USL = 3 = -LSL$. In each case, the values of λ are chosen equally for the compared schemes. An optimization of λ regarding a quick detection of a specific change in one or both process parameters was purposely not done, here. At the bottom of Table 1, the special case $\lambda = 1$ is presented which stands for the Shewhart type of chart. The

numerical ARL computations of the different schemes were done with the R package spc, version 0.5.0 and adequate extensions. The Monte Carlo simulations (10^8 replications) for the alternative EWMA p chart were implemented in C. The EWMA charts under consideration were introduced in Sections 3.1, 4, augmented by the simpler EWMA p chart studied in Steinmetz (2014) that assumes known variance. In Figure 2, the percentage defective is added as function of σ and the four different μ. They gray bullets mark the in-control value $p_0 = h(\mu_0, \sigma_0)$.

Figure 2 and Table 1 quickly leads to the finding that for $\mu \neq \mu_0$ the more general EWMA p scheme behaves worse than the simpler version which implies fixed and known variance. Surprisingly, this can already be observed for small shifts such as $\mu_1 = 0.25$ in Figure 2 (b). However, taking the yield impact of changes in μ and σ seriously means that a yield oriented control chart should not flag for $\mu = \mu_1 = 0.25$ as long as the standard deviation is smaller than 1.05. The two here introduced EWMA p schemes behave quite similarly with some advantages for the alternative one. For $\mu = \mu_0$, the aforementioned two EWMA p charts and the EWMA $\bar{X}$-S^2 (upper) exhibit the best behavior, while the other fall behind. For all other μ values, the first two chart types become worse than the competitors. Their ARL profiles are agreeable, at least for the small shift in Figure 2 (b). Given the patterns in Figure 2, one would recommend the EWMA $\bar{X}$-S^2 (upper) scheme. All other schemes display certain weaknesses. Besides the actual performance of the EWMA p chart, one has to keep in mind the manageable effort of implementing this scheme. This point can still be a good argument for practical application of the EWMA p chart in comparison with more complex approaches.

The EWMA $\bar{X}$, which is not originally set up to detect changes in σ, does not perform unexpectedly. Referring to the results of Steinmetz (2014), EWMA $\bar{X}$ outperforms its rivals in case of isolated changes in mean and almost stable in-control standard deviation. Here, the advantages of a single mean chart prevail which leads to top results in ARL performance considerations. Not surprisingly, the picture turns in case of isolated changes in the standard deviation and stable in-control mean. Here, EWMA $\bar{X}$ performs worst. Basically, the performance of this scheme becomes better as the change in mean increases. These findings are underlined by Figures 2 (c), (d). Here, the EWMA $\bar{X}$ scheme proceeds in a more and more flat curve shape and an excellent change detection regarding mean along both increasing and decreasing σ.

According to the ARL results in Table 1, the two different EWMA $\bar{X}$-S^2 schemes with either one- or two-sided variance chart perform mediocrely compared to the other control charts. The bigger the change in σ, the more an application of the S^2 chart pays



(a) $\mu = \mu_0 = 0$

(b) $\mu = \mu_1 = 0.25$

(c) $\mu = \mu_1 = 0.5$

(d) $\mu = \mu_1 = 1$

Figure 2: ARL profiles for $\lambda = 0.1$ and changes in μ and σ. The legend in the top-left diagram displays the color coding. Note that the bold solid line marks the profile of the more general EWMA p chart, while the thin one represents the EWMA p scheme focusing on changes in μ.

off. In case of isolated changes in the standard deviation and almost stable in-control mean, the scheme with one-sided S^2 chart outperforms its two-sided opponent. Apart from that, both schemes are almost equal in case of big changes in mean – differences in the performance dissolve.

