SAND2016-7395C

An Introduction to the Bernoulli CUSUM

Author 1

Key Words: Bernoulli CUSUM, Run Length

Summary and Conclusions

The Bernoulli CUSUM (BC) provides a moving window of process performance and is the quickest control chart to detect small increases in fraction defective. Bernoulli CUSUM designs have been developed that require 2, 3, or 4 failures in a moving window to produce a signal. The run length distribution provides insight into the properties of the BC beyond the Average or Median Run length. A retrospective analysis of electrical component pass/fail data using the BC suggested that a problem may have been present during prior production. The BC was implemented for ongoing production.

Introduction

In probability and statistics, a Bernoulli Process is a sequence of independent binary random variables $X_1, X_2, X_3, \ldots$, that take on the value 0 or 1. The random variable X_t takes on the value 1 with probability p and takes on the value 0 with probability $(1-p)$. In manufacturing, we can think of a sequence of manufactured parts as being assigned the value 1 if the part is defective, and the value 0 if the part functions properly. Then p represents the manufacturing fraction defective. It is of interest to monitor the fraction defective and provide timely feedback to the process engineers if the fraction defective is believed to have increased.

The Bernoulli Cumulative Sum (CUSUM) is a statistical process monitoring technique that is used to detect changes in the fraction defective p, from a nominal value p_0 to an unacceptable level p_1. It "cumulates" the number of defects that occur in a manufacturing window and provides an ongoing test of whether the fraction defective has increased. The Bernoulli CUSUM chart has appeared in the recent statistical process control literature, used primarily for high quality, high volume processes. Our challenge has been to modify and use the chart for high quality processes with somewhat lower volume.

In this paper, we will give an overview of the Bernoulli CUSUM (BC), discuss the properties of the BC by examination of the run length distribution, and make recommendations to the practitioner regarding the design of the BC. We will also present a case study of the Bernoulli CUSUM applied to a high reliability electrical component.

Several control charts are traditionally recommended for monitoring processes with pass/fail data. The p-chart is most frequently suggested for this problem. The p-chart monitors the fraction defective in successive samples, with a minimum recommendation of 25 to 50 parts per sample. Other control charts suggested for this problem include the Binomial CUSUM, applied to the number of failures per sample, and the Geometric CUSUM, applied to the number of good parts between failures.

A primary advantage of the Bernoulli CUSUM is that the BC statistic is calculated after each part is inspected. Because of this property, it has been shown to have the best statistical properties for detecting increases in fraction defective for high quality processes (Szarka, 2011). By "best statistical properties" it is meant that this type of control chart will detect increases in fraction defective more quickly than competing control charts.

The upper one-sided Bernoulli CUSUM statistics, B_t, t = 1, 2, . . ., are

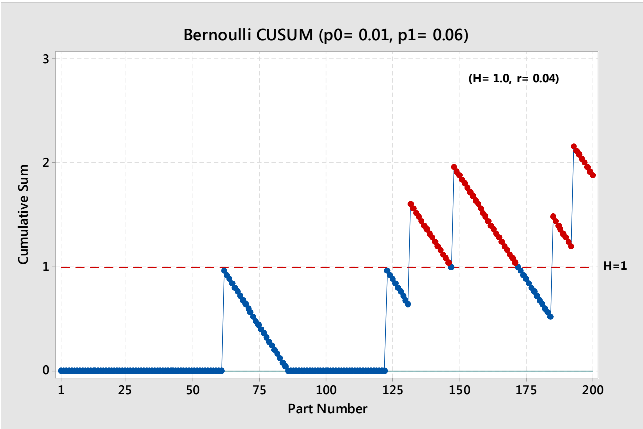
$$B_t = \max(0, B_{t-1} + X_t - r),$$

where $B_0 = 0$ and r is a small constant greater than zero but less than one. The X_t's represent the random Bernoulli sequence of 0's and 1's. An alarm is produced if $B_t \geq H$, a threshold value that is chosen, along with r, as part of the CUSUM design. The Bernoulli CUSUM is related to the Likelihood Ratio Test for testing a simple hypothesis of p_0 vs. p_1. This relationship is discussed in Reynolds and Stoumbos (1999).

