

**Grouped Data Exponentially Weighted
Moving Average Control Charts**

Stefan H. Steiner

RR-96-05

October 1996

Grouped Data Exponentially Weighted Moving Average Control Charts

Stefan H. Steiner

Dept. of Statistics and Actuarial Sciences
University of Waterloo
Waterloo, Ontario N2L 3G1
Canada

Abstract

Methodology is proposed for the design of exponentially weighted moving average (EWMA) control charts when data are obtained by gauging articles into groups. Grouped data occur in industry through the use of step gauges and other similar measurement devices. Run length results for grouped data EWMA charts are compared with similar results previously obtained for Cumulative Sum (CUSUM) schemes based on grouped data and with EWMA charts for variables data. Grouped data EWMA charts are not as affected by the discreteness inherent in grouped data as grouped data CUSUM charts. The methodology is illustrated with an example from a progressive die operation. Grouped data EWMA charts are simple to implement, and are an economical alternative to variables data based EWMA control charts when collecting variables data is prohibitive.

Keywords: Cumulative Sum; CUSUM, EWMA, Grouped Data, Parametric Multinomial.

1. Introduction

In the quality control context, exponentially weighted moving average (EWMA) control charts are used to monitor process quality. EWMA charts, and other sequential approaches like Cumulative Sum (CUSUM) charts, are an alternative to Shewhart control charts especially effective in detecting small persistent process shifts. First introduced by Roberts (1959), EWMA charts have a fairly long history, but only recently have their properties been evaluated analytically (Crowder, 1987; Lucas and Saccucci, 1990). The EWMA also has optimal properties in some forecasting and control applications (Box, Jenkins, and MacGregor, 1974).

For monitoring a process an EWMA control chart consists of plotting:

$$z_t = \lambda x_t + (1 - \lambda)z_{t-1}, \quad 0 < \lambda \leq 1, \quad (1)$$

versus time t , where x_t is an estimate of the process characteristic we wish to monitor, λ is a constant weighting factor, and the starting value z_0 equals an a priori estimate of the monitored process parameter. In (1), x_t may represent the sample mean, sample standard deviation or any other empirically estimated process parameter. Writing out the recursion in (1), the EWMA test statistic z_t equals an exponentially weighted average of all previous observations. In contrast, tabular CUSUM charts assign all past observations, since the last time the CUSUM statistic equaled zero, equal weight (Montgomery, 1991). In quality monitoring applications of EWMA control charts, typical values for the weight λ are between 0.05 and 0.25.

From (1), the mean and variance of z_t , denoted μ_{z_t} and $\sigma_{z_t}^2$ respectively, are easily derived (Montgomery, 1991). Assuming the x_t 's are independent random variables with mean μ_x and variance σ_x^2 gives

$$\begin{aligned} \mu_{z_t} &= \mu_x, \quad \text{and} \\ \sigma_{z_t}^2 &= \sigma_x^2 \left(\frac{\lambda}{2 - \lambda} \right) \left[1 - (1 - \lambda)^{2t} \right] \cong \sigma_x^2 \left(\frac{\lambda}{2 - \lambda} \right) \quad \text{as } t \rightarrow \infty. \end{aligned} \quad (2)$$

Control limits for an EWMA control chart are typically derived based on $\pm L$ sigma limits, where L is usually equal to three as in the design of Shewhart control chart limits. The fact that the z_t 's are not independent is ignored. Thus, the control limits of an EWMA chart used to monitor the process mean are :

$$\mu_{z_t} \pm L\sigma_{z_t} = \mu_x \pm L\sigma_x \sqrt{\left(\frac{\lambda}{2-\lambda}\right) \left[1 - (1-\lambda)^{2t}\right]}, \quad (3)$$

where, in applications, μ_x and σ_x are typically estimated from preliminary data as the sample mean and sample standard deviation respectively. The process is considered out-of-control whenever the EWMA test statistic z_t falls outside the range of the control limits given by (3). Notice that in the limiting case with $\lambda = 1$ the EWMA chart is identical to a Shewhart \bar{X} control chart.

Previous work with EWMA control charts has focused on variables data, where the measurements are continuous random variables. However, it is not always possible or practical to use variables or precise measurement data in quality control. The widespread occurrence of binomial pass/fail attribute data in industry attests to the economic advantages of collecting go/no-go data over exact measurements. Variables data provide more information, but gauging, or classifying observations into one of a number of groups based on a critical dimension, is often preferred since it takes less skill, is faster, is less costly, and is a tradition in certain industries (Schilling, 1981, Ladany, 1976). Grouped data occur in industry because of multiple go/no-go gauges, step gauges, or other similar measurement devices (Steiner, Geyer and Wesolowsky, 1994). A step gauge with $(k-1)$ gauge limits yields k -group data. Note that pass/fail binomial attribute data represents the special case of two group data. For more information on grouped data, see Haitovsky (1982).

The development of control charting methodology for use with grouped data other than binomial data started with Stevens (1948). Stevens proposed two simple ad hoc

Shewhart control charts for simultaneously monitoring the mean and standard deviation of a normal distribution using three-group data. More recently, Beja and Ladany (1974) proposed using three group data to test for one sided shifts in the mean of a normal distribution with known process dispersion. In the realm of sequential quality control methods, Schneider and O'Conneide (1987) proposed a Cumulative Sum (CUSUM) scheme for monitoring the mean of a normal distribution with two-group data. Geyer, Steiner and Wesolowsky (1996) extended this CUSUM to the use of three-group data, with gauges symmetric about the midpoint between the target mean and the out-of-control mean that the chart should detect quickly. Gan (1990) proposed a modified EWMA chart for use with binomial data. The modified form of the EWMA uses (1), but rounds off the EWMA test statistic and calculates the run length properties using a Markov chain. Unfortunately, this solution approach is only appropriate when the expected number of failures in a sample is fairly large. As a result, the solution procedure typically requires large sample sizes, especially when the probability of failure is small. Steiner et al. (1994, 1996A) were the first to consider the general k -group case. They developed methodology for one-sided and two-sided acceptance sampling plans, acceptance control charts and Shewhart type control charts. In addition, Steiner et al. (1996B) considered k -group Sequential Probability Ratio Tests (SPRTs) and CUSUM procedures. These k -groups control charts use the likelihood ratio to derive an efficient test statistic. Steiner et al. (1994, 1996A, 1996B) also give design procedures for the various types of k -group control charts, calculate run length properties, and address the question of optimal gauge design. These articles show that k -group control charts are efficient alternatives to standard variables based techniques.

