CUSUM Procedure for Monitoring Variability

by

M.S. Srivastava
Department of Statistics
University of Toronto

Technical Report No. 9512, June 28, (1995)

TECHNICAL REPORT SERIES

University of Toronto

Department of Statistics

CUSUM Procedure for Monitoring Variability

M.S. Srivastava

University of Toronto

Abstract

In this paper, the CUSUM procedure for monitoring an increase in variability when the observations are independently and normally distributed is studied. The procedure is based on the logarithm of the likelihood ratio of the probability density functions under the two competing hypotheses. Formulae that approximate the average run length of the CUSUM procedure for detecting an increase (or decrease) in the variance of a normal distribution are given. These formulae, when corrected for the overshoot from the boundary, provide a very accurate approximation.

Keywords and Phrases: Average run length; Overshoot correction; Sequential probability ratio test; Wald's approximation.

1 Introduction

Quality control schemes are widely used to improve the quality of a manufacturing process. One of the most popular control schemes is the cumulative sum (CUSUM) chart, which was introduced by Page [7]. A widely used criterion for evaluating and comparing the performance of a control scheme is the average run length, that is, the average number of observations taken before the scheme signals. When the process is in control, the average run length, denoted by ARL_0, is usually fixed and taken to be large to avoid false alarms. This ARL_0 is used to determine the control limit, or decision boundary, of the control procedure. To choose among many available procedures with the same ARL_0, a procedure which signals with the least possible number of observations after the change should be selected. This delay in detection time is defined in many ways. The one most commonly used in the literature is the average number of observations taken until a signal is given, when a change has occurred on the first observation. This is denoted by ARL_1. Another criteria is the conditional average delay time, denoted by $CADT(t)$, which is the average number of observations past the change point t, given that the change has occurred before the control scheme signals. Lorden [5] and Moustakides [6] considered the conditional average delay time over those events before the change point that least favours the detection of the change point. They showed that among all sequential procedures that have the same in-control average run length ARL_0, the CUSUM procedure minimizes the above conditional average delay time. This minimum value turns out to be ARL_1 for the CUSUM scheme. Thus, the CUSUM procedure has the smallest ARL_1 among all procedures having the same ARL_0.

Another attractive feature of the CUSUM procedure is that simple approximating formulae for ARL_0 and ARL_1 can be obtained from Wald's results on sequential probability ratio tests, since CUSUM procedures are based on the logarithm of the likelihood ratios of the densities under the two competing hypotheses. The formulae obtained from Wald's method neglects the excess over the boundary and may not be very accurate, as has been demonstrated many times in the literature. See, for example, Siegmund [9]. These approximating formulae, however, can be made more accurate by applying corrections due to the overshoot. Thus, extensive computations such as simulation or the Markov chain method of Brook and Evans [1] are not required to compute the decision boundary.

Inspite of the above attractive properties of the CUSUM procedure, no systematic study, as far as we know, has been carried out for the CUSUM procedure to monitor an increase (or decrease) in the variance. See the recent paper by Yashchin [14]. Several other procedures have recently been proposed in the literature. See, for example, Tuprah and Ncube [12] and Crowder and Hamilton [2]. All these procedures, however, require extensive computation with apparently no significant gain in performance, as has been shown by Srivastava and Chow [10]. In this paper, we study the CUSUM procedure and provide simple formulae for the average run lengths. The organization of the paper is as follows:

In Section 2, we describe the CUSUM procedure for monitoring an increase, decrease or either an increase or a decrease in the variance. A very accurate approximation for the in control average run length ARL_0 as well as $ARL_1(\sigma_+)$, where σ_+ is the reference value, is given in Section 3. The average run length for general σ^2 is also evaluated in this section. Section 4 considers the case in which more observations are taken at any inspection time. The case of unknown mean is considered in Section 5. An example for monitoring variance components, considered by Yashchin [14] is reconsidered in Section 6. The paper concludes in Section 7.

2 CUSUM Procedure

When the process is in control, the sequence of observations $x_1, x_2, \ldots$ are independently and identically distributed (iid) with probability density function f_0 which we assume to be the normal density with mean zero and variance one, denoted by $N(0,1)$. That is, if the observations have known variance σ_0^2, it is assumed that each observation is divided by σ_0. When the process goes out of control at some unknown point ν, then the observations $x_1, \ldots, x_{\nu-1}$ are iid $N(0,1)$ and the observations $x_\nu, x_{\nu+1}, \ldots$ are iid $N(0,\sigma^2)$. To monitor for an increase in the variance, we shall have $\sigma^2 > 1$. Similarly, to monitor for a decrease in the variance, we shall have $\sigma^2 < 1$. We shall now describe the monitoring scheme.

2.1 Monitoring for an Increase in the Variance

Generally, the variance σ^2 in the alternative may not be known. However, it is assumed that the manufacturer knows the acceptable limit of the variability, beyond which the product will not be acceptable. This value of the variance, called the reference value, will be denoted by σ_+^2. For a given reference value $\sigma_+^2 > 1$, define

$$\lambda^+ = \frac{\sigma_+^2 \log\left(\sigma_+^2\right)}{\sigma_+^2 - 1} > 0 \tag{2.1}$$

and

$$C_n^+ = \max\left\{0, C_{n-1}^+ + \left(x_n^2 - \lambda^+\right)\right\}, \quad n = 1, 2, \ldots \tag{2.2}$$

where $C_0^+ = 0$. It will be shown in Appendix A that

$$1 \leq \lambda^+ \leq \sigma_+^2. \tag{2.3}$$

The CUSUM procedure triggers a signal as soon as $C_n^+ > h^+$, where $h^+ (> 0)$ is chosen so that the in-control average run length ARL_0 is a specified (usually large) number. The process is stopped and checked for an increase in the variance from 1 to $\sigma_+^2 > 1$ at the stopping time

$$T^+ = \min\left\{n \geq 1 : C_n^+ > h^+\right\}. \tag{2.4}$$

2.2 Monitoring for a Decrease in the Variability

In this section, we consider the problem of monitoring for a decrease in the variance from $\sigma^2 = 1$ to $\sigma_-^2 < 1$, where σ_-^2 is the reference value (assumed to be known). Define

$$\lambda^- = \frac{\sigma_-^2 \log\left(\sigma_-^2\right)}{\sigma_-^2 - 1} > 0 \tag{2.5}$$

and

$$C_n^- = \max\left\{0, C_{n-1}^- + \left(\lambda^- - x_n^2\right)\right\}, \quad n = 1, 2, \ldots \tag{2.6}$$

where $C_0^- = 0$. It will be shown in Appendix A that

$$\sigma_-^2 \leq \lambda^- \leq 1. \tag{2.7}$$

The CUSUM procedure triggers a signal as soon as $C_n^- > h^-$, where $h^-(> 0)$ is chosen so that the in-control average run length ARL_0 is a specified (usually large) number. The process is stopped and checked for a decrease in the variance from 1 to $\sigma_-^2 < 1$ at the stopping time

$$T^- = \min\left\{n \geq 1 : C_n^- > h^-\right\}. \tag{2.8}$$

2.3 Monitoring for an Increase or Decrease in Variability

In order to monitor for an increase or decrease in the variability, the control schemes of (2.4) and (2.8) are applied simultaneously. The process will be stopped as soon as either of the two control schemes signals a change. The run length will be

$$T = \min\left\{T^+, T^-\right\}$$

3 Average Run Lengths

In this section, we give expressions for the average run lengths. We begin with the case of detecting an increase in the variance.