The measure of performance often used to evaluate the Bernoulli CUSUM is the Average Run Length (ARL), the number of parts produced until the threshold H is exceeded. This provides a warning alarm that the process fraction defective may have increased. An investigation of the process would follow any such alarm.

Because the Run Length distribution is highly skewed, we will instead use the Median Run Length (MRL) as the primary measure of performance. This same measure is used to determine the best possible design for the CUSUM chart in terms of choice of H and r.

As an example of how the Bernoulli CUSUM works, see the control chart below.



This example uses simulated data with an initial defect rate of p_0= 0.01 for the first 100 observations followed by a defect rate of p_1= 0.06 for the second 100 observations. The control limit is at H= 1.0 and the reference value is r = 0.04. The Bernoulli CUSUM stays at zero until the first defect occurs at part number 62, where the CUSUM increases to the value (1-0.04) = 0.96. From that point forward, the CUSUM decreases by 0.04 for each part that passes until it reaches zero. With two failures at part numbers 123 and 132, the CUSUM signals because two failures have occurred in a relatively small window of parts. With H= 1.0 and r= 0.04, the CUSUM will signal whenever 2 failures occur within a window of 1/r= 25 parts.

Advantages of the Bernoulli CUSUM:

1. The method has been shown to detect increases in the process fraction defective faster than competing methods, measured by Median Run Length. It is used to answer the question: Has the fraction defective increased?
2. The method has the advantage of testing for an increase in fraction defective after each part is tested. There is no need to accumulate parts before testing for an increase.
3. The method provides a moving window of current process performance.
4. The method can be used for process data, product acceptance data, and shelf life data. The ordering of the individual data values must of course be meaningful.
5. The method is relatively easy to explain and implement, and can be plotted by standard statistical packages such as Minitab.

The advantages listed above have led to the development of the Bernoulli CUSUM for the monitoring of the production of a high reliability, high consequence electrical component that will be discussed below. The desire is that it provide an early indication of an increase in fraction defective during product acceptance testing.

Design of a Bernoulli CUSUM and Run Length Distribution

The recommended design of the Bernoulli CUSUM consists of the following steps:

1. Choose the control limit H and the reference value r to set the Median Run Length (MRL) at a desirable level when the fraction defective is at nominal. This corresponds to setting the "false alarm" rate. Choices of (H, r) can be explored via simulation or using tables of ARLs and MRLs.
2. For the choices of (H, r) from Step 1, evaluate the MRL for values of the fraction defective that are greater than nominal. This corresponds to evaluating the "time to detection" of an unacceptable fraction defective.
3. Iterate on the choice of (H, r) if necessary.

Tables of ARLs have been constructed for values of H and r such that $1.0 \leq H \leq 3.0$ and $0.01 \leq r \leq 0.04$. These tables provide a starting point for choosing H and r. Percentiles of the Run Length Distribution, obtained through simulation, are used for a more detailed analysis of the BC performance, and to make probability statements about possible outcomes.

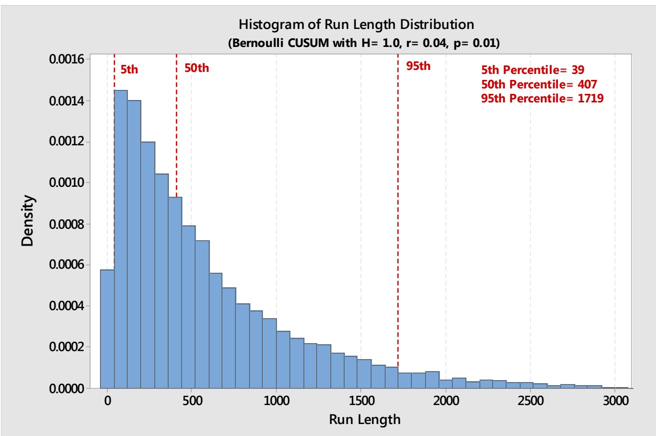
Example Table of Average Run Lengths for various (H, r) combinations