This article addresses the derivation of the general k -group EWMA control chart. Grouped data EWMA procedures bridge the gap between the efficiency of binomial attribute procedures and that of variables based EWMA charts. An important question, addressed later, pertains to how this loss of information in the raw data effects the

performance of the grouped data EWMA chart in comparison to variables based EWMA. For grouped data, EWMA charts may be a better choice than a CUSUM chart since, due to the exponential weighting of past observations, the EWMA smoothes out the inherent discreteness. This is an advantage that allows more flexibility in the design of grouped data EWMA charts as compared with grouped data CUSUM charts.

This article is organized in the following manner. Section 2 discusses two possible grouped data scoring procedures and recommends unbiased estimate scores for EWMA charts. In Section 3 and 4, EWMA control charts for the k -group case are developed; and the run length properties of grouped data and variables data based EWMA charts and grouped data CUSUM charts are compared. Section 5 provides an example of applying this methodology in an industrial situation, and Section 6 briefly discusses optimal gauge placement. The Appendix shows how the run length distribution of grouped data EWMA charts can be approximated using a Markov chain.

2. Sequential Scoring Procedures for Grouped Data

When using grouped data in control charts the need arises to assign the grouped observations a numerical value based on their grouping. For go/no-go gauges, observations are usually treated singly as Bernoulli random variables. However, when observations are grouped into multiple intervals, a number of different scoring/weighting procedures are feasible. This article considers two scoring schemes; namely, midpoint scores and unbiased estimate scores. In industry, group interval midpoints are used most often. However, as will be shown, midpoint weights have some undesirable properties, and if some additional process information is available unbiased estimate scores are a better choice.

Throughout this article, it is assumed that, although the data are grouped, there exists an underlying continuous measurement that is unobservable. Let X represent the underlying measurement, and let $t_1 < t_2 < \dots < t_{k-1}$ denote the $k-1$ endpoints or gauge

limits used to derive the k -group data. We assume, for the moment, $k - 1$ predetermined gauge limits. In many applications, the grouping criterion is fixed since it is based on some standard classification device or procedure. Section 6 addresses the relaxation of this assumption. Assume the random variable X has probability distribution $f(x; \theta)$ and cumulative distribution function $F(x; \theta)$, where θ is the process parameter of interest. Let w_j be the group weight or score assigned to all observations falling into the j th group. Defining $t_0 = -\infty$ and $t_k = \infty$, the probability that an observation falls into the j th interval is given by

$$\pi_j(\theta) = F(t_j; \theta) - F(t_{j-1}; \theta), \quad j = 1, 2, \dots, k. \quad (4)$$

Ideally, the group weights chosen have a physical interpretation. This makes the interpretation of the resulting control charts easier for the industrial personnel. This implies that the weight for each group should lie somewhere between the group gauge limits. Strictly applying this criterion precludes the use of likelihood ratio weights as suggested by Steiner et al. (1996A) in the Shewhart control chart context.

Furthermore, often the group weights are used not only in a control chart, but also to estimate the current process mean and variance so that we can calculate process capability measures. Typically the process mean and variance are estimated as the sample mean and variance of the group weights. As a result, it is of interest to consider the properties of these estimates. Ideally the sample mean and variance are unbiased estimates of the true process parameters. However, this is not possible with group data for all true parameter values. For any weighting scheme w_j , the expected value of the process mean estimate and process standard deviation estimate at the parameter value θ are respectively:

$$\begin{aligned} E(w) &= \hat{\mu} = \sum_{j=1}^k w_j \pi_j(\theta), \quad \text{and} \\ \text{Var}(w) &= \hat{\sigma}^2 = \sum_{j=1}^k (w_j)^2 \pi_j(\theta) - E(\hat{\mu})^2 \end{aligned} \quad (5)$$

where $\pi_j(\theta)$ is given by (4).

These parameter estimates can be substantially different from the true process values. Naturally, any bias in parameter estimation adversely affects process capability calculations and our understanding of the process.

The midpoint approach is a very simple scoring procedure and is often used in industry. Each observation falling into a particular group is assigned a score equal to the group interval midpoint. Assuming gauge limits t , Midpoint group weights are given as:

$$w_j^{(m)} = \begin{cases} (3t_1 - t_2)/2 & \text{for } j = 1 \\ (t_{j-1} + t_j)/2 & \text{for } 2 \leq j \leq k-1 \\ (3t_{k-1} - t_{k-2})/2 & \text{for } j = k \end{cases} \quad (6)$$

The midpoint scheme is attractive because it is very simple, and the scores retain a clear physical meaning. In addition, the midpoint scores can be determined without knowledge of the underlying process distribution. However, calculating the sample mean and variance of the midpoint scores can yield biased estimates of the true process mean and variance. Using (5) with group weights defined in (6) we may derive the expected bias in the estimates of the process mean and variance. Figure 1 shows the results for a range of true process mean values and $t = [-2, -1, 0, 1, 2]$. The Figure illustrates that, using the midpoint weights when the process is in-control, the sample mean is an unbiased estimate of the process mean (when the gauge limits are placed symmetrically), but the standard deviation is typically overestimated.