The newly introduced alternative EWMA p chart instantly surprises with its suitable performance. Table 1 shows that the alternative EWMA p chart performs well in case of isolated and big changes in the standard deviation and stable in-control mean. Here,

the new scheme provides better ARL values than all of its opponents. The advantages of quickly detecting deteriorating standard deviation decrease by an increasing change in mean. When it comes to increasing changes in mean, the classical EWMA $\bar{X}$ and the two EWMA $\bar{X}$-S^2 schemes have better ARL performances. In this case, the two EWMA p charts do not differ significantly – see Figure 2 for increasing and decreasing standard deviation, specific changes in mean and fixed $\lambda = 0.1$. For increasing EWMA weight parameter λ, the ARL values of the two EWMA p charts converge. Thus, it follows from Section 3.1 and Section 4.3 that both schemes are equivalent in the Shewhart case ($\lambda = 1$) – slight inaccuracies in case of Monte Carlo simulations excepted. Note that only for the alternative EWMA p Monte Carlo simulations were utilized, while for the Shewhart chart based on $h(\bar{X}_i, S_i)$ the cdf $F_{\hat{p}}()$ given in (2) was applied.

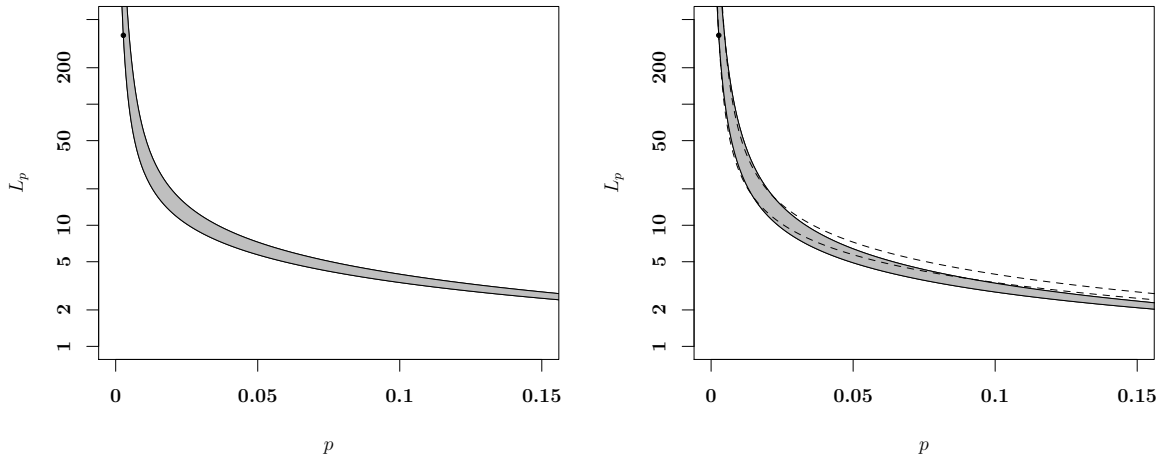
Apart from the mentioned relationship between the two EWMA p schemes and an increasing value of λ, Table 1 carves out that the choice of λ does not have a big impact on the (relative) performance of the different schemes. Thus, the differences of the applied schemes regarding their ARL performance is predominantly independent from the chosen value of λ.

Following Krumbholz and Zöller (1995) and their OC bands, we want to illustrate the issue that different μ-σ combinations fulfill $p = h(\mu, \sigma)$, so that the ARL as function of the actual percentage defective p is not uniquely defined anymore. For $\lambda = 0.1$ and in-control ARL 370.4, the resulting ARL band of the EWMA p chart is shown in Figure 3. The lower limit of the band relates to the largest possible $\sigma = \sigma_p^*$ depending on p, while the upper one corresponds to $\sigma \approx 0$. Hence, the smaller σ, the larger the detection delay, if p is fixed. In practice however, one would not observe standard deviations that are close to zero. Lastly, the bullet at $(p_0, 370.4)$ marks the calibration point of the EWMA p chart.

The graph in Figure 3 (d) corresponds directly to Krumbholz and Zöller (1995), Figure 1, where OC bands for various samples sizes n were drawn. Recall the simple relationship $ARL = 1/(1-OC)$ for Shewhart charts. From the shapes in Figure 3 we might conclude that there is a considerable spread of the ARL for all considered values of p. In terms of the out-of-control ARL level, the EWMA p chart with $\lambda = 0.2$ provides the most balanced performance. Eventually, the EWMA p performance is much less sensitive to the choice of λ than the one of the more popular EWMA $\bar{X}$ charts.

