Journal of the KSQC
Vol. 20. No. 1. June. 1992

1992년 6월 韓國品質管理學會誌 제20권 제1호

Determining the Decision Limit of CUSUM Chart for A Fixed Sample Size

Chang Wook Kang*
Douglas M. Hawkins**

Abstract

When we compare different control charting schemes. the average run length of each control chart is usually used. The use of the average run length implies that there is unbounded number of samples or observations. The regression recursive residuals. however. have been applied to the cumulative sum chart to detect whether the mean or variance changes. To implement choice of decision interval. we calculate the probability that certain fixed number of control statistics stay in the in-control state. This probability can be used as the significance level of a test for detecting the change in the residual mean or variance of the data with a finite number of observations.

1. Introduction

Since their introduction, the cumulative sum(CUSUM) control chart have found their widespread applications in monitoring manufacturing processes. It implies that the CUSUM control schemes are used with unbounded number of samples from the process. After Brwon, Durbin, and Evans(1975) used first the

* Hanyang University
* * University of Minnesota

CUSUM of recursive residuals to test the constancy of regression relationships over time, the applications of CUSUMs of the recursive residuals in the ordinary regression problems have been studied in Galpin and Hawkins(1984), Kang(1990), and Hawkins(1991). In situations when we apply the CUSUM of recursive residuals to test outlyingness of the data, the design procedures for the CUSUM control chart may be slightly different from the standard design steps described in Lucas(1976). There are two parameters, k and h, in the CUSUM chart. k is usually set to be $\triangle/2$ where $\triangle$ is the shift to be detected and is a multiple of standard deviation. The h value of CUSUM control chart for the industrial applications can be chosen based on the average run length (ARL), while it can be chosen based on the false alarm rate for the CUSUM chart of a fixed sample size.

In this article, we discuss the calculation of the probability that the fixed number of statistics remain in control. This probability is considered as the significance level of the test to check whether there exists any oulier in the data. In section 2, we describe the procedures which we calculate the probability. The results are summarized in section 3.

2. Calculation

Let the Xi be the control statistic and assume to have the independent and identical normal distribution with mean 0 and variance 1 without loss of generality for $i = 1, \cdots, n$. To implement the decision interval form of cusum, we initially set up the starting value :

$$S_0^+ = 0$$

As values X_i become available, we update the cusums :

$$S_i^+ = \max(0,\ S_{i-1}^+ + x_i - k) \qquad (1)$$

where k is the allowance value for each observation and is usually half of the shift to detect. While the process is in control, Si^+ will characteristically 'bump along' on the axis at the value zero, then makes an excursion away from zero. These excursions away from zero terminate, either in the cusum going back to zero or in exceeding the value h. In the CUSUM chart, h denotes the decision limit instead as UCL or LCL in other control charts. The in-control state means that the Si^+ remains under h. Therefore if Si^+ exceeds h, then the out-of-control signal will be given. To test the outlyingness of observations in the data, we need to predetermine the significance level of the test. It makes that we need to calculate the probability that all the data remain under control. The decision limit, h, will be chosen based on one minus this probability.

We can write the probability that the ith observation is under control given all the observations before the ith one are also under control as

$Pr[S_i^- < h \mid S_j^- < h \text{ for all } j < i]$ .... (2)

Being based on the definition of S_i^+, we can think of four different situations to trace the paths from the (i-1)th state to the ith state. They follow :

(i) $0 < S_i^- < h$, $S_{i-1}^+ = 0$
(ii) $0 < S_i^- < h$, $0 < S_{i-1}^- < h$
(iii) $S_i^- = 0$, $S_{i-1}^+ = 0$ .... (3)
(iv) $S_i^- = 0$, $0 < S_{i-1}^+ < h$

And we difine $P1_i$, $P2_i$, $P3_i$, and $P4_i$ as the probability of (i), (ii), (iii) and (iv) respectively.
If we compute the probabilities of the above four events, the sum of four probabilities is the probability that all the observations upto and including the ith case are in control. To compute the exact probability, we have to solve the integration. We apply the trapezoidal rule to numerical integration and get the approximate results. Let define ds the equally spaced subinterval between 0 and h. Then the trapezoidal rule is

$$\int_a^b f(x)\,dx \cong ds/2(f_0 + 2f_1 + \cdots + 2f_{n-1} + f_n)$$

where $f_0 = f(a)$, $fn = f(b)$, $ds = (b-a)/n$, and $f_i = f(a + i \cdot ds)$.
we define the event $E_i(s)$,

$E_i(s) = \{S_i^- \in (s, s+ds)$ and $S_j^+ < h$ for all $j<i\}$ .... (4)

where $s > 0$. And we define another event D_i.