The midpoint score approach has an additional difficulty because intervals that extend to $-\infty$ or ∞ do not have true midpoints. In the definition (6), end-groups are assigned scores based on the most extreme gauge limits and the distance to the second most extreme scores on either side. Clearly, although this approach seems reasonable if the groups are of equal width, other definitions are possible.

The unbiased estimate weights are derived so that, when the process is in-control, the expected sample mean and variance are unbiased estimates of the process mean and variance. However, these two conditions do not specify unique weights. As a result, the recommended unbiased estimate weights also have the smallest sum of squared bias terms at μ_1 and μ_{-1} , where μ_1 and μ_{-1} are process mean values on each side of the null that we wish to detect quickly. Thus, the unbiased estimate weights are derived as the $w^{(u)}$ weights that

$$\begin{aligned} & \text{minimize} \left\{ \left(E(w^{(u)} | \mu_1) - \mu_1 \right)^2 + \left(E(w^{(u)} | \mu_{-1}) - \mu_{-1} \right)^2 \right\} & (7) \\ & \text{subject to } E(w^{(u)} | \mu_0) = \mu_0, \text{Var}(w^{(u)} | \mu_0) = \sigma_0^2. \end{aligned}$$

For example, when $t = [-2, -1, 0, 1, 2]$ and choosing $\mu_1 = .5$ the unbiased estimate weights are $-2.8, -1.4, -.4, .4, 1.4, 2.8$. These weights are different than the midpoint weights $-2.5, -1.5, -.5, .5, 1.5, 2.5$.

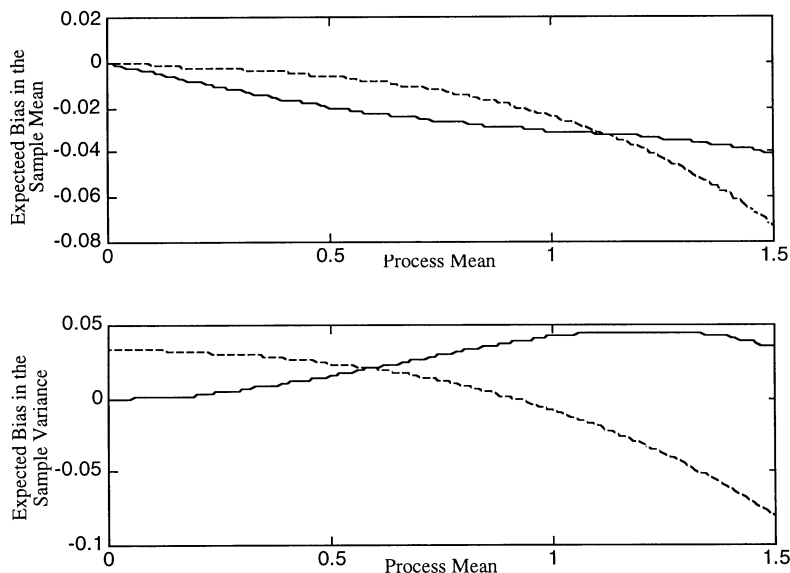


Figure 1: Expected Bias in the Parameter Estimates
 solid line - unbiased estimate weights with $\mu_1 = .5$, $\mu_{-1} = -.5$
 dashed line - midpoint weights
 $t = [-2, -1, 0, 1, 2]$

Figure 1 shows the expected bias of the midpoint and unbiased estimate weights for various values of the true process mean μ when the underlying process is normally distributed with variance equal to unity. In-control, i.e. at $\mu = 0$, both methods yield unbiased estimates of the process mean, however the estimate of the process variance is biased for the midpoint approach.

Group weights defined as the conditional expected value of an observation that falls into a particular group, as given by $w_j^{(c)} = E(X | X \in j^{\text{th}} \text{ group}, \mu = \mu_0)$ were also considered. However, conditional expected value weights have large negative bias when estimating both the process mean and process variance. As a result, conditional expected value weights are not considered further in this article.

Calculating the unbiased estimate weights requires knowledge of the underlying process distribution, or at least information about the group probabilities at both μ_0 and μ_1 . The unbiased estimate weights are desirable when the group weights are used to directly estimate process measures such as process capability. Figure 2 gives an example of how the optimal weights derived according to (7) change with different values of μ_1 ($= -\mu_{-1}$) when $t = [-2, -1, 0, 1, 2]$.

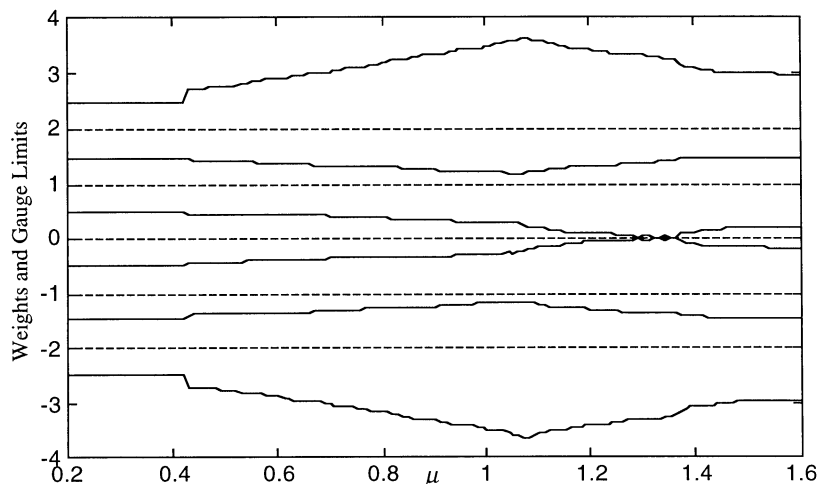


Figure 2: Unbiased Estimate Weights for $t = [-2, -1, 0, 1, 2]$

Since group weights are often used in both control charts and in process capability calculations the unbiased estimate scoring approach is recommended if a good understanding of the underlying distribution is available. Otherwise, the group midpoint approach provides reasonably good results. However, the solution methodology that will be used to derive group data EWMA control charts works with any scoring procedure so long as each observation that falls into a particular group is assigned the same weight.