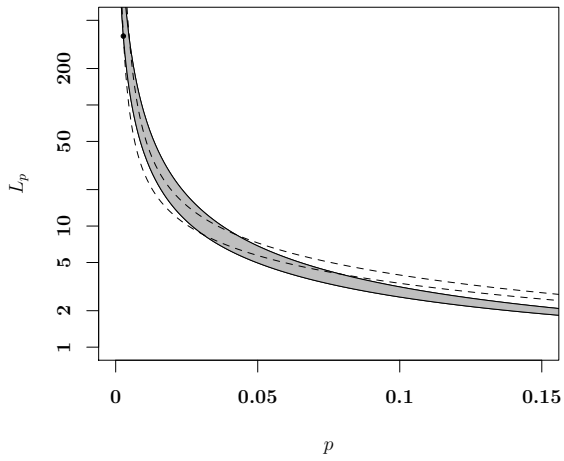
Table 1: ARL for various values of λ and changes in μ and σ.

λ	0.05															
μ	0				0.5				1.0				2.0			
σ	0.9	1.0	1.1	1.2	0.9	1.0	1.1	1.2	0.9	1.0	1.1	1.2	0.9	1.0	1.1	1.2
EWMA p	60879	370.4	54.7	24.3	1225	84.0	30.3	17.3	36.3	18.8	12.3	9.03	4.04	3.49	3.11	2.84
EWMA p, μ only	2786	370.4	117	59.6	27.3	22.2	18.6	15.8	5.60	5.36	5.13	4.92	1.36	1.39	1.42	1.44
EWMA $\bar{X}$	733	370.4	222	149	9.38	9.41	9.43	9.45	4.43	4.45	4.47	4.49	2.27	2.30	2.31	2.33
EWMA $\bar{X}$-S^2	1686	370.4	54.3	21.4	10.6	10.6	10.4	9.68	4.90	4.92	4.93	4.91	2.55	2.55	2.55	2.55
EWMA $\bar{X}$-S^2 (two)	71.1	370.4	78.6	27.5	10.6	10.6	10.6	10.1	4.90	4.92	4.93	4.94	2.54	2.55	2.55	2.55
EWMA p (alt)	113589	370.4	43.4	17.8	440	53.8	22.8	13.6	27.6	17.3	12.3	9.28	8.37	7.22	6.28	5.50
λ	0.1															
μ	0				0.5				1.0				2.0			
σ	0.9	1.0	1.1	1.2	0.9	1.0	1.1	1.2	0.9	1.0	1.1	1.2	0.9	1.0	1.1	1.2
EWMA p	16942	370.4	54.0	21.0	979	86.6	27.0	14.2	33.4	15.6	9.76	7.02	3.10	2.69	2.40	2.20
EWMA p, μ only	1958	370.4	122	58.4	24.9	19.3	15.6	13.1	4.33	4.17	4.03	3.90	1.18	1.21	1.24	1.27
EWMA $\bar{X}$	846	370.4	201	125	8.39	8.38	8.37	8.34	3.69	3.71	3.73	3.75	2.00	2.00	2.00	2.00
EWMA $\bar{X}$-S^2	1929	370.4	61.4	22.0	9.50	9.44	9.17	8.42	4.04	4.06	4.07	4.04	2.07	2.09	2.10	2.11
EWMA $\bar{X}$-S^2 (two)	87.2	370.4	99.6	30.5	9.46	9.46	9.33	8.85	4.04	4.06	4.08	4.08	2.07	2.09	2.10	2.12