3.1 Average of T^+

Let

$$S_n = \sum_{i=1}^{n} \left(x_i^2 - \lambda^+\right),$$

and

$$N^+ = \min\left\{n \geq 1 : S_n \not\in (0, h^+)\right\}.$$

Then, following Page [7], it can be shown that

$$E(T^+) = \frac{E(N^+)}{P\left(S_{N^+} \geq h^+\right)}.$$

Since S_n is a scalar multiple of the log of the likelihood ratio of the n observations, the expression on the right side can be approximated using Wald's results for sequential probability ratio tests. These approximations, however, neglect the excess over the boundary and usually do not provide accurate approximations of the average run lengths. Many attempts have been made in the literature to correct for this overshoot. See, for example, Reynolds [8], Khan [4], and Siegmund [9]. It has been found, however, that the correction used in Siegmund [9] provides a very accurate approximation. Based on this theory, Srivastava and Chow [10] computed the corrections required for this case. See Table 5. Thus, using Wald's theory and applying the correction for the overshoot, we find that a very accurate approximation for the 'in-control' average run length is given by

$$ARL_0^+ \equiv E_0\left(T^+\right) \simeq \frac{e^{a_1 h_1} - a_1 h_1 - 1}{a_1\left(\lambda^+ - 1\right)}, \tag{3.9}$$

where

$$a_1 = \frac{\sigma_+^2 - 1}{2\sigma_+^2}, \text{ and } h_1 = h^+ + 1.4874\sqrt{2}\lambda^+. \tag{3.10}$$

Similarly, when the process goes out of control at the first observation and $\sigma = \sigma_+$, then the 'out of control' average run length is given by

$$ARL_1^+ \equiv E_{1,\sigma_+}\left(T^-\right) \simeq \frac{e^{-a_1 h_1} + a_1 h_1 - 1}{a_1\left(\sigma_+^2 - \lambda^+\right)}. \tag{3.11}$$

In Tables 1a and 1b, we compare these values with those obtained through simulation.

h^+	σ_+								
	1.10	1.30	1.50	1.70	1.90	2.10	2.30	2.40	2.50
5	31(0.33)	40(0.43)	49(0.55)	60(0.66)	69(0.75)	77(0.85)	90(1.02)	93(1.03)	99(1.10)
	29	39	48	58	69	80	91	97	103
7	53(0.56)	76(0.83)	100(1.09)	125(1.39)	159(1.75)	187(2.05)	213(2.35)	225(2.51)	238(2.65)
	51	73	99	126	155	186	217	234	250
9	81(0.85)	129(1.38)	191(2.08)	259(2.88)	328(3.58)	413(4.65)	493(5.52)	540(5.90)	588(6.49)
	80	129	190	260	336	418	504	548	593
11	122(1.27)	218(2.35)	363(3.96)	524(5.88)	707(8.06)	915(10.1)	1138(12.7)	1266(14.0)	1359(15.3)
	119	217	352	519	711	923	1150	1269	1390
13	170(1.76)	348(3.80)	638(7.04)	998(11.1)	1486(16.5)	2040(22.8)	2610(29.6)	2879(31.8)	3181(35.6)
	170	352	637	1021	1487	2020	2607	2918	3238
14	202(2.09)	445(4.84)	861(9.45)	1423(15.8)	2165(24.0)	2956(33.0)	4034(45.9)	4413(49.9)	4859(53.7)
	200	444	853	1426	2144	2983	3920	4419	4937
15	232(2.45)	555(6.05)	1163(12.8)	1966(22.1)	3127(34.8)	4487(50.6)	5773(65.7)	6580(72.8)	7416(82.4)
	233	558	1138	1989	3088	4400	5888	6688	7522

Table 1a. Simulated and approximated values of ARL_0 for various values of h^+ and σ_+. Simulated values are based on 8000 replications. Standard errors are given in parentheses.

h^+	σ_+								
	1.10	1.30	1.50	1.70	1.90	2.10	2.30	2.40	2.50
5	19(0.19)	11(0.11)	8(0.07)	6(0.05)	5(0.04)	4(0.04)	4(0.03)	3(0.03)	3(0.03)
	17	9	6	4	3	3	2	2	2
7	28(0.27)	15(0.14)	10(0.09)	7(0.07)	6(0.05)	5(0.04)	4(0.04)	4(0.03)	4(0.03)
	26	14	8	6	4	3	3	2	2
9	38(0.37)	20(0.18)	12(0.11)	9(0.08)	7(0.06)	6(0.05)	5(0.04)	5(0.04)	4(0.04)
	37	18	11	7	5	4	3	3	3
11	50(0.48)	24(0.22)	15(0.13)	10(0.09)	8(0.07)	7(0.06)	5(0.05)	5(0.04)	5(0.04)
	49	22	13	9	7	5	4	4	3
13	62(0.59)	29(0.25)	17(0.15)	12(0.10)	9(0.08)	7(0.06)	6(0.05)	6(0.05)	5(0.04)
	61	27	16	11	8	6	5	4	4
14	70(0.65)	31(0.26)	19(0.15)	13(0.11)	10(0.08)	8(0.06)	6(0.05)	6(0.05)	5(0.04)
	68	30	17	11	8	6	5	4	4
15	76(0.69)	34(0.29)	20(0.17)	14(0.11)	10(0.08)	8(0.06)	7(0.05)	6(0.05)	6(0.05)
	75	32	18	12	9	7	5	5	4

Table 1b. Simulated and approximated values of ARL_1 for various values of h^+ and σ_+. Simulated values are based on 8000 replications. Standard errors are given in parentheses.

