

# Guide to data analysis, quality control and improvement using cusum techniques —

## Part 4: Cusum methods for discrete (count/classified) data

ICS 03.120.30; 07.020

# Committees responsible for this British Standard

The preparation of this British Standard was entrusted to Technical Committee SS/4, Statistical process control, upon which the following bodies were represented:

- Association for Road Traffic Safety and Management (ARTSM)
- BAE Systems
- British Standards Society (BSS)
- Clay Pipe Development Association (CPDA)
- Federation of Small Businesses (FSB)
- Gauge and Tool Makers Association (GTMA)
- General Domestic Appliances Ltd.
- Institute of Quality Assurance
- National Physical Laboratory
- Royal Statistical Society

This British Standard was published under the authority of the Standards Policy and Strategy Committee on 17 March 2003

© BSI 17 March 2003

First published October 1982

The following BSI references relate to the work on this British Standard:  
Committee reference SS/4  
Draft for comment 02/400136 DC

ISBN 0 580 41000 5

## Amendments issued since publication

| Date | Amd. No. | Comments |
|------|----------|----------|
|      |          |          |
|      |          |          |
|      |          |          |
|      |          |          |
|      |          |          |

# Contents

|                                                                                                                                     | Page               |
|-------------------------------------------------------------------------------------------------------------------------------------|--------------------|
| Committees responsible                                                                                                              | Inside front cover |
| Foreword                                                                                                                            | iii                |
| Introduction                                                                                                                        | 1                  |
| 1 Scope                                                                                                                             | 1                  |
| 2 Normative references                                                                                                              | 1                  |
| 3 Terms, definitions, abbreviations and symbols                                                                                     | 1                  |
| 4 Discrete data: cusum basics                                                                                                       | 3                  |
| 5 Cusum decision schemes for discrete data                                                                                          | 9                  |
| 6 Cusum decision schemes for count data                                                                                             | 13                 |
| 7 Cusum decision schemes for classified data                                                                                        | 23                 |
| 8 Design of cusum schemes for low rates of occurrence of events                                                                     | 29                 |
| Annex A (normative) Tests for validity of Poisson or binomial distribution assumptions                                              | 35                 |
| Annex B (normative) Cusum design schemes for count (Poisson) data                                                                   | 38                 |
| Annex C (informative) Tabular cusum design schemes for binomial data                                                                | 43                 |
| Bibliography                                                                                                                        | 51                 |
| Figure 1 — Basic cusum plot of spot weld data                                                                                       | 5                  |
| Figure 2 — Standard shewhart $np$ control chart for spot weld data                                                                  | 6                  |
| Figure 3 — Cusum plot of magazine printing errors by issue number                                                                   | 8                  |
| Figure 4 — Standard Shewhart $c$ chart for magazine printing errors                                                                 | 9                  |
| Figure 5 — Outline features of one-sided V-mask for discrete data                                                                   | 12                 |
| Figure 6 — Cusum chart with V-mask for stock-outs ( $H = 8$ ; $K = 6$ ; target = $T = 4$ ; $F = K - T = 2$ )                        | 16                 |
| Figure 7 — Cusum plot of fabric data with “up” and “down” truncated V-masks                                                         | 22                 |
| Figure 8 — Decision interval plot of fabric data with upper and lower control limits                                                | 22                 |
| Figure 9 — Cusum plot of spot weld data with V-mask having $H = 3$ ; $F = 1$                                                        | 24                 |
| Figure 10 — Audited product cusum plot with asymmetrical truncated V-mask for up and down movement                                  | 28                 |
| Figure 11 — Cusum plot of audited product with upper and lower control limits                                                       | 29                 |
| Table 1 — Nonconforming spot weld production figures and cusum table                                                                | 5                  |
| Table 2 — Printing errors and basic cusum tabulation for magazine issues                                                            | 8                  |
| Table 3 — Table for cusum stock-out plot with V-mask                                                                                | 15                 |
| Table 4 — Cusum DI (decision interval) tabulation: ( $H = 8$ ; $K = 6$ )                                                            | 17                 |
| Table 5 — Warp knitting plant data: cusum tabulation                                                                                | 21                 |
| Table 6 — Decision interval (DI) tabulation of spot weld data                                                                       | 24                 |
| Table 7 — ARL performance for a binomial cusum scheme with a sample size = 50 and target value, $p_a = 0.2$ ( $H = 12$ ; $K = 12$ ) | 25                 |
| Table 8 — Audited product cusum                                                                                                     | 27                 |
| Table 9 — ARLs for 1 % and 4 % proportions in terms of $K$ for $H = 1$                                                              | 30                 |
| Table 10 — ARL performance of two $H = 1$ cusum schemes in terms of failure                                                         | 30                 |
| Table 11 — Cusum schemes in terms of selected ARLs and in- and out-of-control probabilities                                         | 31                 |
| Table 12 — ARL performance of two $K = 0.02$ schemes in terms of failure rate                                                       | 31                 |
| Table A.1 — Critical values of variance ratio for testing binomial or Poisson distribution assumptions                              | 36                 |