3. EWMA Control Charts with Grouped Data

The proposed EWMA control charts for grouped data are based on Expression (1), where x_t equals the average group weight assigned a sample. When monitoring the process for mean shifts the group weights can be given by any weighting procedure that retains the group ordering, such as one of the possibilities discussed in Section 2. Using grouped data there are only a finite number of possible average group weights, where the number is based on the number of groups utilized and the sample size. As such there are a finite number of possibilities for x_t . The test statistic z_t in (1), however also depends on the previous value z_{t-1} . As a result, the repeated use of formula (1) smoothes out the discreteness inherent in the observed values. This section addresses the questions of grouped data EWMA control chart design and performance.

The performance of EWMA control charts is usually discussed in terms of the run length. Crowder (1987) used an integral equation approach to derive the run length properties of EWMA control charts based on variables data. Crowder gives tables of run length results for various combinations of the parameters λ and L in (1) and (3). Unfortunately, Crowder's approach can not handle the discreteness inherent in the grouped data EWMA case. An alternative solution procedure, presented in the Appendix, involves modeling the situation with a Markov chain. In the control chart context, the Markov chain solution approach was first developed by Page (1954) to evaluate the run length properties of a cumulative sum (CUSUM) chart. Grouped data with its inherent discreteness appears

well suited to the Markov approach, since the Markov framework requires a discretization of the state space. The proposed Markov chain solution methodology can also provide approximate solutions for EWMA control charts based on variables data, and the proposed method was used to verify the results reported in Crowder (1987).

Table 1 gives Average Run Length (ARL) values for the EWMA charts based on variables (continuous) data, and EWMA control charts for different grouping criterion, given by t . The EWMA charts are designed to detect shifts in the mean of a normal process with in-control mean of zero and variance equal to unity. The process shifts are given in standard deviation units ($\sigma_x/\sqrt{n} = 1/\sqrt{n}$). The group data EWMA charts are designed to match the in-control ARL of the variables based EWMA as closely as possible, but due to the discreteness of the group weights an exact match is not always possible. The run lengths shown in Table 1 are all derived assuming the EWMA starts in the zero state when the process shift occurs. Steady state results provide a more realistic approximation. However, as shown in Lucas and Saccucci (1990), the zero state and steady state run lengths are very similar. Figure 3 plots the results from Table 1 on a log scale. Note that the results in Table 1 are generated for the unit sequential case, i.e., $n = 1$. For larger sample sizes the grouped data results are slightly better, since there is less discreteness, but the difference is not substantial for small sample sizes.

Table 1: Average Run Length for Two-sided Grouped Data EWMA Charts, $X \sim N(0,1)$

$\sqrt{n}\mu_x$	continuous		$t = [-2, -1, 0, 1, 2]$		$t = [-1, 0, 1]$		$t = [-1, 1]$	
	$\lambda = .25$ $L=2.998$	$\lambda = .10$ $L=2.814$	$\lambda = .25$ $L=2.991$	$\lambda = .10$ $L=2.802$	$\lambda = .25$ $L=2.821$	$\lambda = .10$ $L=2.763$	$\lambda = .25$ $L=2.981$	$\lambda = .10$ $L=2.837$
.00	500	500	498	500	511	498	515	487
$\pm .50$	48.2	31.3	51.9	33.7	53.4	35.4	63.3	40.7
± 1.00	11.1	10.3	12.1	11.0	13.1	12.1	14.9	13.0
± 1.50	5.46	6.09	6.01	6.55	6.96	7.66	7.42	7.81
± 2.00	3.61	4.36	4.14	4.79	5.10	6.06	5.25	6.06
± 3.00	2.26	2.87	3.11	3.45	4.13	5.12	4.14	5.12
± 4.00	1.73	2.19	3.00	3.07	4.01	5.01	4.01	5.01

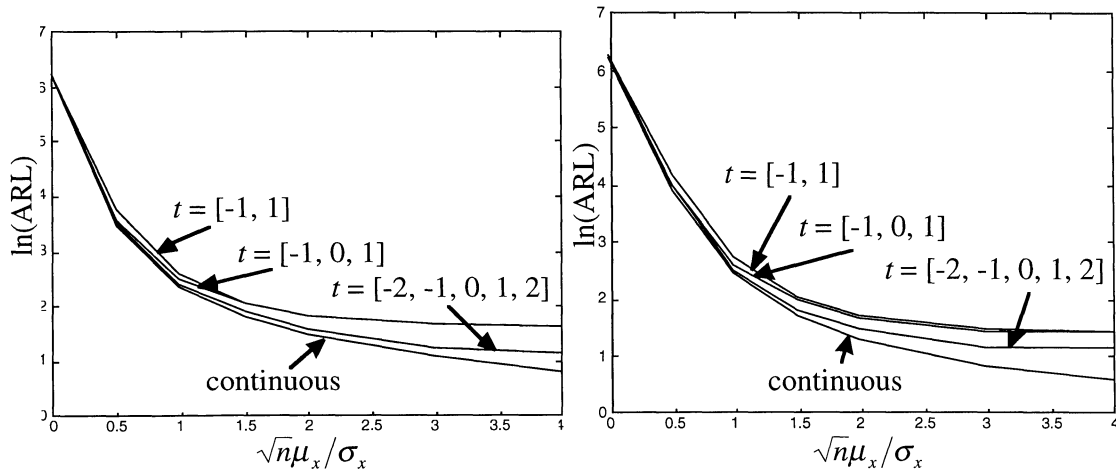


Figure 3: ARL plots comparing grouped data EWMA with variables data EWMA $\lambda = .10$ on the left, $\lambda = .25$ on the right

Table 1 and Figure 3 show, as expected, that the grouped data EWMA charts are not as effective as a variables based EWMA for detecting process mean shifts. This decrease in efficiency is due to the information lost when grouping the data. However, the loss of efficiency in detecting fairly small process mean shifts (say of the order of one standard deviation unit) is quite small. For very large process shifts, on the other hand, the grouped data charts perform poorly since there is a maximum weight value that any observation can take. In applications, EWMA charts are typically used to detect fairly small process shifts. This suggests that grouped data EWMA charts are a viable alternative when collecting variables data is prohibitively expensive or impossible.