To obtain $ARL_1(\sigma)$ for a general alternative σ^2 with reference value σ_+^2, let

$$\phi^+(\sigma^2) = \frac{\lambda^+}{\sigma^2} + \log(\sigma^2)$$

It will be shown in Appendix B that for a given σ^2, there exists another value of σ^2, say σ_1^2, such that

$$\phi^+(\sigma^2) = \phi^+(\sigma_1^2)$$

The value of σ_1^2 so obtained will be less than λ^+ if $\sigma^2 > \lambda^+$ and greater than λ^+ if $\sigma^2 < \lambda^+$. An approximation of $ARL_1^+(\sigma)$ is given by

$$ARL_1^+(\sigma) \simeq \frac{e^{-a_1(\sigma)h_1} + a_1(\sigma)h_1 - 1}{|a_1(\sigma)(\sigma^2 - \lambda^+)|}, \tag{3.12}$$

where

$$a_1(\sigma) = \frac{\sigma^2 - \sigma_1^2}{2\sigma^2\sigma_1^2}, \text{ and } h_1 = h^+ + 1.4874\sqrt{2}\lambda^+.$$

To calculate $ARL_1^+(\sigma)$, we need to obtain σ_1 for a given value of σ, which is given in Table 2 below. A comparison of the theoretical values with the simulated values is presented in Table 3.

σ_+	λ^+	σ							
		1.1	1.3	1.5	1.7	1.9	2.1	2.3	2.5
1.1	1.04802	1.	0.868033	0.784569	0.72671	0.684011	0.651043	0.624705	0.603099
1.2	1.09242	1.08492	0.934722	0.840772	0.77615	0.728738	0.692297	0.663292	0.63957
1.3	1.13367	1.16909	1.	0.895337	0.823873	0.771727	0.731818	0.700162	0.674344
1.4	1.17215	1.25254	1.06395	0.948379	0.870009	0.813118	0.769752	0.735463	0.707573
1.5	1.20817	1.3353	1.12666	1.	0.914675	0.853035	0.806227	0.769328	0.739389
1.6	1.24200	1.41741	1.1882	1.05029	0.957974	0.89159	0.841357	0.801872	0.76991
1.7	1.27388	1.4989	1.24862	1.09934	1.	0.928878	0.875243	0.833197	0.799237
1.8	1.30399	1.57978	1.308	1.14722	1.04084	0.964989	0.907974	0.863393	0.827461
1.9	1.33250	1.66009	1.36637	1.194	1.08056	1.	0.93963	0.892541	0.854663
2.0	1.35956	1.73985	1.4238	1.23973	1.11923	1.03398	0.970284	0.920714	0.880916
2.1	1.38529	1.81908	1.48033	1.28449	1.15692	1.067	1.	0.947976	0.906284
2.2	1.40981	1.8978	1.536	1.32831	1.19367	1.09911	1.02884	0.974387	0.930826
2.3	1.43322	1.97602	1.59085	1.37125	1.22955	1.13037	1.05685	1.	0.954594
2.4	1.45560	2.05378	1.64492	1.41335	1.26461	1.16082	1.08408	1.02486	0.977638
2.5	1.47704	2.13107	1.69823	1.45466	1.29887	1.19051	1.11059	1.04902	1.

Table 2. Values of σ_1 used to obtain $ARL_1^+(\sigma)$ for various values of σ_+ and σ.

ARL_0	σ_+	h^+	σ							
			1.10	1.30	1.50	1.70	1.90	2.10	2.30	2.50
125	1.1	11.248	52(0.49)	20(0.16)	12(0.09)	8(0.06)	6(0.05)	5(0.04)	5(0.03)	4(0.03)
			49	18	10	7	5	4	3	2
	1.3	8.881	53(0.54)	19(0.18)	11(0.09)	8(0.06)	6(0.05)	5(0.04)	4(0.03)	4(0.03)
			43	16	9	6	4	3	2	2
	1.5	7.706	54(0.57)	19(0.19)	11(0.10)	7(0.06)	6(0.05)	5(0.04)	4(0.03)	3(0.03)
			44	16	8	5	4	3	2	2
	1.7	6.973	56(0.61)	20(0.20)	11(0.10)	7(0.07)	6(0.05)	4(0.04)	4(0.03)	3(0.03)
			41	15	8	5	4	3	2	2
	1.9	6.451	56(0.61)	20(0.21)	11(0.11)	7(0.07)	6(0.05)	5(0.04)	4(0.03)	3(0.03)
			48	17	9	6	4	3	2	2
	2.1	6.046	57(0.62)	20(0.21)	11(0.11)	7(0.07)	6(0.05)	4(0.04)	4(0.03)	3(0.03)
			55	19	10	6	4	3	2	2
	2.3	5.714	57(0.63)	21(0.22)	11(0.11)	7(0.07)	6(0.05)	4(0.04)	4(0.03)	3(0.03)
			44	16	8	5	4	3	2	2
	2.5	5.431	57(0.64)	20(0.22)	11(0.12)	7(0.07)	6(0.05)	4(0.04)	4(0.03)	3(0.03)
			49	18	9	6	4	3	2	2
250	1.1	15.464	81(0.74)	27(0.21)	15(0.11)	11(0.07)	8(0.06)	7(0.05)	5(0.04)	5(0.03)
			75	25	14	9	6	5	4	3
	1.3	11.582	84(0.86)	25(0.23)	14(0.11)	9(0.07)	7(0.05)	6(0.04)	5(0.04)	4(0.03)
			74	22	12	7	5	4	3	3
	1.5	9.882	89(0.95)	26(0.26)	14(0.12)	9(0.08)	7(0.05)	5(0.04)	4(0.04)	4(0.03)
			71	22	11	7	5	4	3	2
	1.7	8.893	92(0.99)	27(0.27)	14(0.13)	9(0.08)	7(0.06)	5(0.04)	4(0.04)	4(0.03)
			72	22	11	7	5	3	3	2
	1.9	8.224	95(1.03)	28(0.29)	14(0.13)	9(0.08)	7(0.06)	5(0.04)	4(0.04)	4(0.03)
			87	25	12	7	5	4	3	2
	2.1	7.725	94(1.05)	28(0.30)	14(0.14)	9(0.08)	7(0.06)	5(0.05)	4(0.04)	4(0.03)
			76	23	11	7	5	3	3	2
	2.3	7.329	98(1.08)	29(0.31)	14(0.14)	9(0.09)	7(0.06)	5(0.05)	4(0.04)	4(0.03)
			87	26	12	7	5	4	3	2
	2.5	6.999	96(1.08)	30(0.32)	14(0.15)	9(0.09)	7(0.06)	5(0.05)	4(0.04)	4(0.03)
			70	23	11	7	4	3	3	2
500	1.1	20.514	119(1.05)	35(0.24)	20(0.13)	14(0.08)	10(0.07)	8(0.05)	7(0.05)	6(0.04)
			111	33	18	12	8	6	5	4
	1.3	14.517	126(1.30)	32(0.28)	17(0.13)	11(0.08)	8(0.06)	6(0.05)	6(0.04)	5(0.03)
			117	30	15	9	7	5	4	3
	1.5	12.177	139(1.47)	33(0.31)	16(0.14)	11(0.09)	8(0.06)	6(0.05)	5(0.04)	4(0.03)
			133	31	15	9	6	5	4	3
	1.7	10.891	148(1.60)	36(0.36)	17(0.15)	10(0.09)	7(0.06)	6(0.05)	5(0.04)	4(0.03)
			119	30	14	8	6	4	3	3
	1.9	10.054	157(1.73)	37(0.38)	17(0.16)	11(0.10)	7(0.07)	6(0.05)	5(0.04)	4(0.03)
			152	35	15	9	6	4	3	3
	2.1	9.450	157(1.73)	39(0.41)	17(0.16)	10(0.10)	7(0.07)	6(0.05)	5(0.04)	4(0.03)
			140	34	15	9	6	4	3	3
	2.3	8.982	166(1.82)	41(0.45)	18(0.18)	11(0.10)	8(0.07)	6(0.05)	5(0.04)	4(0.04)
			121	32	14	8	5	4	3	2
	2.5	8.601	165(1.86)	41(0.45)	19(0.19)	11(0.11)	8(0.07)	6(0.05)	5(0.04)	4(0.04)
			138	35	15	9	6	4	3	3