|                                                                                                    | Page |
|----------------------------------------------------------------------------------------------------|------|
| Table B.1 — C1 and C2 Schemes for count (Poisson) data in terms of $m$ , $H$ and $K$               | 38   |
| Table B.2 — ARL characteristics for cusum schemes in terms of $H$ and $K$ for count (Poisson) data | 39   |
| Table B.3 — ARLs for low rate schemes with $H = 1$ for count (Poisson) data                        | 40   |
| Table B.4 — ARLs for schemes with $K = 1$ : for count (Poisson) data                               | 41   |
| Table B.5 — ARL data for schemes with $K = 0.25$ : for count (Poisson) data                        | 42   |
| Table C.1 — C1 and C2 schemes for binomial data                                                    | 43   |
| Table C.2 — ARL data for binomial variables design s                                               | 44   |
| Table C.2a) — C1 design schemes with $p_a = 0.1$                                                   | 44   |
| Table C.2b) — C2 design schemes with $p_a = 0.1$                                                   | 45   |
| Table C.2c) — C1 design schemes with $p_a = 0.2$                                                   | 45   |
| Table C.2d) — C2 design schemes with $p_a = 0.2$                                                   | 46   |
| Table C.2e) — C1 design schemes with $p_a = 0.3$                                                   | 46   |
| Table C.2f) — C2 design schemes with $p_a = 0.3$                                                   | 47   |
| Table C.2g) — C1 design schemes with $p_a = 0.4$                                                   | 47   |
| Table C.2h) — C2 design schemes with $p_a = 0.4$                                                   | 48   |
| Table C.2i) — C1 design schemes with $p_a = 0.5$                                                   | 48   |
| Table C.2j) — C2 design schemes with $p_a = 0.5$                                                   | 49   |
| Table C.3 — ARLs for low rate schemes with $H = 1$ for binomial events                             | 49   |

## Foreword

This part of BS 5703 was prepared by Technical Committee SS/4. It supersedes BS 5703-4:1982, which is withdrawn.

It was deemed necessary to revise BS 5703 for the following reasons:

- to present the text from the ordinary users view point rather than that of the statistician;
- to incorporate modern cusum computer software routines;
- to reduce the quantity of text and eliminate repetition;
- to improve the diagrams;
- to bring statistical symbols into line with standard practice.

BS 5703 is published in four parts:

- *Part 1: Uses and value of cusum charts in business, industry, commerce and public service;*
- *Part 2: Introduction to decision-making using cusum techniques;*
- *Part 3: Cusum methods for process/quality measured data;*
- *Part 4: Cusum control using for discrete (count/attribute) data methods.*

This publication does not purport to include all the necessary provisions of a contract. Users are responsible for its correct application.

