ELSEVIER

Statistics & Probability Letters 24 (1995) 219-224

STATISTICS & PROBABILITY LETTERS

Alarm rates for quality control charts

Thomas M. Margavio[a], Michael D. Conerly[b], William H. Woodall[b,*], Laurel G. Drake[b]

[a] Southwest Missouri State University, Springfield, MO 65804, USA
[b] University of Alabama, 300 Alston, Box 870226 Tuscaloosa, AL 35487-0226, USA

Received March 1994; Revised July 1994

Abstract

There is a direct relationship between a single alarm probability and the average run length only for basic Shewhart charts such as the $\bar{X}$-chart. Alarm rates are defined in this paper that can be applied with charts such as the cumulative sum (CUSUM) chart and the exponentially weighted moving average (EWMA) chart that base decisions on several observations, not just the most recent one. Methods for determining EWMA chart limits are compared on the basis of their false alarm rates. It is shown how control charts can be more flexibly and carefully defined by considering a desired pattern of in-control false alarm rates in conjunction with a desired in-control average run length.

Keywords: Statistical process control; Exponentially weighted moving average control chart; Average run length

---

1. Introduction

In this paper we consider various control charts used to monitor the mean level of a quality characteristic of interest. It is assumed that one observes successively the independent sample means $\bar{X}_1, \bar{X}_2, \ldots$, where $\bar{X}_i \sim N(\mu_i, \sigma^2/n)$, $i = 1, 2, \ldots$. For simplicity it is assumed that when the process is in statistical control we have $\mu_i = \mu_0$, $i = 1, 2, \ldots$, and that σ^2 is known and remains constant.

The random number of samples required for a control chart to give an out-of-control signal, N, is referred to as the run length of the chart. Control chart performance is usually measured in terms of $E(N)$, the average run length (ARL).

For any type of control chart, we propose that the alarm rate at sample i be defined as the probability of an alarm at sample i given no alarm prior to the ith sample, i.e.,

$$r_i = P(N = i \mid N > i - 1), \quad i = 1, 2, \ldots. \tag{1}$$

Although alarm rates can be calculated using run length probabilities, they provide insight into chart performance that cannot be obtained directly from the run length distribution itself. The definition of an alarm rate is similar to that of a mortality rate in actuarial science (Jordan, 1967) and a hazard rate in

---

[*] Corresponding author.

SSDI 0167-7152(94)00174-X

reliability theory (Barlow and Proschan, 1975). The mortality rate q_i, for example, is the probability of a person dying within one year given survival to age, i, $i = 0, 1, \ldots$.

The widely used Shewhart (1931) $\bar{X}$-chart signals that the process is out-of-control at sample N if $\bar{X}_N$ falls outside the control limits

$$\mu_0 \pm k\sigma/\sqrt{n}, \tag{2}$$

where $k > 0$. For the $\bar{X}$-chart the alarm rate when $\mu_i = \mu$, $i = 1, 2, \ldots$, is constant, i.e., $r_i = p$, for $i = 1, 2, \ldots$, where

$$p = 1 + \Phi(-k - \delta) - \Phi(k - \delta), \tag{3}$$

$\delta = \sqrt{n}(\mu - \mu_0)/\sigma$ and $\Phi(\cdot)$ is the standard normal cumulative distribution function. In fact, the run length N of the $\bar{X}$-chart is a geometric random variable when $\mu_i = \mu$, $i = 1, 2, \ldots$, and the ARL is determined by the direct relationship

$$\mathrm{ARL}_0 = 1/p. \tag{4}$$

In most texts, p is referred to by α when $\mu = \mu_0$ and by $1 - \beta$ when $\mu \neq \mu_0$. For example, in the usual case where $k = 3$, we have $\alpha = 0.0027$ and $\mathrm{ARL}_0 = 370.4$. Thus, an alarm probability and the ARL can be directly related for the $\bar{X}$-chart.