p	H=1.0 r=0.01	H=1.2 r=0.01	H=1.4 r=0.01	H=1.0 r=0.04	H=1.2 r=0.04	H=1.4 r=0.04
0.01	260	285	310	583	720	885
0.02	108	113	114	185	215	263
0.03	68	70	76	99	108	139
0.04	51	51	51	66	70	81
0.05	40	40	43	49	53	57
0.06	33	33	34	38	40	46
0.07	29	29	30	32	33	37
0.08	25	25	25	27	29	31
0.09	22	22	22	24	25	27
0.10	20	20	20	21	22	24

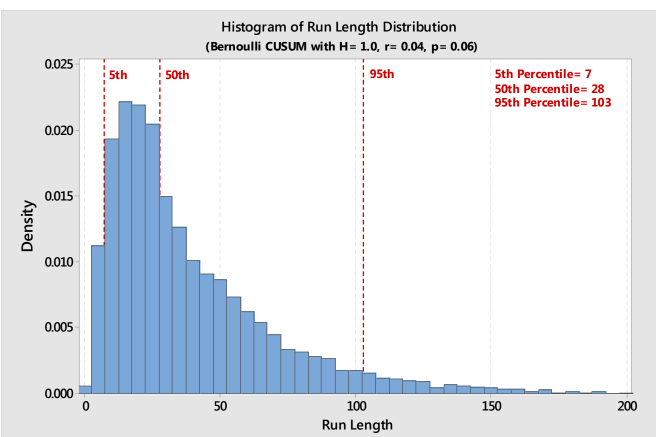
This table shows that the ARLs increase as H increases (with r fixed), and as r increases (with H fixed). The MRLs for various choices of (H, r) can also be determined from the same simulations. Because the run length distribution is skewed, the Median Run Length is a better estimate of central tendency. Percentiles of the run length distribution can also be used to determine probability bounds on run lengths.

For the special case (H, r)= (1.0, 0.04) the run length distributions with p= 0.01 and p= 0.06 appear below.

Run Length Distribution with p= 0.01



And Run Length Distribution with p= 0.06



These histograms, each based on 10,000 simulations each, show how skewed the run length distributions are for p= 0.01 and p= 0.06. The median value provides the best estimate of the central tendency, and the 5th and 95th percentiles provide a 90% probability interval for the run length outcome.

Bernoulli CUSUM for an Electrical Component

It is desired to monitor the production of an electrical component that is a high reliability, high consequence, expensive component. The component requires some operator assembly, so a nominal value of p_0= 0.005 is considered the lowest reasonably attainable fraction defective. One hundred percent inspection is performed and the most common failure mode is high voltage breakdown (HVB). A single component is also very expensive, so a timely feedback regarding any process problem is critical. False alarms are also costly, so a median run length when p= 0.005 is desired to be at least 8000.

Because of the skewness of the run length distributions, the proposed Bernoulli CUSUM design strategy uses the Median Run Length (MRL) in the following way:

Subject to MRL ≥ 8000 when p_0= 0.005,

Investigate CUSUM performance when p_1= 0.01, 0.02, 0.03, 0.04, and 0.05.

Choose the best overall combination of (H, r).

The value p_0 is the greatest allowable fraction defective. When the process is operating at this level or better, it is desirable to have a large MRL, to minimize false alarms. The value p_1 is the fraction defective that is unacceptable and must be detected quickly. When the process is operating at this level or worse, it is desirable to have a small MRL. Various combinations of (H, r) that produce an MRL of approximately 8000 when p_0= 0.005 were found via simulation techniques. These combinations appear in the table below.