Table 3. Simulated and approximated values of $ARL_1^+(\sigma)$ for various values of h^+ and σ_+. Simulated values are based on 8000 replications. Standard errors are given in parentheses.

3.2 Average Value of T^-

Following as in Section 3.1, we obtain

$$ARL_0^- \simeq \frac{e^{a_2 h_2} - a_2 h_2 - 1}{a_2\,(1-\lambda^-)}, \tag{3.13}$$

$$ARL_1^- \simeq \frac{e^{-a_2 h_2} + a_2 h_2 - 1}{a_2\,(\lambda^- - \sigma_-^2)}, \tag{3.14}$$

where

$$a_2 = \frac{1-\sigma_-^2}{2\sigma_-^2}, \text{ and } h_2 = h^- + 1.4874\sqrt{2}\lambda^-. \tag{3.15}$$

Similarly, for general σ,

$$ARL_1^-(\sigma) \simeq \frac{e^{-a_2(\sigma)h_2} - a_2(\sigma)h_2 - 1}{|a_2(\sigma)\,(\sigma^2 - \lambda^-)|}, \tag{3.16}$$

where

$$a_2(\sigma) = \frac{\sigma_2^2 - \sigma^2}{2\sigma^2\sigma_2^2}$$

and σ_2^2 is the solution of

$$\phi^-(\sigma^2) = \phi^-(\sigma_2^2)$$

for ϕ^- defined by

$$\phi^-(\sigma^2) = \frac{\lambda^-}{\sigma^2} + \log(\sigma^2).$$

σ_-	λ^-	σ							
		0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2
0.9	.898337	1.	1.14742	1.38536	1.82606	2.85291	6.55722	44.1043	15059.7
0.8	.793399	0.881617	1.	1.18682	1.52163	2.26165	4.68812	24.6098	4056.41
0.7	.685375	0.764891	0.857405	1.	1.24693	1.76375	3.30041	13.4878	1051.27
0.6	.574679	0.649875	0.719689	0.824649	1.	1.34693	2.28033	7.26629	263.497
0.5	.462098	0.536631	0.586904	0.660511	0.778937	1.	1.53725	3.84832	64.5044
0.4	.349063	0.425223	0.459104	0.50732	0.581863	0.712855	1.	1.99669	15.6919
0.3	.238148	0.315727	0.336356	0.364811	0.406959	0.476481	0.613722	1.	3.89451
0.2	.134120	0.208235	0.218744	0.232738	0.252505	0.282965	0.336946	0.460937	1.
0.1	.0465169	0.102877	0.106424	0.110948	0.116985	0.125588	0.139142	0.164547	0.233516

Table 4. Values of σ_2 used to obtain $ARL_1^-(\sigma)$ for various values of σ_- and σ.

3.3 Average Run Lengths of T

Let

$$\theta_1 = \frac{1}{\sqrt{2}}\left(1 - \frac{\lambda^+}{\sigma_+^2}\right), \quad \theta_2 = \frac{1}{\sqrt{2}}\left(1 - \frac{\lambda^-}{\sigma_-^2}\right).$$

Then, $\theta_2 < 0 < \theta_1$. Let

$$\psi(\theta_i) = -\frac{\theta_i}{\sqrt{2}} - \frac{1}{2}\log\left(1 - \theta_i\sqrt{2}\right), \quad i = 1, 2.$$

We shall assume that

$$\theta_1^{-1}\psi(\theta_1) - \theta_2^{-1}\psi(\theta_2) \geq \theta_1^{-1}a_1h^+ - \theta_2^{-1}a_2h^-.$$

Then, it follows from van Dobben de Bruyn [13] that

$$ARL_i^{-1} = \left(ARL_i^+\right)^{-1} + \left(ARL_i^-\right)^{-1}, \quad i = 0, 1.$$

4 Several Observations at Each Inspection

Often we have more than one observation to monitor for an increase or decrease in the variance. For example, in monitoring variance components in a nested model, such as considered by Yashchin [14], we have more than one observation at each stage. Also, it may be desireable to take more observations at each inspection or to apply the control scheme after collecting several observations, since the chi square distribution with one degree of freedom is more skewed than a chi square with, say, five degrees of freedom. This is reflected in the overshoot correction, as shown in Table 5, which approximately remains the same when we have five or more observations. Now, suppose at time n we observe $s \geq 1$ observations $x_{n1}, \ldots, x_{ns}$, which are independent and identically distributed. Let $\tilde{C}_{0,1} = 0$ and

$$\tilde{C}_{n,1} = \max\left\{0, \tilde{C}_{n-1,1} + \sum_{j=1}^{s} x_{nj}^2 - s\lambda^+\right\}, \quad n = 1, 2, \ldots$$

Then, the off-target signal for an increase in the variance from $\sigma^2 = 1$ to $\sigma^2 = \sigma_+^2$ is given at the $\tilde{T}^{+\,th}$ sample, where

$$\tilde{T}^+ = \min\left\{n : \tilde{C}_{n,1} \geq \tilde{h}^+\right\}.$$

The values of $(OSC)_s$ are obtained from the theoretical formula given in Srivastava and Chow [10]. It should be mentioned that for $s = 2$, this correction has already been evaluated by Siegmund [9].