$D_i = \{S_i^+ = 0$ and $S_j^+ < h$ for all $j<i\}$ .... (5)

We can rewrite $E_i(s)$ as

$E_i(s) = \{\bigcup_{t=0}^{h} [E_{i-1}(t)$ and $E_i(s)]\} \cup \{D_{i-1}$ and $E_i(s)\}$ .... (6)

Also D_i can be rewritten as

$D_i(s) = \{\bigcup_{t=0}^{h} [E_{i-1}(t)$ and $D_i(s)]\} \cup \{D_{i-1}$ and $D_i(s)\}$ .... (7)

From(4), given $E_{i-1}(t)$, $S_i^- = \max(0, t+x_i-k)$ and $S_i^+ = 0$ if $t+x_i-k < 0$.

Therefore

$$P_r[S_i^+=0 \mid E_{i-1}(t)]=P_r[x_i<k-t]=\Phi(k-t) \quad \cdots (8)$$

Where $\Phi(\cdot)$ is the cumulative function of the standard normal distribution.
We define the probability of $E_i(s)$ and D_i as

$$Pr[E_i(s)]=f_i(s)\cdot ds \text{ and} \quad \cdots (9)$$

$$Pr[D_i]=P_i \quad \cdots (10)$$

Where

$$f_i(s)=\int_0^h f_{i-1}(t)\cdot\phi(s-t+k)\,dt+P_{i-1}\cdot\phi(s+k) \quad \cdots (11)$$

$$P_i=\int_0^h f_{i-1}(t)\cdot\Phi(k-t)\,dt+P_{i-1}\cdot\Phi(k) \quad \cdots (12)$$

Where $\phi(\cdot)$ is the standard normal density function.
Therefore

$$\begin{aligned}P_r[S_1^+<h] &= P_r[S_1^+=0 \text{ and } S_0^+=0]+P_r[S_1^+=0 \text{ and } 0<S_0^+<h]\\ &+P_r[0<S_1^+<h \text{ and } S_0^+=0]\\ &+P_r[0<S_1^+<h \text{ and } S_0^+<h]\\ &=P1_1+P2_1+P3_1+P4_1\\ &=\Phi(k)+0+\int_0^h \phi(s+k)\,ds+0 \end{aligned} \quad \cdots (13)$$

since $S_0^+=0$ by definition. Thus $f_1(s)=\phi(s+k)$.

$$\begin{aligned}P_r[S_2^+<h] &= P_r[S_2^+=0 \text{ and } S_1^+=0]+P_r[S_2^+=0 \text{ and } 0<S_1^+<h]\\ &+P_r[0<S_2^+<h \text{ and } S_1^+=0]\\ &+P_r[0<S_2^+<h \text{ and } 0<S_1^+<h]\\ &=P1_2+P2_2+P3_2+P4_2 \end{aligned} \quad \cdots (14)$$

for $P1_1=\Phi(k)\cdot(P1_1+P2_1)=\Phi^2(k)$

$$P2_2=\int_0^h \phi(t+k)\cdot\Phi(k-t)dt$$

$$P3_2 = (P1_1 + P2_1) \cdot \int_0^h \phi(s+k)\,ds$$

$$P4_2 = \int_0^h \int_0^h \phi(t+k) \cdot \phi(s-t+k)\,dtds$$

Thus $f_2(s) = \int_0^h f_1(t) \cdot \phi(s-t+k)\,dt + (P1_1 + P2_1) \cdot \phi(s+k)$

$$\begin{aligned} P_r[S_3^+ < h] &= P_r[S_3^+ = 0 \text{ and } S_2^+ = 0] + P_r[S_3^+ = 0 \text{ and } 0 < S_2^+ < h] \\ &\quad + P_r[0 < S_3^+ < h \text{ and } S_2^+ = 0] \\ &\quad + P_r[0 < S_3^+ < h \text{ and } 0 < S_2^+ < h] \\ &= P1_3 + P2_3 + P3_3 + P4_3 \end{aligned} \quad \cdots\cdots (15)$$

For $P1_3 = \Phi(k) \cdot (P1_2 + P2_2)$

$$P2_3 = \int_0^h f_2(t) \cdot \Phi(k-t)\,dt$$

$$P3_3 = (P1_2 + P2_2) \cdot \int_0^h \phi(s+k)\,ds$$

$$P4_3 = \int_0^h \int_0^h f_2(t) \cdot \phi(s-t+k)\,dtds$$

Thus $f_3(s) = \int_0^h f_2(t) \cdot \phi(s-t+k)\,dt + (P1_2 + P2_2) \cdot \phi(s+k)$

.