EWMA p (alt)	4171	370.4	48.8	18.3	482	58.2	22.0	12.5	24.4	14.4	10.1	7.65	6.12	5.34	4.70	4.18
λ	0.2															
μ	0				0.5				1.0				2.0			
σ	0.9	1.0	1.1	1.2	0.9	1.0	1.1	1.2	0.9	1.0	1.1	1.2	0.9	1.0	1.1	1.2
EWMA p	6836	370.4	61.8	21.6	798	98.8	28.8	13.7	36.5	15.2	8.69	6.15	2.60	2.26	2.03	1.86
EWMA p, μ only	1516	370.4	134	64.2	29.1	20.6	15.7	12.7	3.78	3.65	3.54	3.43	1.10	1.13	1.15	1.18
EWMA $\bar{X}$	970	370.4	182	106	8.30	8.14	7.98	7.83	3.12	3.15	3.17	3.19	1.63	1.62	1.61	1.61
EWMA $\bar{X}$-S^2	2067	370.4	74.1	25.2	9.71	9.39	8.86	7.90	3.41	3.43	3.44	3.41	1.78	1.76	1.75	1.74
EWMA $\bar{X}$-S^2 (two)	120	370.4	138	41.7	9.58	9.41	9.08	8.44	3.41	3.43	3.45	3.45	1.78	1.76	1.75	1.74
EWMA p (alt)	8419	370.4	59.1	20.6	5550	73.4	24.4	12.6	25.9	13.4	8.91	6.61	4.59	4.02	3.57	3.21
λ	1.0															
μ	0				0.5				1.0				2.0			
σ	0.9	1.0	1.1	1.2	0.9	1.0	1.1	1.2	0.9	1.0	1.1	1.2	0.9	1.0	1.1	1.2
EWMA p	2312	370.4	99.3	37.6	595	140	48.9	22.3	60.5	25.0	13.1	8.07	2.55	2.13	1.88	1.70
EWMA p, μ only	1165	370.4	157	80.5	54.8	33.4	22.9	17.0	5.05	4.50	4.10	3.81	1.05	1.08	1.10	1.12
EWMA $\bar{X}$	1165	370.4	157	80.5	54.8	33.4	22.9	17.0	5.05	4.50	4.10	3.81	1.05	1.08	1.10	1.12
EWMA $\bar{X}$-S^2	1748	370.4	112	44.8	96.1	50.6	29.3	18.0	7.09	5.97	5.17	4.52	1.09	1.11	1.14	1.17
EWMA $\bar{X}$-S^2 (two)	415	370.4	222	122	81.8	50.6	33.5	23.9	7.01	5.97	5.26	4.76	1.09	1.11	1.14	1.17
EWMA p (alt)	2315	370.4	99.2	37.6	595	140	48.8	22.3	60.6	25.0	13.1	8.07	2.55	2.13	1.88	1.70