The expression in (4) does not hold for other control charts such as the cumulative sum (CUSUM) chart, the exponentially weighted moving average (EWMA) chart, the moving average chart and the $\bar{X}$-chart with supplementary runs rules. These charts base decisions regarding whether or not the process is in-control on several past values of the sample mean, not just the most recent one. It can be shown that the usual methods of designing these charts result in charts with alarm rates that vary over time when $\mu_i = \mu_0$, $i = 1, 2, \ldots$ We refer to these in-control alarm rates as false alarm rates $r_i = \alpha_i$, $i = 1, 2, \ldots$ Varying false alarm rates may be desirable in some cases, e.g. to obtain extra sensitivity to initial out-of-control conditions, but we recommend that the pattern of these alarm rates be considered explicitly in control chart design. Furthermore, Adams et al. (1992) show that single false alarm probabilities are often used inappropriately to design CUSUM charts and to evaluate the performance of $\bar{X}$-charts with runs rules.

The alarm rates in Eq. (1) can be applied to any control chart. For the remainder of this paper, however, we focus on the false alarm rates for the EWMA charts because the required calculations are the most tractable. The design of EWMA charts is discussed in Section 2 and the false alarm rates of three types of EWMA control chart limits are compared in Section 3. Section 4 contains our conclusions.

2. EWMA chart design

The EWMA control chart is based on the statistics

$$Z_i = \lambda \bar{X}_i + (1 - \lambda) Z_{i-1}, \quad i = 1, 2, \ldots, \tag{5}$$

where $Z_0 = \mu_0$ and $0 < \lambda \leqslant 1$. A common approach to determining the control limits (see, e.g. Montgomery, 1991, p. 300) is to note that Z_i has variance

$$\sigma_i^2 = [\sigma^2/n][\lambda/(2 - \lambda)][1 - (1 - \lambda)^{2i}] \tag{6}$$

and consequently to use the limits

$$\mu_0 \pm k\sigma_i, \quad i = 1, 2, \ldots, \tag{7}$$

where $k > 0$. We refer to the limits in (7) as the "variance-adjusted" limits. The limiting value of the variance in (6), as i increases, is

$$\sigma_{\infty}^2 = \sigma^2/n\left(\frac{\lambda}{2-\lambda}\right). \tag{8}$$

Thus, a number of authors recommend the control limits

$$\mu_0 \pm k\sigma_{\infty}. \tag{9}$$

We refer to these limits as the "constant" control limits. Lowry et al. (1992) use an analogous argument to determine control limits for a multivariate EWMA chart.

The EWMA chart was originally proposed by Roberts (1959). The run length properties of the EWMA chart with the constant control limits in (9) have been studied by Robinson and Ho (1978), Waldmann (1986), Crowder (1987a, 1987b, 1989), Lucas and Saccucci (1990) and Gan (1991). Chandrasekaran et al. (1993) consider the properties of the EWMA chart with the variance-adjusted control limits in (7), referring to these as the "correct" limits.

It is usually recommended that EWMA charts be designed on the basis of ARL performance. The value of the smoothing constant and the value of k in (7) or (9) are determined by the size of the shift in the mean considered important enough to be detected quickly and a specified in-control ARL. It is important to note, however, that the expression for the variance of Z_i in (6) for $i > 1$ is derived by ignoring the truncation effects of the control limits employed at the previous $i - 1$ samples. Thus, using a constant multiple of the standard deviation to give the control limits in (7) is done somewhat arbitrarily, i.e., for convenience rather than to meet any additional well-defined statistical criterion on chart performance. Contrary to the impression usually given, there is no clear theoretical justification based on statistical performance for using either the variance-adjusted limits in (7) or the constant control limits in (9).