All EWMA control charts have two design parameters; namely λ and L , as defined in (1) and (3). Often EWMA charts are designed by specifying a desired in-control run length and the magnitude of the shift in the process that is to be detected quickly. Lucas and Saccucci (1990) provide a lookup table of optimal parameter values for the variables data case. The same general procedure is suggested for grouped data charts. However, due to the inherent discreteness, the desired ARLs may not be precisely attainable. Changes to L do not necessarily change the ARL of the EWMA, since not all state values are attainable. Using Table 4 in Lucas and Saccucci (1990) good initial values for λ and L can be found. Generally, small λ values are good for detecting small process shifts, but

are poor for larger shifts, and vice versa for large λ . Using the solution methodology presented in the Appendix, n and L are adjusted until the desired in-control and out-of-control ARLs are closely met. Large values of L lead to large ARLs, while increasing the sample size n decreases the out-of-control ARL and the problem discreteness. A step-by-step design procedure is given below:

Design Procedure for Grouped Data EWMA Control Charts

1. Find the suggested optimal λ and L values for the variables data based EWMA from Crowder (1987). Set the sample size n equal to unity.
2. Keeping λ fixed, adjust L until the desired in-control ARL is attained. The methodology presented in the appendix can be used to find the in-control ARL for each combination of λ and L .
3. Determine the out-of-control ARL at the current values of n , λ and L .
4. If the desired out-of-control ARL is exceeded, increment n , and repeat this procedure starting at Step 2.

4. Comparison with CUSUM Procedures for Grouped Data

Both EWMA charts and CUSUM charts are designed to detect small persistent process shifts. Past researchers (Lucas and Saccucci, 1990) found that there is very little difference between EWMA and CUSUM procedures in terms of ARL for detecting persistent process mean shifts. In this section, the performance of grouped data and variables based EWMA charts and CUSUM charts for detecting a process mean shifts are compared. The in-control process is assumed to be normally distributed, and without loss of generality, the in-control process mean and variance are set to zero and unity respectively. Since EWMA charts are inherently two-sided, they are compared with a two-sided tabular CUSUM. A tabular CUSUM procedure to detect increases in the process mean consists of plotting: $Y_t = \max(0, Y_{t-1} + x_t - k)$, where $Y_0 = 0$, and k is a design

parameter that specifies an indifference region (Montgomery, 1991). The CUSUM concludes the process mean has shifted upwards whenever $Y_i \geq h$, where h is another design parameter. A two-sided tabular CUSUM is created by simultaneously monitoring two one-sided CUSUM, where the aim of one is to detect upward mean shifts, while the aim of the other is to detect downward shifts (Montgomery, 1991). For both the EWMA and CUSUM charts based on grouped data we used the conditional expectation weights as discussed in Section 2, though similar qualitative results have been derived for other weighting schemes.

The ARL results are given in Table 2. The variables based CUSUM chart with $h=5$ and $k=.5$ has an in-control ARL of 430, and an out-of-control ARL at a one sigma unit shift in the mean of 10.2. These ARL values are used as the standard. The variables based EWMA chart is designed to match these standard ARL values. The grouped data cases are designed so that their in-control run lengths match the target 430. For the grouped data CUSUM charts, ARL values are determined using the methodology presented in Steiner, Geyer and Wesolowsky (1996). The run length results are matched by altering the value of h . However, due to the inherent discreteness of grouped data, the desired in-control ARL of 430 is not precisely obtainable. To make the ARLs easier to compare, the grouped data CUSUM ARLs, presented in Table 2, are theoretical values estimated using linear interpolation between the two closest cases. In the EWMA grouped data case the value of L was altered to yield the desired in-control run length. For the EWMA grouped data chart more flexibility is available and the desired in-control run length was obtained without using interpolation. This design advantage of grouped data EWMA charts is discussed in more detail later.

Table 2: Average Run Length Comparison between Two-sided Grouped Data EWMA Charts and Two-sided Grouped Data CUSUM

$\sqrt{n} \mu_x / \sigma_x$	continuous		$t = [-2, -1, 0, 1, 2]$		$t = [-1, 0, 1]$		$t = [-1, 1]$	
	CUSUM	EWMA	CUSUM	EWMA	CUSUM	EWMA	CUSUM	EWMA
	$k=.5$ $h=5.0$	$\lambda=.2045$ $L=2.915$	$k=.5$ $h \cong 4.48$	$\lambda=.2045$ $L=2.897$	$k=.5$ $h \cong 4.074$	$\lambda=.2045$ $L=2.8$	$k=.5$ $h \cong 3.691$	$\lambda=.2045$ $L=2.78$
.00	430	430	430	430	430	430	430	430
$\pm .50$	37.0	39.3	41.6	42.3	42.3	43.6	47.8	51.5
± 1.00	10.2	10.2	11.1	11.0	11.7	12.0	13.2	13.5
± 1.50	5.7	5.4	6.0	5.9	6.8	6.7	7.5	7.1
± 2.00	4.0	3.7	4.2	4.1	5.3	5.0	5.7	5.2
± 3.00	2.5	2.3	3.1	3.1	4.5	4.1	4.7	4.1
± 4.00	2.0	1.8	3.0	3.0	4.4	4.0	4.5	4.0

Results in Table 2 show that for grouped data as well as variables data there is very little difference between the EWMA and CUSUM charts in terms of run length performance. It appears that the CUSUM chart is slightly better at detecting process shifts smaller than the shift the chart was designed to detect, while EWMA charts are slightly better for larger process shifts. However, this pattern is reversed for smaller values of λ .