.

.

$$\begin{aligned} P_r[S_i^+ < h] &= P_r[S_i^+ = 0 \text{ and } S_{i-1}^+ = 0] + P_r[S_i^+ = 0 \text{ and } 0 < S_{i-1}^+ < h] \\ &\quad + P_r[0 < S_i^+ < h \text{ and } S_{i-1}^+ = 0] \\ &\quad + P_r[0 < S_i^+ < h \text{ and } 0 < S_{i-1}^+ < h] \\ &= P1_i + P2_i + P3_i + P4_i \end{aligned} \quad \cdots\cdots (16)$$

For $P1_i = \Phi(k) \cdot (P1_{i-1} + P2_{i-1})$

$$P2_i = \int_0^h f_{i-1}(t) \cdot \Phi(k-t)\,dt$$

$$P3_i = (P1_{i-1} + P2_{i-1}) \cdot \int_0^h \phi(s+k)\,ds$$

$$P4_i = \int_0^h \int_0^h f_{i-1}(t) \cdot \phi(s-t+k)\,dtds$$

Thus $f_i(s) = \int_0^h f_{i-1}(t) \cdot \phi(s-t+k)\, dt + (P1_{i-1} + P2_{i-1}) \cdot \phi(s+k)$

By (13) to (16), we calculated the probability that the process is under control upto and including the ith case. We summarized the results in the following section.

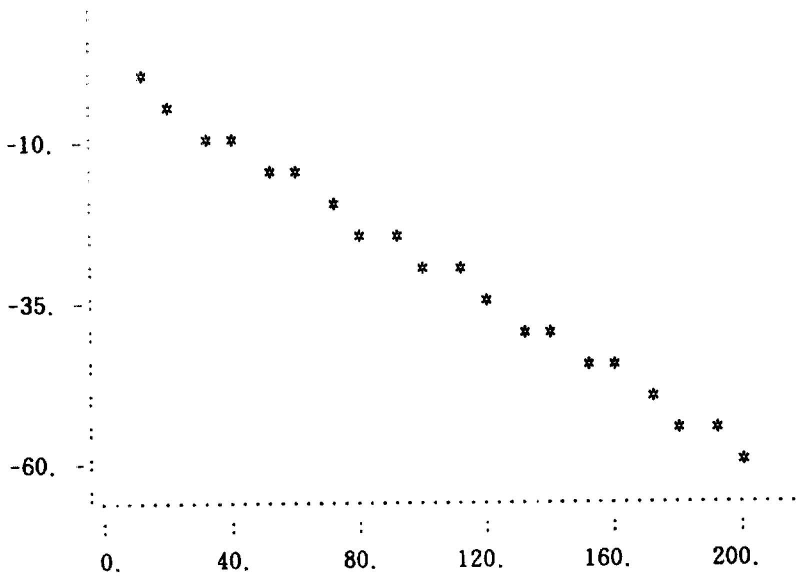
3. Summary

We present the probabilities in the tables 1 to 4. Table 1 shows the probabilities taking k=0.1 with varying values of h and sample size. Tables 2 to 4 have the same structure except for different k values, 0.25, 0.5, and 1.0 respectively. From Tables 1 to 4, we can see that the smaller shift detection we want, the larger h value should be used for the desired significance level. From the probabilities in Tables 1 to 4, we looked into the relationship between probability and the sample size n. Figure 1 shows the scatter plot of sample size n versus the logarithmic transformation of the probability for k=0.5 and h=4.0. We found that the linear regression model 100 · log(probability) = a + b · n is very good fit of the data(first column vs each other column). That is, no matter what the values of k and h are, the above regression is excellently fitted. Table 5 gives the result of regession analysis for k=0.5 and h=4.0. All of other regression results for different k and h values are pretty similar. Therefore we can get the predicted probability for the sample size which is not in the table.

Figure 1. The scatter plot of sample size n versus the logarithmic transformation of probability for k=0.5 and h=4.0

Plot : x-axis=n

y-axis=100 · ln(probability)