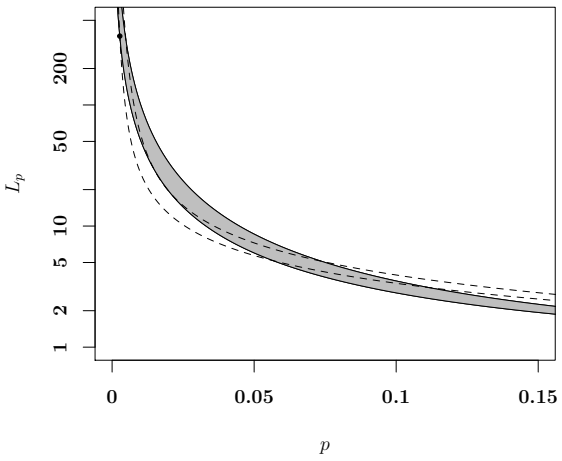


(a) $\lambda = 0.1$

(b) $\lambda = 0.2$



(c) $\lambda = 0.5$



(d) $\lambda = 1$

Figure 3: ARL as function of percentage defective p for four different λ values. The $\lambda = 0.1$ contours are added to the other three plots as benchmark.

6 Conclusions

Based on the basic idea of monitoring the percent defective p with a control chart for variables (Krumbholz and Zöller, 1995), an EWMA p chart for variables was introduced by Knoth and Steinmetz (2013) and studied for changing σ by Steinmetz (2014). The present paper applies the EWMA p idea to a more realistic scenario. In contrast to previous papers, the standard deviation is now in equal measure under the risk of change like the mean. Therefore, both process parameters μ and σ are estimated and used for the monitoring design.

For performance comparisons, the ARL concept is chosen. Here, the EWMA p chart is compared to the ordinary EWMA $\bar{X}$ chart, an EWMA $\bar{X}$-S^2 scheme with one-/two-sided variance chart and a newly introduced alternative EWMA p chart.

It turns out that both the EWMA p chart and the alternative EWMA p chart provide moderate ARL values in case of changing location and scale, even though the results of Knoth and Steinmetz (2013); Steinmetz (2014) promised a higher applicability. For the case of a separate change in the standard deviation σ, the two EWMA p charts perform well while the alternative has slight performance advantages. As the change in the mean μ becomes bigger, it is no surprise that the EWMA $\bar{X}$ and the combined EWMA $\bar{X}$-S^2 schemes outperform the EWMA p approaches. Thus, for sole changes in μ, the EWMA $\bar{X}$ chart clearly excels the other schemes. Finally, the results are irrespective of the chosen value of the EWMA weight parameter λ which leads to almost equal results within the performance comparisons.

The bottom line is that the application of a single EWMA p chart (regular or alternative) monitoring the percent defective $p = f(\mu, \sigma)$ still pays off. Because of the minor effort of applying a single chart in contrast to applying two separate charts monitoring mean and standard deviation, the EWMA p concept turns out to be an option for well established schemes.

In order to compute the ARL of the EWMA p chart, the collocation method is exploited. For the so-called alternative EWMA p chart, only Monte Carlo simulations seemed to be feasible.

References

D. Brook and D. A. Evans (1972). An approach to the probability distribution of CUSUM run length. *Biometrika* **59**(3), 539-549. doi: 10.1093/biomet/59.3.539

M. Bruhn-Suhr and W. Krumbholz (1990). A new variables sampling plan for normally distributed lots with unknown standard deviation and double specification limits. *Statistical Papers* **31**(1), 195-207. doi: 10.1007/BF02924690

S. V. Crowder (1987). A simple method for studying run-length distributions of exponentially weighted moving average charts. *Technometrics* **29**(4), 401-407. doi: 10.1080/00401706.1987.10488267

R. Domangue and S. C. Patch (1991). Some omnibus exponentially weighted moving average statistical process monitoring schemes. *Technometrics* **33**(3), 299-313. doi: 10.1080/00401706.1991.10484836

F. F. Gan (1995). Joint monitoring of process mean and variance using exponentially weighted moving average control charts. *Technometrics* **37**(4), 446-453. doi: 10.1080/00401706.1995.10484377

D. M. Hawkins (1992). Evaluation of average run lengths of cumulative sum charts for an arbitrary data distribution. *Communications in Statistics – Simulation and Computation* **21**(4), 1001-1020. doi: 10.1080/03610919208813063

S. Knoth (2005). Accurate ARL computation for EWMA-S^2 control charts. *Statistics and Computing* **15**(4), 341-352. doi: 10.1007/s11222-005-3393-z

S. Knoth (2007). Accurate ARL calculation for EWMA control charts monitoring simultaneously normal mean and variance. *Sequential Analysis* **26**(3), 251-263. doi: 10.1080/07474940701404823

S. Knoth and W. Schmid (2002). Monitoring the mean and the variance of a stationary process. *Statistica Neerlandica* **56**(1), 77-100. doi: 10.1111/1467-9574.03000

S. Knoth and S. Steinmetz (2013). EWMA p charts under sampling by variables. *International Journal of Production Research* **51**(13), 3795-3807. doi: 10.1080/00207543.2012.746799

W. Krumbholz and A. Zöller (1995). p-Karten vom Shewhartschen Typ für die messende Prüfung. *Allgemeines Statistisches Archiv* **79**, 347-360.