Although there is little difference between grouped data EWMA charts and grouped data CUSUM charts in terms of ARL, there are other reasons why an EWMA chart may be preferable. First, EWMA charts are two-sided by design, whereas a two-sided CUSUM chart requires either the use of the awkward V-mask, or two one-sided tubular CUSUM charts (Montgomery, 1991). As a result, if two-sided monitoring of the process is required, variables based or grouped data based EWMA charts are easier to implement. Second, in the grouped data case, over time the EWMA test statistic smoothes out the inherent discreteness in the average group weight, whereas a grouped data CUSUM test statistic remains a simple linear combination of the initial group weights. As a result, grouped data EWMA charts are more flexible in their design than grouped data CUSUM charts, especially when the sample size is small. Figure 4 illustrates this point very effectively. Figure 4 shows the discreteness in the resulting in-control ARL for grouped data CUSUM and EWMA charts when the design parameters h and L are changed for fixed parameters $k = .5$ and $\lambda = .2$. As the parameter values h and L increase, the ARL of the

chart should also increase. Figure 4 shows that the number of possible ARL values is much greater for an EWMA grouped data chart. This is a clear advantage when designing grouped data EWMA charts since typically sequential control charts are designed to have certain ARL characteristics. In Figure 4, for the grouped data CUSUM, the slight decreases observed in the ARL of the grouped data CUSUM as h increase represent some small errors in the approximation of the ARL.

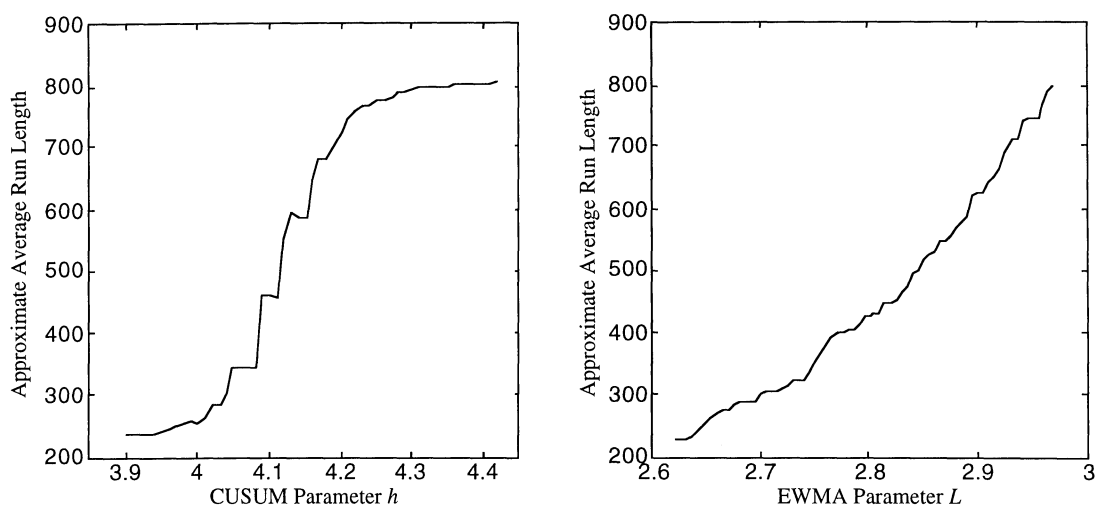


Figure 4: Comparison of the Discreteness in the In-control ARL of Grouped Data CUSUM and EWMA Charts with $t = [-1, 0, 1]$

5. Metal Fasteners Example

This section illustrates the application of a grouped data based EWMA chart to an industrial situation. In the manufacture of metal fasteners in a progressive die environment, the opening gap dimension of a metal clamp, called a robotics clamp, was considered critical. See Figure 5. This problem was previously considered in Steiner, Geyer and Wesolowsky (1994 & 1996A).

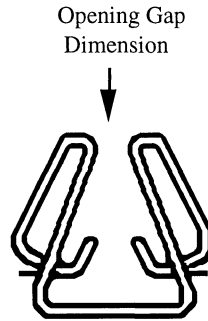


Figure 5: Robotics Clamp

Obtaining exact measurements of the gap dimension on the shop floor was prohibitively difficult and expensive. The metal used in the clamp is fairly pliable, and as a result, using calipers distorts the opening gap dimension. Another alternative, an optical measuring device, is expensive and not practical for on-line quality monitoring. As a result, the only economical alternative on the shop floor is to use step gauges, where clamps are classified into different groups based on the smallest diameter pin that the clamp's opening gap does not fall through. The step gauge employed consisted of three pins of diameters 53, 54, and 55 thousandths of an inch. Using the given step-gauge, units are classified into 4 intervals with corresponding interval midpoint weights of 52.5, 53.5, 54.5, 55.5.

From previous measurements it is known that the process mean is currently stable, producing clamps with an average open gap dimension of 54.2 thousandths of an inch (μ_x) and standard deviation of 1.3 (σ_x). We wish to monitor the stability of the mean width of opening gap. In Steiner et al. (1996A) a grouped data Shewhart control chart was proposed that has an in-control ARL approximately equal to 370, and an out-of-control ARL, at a mean shift of one half a standard deviation unit, of approximately 15.5 and 12.7 for positive shifts and negative shifts respectively. This was accomplished with a sample of size 12 units. Since this is a fairly small process shift we would expect to do better with an EWMA chart.