As in Section 2, by Wald's sequential probability ratio test and applying corrections for the overshoot, we obtain

$$E_0\left(\tilde{T}^+\right) \simeq \frac{e^{a_1\tilde{h}_1} - a_1\tilde{h}_1 - 1}{sa_1\left(\lambda^+ - 1\right)}$$

$$E_{1,\sigma_+}\left(\tilde{T}^+\right) \simeq \frac{e^{-a_1\tilde{h}_1} + a_1\tilde{h}_1 - 1}{sa_1\left(\sigma_+^2 - \lambda^+\right)},$$

where

$$\tilde{h}_1 = \tilde{h}^+ + \lambda^+\sqrt{2s}(OSC)_s, \quad a_1 = \frac{\sigma_+^2 - 1}{2\sigma_+^2},$$

and $(OSC)_s$ denotes the overshoot correction for the sample size s.

s	1	2	3	4	5	6	7	8	9	10
$(OSC)_s$	1.4874	1.3333	1.2785	1.2490	1.2339	1.2225	1.2144	1.2081	1.2035	1.1996
s	11	12	13	14	15	16	17	18	19	20
$(OSC)_s$	1.1965	1.1939	1.1917	1.1898	1.1882	1.1867	1.1855	1.1843	1.1833	1.1824

Table 5. Overshoot Correction for various sample sizes.

Thus, the 'in control' and 'out of control' average run lengths are given by

$$ARL_{0s}^+ = sE_0^+\left(\tilde{T}_1\right),$$

$$ARL_{1s}^+(\sigma_+) = sE_{1,\sigma_+}^+\left(\tilde{T}_1\right)$$

respectively. It is clear from the above results that the only difference between the results for $s = 1$ and general s is in the correction term applied for the overshoot. Since the observations are taken s at a time, a larger overshoot is expected. Although OSC_s gets smaller for larger s, it is multiplied by $\sqrt{s}$, reflecting the larger correction that is needed. Expressions for general σ can be found in Srivastava and Chow [10].

5 Mean Unknown

When the mean is unknown, we need at least two observations. Thus, by letting $\tilde{s} = s - 1 \geq 1$, the $\tilde{C}_n$ of the last section becomes $\tilde{C}_0 = 0$ and

$$\tilde{C}_n = \max\left\{0, \tilde{C}_{n-1} + \sum_{j=1}^{s}(x_{nj} - \bar{x}_n)^2 - \tilde{s}\lambda^+\right\}$$

and so the results of the previous section are applicable here with $\tilde{s} = s - 1$.

6 Monitoring Variance Components

Yashchin [14] considered the problem of monitoring oxide thickness parameters based on measurements obtained on a lot-by-lot basis. From each lot, a sample of R wafers is selected and m measurements of oxide film thickness are taken from each wafer. It is assumed that the measurements follow a nested random effects model

$$\begin{aligned} &X_{irn} = \mu + L_i + W_{r(i)} + E_{irn} \\ &i = 1, 2, \ldots; \;\; r = 1, 2, \ldots, R; \;\; n = 1, 2, \ldots, N, \end{aligned} \tag{6.1}$$

where μ is the process mean, L_i is the random effect on the i^{th} lot, $W_{r(i)}$ is the nested random effect of the r^{th} wafer in the i^{th} lot, and E_{irn} is the random noise representing the effect of the n^{th} measurement taken from the r^{th} wafer of the i^{th} lot. It is assumed hat L_i, $W_{r(i)}$, and E_{irn} are independently distributed with mean zero and variances σ_b^2, σ_w^2, and σ^2 respectively. The target values of μ, σ_b^2, σ_w^2, and σ^2 are assumed known. We shall use the CUSUM procedure, since it is rarely beaten by other procedures (see, e.g., Pollak and Siegmund [9], Srivastava and Chow [10], and Srivastava and Wu [11]). The monitoring of the process mean is based on the sample mean of each lot,

$$\bar{X}_i = (RN)^{-1}\sum_{r=1}^{R}\sum_{n=1}^{N} X_{irn}, \;\; i = 1, 2, \ldots,$$

the monitoring of which has been carried out by Yashchin [14] by the CUSUM procedure. We shall therefore confine ourselves to monitoring for an increase in the variability from the given target values of σ_b^2, σ_w^2, and σ^2.

The monitoring of the process for an increase in σ^2 is based on the sample variance of each lot. Equivalently, it is based on

$$S_{\sigma i}^2 = \frac{1}{\sigma^2} \sum_{r=1}^{R} \sum_{n=1}^{N} \left(X_{rin} - \bar{X}_{ir\cdot} \right)^2, \quad i = 1, 2, \ldots$$

where

$$\bar{X}_{ir\cdot} = N^{-1} \sum_{n=1}^{N} X_{irn}.$$

This is equivalent to monitoring for an increase in the variance from 1 to σ_+^2 based on $S = R(N-1)$ observations considered in Section 4. Thus, the formulae for ARL_0 and $ARL_1(\sigma_+)$ given there can be used to obtain the decision boundary and the performance of the procedure.

It remains to monitor for changes in σ_b^2 and σ_w^2. Although traditionally the process is monitored for a change from a target value of [illegible] to an increase in the variance, Yashchin [14] has given an example where the problem involves monitoring these variances around a large target value. The monitoring by Yashchin is based on the difference of two statistics which may be subject to sudden abrupt change. Thus, our monitoring of σ_w^2 will be based on

$$S_{wi}^2 = \frac{R}{\sigma_w^2 + \frac{\sigma^2}{N}} \sum_{r=1}^{R} \left(\bar{X}_{ir\cdot} - \bar{X}_i^2 \right)^2, \quad i = 1, 2, \ldots$$

which has a chi-square distribution on $R-1$ degrees of freedom. Again the results of Section 4 are applicable.

The monitoring of σ_b^2 should be based on the statistics

$$S_{bi}^2 = \left[\sigma_b^2 + \frac{\sigma_w^2}{R} + \frac{\sigma^2}{RN} \right]^{-1} \frac{\left[(\bar{X}_1 + \cdots + \bar{X}_i) - i\bar{X}_{i+1} \right]^2}{i(i+1)}, \quad i = 1, 2, \ldots$$

which are independently distributed as chi-squares, each with one degree of freedom. Thus, the results of Sections 2 and 3 are applicable.