It is shown in the next section that EWMA charts with control limits chosen as in (7) or (9) have false alarm rates which vary over time. Another choice could be to design EWMA charts with "time-varying" control limits defined by

$$\mu_0 \pm k_i\sigma_i, \quad i = 1, 2, \ldots, \tag{10}$$

where the values k_i, $i = 1, 2, \ldots$, are determined so that the false alarm rate at each sample is constant, i.e.,

$$\alpha_i = \alpha_0, \quad i = 1, 2, \ldots,$$

where α_0 is a specified false alarm rate. This approach results in a geometric in-control run length distribution and maintains the direct relationship in (4) between the false alarm rate and the in-control ARL. We use the constant false alarm rate chart here only as an illustration. Any specific choice of false alarm rates α_i, $i = 1, 2, \ldots$, however, leads to a unique (within numerical precision) set of symmetric control limits. The user may specify false alarm rates which vary over time according to a specified pattern. For example, Lucas and Crosier (1982) argue that one may prefer a control chart with the fast initial response (FIR) feature for the increased sensitivity to initial out-of-control conditions. Our approach can also be modified for use with other types of control charts such as the CUSUM chart or charts for monitoring some parameter other than the process mean. If one uses an existing, ad hoc method to design a control chart, then we recommend that the false alarm rate structure be considered in evaluating chart performance.

3. False alarm rate comparisons

In this section we compare the false alarm rates of three types of EWMA charts which are all designed to have an in-control ARL of 500. Fig. 1 shows the false alarm rates for the EWMA charts with the

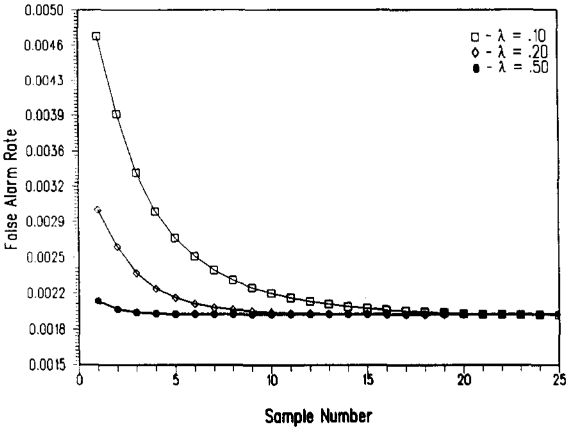

Fig. 1. False alarm rates for the EWMA chart with variance-adjusted limits of Eq. (7).

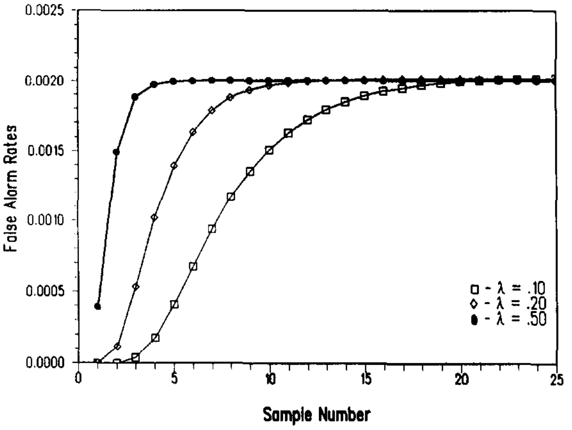

Fig. 2. False alarm rates for the EWMA chart with constant control limits given in Eq. (9).

variance-adjusted control limits in (7) for $\lambda = 0.1, 0.2$, and 0.5. The false alarm rates with the constant control limits in (9) are shown in Fig. 2. For the time-varying control limits given in Fig. 3, the false alarm rate is constant at $\alpha_0 = 0.002$.