Using the design methodology suggested in Section 3, with a fixed sample size of 12, we derive an EWMA chart with $\lambda = .1$, $L = 2.54$, and $n = 12$. This grouped data EWMA chart has an in-control ARL of 370 and out-of-control ARLs of around 7.8 and 5.6 for positive and negative mean shifts of half a standard deviation unit respectively. These values are significantly better than the corresponding Shewhart chart with the same sample size. Figure 6 shows the resulting EWMA chart using the Steiner et al. (1996A) data. The process was in-control for the first 10 samples, and shifted down approximately one standard deviation unit starting at observation 11. The EWMA chart shown in Figure 6 signals at observation 12.

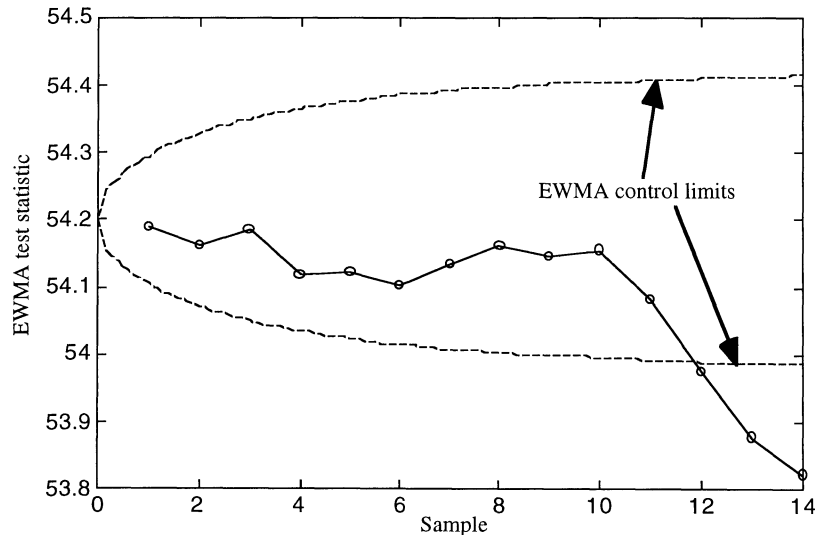


Figure 6: EWMA Control Chart

6. Optimal Grouping Criterion

The run length results and comparisons presented in this article assume fixed group intervals. This is often a reasonable assumption due to the use of standardized gauges. However, in some circumstances the placement of the group limits may be under our control. In that case, the question of optimal group intervals for the grouped data EWMA arises. Finding the best gauge limits requires a definition of optimal. One possibility is to find the gauge limits that yield the shortest out-of-control ARL at a given mean shift while

having an in-control ARL of at least ARL_0 . This is an attractive definition of optimal gauge limit, but requires a solution for different in-control ARLs and different out-of-control shifts. Another approach to finding the best gauge limits is to determine the grouping criterion that gives the best estimate of the process mean while the process is in-control. This is an attractive option since usually the process will remain in-control most of the time, and there is connection between good estimation and effective hypothesis testing. The best gauge limits for estimation are found by maximizing the expected Fisher information available in the grouped data. The expected Fisher information provides a measure of the efficiency of the grouped data compared with variables data. Steiner et al. (1996A) derive the best estimation gauge limits for grouped data in the normal mean and standard deviation cases, and Steiner (1994) derives the optimal limits for the Weibull parameter cases.

Summary

EWMA control charts are widely applied in industry. However, all previous analysis of their run length properties considered only variables data or binomial data. This article discusses the design and performance of EWMA control charts for grouped data. Grouped data are a natural compromise between the low data collection and implementation costs of binomial data and the high information content of variables data. EWMA control charts based on grouped data are shown to be nearly as efficient as variables based EWMA charts and are thus an attractive alternative when collection of variables data is not feasible or very expensive. Also, EWMA grouped data charts are more flexible in their design than grouped data CUSUM charts.

Appendix

In this appendix the expected value and variance of the run length of grouped data EWMA control charts are derived. The solution procedure utilizes a Markov chain where the state space between the control limits is divided into $g-1$ distinct discrete states and the out-of-control condition corresponds to the g th state. The states are defined as:

$$\mathbf{s} = (s_1, s_2, \dots, s_{g-1}) = (LCL + y, LCL + 2y, \dots, UCL - 2y, UCL - y),$$

where $y = (UCL - LCL)/g$ and UCL and LCL are the control limits as given by (3).

The transition probability matrix is given by

$$P = \begin{bmatrix} p_{11}, p_{12}, \dots, p_{1g} \\ p_{21}, \dots, p_{2g} \\ \vdots \\ p_{g1}, \dots, p_{gg} \end{bmatrix} = \begin{bmatrix} R & , (I - R) \mathbf{1} \\ 0, \dots, 0, & 1 \end{bmatrix}, \quad (\text{A1})$$

where I is the g by g identity matrix, $\mathbf{1}$ is a g by 1 column vector of ones, and p_{ij} equals the transition probability from state s_i to state s_j . The last row and column of the matrix P correspond to the absorbing state that represents an out-of-control signal. The R matrix equals the transition probability matrix with the row and column that correspond to the absorbing (out-of-control) state deleted.