Tabel 1. Probability of n control statistics ~ N(0, 1) to stay in control for k=0.1

h / n	2.0	2.5	3.0	3.5	4.0	4.5	5.0	5.5	6.0
10	.4462	5885	.7014	.7876	.8519	.8937	.9321	.9554	.9713
20	.1742	3007	.4266	.5398	.6363	.7160	.7805	.8320	.8727
30	.0680	1536	.2593	.3692	.4732	.5660	.6461	.7138	.7701
40	.0265	0785	.1576	.2526	.3519	.4474	.5345	.6114	.6779
50	.0104	0401	.0958	.1728	.2617	.3535	.4421	.5236	.5965
60	.0040	0205	0582	.1182	.1946	.2704	.3657	.4484	.5249
70	.0016	0105	.0354	.0808	.1447	.2208	.3025	.3840	.4618
80	0006	0053	.0215	.0553	.1076	.1745	.2502	.3289	.4063
90	.0002	0027	.0131	.0378	.0800	.1379	.2069	.2817	.3575
100	.0001	0014	.0079	.0259	.0595	.1000	.1712	.2412	.3146
110	.0000	0007	0048	.0177	.0442	.0861	.1416	.2066	.2768
120	.0000	0004	.0029	.0121	.0329	.0681	.1171	.1769	.2435
130	0000	0002	.0018	.0083	.0245	.0538	.0969	.1515	.2143
140	.0000	0001	.0011	.0057	.0182	.0425	.0801	.1298	.1885
150	.0000	0000	.0007	.0039	.0135	.0336	.0663	.1111	.1659
160	.0000	0000	0004	.0026	.0101	.0266	.0548	.0952	.1460
170	.0000	0000	.0002	.0018	.0075	.0210	.0453	.0815	.1284
180	.0000	0000	.0001	.0012	.0056	.0166	.0375	.0698	.1130
190	.0000	0000	.0001	.0008	.0041	.0131	.0310	.0598	0994
200	.0000	0000	.0001	.0006	.0031	.0104	.0257	.0512	.0875

Table 2. Probability of n control statistics ~ N(0, 1) to stay in control for k=0.25

h / n	2.0	2.5	3.0	3.5	4.0	4.5	5.0	5.5	6.0
10	.5910	7265	.8208	.8847	.9270	.9546	.9723	.9834	.9902
20	.3207	4853	.6231	.7296	.8086	.8658	.9067	.9356	9560
30	.1740	3241	.4729	.6012	.7040	.7829	.8421	.8859	.9179
40	.0944	2165	.3589	.4954	.6129	.7079	.7819	.8384	.8808
50	.0512	1446	.2723	.4082	.5335	.6400	.7260	.7934	.8452
60	.0278	0966	.2067	.3363	.4645	.5787	.6741	.7508	.8109
80	.0082	0431	.1190	.2283	.3521	.4731	5812	.6724	.7466
90	.0044	0288	.0903	.1881	.3065	.4277	5396	.6363	.7163
100	.0024	0192	.0686	.1550	.2668	.3867	.5011	.6022	.6873
110	.0013	0128	.0520	.1277	.2323	.3497	.4652	.5699	.6595
120	.0007	0086	.0395	.1053	.2022	.3161	.4320	.5393	.6328
130	.0004	0057	.0300	.0867	.1761	.2858	.4011	.5104	.6071
140	.0002	0038	.0227	.0715	.1533	.2584	.3724	.4830	.5826
150	.0001	0026	.0173	.0589	.1334	.2337	.3458	.4571	.5590
160	.0001	0017	.0131	.0485	.1162	.2113	.3211	.4325	.5363
170	.0000	0011	.0099	.0400	.1011	.1910	.2981	.4093	.5146
180	.0000	0008	.0075	.0329	.0880	.1727	.2768	.3874	.4938
190	.0000	0005	.0057	.0271	.0767	.1562	.2570	.3666	.4738
200	.0000	0003	.0043	.0224	.0667	.1412	.2386	.3469	.4546

Tabel 3. Probability of n control statistics ~ N(0, 1) to stay in control for k=0.5

n \ h	2.0	2.5	3.0	3.5	4.0	4.5	5.0	5.5	6.0
10	.7898	.8851	.9380	.9669	.9826	.9909	.9954	.9977	.9988
20	.6021	.7605	.8598	.9191	.9536	.9736	.9850	.9916	.9953
30	.4590	.6535	.7881	.8734	.9254	.9563	.9745	.9852	.9914
40	.3499	.5616	.7224	.8301	.8979	.9393	.9641	.9788	.9876
50	.2667	.4825	.6622	.7889	.8713	.9226	.9538	.9725	.9837
60	.2033	.4146	.6069	.7497	.8454	.9062	.9436	.9663	.9799
70	.1550	.3563	.5563	.7125	.8203	.8900	.9335	.9600	.9761
80	.1182	.3062	.5099	.6771	.7960	.8742	.9235	.9538	.9723
90	.0901	.2631	.4674	.6435	.7724	.8586	.9136	.9477	.9685
100	.0687	.2261	.4284	.6116	.7495	.8434	.9038	.9416	.9647
110	.0524	.1943	.3927	.5812	.7272	.8284	.8942	.9355	.9610
120	.0399	.1669	.3599	.5524	.7057	.8136	.8846	.9295	.9572
130	.0304	.1434	.3299	.5249	.6847	.7992	.8751	.9235	.9535
140	.0232	.1232	.3024	.4989	.6644	.7849	.8658	.9175	.9498
150	.0177	.1059	.2772	.4741	.6447	.7710	.8565	.9116	.9461
160	.0135	.0910	.2541	.4506	.6256	.7573	.8473	.9057	.9424
170	.0103	.0782	.2329	.4282	.6070	.7438	.8383	.8999	.9388
180	.0078	.0672	.2134	.4070	.5890	.7306	.8293	.8941	.9351
190	.0060	.0577	.1956	.3868	.5715	.7176	.8204	.8883	.9315
200	.0046	.0496	.1793	.3676	.5546	.7048	.8116	.8826	.9278