**Compliance with a British Standard does not of itself confer immunity from legal obligations.**

### Summary of pages

This document comprises a front cover, an inside front cover, pages i to iv pages 1 to 51 and a back cover.

The BSI copyright notice displayed in this document indicates when the document was last issued.



## Introduction

BS 5703-1 demonstrates the versatility and usefulness of the very simple, yet powerful, pictorial cusum approach to interpreting data arranged in any meaningful sequence. BS 5703-2 provides a general introduction to the quantification of decision-making in the design and application of cusum schemes for normally distributed measured data. BS 5703-3 deals with the application of cusum methods to measured data.

This part of BS 5703 concerns itself with the application of cusum methods to discrete data.

It represents a revision of BS 5703-4:1982, which it replaces. The cusum designs in the 1982 version of BS 5703-4 were constrained in two senses. They were limited to tabulations of:

- a) two procedures, termed C1 and C2, having average run lengths (ARLs) at the target level of at least 1 000 and 200 respectively;
- b) schemes to detect shifts in the upward direction only.

These restrictions were introduced to avoid the complications arising from the use of the Poisson distribution for count data and binomial distribution for binomial (yes/no) data.

While retaining the C1 and C2 tabular approach for those who prefer manual access to design specifications this 2003 version of BS 5703-4 is extended in respect of both a) and b). It provides accessibility to more comprehensive cusum schemes than just the C1 and C2 ones. It also includes the use of schemes to detect drift in either direction.

This has been made possible by advances in computer-based software that is now freely available on the world-wide-web. These features considerably enhance the use of the standard to provide optimal cusum designs for specific applications.

## 1 Scope

BS 5703-4 describes cusum methods applicable to binomial (yes/no) data and to counts of attributes and events arranged in some meaningful sequence.

## 2 Normative references

The following referenced documents are indispensable for the application of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

BS ISO 3534-1, *Statistics — Vocabulary and symbols — Part 1: Probability and general statistical terms*.

BS ISO 3534-2, *Statistics — Vocabulary and symbols — Part 2: Statistical quality control*.

BS EN ISO 9000, *Quality management systems — Fundamentals and vocabulary*.

## 3 Terms, definitions, abbreviations and symbols

### 3.1 General

For the purposes of this part of BS 5703, the terms, definitions, abbreviations and symbols of BS ISO 3534-1, BS ISO 3534-2, BS EN ISO 9000:2000, Clause 3, and those in 3.2 apply.

### 3.2 Additional terms, definitions, abbreviations and symbols

#### 3.2.1

##### target value

*T*

value for which departure in average level requires to be detected

NOTE 1 With a charted cusum, deviations from the reference value are cumulated.

NOTE 2 The target value is often referred to as the reference value or the nominal control value. If so it should be acknowledged that it is not necessarily the most desirable or preferred value. It is simply a convenient target value for constructing a cusum chart.

### 3.2.2

#### datum value

**$K, k$**

value chosen to provide an optimal response to a shift of a specified size in the mean

NOTE 1 For normal data the theoretical datum value is one half the shift it is required to detect. Values of the theoretical datum value are often rounded off to the nearest integer, or half integer, for ease of calculations both in establishing ARL performance and in subsequent cusum application.

NOTE 2 With a tabulated cusum, deviations from the datum value are cumulated whereas with a conventional cusum graph deviations from the target value,  $T$ , are cumulated.

NOTE 3 It is necessary to distinguish between  $k$  for a standardized variable and  $K$  for an observed variable.  $K = k\sigma_e$ .

NOTE 4 Upper datum value =  $K_U = T + F$ . Lower datum value =  $K_L = T - F$ .

### 3.2.3

#### reference shift

**$F, f$**

<tabulated cusum> difference between the target value (3.2.1) and datum value (3.2.2)

NOTE It is necessary to distinguish between  $f$  that relates to a standardized variable and  $F$  to an observed variable.  $F = f\sigma_e$ .