The group probabilities π_a , $a = 1, \dots, k$, as defined by (4), and the group weights w_a , $a = 1, \dots, k$, as given by (5) or the solution to (6), set the transition probabilities in the matrix R . Using the defined states \mathbf{s} as a discretization, the transition probabilities p_{ij} are:

$$p_{ij} = \begin{cases} \pi_a & \text{if } s_j - \frac{y}{2} < \lambda w_a + (1 - \lambda)s_i < s_j + \frac{y}{2}, & \text{for } j = 1, 2, \dots, g-1 \\ 0 & \text{otherwise} \end{cases} \quad (\text{A2})$$

$$p_{ig} = \begin{cases} \sum \pi_a & \text{for all } a \text{ such that } \lambda w_a + (1 - \lambda)s_i \geq s_{g-1} + \frac{w}{2} \text{ OR } \lambda w_a + (1 - \lambda)s_i \leq s_1 - \frac{w}{2} \\ 0 & \text{if no such } a \text{ exists} \end{cases}$$

The expected run length and the variance of the run length are found using the matrix R . Letting γ denote the run length of the EWMA, we have

$$\begin{aligned} \Pr(\gamma \leq t) &= (I - R^t) \mathbf{1}, \text{ and thus} \\ \Pr(\gamma = t) &= (R^{t-1} - R^t) \mathbf{1} \quad \text{for } t \geq 1. \end{aligned} \quad (\text{A3})$$

Therefore,
$$E(\gamma) = \sum_{t=1}^{\infty} t \Pr(\gamma=t) = \sum_{t=1}^{\infty} (R^t \mathbf{1}) = (I-R)^{-1} \mathbf{1}. \quad (\text{A4})$$

Similarly, the variance of the run length
$$\text{Var}(\gamma) = 2R(I-R)^{-2} \mathbf{1}, \quad (\text{A5})$$

where (A4) and (A5) are g by 1 vectors that give the mean and variance of the run length from any starting value or state s_i . The mean and variance of the run length that correspond to the starting at z_0 are easily found by finding the i such that $s_i - y/2 \leq z_0 \leq s_i + y/2$. If the control limits are symmetric about z_0 the corresponding state is $s_{g/2}$.

This Markov chain solution approach approximates the solution, with the accuracy of the approximation depending on the number of states (g) used. Larger values of g tend to lead to better approximations. However, unfortunately due to discreteness, the ARL value does not smoothly approach the true value as g increases. As a result, the regression extrapolation technique suggested by Brook and Evans (1972), to find the average run length of a variables data based CUSUM scheme approximated by a Markov chain, is not applicable here. However, fairly close approximations of the true run length properties can be obtained by taking the average result obtained using a few fairly large values of g . For example, the results presented in this article estimate the true value $E(\gamma)_{g=\infty}$ and $\text{Var}(\gamma)_{g=\infty}$ by averaging the results derived with $g = 100, 110, 120, 130, 140, 150$. Verification of this approach using simulation suggests that the derived estimates for the mean and variance of the run length differ from the true value by less than 2-3% in most cases of interest, with the approximation becoming worse as the size of the process shift increases.

Acknowledgments

This research was supported, in part, by the Natural Sciences and Engineering Research Council of Canada.

References

- Beja, A., Ladany, S.P., (1974), "Efficient Sampling by Artificial Attributes," *Technometrics*, Vol. 16, No. 4, pp. 601-611.
- Box, G.E.P., Jenkins, G.M., and MacGregor, J.F. (1974), "Some Recent Advances in Forecasting and Control," *Applied Statistics*, 23, 158-179.
- Brook, D., Evans, D.A. (1972), "An approach to the probability distribution of cusum run length," *Biometrika*, 59, 539-549.
- Crowder, S.V. (1987), "Run-Length Distributions of EWMA Charts," *Technometrics*, 29, 401-407.
- Gan, F.F. (1990), "Monitoring Observations Generated from a Binomial Distribution using Modified Exponentially Weighted Moving Average Control Chart," *Journal of Statistical Computing and Simulation*, 37, pp. 45-60.
- Geyer, P.L., Steiner, S.H., Wesolowsky, G.O. (1995), "Optimal SPRT and CUSUM Procedures Using Compressed Limit Gauges," to appear in the *IIE Transactions*.
- Haitovsky, Y. (1982), "Grouped Data," in the *Encyclopedia of Statistical Sciences*, 3, Kotz S., Johnson N.L. and C.B. Read (editors), Wiley, New York.
- Ladany, S.P. (1976), "Determination of Optimal Compressed Limit Gaging Sampling Plans," *Journal of Quality Technology*, 8, 225-231.
- Lucas, J.M. and Saccucci, M.S. (1990), "Exponentially Weighted Moving Average Control Schemes: Properties and Enhancements," with discussion, *Technometrics*, 32, 1-29.
- Montgomery, D.C. (1991), *Introduction to Statistical Quality Control*, Second Edition, John Wiley and Sons, New York.
- Page, E. S. (1954), "Continuous Inspection Schemes," *Biometrika*, Vol. 41, 100-114.
- Roberts, S.W. (1959), "Control Chart Tests Based on Geometric Moving Averages," *Technometrics*, 1, 239-250.

- Schilling, E.G. (1981), *Acceptance Sampling in Quality Control*, Marcel Dekker, New York.
- Schneider, H. and O'Connell, C. (1987). "Design of CUSUM Control Charts Using Narrow Limit Gauges," *Journal of Quality Technology*, 19, 63-68.
- Steiner, S.H. (1994), "Quality Control and Improvement Based on Grouped Data," unpublished Ph.D. Thesis, McMaster University, Hamilton, Ontario.
- Steiner, S.H., Geyer, P.L., Wesolowsky, G.O. (1994), "Control Charts based on Grouped Data," *International Journal of Production Research* 32, 75-91.
- Steiner, S.H., Geyer, P.L., Wesolowsky, G.O. (1996A), "Shewhart Control Charts to Detect Mean and Standard Deviation Shifts based on Grouped Data," to appear in the *Journal of Quality and Reliability Engineering International*.
- Steiner, S.H., Geyer, P.L., Wesolowsky, G.O. (1996B), "Grouped Data Sequential Probability Ratio Tests and Cumulative Sum Control Charts," *Technometrics*, Vol. 38, August, pp. 230-237.
- Stevens, W.L. (1948), "Control by Gauging," *Journal of the Royal Statistical Society: Series B*, No. 1, 54-108.