To obtain these figures we used the results of Waldmann (1986) and Crowder (1987a) which show that under the assumption of independence between observations, the EWMA chart can be represented as a continuous state Markov process. If we assume normality and, without loss of generality, that $\mu_0 = 0$ and $\sigma^2/n = 1$, then the density of Z_i using the time-adjusted limits when the process is in-control is given by

$$g_i(z) = \int_{-k_{i-1}}^{k_{i-1}} \frac{\sigma_i}{\lambda} g_{i-1}(y) f_i\left(\frac{\sigma_i z - (1-\lambda)\sigma_{i-1} y}{\lambda}\right) \mathrm{d}y, \quad i = 2, 3, \ldots, \tag{11}$$

where $f_i(\cdot)$ represents the p.d.f. of X_i. For our in-control comparisons we took $f_i(\cdot)$ to be the standard normal p.d.f. For the out-of-control case, $f_i(\cdot)$ would be a normal p.d.f. with a non-zero mean. The limits of

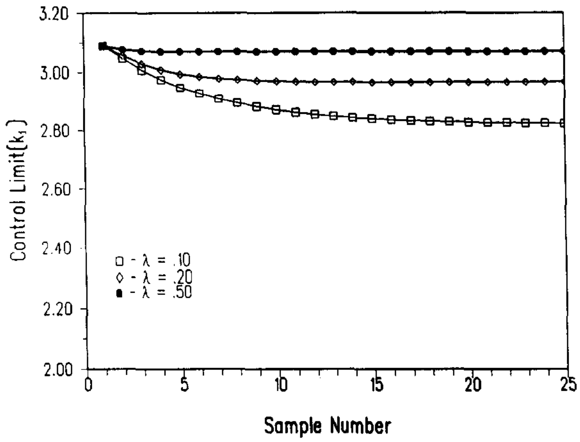

Fig. 3. Control chart constants k_i for EWMA chart with time-varying control limits in Eq. (10).

integration in (11) are the control limits, which may be specified as in (7) or (9) for the other types of control limits. Note that $g_1(z)$ is a normal p.d.f. with mean zero and variance λ^2 and

$$\int_{-\infty}^{\infty} g_i(z)\,\mathrm{d}z = P(N > i-1), \quad i = 1, 2, \dots,$$

where N is the run length of the EWMA chart. If we let $p_i = P(N = i)$, $i = 1, 2, \dots$, represent the run length probabilities, then

$$\sum_{j=1}^{i} p_j = 1 - \int_{-k_i}^{k_i} g_i(z)\,\mathrm{d}z. \tag{12}$$

Eqs. (11) and (12) can be evaluated using numerical techniques for a given value of λ and fixed control limits. The alarm rates in (1) can then be approximated using the relationship

$$r_i = p_i \Bigg/ \left(1 - \sum_{j=1}^{i-1} p_j\right).$$

The expressions in (11) and (12) are straightforward generalizations of the results given by Waldmann (1986) for the case $k_i\sigma_i = k\sigma_\infty$, $i = 1, 2, \dots$. To calculate the false alarm rates in our paper, 24-point Gaussian quadrature was used to evaluate the integrals in Eqs. (11) and (12). Gan (1991) also used this numerical method. All computations were performed in double precision using an IBM 3081D computer.

Determining the time-varying control limits in (10) requires iteratively finding the control limit constants k_i such that $p_1 = \alpha_0$ and

$$p_i = \prod_{j=1}^{i-1} (1-\alpha_0)^j \alpha_0, \quad i = 2, 3, \dots$$

For each value of i, we used the IMSL (1987) subroutine DZREAL for numerically solving nonlinear equations.

From Fig. 1, one can see that the false alarm rate using the variance-adjusted limits can be much higher than 0.002 initially, but the values decrease more rapidly to limiting values slightly less than 0.002 as λ increases. From Fig. 2, the false alarm rates are quite low initially when the constant control limits are used. The false alarm rates converge more quickly to limiting values slightly greater than 0.002 when λ is large.