Table 1. Combinations of (H, r) that Produce an MRL of Approximately 8000.

p	H=2.0 r= 0.024	H=2.2 r= 0.020	H=2.4 r= 0.017	H=2.6 r= 0.014	H=2.8 r= 0.012	H=3.0 r= 0.0105
0.005	8000	8000	8000	8000	8000	8000
0.01	1256	1274	1153	1013	971	881
0.02	258	255	244	232	233	231
0.03	123	123	123	123	130	134
0.04	78	80	83	85	89	93
0.05	58	58	62	66	71	74

From this table we can see that the MRLs vary for each combination of (H, r) across the various values of fraction defective p. The larger H value has faster detection for p= 0.01 and p= 0.02, but slightly slower detection for p=0.04 and p=0.05. Since we are interested in fast detection at p= 0.01 and p= 0.02, the recommended choice is to use (H, r)= (3.0, 0.0105) for the Bernoulli CUSUM. More percentiles of the associated Run Length distribution appear in the table below.

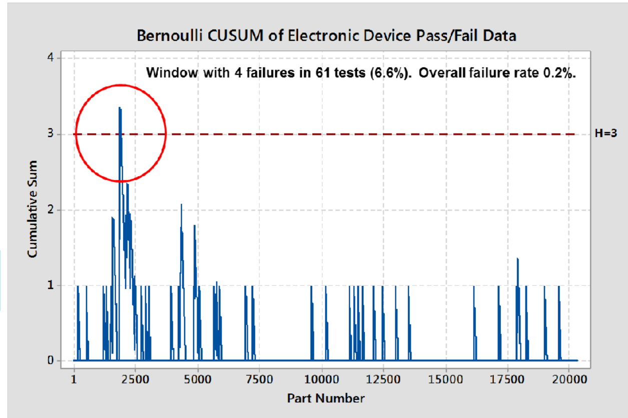
Table 2. Run Length Distribution of Bernoulli CUSUM with (H, r)= (3.0, 0.0105)

p	ARL	5th	25th	50th	75th	95th
0.005	11500	650	3350	8000	15500	34000
0.01	1220	164	454	881	1631	3461
0.02	285.8	69	143	231	373	687
0.03	155.9	47	85	134	200	342
0.04	108.2	35	64	93	140	227
0.05	82.9	29	52	74	103	176

This table gives the 5th, 25th, 50th (Median), 75th, and 95th percentiles for the Run Length distribution of the Bernoulli CUSUM using (H, r)= (3.0, 0.0105). Looking at the row with fraction defective p= 0.01, we see that the Median Run Length (50th percentile) is 881. The 5th percentile is 164 and the 95th percentile is 3461. These values provide a "best case" and "worst case" number of parts that will be needed to detect an increase in fraction defective to p= 0.01. To lower these numbers, the MRL when p= 0.005 would also have to be lowered, resulting in an increased false alarm rate. The choice of (H, r)= (3.0, 0.0105) is an attempt to balance the desire to quickly detect an increase in fraction defective with the desire to keep the false alarm rate very low.

This design was used in a retrospective analysis of electrical component pass/fail data.

Bernoulli CUSUM of Electronic Component Pass/Fail Data



The CUSUM analysis suggested a process problem occurred around test number 1800. The Bernoulli CUSUM was implemented to monitor ongoing production.

References

Bourke, Patrick (1991). Detecting a Shift in Fraction Nonconforming Using Run-Length Control Charts With 100% Inspection. *Journal of Quality Technology*, Vol. 23, No. 3, pp 225-238.

Bourke, Patrick (2001). Sample Size and the Binomial CUSUM Control Chart: The Case of 100% Inspection. *Metrika*, 53: 51-70.

Lucas, James (1985). Counted Data CUSUM's. *Technometrics*, Vol. 27, No. 2, pp. 129-144.

Lucas, James (1989). Control Schemes for Low Count Levels. *Journal of Quality Technology*, Vol. 21, No. 2, pp 199-201.

Reynolds, Marion and Zachary Stoumbos (1999). A CUSUM Chart for Monitoring a Proportion When Inspecting Continuously. *Journal of Quality Technology*, Vol. 31, No. 1, pp 87-108.