We shall now apply the above CUSUM methods to monitor the data of Example 3.2 given in Yashchin [14] with target values

$$\sigma = 20, \;\; \sigma_w = 30, \;\; \sigma_b = 60,$$

and where

$$R = 2 \;\text{ and }\; N = 4.$$

Choosing the reference value σ_+ as $1 + \sqrt{2}$, which is one plus one standard deviation, we obtain

$$\lambda^+ = 2.128$$

The monitoring of σ^2, σ_w^2, and σ_b^2 are based on

$$\begin{aligned}\tilde{C}_{n,1} &= \max\left\{0, \tilde{C}_{n-1,1} + S_{\sigma n}^2 - 6\lambda^+\right\}, \;\; \tilde{C}_{0,1} = 0, \\ \tilde{C}_{n,2} &= \max\left\{0, \tilde{C}_{n-1,2} + S_{wn}^2 - \lambda^+\right\}, \;\; \tilde{C}_{0,2} = 0,\end{aligned}$$

and

$$\tilde{C}_{n,3} = \max\left\{0, \tilde{C}_{n-1,3} + S_{bn}^2 - \lambda^+\right\}, \;\; \tilde{C}_{0,3} = 0,$$

respectively. Figure I shows the CUSUM charts. We arrive at the same conclusion as Yashchin [14], but without the intricacies involved in choosing reference values and the need for numerical methods to obtain the decision boundary.

Please place Fig. 1 about here

We also plotted Shewhart processes in Fig. 2, which is a plot of the standardized variance components (standardized by degrees of freedom and target variance). The Wafer to Wafer process reaches approximately 6.81 at the 19^{th} observation, but is not detected by the Shewhart chart. However, the CUSUM process detects this shift at the 24^{th} observation, a delay of only 5 observations. Yashchin's method also detects it, but at the 30^{th} observation, which is a delay of 11 observations.

Please place Fig. 2 about here

Finally, it should be added that we arrive at the same conclusion if the reference value σ_+ is chosen to be $1 + \frac{1}{\sqrt{2}}$, which is one plus one half of the standard deviation. Thus, it appears that the CUSUM process for the variance may not be as sensitive to the selection of reference value as it is when used for detecting a shift in the mean – a nice feature.

7 Conclusion

In this paper, I have provided analytical results for the CUSUM procedure for detecting an increase/decrease in the variance. For example, the decision boundary can be obtained in a matter of seconds without resorting to extensive numerical methods. The method has also been applied to an example considered earlier by Yashchin. By applying the CUSUM method directly, it has been shown that it performs better than the method followed by Yashchin.

A comparison of the CUSUM method with other methods proposed in the literature including the Shiryayev-Roberts procedure was carried out by Srivastava and Chow (1992). It is shown there that the CUSUM procedure is rarely beaten by any of these procedures in terms of ARL_1 and the conditional average delay time.

Thus, in conclusion, the CUSUM procedure is recommended for detecting an increase/decrease in the variance.

Appendix A

Here, we show that $1 \leq \lambda^+ \leq \sigma_+^2$, where

$$\lambda^+ = \frac{\sigma_+^2 \log(\sigma_+^2)}{\sigma_+^2 - 1}$$

First, we note that

$$e^{\frac{1}{\sigma_+^2} - 1} \geq 1 + \left(\frac{1}{\sigma_+^2} - 1 \right)$$

Hence,

$$\log\left(\frac{1}{\sigma_+^2}\right) \leq \frac{1}{\sigma_+^2} - 1,$$

which gives

$$\sigma_+^2 \log(\sigma_+^2) \geq \sigma_+^2 - 1$$

Thus,

$$\lambda^+ = \frac{\sigma_+^2 \log(\sigma_+^2)}{\sigma_+^2 - 1} \geq 1$$

Again, we note that

$$e^{\sigma_+^2 - 1} \geq 1 + \left(\sigma_+^2 - 1\right)$$

Hence

$$\log(\sigma_+^2) \leq \left(\sigma_+^2 - 1\right)$$

Thus,

$$\lambda^+ = \frac{\sigma_+^2 \log(\sigma_+^2)}{\sigma_+^2 - 1} \leq \sigma_+^2, \quad \text{since } \sigma_+^2 - 1 > 0.$$

Next, we show that

$$\sigma_-^2 \leq \lambda^- \leq 1.$$

We have

$$e^{\sigma_-^{-2} - 1} \geq 1 + \left(\sigma_-^{-2} - 1\right),$$

and hence,

$$\log(\sigma_-^{-2}) \leq \left(\sigma_-^{-2} - 1\right), \quad \text{where } \sigma_-^{-2} - 1 > 0$$

giving

$$\frac{\log\left(\sigma_-^{-2}\right)}{\sigma_-^{-2}-1} \leq 1.$$

But,

$$\frac{\log\left(\sigma_-^{-2}\right)}{\sigma_-^{-2}-1} = \frac{\sigma_-^2 \log\left(\sigma_-^{-2}\right)}{1-\sigma_-^2} = \frac{\sigma_-^2 \log\left(\sigma_-^2\right)}{\sigma_-^2-1} = \lambda^-$$

Hence

$$\lambda^- \leq 1.$$

Similarly,

$$e^{\sigma_-^2-1} \geq 1 + \left(\sigma_-^2 - 1\right)$$

giving

$$\log\left(\sigma_-^2\right) \leq \sigma_-^2 - 1$$

Thus,

$$\log\left(\sigma_-^{-2}\right) \geq 1 - \sigma_-^2 > 0$$

Hence,

$$\lambda^- = \frac{\sigma_-^2 \log\left(\sigma_-^{-2}\right)}{1-\sigma_-^2} \geq \sigma_-^2.$$

Appendix B
Average Run Length Under General Alternatives

In this section, we derive the average run length when the variance is $\sigma^2 > \lambda^+$. From Section 2, we have

$$\begin{aligned}
\log \frac{f_1(x)}{f_0(x)} &= a_1(x^2 - \lambda^+), \;\; a_1 = \frac{\sigma_+^2 - 1}{2\sigma_+^2} \\
&= a_1 \lambda^+ \sqrt{2} \frac{\left[\frac{x^2}{\lambda^+} - 1\right]}{\sqrt{2}} \\
&= a_1 \lambda^+ \sqrt{2} W.
\end{aligned}$$