Lucas and Saccucci (1990) have shown, however, that values of λ in the range from 0.1 to 0.2 are most effective in detecting small shifts in the process mean. For these cases there are large initial differences in the false alarm rates. The false alarm rate patterns have a more pronounced effect if the production run is short since in the long run the rates converge to relatively close limiting values. Using the variance-adjusted limits in (7) gives an increased number of false alarms initially and the constant control limits in (9) are much less sensitive to early shifts in the process mean compared to the use of the time-varying control limits with a constant false alarm rate. If a process shift is delayed for some time, then there will be little difference in the out-of-control statistical performance of the three types of EWMA control limits.

4. Conclusion

Crowder (1989) and Lucas and Saccucci (1990), among others, recommend designing EWMA charts on the basis of a shift in the mean considered important enough to detect quickly and the desired in-control ARL. Similar recommendations have been made for CUSUM charts. Because specifying the in-control ARL does not lead to a unique set of control limits, we recommend that control chart limits be based on the desired in-control ARL and a desired pattern of false alarm rates. We use the EWMA chart with a constant false alarm rate only as an illustration. If one uses a simpler, ad hoc method leading to limits such as those in (7) or (9) for the EWMA chart, then the false alarm rate pattern should be considered in evaluating chart performance. In addition, as Crowder (1987b) and others recommend, the out-of-control ARLs should be calculated over a range of process shifts in order to check a proposed chart's sensitivity to out-of-control conditions.

References

Adams, B.M., C.A. Lowry and W.H. Woodall (1992), The use (and misuse) of false alarm probabilities in control chart design, in: *Frontiers in Statistical Quality Control 4*, H.-J. Lenz, G.B. Wetherill and P.-Th. Wilrich, eds. (Physica, Heidelberg) 155–168.

Barlow, R.E. and F. Proschan (1975), *Statistical Theory of Reliability and Life Testing* (Holt, Rinehart and Winston, New York).

Chandrasekaran, S., J.R. English and R.L. Disney (1993), Modeling and analysis of exponentially weighted moving average control schemes, to appear in: *IIE trans.*

Crowder, S.V. (1987a), A simple method for studying run-length distributions of exponentially weighted moving average charts, *Technometrics* **29**, 401–407.

Crowder, S.V. (1987b), Average run lengths of exponentially weighted moving average control charts, *J. Quality Technol.* **19**, 161–164.

Crowder, S.V. (1989), Design of exponentially weighted moving averages, *J. Quality Technol.* **21**, 155–162.

Gan, F.F. (1991), Computing the percentage points of the run length distribution of an exponentially weighted moving average control chart, *J. Quality Technol.* **23**, 359–362.

IMSL. (1987), User's Manual Math/Library v.1.

Jordan, C.W. Jr. (1967), *Life Contingencies* (Society of Actuaries, Chicago, Il).

Lowry, C.A., W.H. Woodall, C.W. Champ and S.E. Rigdon (1992), A multivariate exponentially weighted moving average control chart, *Technometrics* **34**, 46–53.

Lucas, J.M. and M.S. Saccucci (1990), Exponentially weighted moving average control schemes: properties and enhancements (with discussion), *Technometrics* **32**, 1–29.

Lucas, J.M. and R.B. Crosier (1982), Fast initial response for CUSUM quality control schemes, *Technometrics* **24**, 199–205.

Montgomery, D.C. (1991), *Introduction to Statistical Quality Control* (Wiley, New York, 2nd ed.).

Roberts, S.W. (1959), Control chart tests based on geometric moving averages, *Technometrics* **1**, 239–250.

Robinson, P.B. and T.Y. Ho (1978), Average run lengths of geometric moving average charts by numerical methods, *Technometrics* **20**, 85–93.

Shewart, W.A. (1931), *Economic Control of Quality of Manufactured Product* (Van Nostrand-Reinhold, New York) (Republished in 1980 by the American Society for Quality Control).

Waldmann, K.H. (1986), Bounds for the distribution of the run length of geometric moving average charts, *J. Roy. Statist. Soc. Ser. B (Appl. Statist.)* **35**, 151–158.