Table 4. Probability of n control statistics ~ N(0, 1) to stay in control for K=1.0

n \ h	2.0	2.5	3.0	3.5	4.0	4.5	5.0
10	.9656	.9882	.9959	.9986	.9995	.9998	1.0000
20	.9289	.9745	.9909	.9968	.9989	.9996	.9999
30	.8937	.9610	.9859	.9949	.9982	.9994	.9998
40	.8597	.9477	.9809	.9931	.9975	.9991	.9997
50	.8271	.9346	.9760	9912	.9968	.9989	.9997
60	.7957	.9217	.9710	9894	.9962	.9987	.9996
70	.7654	.9089	.9661	9876	.9955	.9984	.9995
80	.7364	.8964	.9612	9857	.9948	.9982	.9994
90	.7084	.8840	.9564	9839	.9942	.9980	.9994
100	.6815	.8717	.9516	.9821	.9935	.9977	.9993
110	.6556	.8597	.9468	.9803	.9928	.9975	.9992
120	.6307	.8478	.9420	.9785	.9922	.9973	.9991
130	.6068	.8361	.9372	.9767	.9915	.9970	.9991
140	.5837	.8245	.9325	.9749	.9908	.9968	.9990
150	.5616	.8131	.9278	.9731	.9902	.9965	.9989
160	.5402	.8018	.9231	.9713	.9895	.9963	.9989
170	.5197	.7908	.9184	.9695	.9888	.9961	.9988
180	.5000	.7798	.9138	.9677	.9882	.9958	.9987
190	.4810	.7690	.9092	.9659	.9875	.9956	.9986
200	.4627	.7584	.9046	9641	.9868	.9954	.9986

Table 5. The result of regression analysis for k=0.5 and h=4.0

Variable	Estimate	Std. Error	t value
intercept	1.273050	2.7053658E-03	470.56
n	-.3011329	2.2583957E-05	-13333.93

Degress of freedom = 18
Residual mean square = 3.3917336E-5
Root mean square = 5.8238592E-3
R-squared = 1.0000

The following steps lead to the construction of the CUSUM chart that we suggest in this article.

STEP 1 : set k=△/2 where △ is the shift (multiple of standard deviation) to be detected
STEP 2 : set the significance level of the test
STEP 3 : find the probability which equals to one minus significance level of STEP 2 from the row of the sample size in the problem
STEP 4 : determine the h value which corresponds to the probability of STEP 3

For example, suppose that we have 50 control statistics which are assumed to have standard normal distribution. We wish to detect one standard deviation shift of the mean with $\alpha=0.05$. Since △ is one standard deviation, k=0.5. From Table 3, h=5 gives the significance level of 0.05 for n=50.
If we do a multiple test then we use Bonferroni's significance level, that is, α/m where m is the number of independent tests.

References

1. Brown, R. L., Durbin, J. and Evans, J. M. (1975), "Techniques for Testing the Constancy of Regression Relationships Over Time" (with discussion), Journal of the Royal Statistical Society, 37, 149-192.
2. Galpin, J. S. and Hawkins, D. M. (1984), "The Use of Recursive Residuals in Checking Model Fit in Linear Regression", The American Statistician, 38, 94-105.
3. Hawkins, D. M. (1991), "Diagnostics for Use With Regression Recursive Residuals", Technometrics, 33, 221-234.
4. Kang, C. W. (1990), Diagnostic Methods Using Recursive Analysis in Nonlinear Regression, Ph. D. Thesis, University of Minnesota.
5. Lucas, J. M. (1976), "The Design and Use of V-Mask Control Schemes", Journal of Quality Technology, 8, 1-11.