Since $x \sim N(0, \sigma^2)$, the random variable (x^2/σ^2) has a chi-square distribution with one degree of freedom. Hence, the pdf of

$$W = \frac{\left[\frac{x^2}{\lambda^+} - 1\right]}{\sqrt{2}}$$

is given by

$$g_\theta(w) = e^{\theta w - \psi(\theta)} f_0(w),$$

where

$$\begin{aligned}
f_0(w) &= \frac{1}{\sqrt{2}\Gamma\left(\frac{1}{2}\right)} \left(\frac{w\sqrt{2}+1}{\sqrt{2}}\right)^{\frac{1}{2}-1} e^{-\frac{1}{2}(w\sqrt{2}+1)}, \;\; w > -\frac{1}{\sqrt{2}}, \\
\theta &= \frac{1}{\sqrt{2}}\left(1 - \frac{\lambda^+}{\sigma^2}\right), \\
\psi(\theta) &= -\frac{\theta}{\sqrt{2}} - \frac{1}{2}\log\left(1 - \theta\sqrt{2}\right) = \frac{1}{2}\left[\frac{\lambda^+}{\sigma^2} + 2\log\sigma - (1 + \log\lambda^+)\right].
\end{aligned}$$

It may be noted that $f_0(w)$ is a pdf with mean zero and variance one. Hence, $\psi'(0) = 0$, where $\psi'(0)$ is the derivative of $\psi(\theta)$ evaluated at $\theta = 0$. Since $\psi(\cdot)$ is a convex function and $\theta > 0$ (as $\sigma^2 > \lambda^+$) there exists $\theta_0 < 0 < \theta$ such that $\psi(\theta_0) = \psi(\theta)$. For a $\theta_0 < 0 < \theta$ to exist, there must exist a $\sigma_1^2 < \lambda^+ < \sigma^2$ such that

$$\frac{\lambda^+}{\sigma_1^2} + 2\log(\sigma_1) = \left(\frac{\lambda^+}{\sigma^2}\right) + 2\log(\sigma)$$

This can be solved iteratively. Define θ_0 by

$$\begin{aligned}\theta_0 &= \frac{1}{\sqrt{2}}\left(1-\frac{\lambda^+}{\sigma_1^2}\right) < 0,\\ a_1(\sigma) &= \frac{\theta-\theta_0}{\lambda^+\sqrt{2}} = \frac{\sigma^2-\sigma_1^2}{2\sigma_1^2\sigma^2}\end{aligned}$$

Thus, we can write (since $\psi(\theta)=\psi(\theta_0)$)

$$\begin{aligned}\log\left(\frac{g_\theta(w)}{g_{\theta_0}(w)}\right) &= (\theta-\theta_0)w\\ &= \frac{\theta-\theta_0}{\lambda^+\sqrt{2}}\lambda^+\sqrt{2}w\\ &= \frac{a_1(\sigma)}{a_1}a_1\lambda^+\sqrt{2}w\\ &= \frac{a_1(\sigma)}{a_1}\log\left(\frac{f_1(x)}{f_0(x)}\right)\end{aligned}$$

Hence, after applying the correction for the overshoot, we get

$$ARL_1(\sigma) = \frac{e^{-a_1(\sigma)h_1} + a_1(\sigma)h_1 - 1}{a_1(\sigma)(\sigma^2-\lambda^+)}.$$

Similarly, we can obtain results for the case when $\sigma^2 < \lambda_1$.

References

[1] Brook, D. and Evans, D.A. (1972). "An Approach to the Probability Distribution of CUSUM Run Length", *Biometrika*, **59**, 539-549.

[2] Crowder, S.V. and Hamilton, M.D. (1992). "An EWMA for monitoring a Process Standard Deviation", *Journal of Quality Technology*, **24**, 12-21.

[3] Domangue, R. and Patch, S.C. (1991). "Some Omnibus Exponentially Weighted Moving Average Statistical Process Monitoring Schemes", *Technometrics*, **33**, 299-313.

[4] Khan, R.A. (1978). "Wald's Approximations to the Average Run Length in CUSUM Procedures", *J. Statist. Planning and Inference*, **2**, 63-77.

[5] Lorden, G. (1971). "Procedures for Reacting to a Cahnge in Distribution", *Ann. Math. Statist.*, **42**, 1897-1908.

[6] Moustakides, G.V. (1986). "Optimal Stopping Time for Detecting Change in Distribution", *Ann. Statist.*, **14**, 1379-1387.

[7] Page, E.S. (1954). "Continuous Inspection Schemes", *Biometrika*, **41**, 100-114.

[8] Reynolds, M. (1975). "Approximations to the Average Run Length In Cumulative Sum Control Charts", *Technometrics*, **17**, 65-71.

[9] Siegmund, D. (1985). "Sequential Analysis: Tests and Confidence Intervals", Springer, Berlin.

[10] Srivastava, M.S. and Chow, W. (1992). "Comparison of the CUSUM Procedure with Other Procedures that Detect an Increase in the Variance and a Fast Accurate Approximation for the ARL of the CUSUM Procedure", *Tech. Report Series*, **No. 9122**, Department of Statistics, University of Toronto.