J. M. Lucas and M. S. Saccucci (1990). Exponentially weighted moving average control schemes: Properties and enhancements. *Technometrics* **32**(1), 1-12. doi: 10.1080/00401706.1990.10484583

E. S. Page (1954). Continuous inspection schemes. *Biometrika* **41**(1-2), 100-115. doi: 10.1093/biomet/41.1-2.100

M. R. Reynolds Jr. and Z. G. Stoumbos (2001a). Individuals control schemes for monitoring the mean and variance of processes subject to drifts. *Stochastic Analysis and Applications* **19**(5), 863-892. doi: 10.1081/SAP-120000226

M. R. Reynolds Jr. and Z. G. Stoumbos (2001b). Monitoring the process mean and variance using individual observations and variable sampling intervals. *Journal of Quality Technology* **33**(2), 181-205.

M. R. Reynolds Jr. and Z. G. Stoumbos (2004). Control charts and the efficient allocation of sampling resources. *Technometrics* **46**(2), 200-214. doi: 10.1198/004017004000000257

S. W. Roberts (1959). Control-charts-tests based on geometric moving averages. *Technometrics* **1**(3), 239-2s50. doi: 10.1080/00401706.1959.10489860

S. Steinmetz (2014). EWMA charts: ARL considerations in case of changes in location and scale. *Advances in Statistical Analysis* **98**(4), 371-387. doi: 10.1007/s10182-013-0221-1

A Numerics for calculating the *ARL* of EWMA *p*

Extending the Markov chain approach from Knoth and Steinmetz (2013) to the here introduced EWMA p chart is straightforward. One simply utilizes the more complex formula for $\hat{F}(p)$, namely (2). The resulting performance including the usual accelerating techniques like in Knoth and Steinmetz (2013) is illustrated in Figure 4. The changes to the ARL integral equation are two-fold. Of course, the formula behind $\hat{f}(p)$, (3), is more complex as well. And the original lower integral limit has to be replaced. Hence, the ARL integral equation is given by:

$$\mathcal{L}(z) = 1 + \int_{(1-\lambda)z}^{c_{E,p}} \mathcal{L}(x) \frac{1}{\lambda} f_{\hat{p}} \left(\frac{x - (1-\lambda)z}{\lambda} \right) dx \quad , \; z \in [0, c_{E,p}] \, .$$

Following the original collocation setup from Knoth and Steinmetz (2013), however, did not provide a suitable numerical procedure. The awkward behavior of $\hat{f}(p)$ for $p \to 0$ could not be counterbalanced by applying reasonable changes of variables. Nonetheless,

using partial integration is the key to get an applicable procedure. The definite integrals of the collocation procedure are replaced as follows (the notation is taken from Knoth and Steinmetz, 2013):

$$\begin{aligned}\sum_{j=1}^{N} c_j T_j^*(z_i) &= 1 + \sum_{j=1}^{N} c_j \int_{(1-\lambda)z_i}^{c_u} T_j^*(x)\frac{1}{\lambda} f_{\hat{p}}\left(\frac{x-(1-\lambda)z_i}{\lambda}\right) dx \\ &= 1 + F_{\hat{p}}\left(\frac{c_{E,p}-(1-\lambda)z_i}{\lambda}\right) - \int_{(1-\lambda)z_i}^{c_{E,p}} \frac{2}{c_{E,p}} t_j^*(x) F_{\hat{p}}\left(\frac{x-(1-\lambda)z_i}{\lambda}\right) dx\,.\end{aligned}$$

Thereby, $T_j^*()$ denotes modified Chebyshev polynomials, the nodes z_i are given by the roots of the corresponding Chebyshev polynomial of order $N+1$, and $t_j^*()$ marks the derivative of $T_j^*()$. The latter can be determined easily by applying the rule $t_j(x) = j/(1-x^2)\Big(T_{j-1}(x) - xT_j(x)\Big)$ for $j > 1$ (plus $t_0(x) \equiv 0$ and $t_1(x) \equiv 1$). The versions without "$*$" correspond to the ordinary Chebyshev polynomials on $[-1, 1]$. It holds that $t_j^*(z) = 2/c_{E,p}\, t_{j-1}(2z/c_{E,p} - 1)$ for $z \in [0, c_{E,p}]$ and, of course, $T_j^*(z) = T_{j-1}(2z/c_{E,p} - 1)$ (compare Knoth and Steinmetz, 2013) for $j = 1, 2, \ldots, N$. The parameter N stands for the number of Chebyshev polynomials involved in the approximation. The integrand of the new integral behaves well, so that one obtains sufficiently high accuracy with already 20 quadrature nodes.