Sego, Landon (2006). Applications of Control Charts in Medicine and Epidemiology. Unpublished Dissertation, Virginia Polytechnic Institute and State University. (Available Online free)

Szarka, John (2011). Surveillance of Negative Binomial and Bernoulli Processes. Unpublished Dissertation, Virginia Polytechnic Institute and State University. (Available Online free)

Szarka, John and William Woodall (2011). A Review and Perspective on Surveillance of Bernoulli Processes. *Quality and Reliability Engineering International*, Vol. 27, No. 6, pp 735-752.

Biographies

Appendix- Tables of ARLs for the Bernoulli CUSUM

H= 1.0 to 3.0 in increments of 0.20

r= 0.01, 0.02, 0.03, and 0.04

p= 0.01, 0.02, …, 0.09, 0.10

p	H=1.0 r=0.02	H=1.2 r=0.02	H=1.4 r=0.02	H=1.6 r=0.02	H=1.8 r=0.02
0.01	362	415	500	620	850
0.02	131	145	159	185	220
0.03	77	81	90	100	118
0.04	54	54	61	67	76
0.05	42	45	45	50	58
0.06	34	36	36	40	45
0.07	29	29	31	33	38
0.08	25	25	26	27	33
0.09	22	23	24	24	28
0.10	20	20	21	21	24

p	H=1.0 r=0.01	H=1.2 r=0.01	H=1.4 r=0.01	H=1.6 r=0.01	H=1.8 r=0.01
0.01	260	285	310	368	446
0.02	108	113	114	135	153
0.03	68	70	76	79	90
0.04	51	51	51	55	63
0.05	40	40	43	43	49
0.06	33	33	34	35	39
0.07	29	29	30	29	32
0.08	25	25	25	25	29
0.09	22	22	22	23	24
0.10	20	20	20	20	21

p	H=1.0 r=0.03	H=1.2 r=0.03	H=1.4 r=0.03	H=1.6 r=0.03	H=1.8 r=0.03
0.01	463	540	680	890	1360
0.02	155	170	210	248	330
0.03	87	93	110	128	157
0.04	59	67	70	78	93
0.05	45	48	52	58	70
0.06	36	38	41	45	51
0.07	30	32	35	36	43
0.08	26	27	29	31	36
0.09	23	24	24	27	31
0.10	20	21	22	24	27

p	H=1.0 r=0.04	H=1.2 r=0.04	H=1.4 r=0.04	H=1.6 r=0.04	H=1.8 r=0.04
0.01	586	720	885	1235	2190
0.02	185	215	263	330	480
0.03	99	108	139	162	210
0.04	66	70	81	98	119
0.05	49	53	57	64	81
0.06	39	40	46	54	63
0.07	32	33	37	40	48
0.08	27	29	31	34	42
0.09	24	25	27	31	35
0.10	21	22	24	25	30

p	H=2.0 r=0.02	H=2.2 r=0.02	H=2.4 r=0.02	H=2.6 r=0.02	H=2.8 r=0.02
0.01	1355	1730	2150	3015	4100
0.02	300	325	400	440	518
0.03	150	155	167	197	217
0.04	93	97	109	119	126
0.05	67	70	77	86	93
0.06	55	58	61	68	74
0.07	44	48	50	55	59
0.08	38	40	41	45	49
0.09	35	35	36	40	42
0.10	30	31	32	35	39

p	H=1.0 r=0.04	H=1.2 r=0.04	H=1.4 r=0.04	H=1.6 r=0.04	H=1.8 r=0.04
0.01	586	720	885	1235	2190
0.02	185	215	263	330	480
0.03	99	108	139	162	210
0.04	66	70	81	98	119
0.05	49	53	57	64	81
0.06	39	40	46	54	63
0.07	32	33	37	40	48
0.08	27	29	31	34	42
0.09	24	25	27	31	35
0.10	21	22	24	25	30