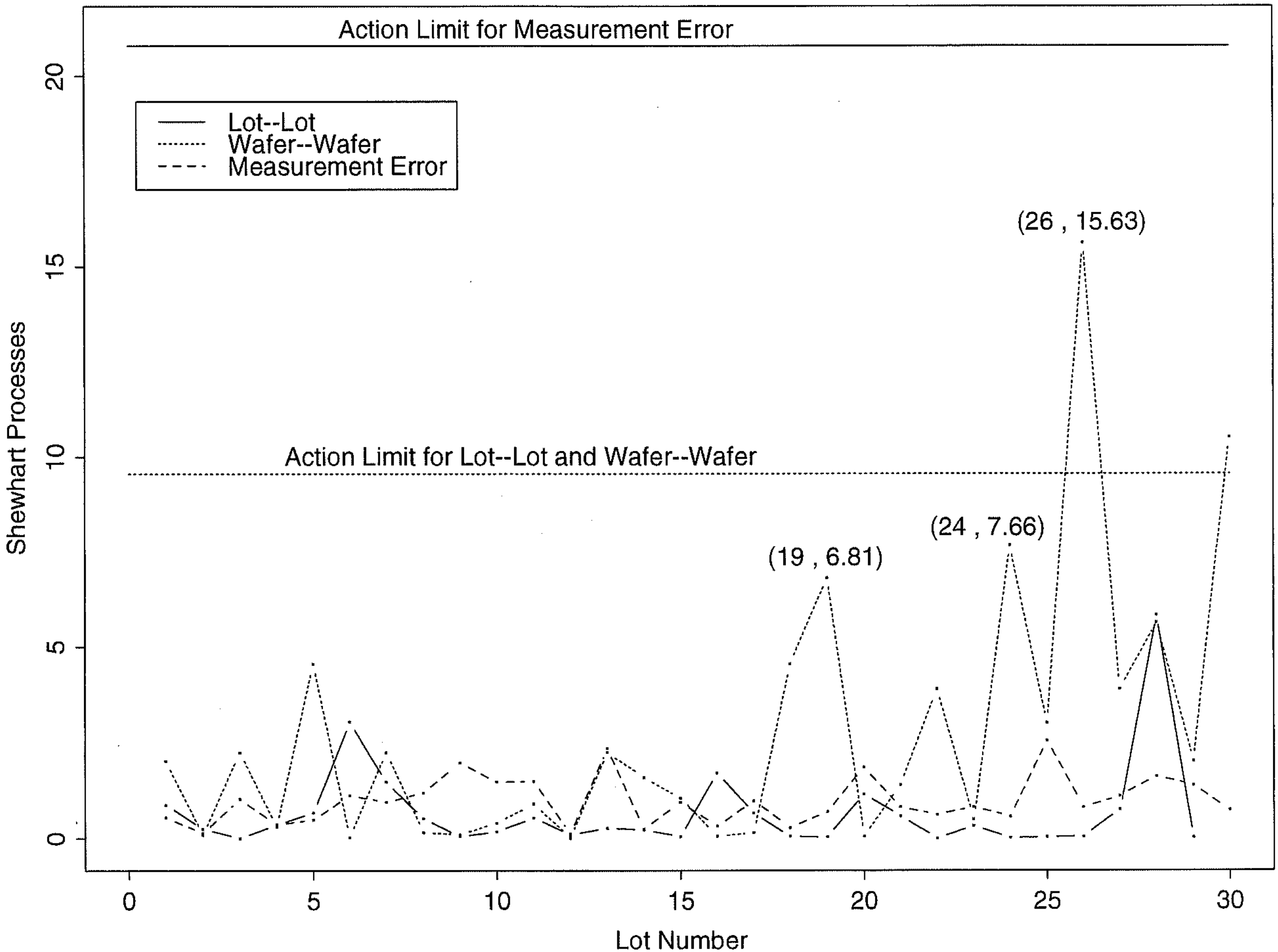
[11] Srivastava, M.S. and Wu, Y. (1993). "Comparison of EWMA, CUSUM, and Shiryayev-Roberts Procedures for Detecting a Shift in the Mean", *Ann. Statist.*, **21**, 645-670.

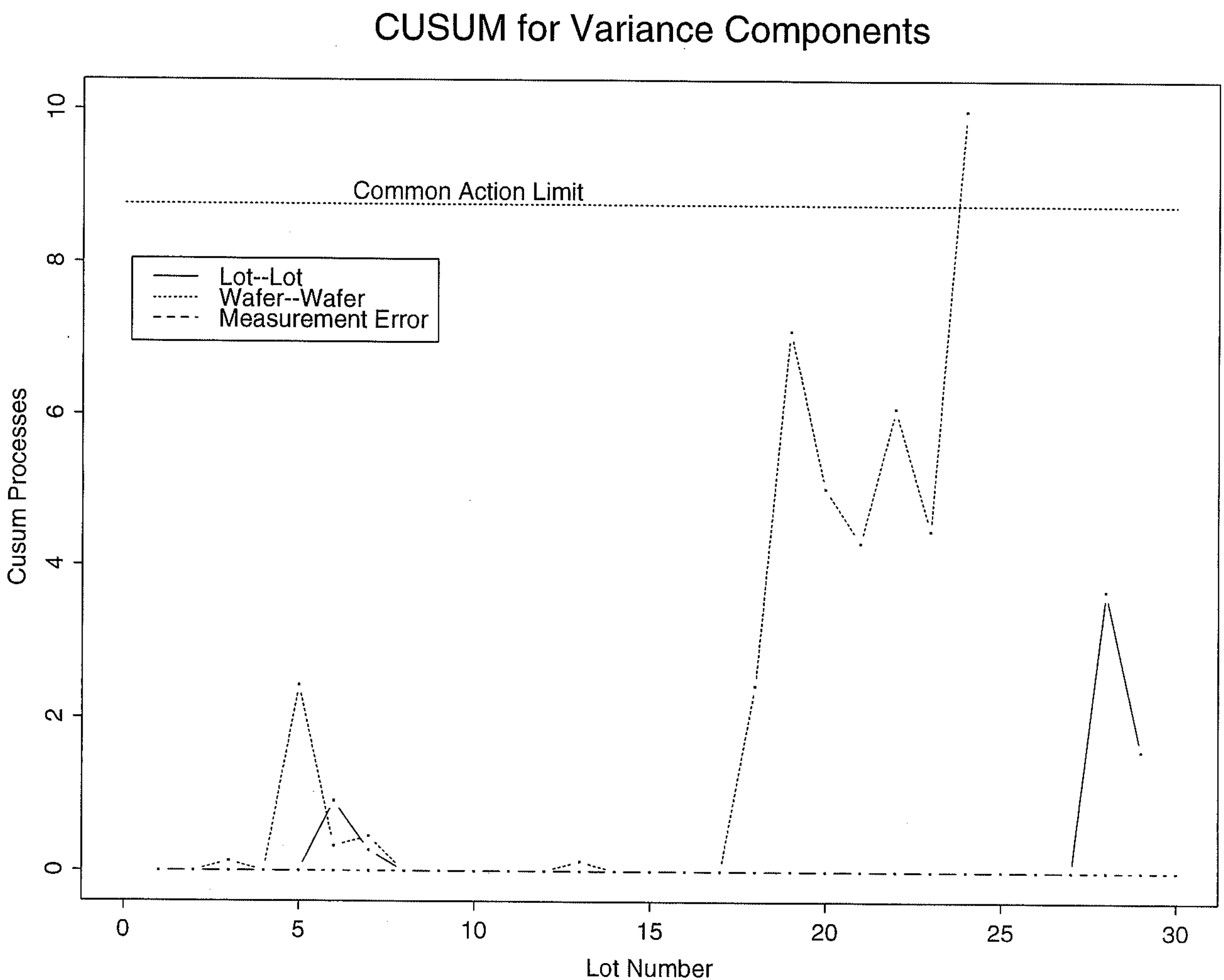
[12] Tuprah, K. and Ncube, M. (1987). "A comparison of dispersion quality control charts", *Sequential Analysis*, **6**, 155-163.

[13] van Dobben de Bruyn, C.S. (1968). "Cumulative Sum Tests", Griffith, London.

[14] Yashchin, E. (1994). "Monitoring Variance Components", *Technometrics*, **4**, 379-393.

Variance Components (Standardized)




CUSUM for Variance Components
Common Action Limit
Lot--Lot
Wafer--Wafer
Measurement Error
Cusum Processes
0
2
4
6
8
10
Lot Number
0
5
10
15
20
25
30