See the following Figure 4 for illustration of the competing numerical algorithms. The EWMA smoothing parameter is set to $\lambda = 0.2$ which ends in $c_{E,p} = 0.044\,655$ yielding the common in-control ARL 370.4. Figure 4 displays numerical performance profiles for the ordinary Markov chain approximation, two improvements – polynomial extrapolation following Brook and Evans (1972) and the Richardson extrapolation due to Hawkins (1992) – and collocation as described above. For increasing matrix dimension N, the resulting approximation ARL_N is given. The number N is chosen as a very basic proxy for the computational amount aka time. Of course, the quadratures needed for collocation increase the time considerably. However, for $N = 15$, which ensures already high accuracy, the CPU time is less than 2 seconds on an Intel Core 2 Duo with 1.4GHz. The Markov chain approximation needs about the same time for $N = 64$ with much lower accuracy.

To validate the results, a Monte Carlo study with 10^9 replications was performed. It provides 370.415 with a standard error 0.012. A corresponding stripe with width of four times the standard errors is added to Figure 4 as well. The Monte Carlo result confirms the collocation numbers.

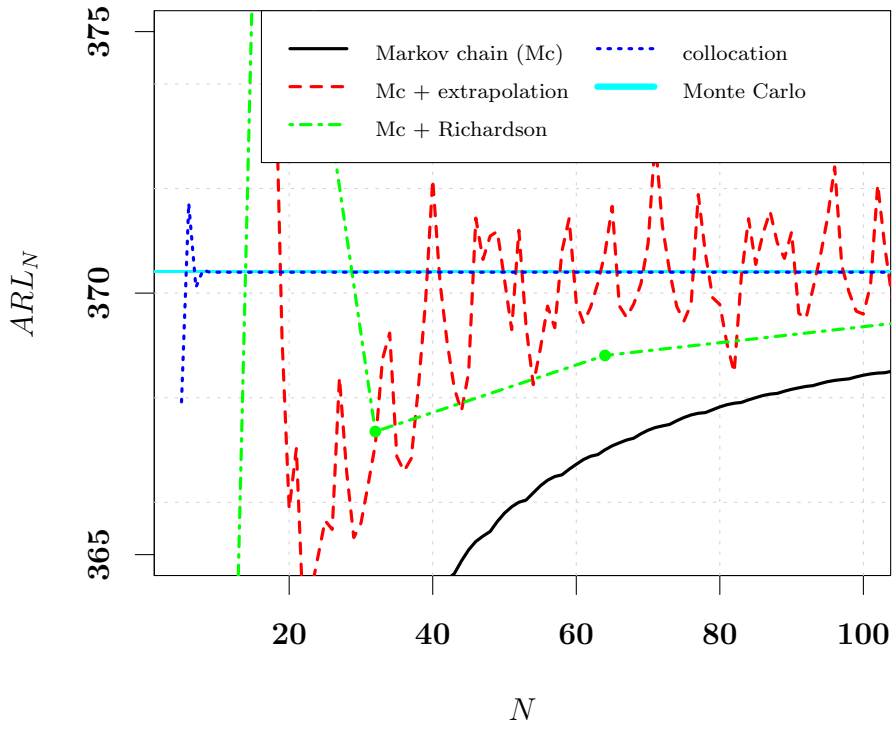


Figure 4: ARL approximation vs. matrix dimension N for the Markov chain method, two accelerators, and the collocation procedure.