### 3.2.4

#### reference shift

**$F, f$**

<truncated V-mask> slope of the arm of the mask (tangent of the mask angle)

NOTE It is necessary to distinguish between  $f$  that relates to a standardized variable and  $F$  to an observed variable.  $F = f\sigma_e$ .

### 3.2.5

#### decision interval

**$H, h$**

<tabulated cusum> cumulative sum of deviations from a datum value (3.2.2) required to yield a signal

NOTE It is necessary to distinguish between  $h$  that relates to a standardized variable and  $H$  to an observed variable.  $H = h\sigma_e$ .

### 3.2.6

#### decision interval

**$H, h$**

<truncated V-mask> half height at the datum of the mask

NOTE It is necessary to distinguish between  $h$  that relates to a standardized variable and  $H$  to an observed variable.  $H = h\sigma_e$ .

### 3.2.7

#### average run length

**ARL**

average number of samples taken up to the point at which a signal occurs for a particular process level

NOTE ARL is normally related to a particular process level in which case it carries an appropriate subscript, as, for example,  $ARL_{in-control}$  and  $ARL_{out-of-control}$ ,  $ARL_{0.1}$  and  $ARL_{0.3}$ ,  $ARL_{acceptable}$  and  $ARL_{rejectable}$ .

## 3.3 Symbols

- $m$  parameter (rate of occurrence, per unit of observations, of an event of interest) of a Poisson distribution;
- $n$  the number of individual items comprising a sample;
- $p$  parameter (proportion of items having an attribute of interest) of a binomial distribution;
- $\sigma$  population parameter representing the standard deviation of a population;
- $\sigma_e$  standard deviation, or standard error, of the variable of interest;
- $S$  sample statistic, representing the sample estimate of the standard deviation;
- $s$  realized, or measured, value of the sample statistic,  $S$ ;
- $\mu$  population parameter representing the arithmetic mean of the population;

- $X$  random variable;  
 $x$  realized value of the random variable  $X$ ;  
 $\bar{X}$  sample statistic, representing the sample estimate of the arithmetic mean;  
 $\bar{x}$  realized, or measured, value of the sample statistic,  $\bar{X}$ .

## 4 Discrete data: cusum basics

### 4.1 Data classification

Two general classes of data are distinguished here: *continuous data* and *discrete data*. These classes are also known as, respectively, measured data and attribute data, or quantitative data and qualitative data. Whereas the former is measured on a numerical scale with a continuum of possible values, the latter is referenced on a scale with only a set or sequence of distinct values.

The cusum aspects of continuous data are discussed in detail in BS 5703-3. In this part of BS 5703 the focus is on the application of cusum techniques to discrete data. Discrete data is divided, for convenience, into two categories: *classified data* and *countable data*.

In conventional statistical process control, classified and countable data are handled using:

- " $p$ " charts for proportion of items, or units, having a particular attribute (e.g. nonconforming), particularly when the sample size is variable;
- " $np$ " charts for numbers of items, or units, having a particular attribute (e.g. nonconforming), from samples of constant size;
- " $u$ " charts for number of events (e.g. nonconformities) per sample, particularly when the sample size is variable;
- " $c$ " charts for number of events (e.g. nonconformities) for samples of constant size.

Cusum charts are used in a more general sense than just dealing with nonconforming items and nonconformities. They also focus more on persistent rather than transient changes. For this reason the role of cusum is considerable in quality and business performance improvement, where one is concerned primarily with changing the level of discrete data, e.g. improving yields, decreasing faults, increasing output, growing sales leads, reducing lead times and enhancing profits.

### 4.2 Classified data

#### 4.2.1 Data with two classes: binomial data

With classified data each item of data is classified as belonging to a number of categories. Frequently the number of categories is two, namely a binomial situation where, for instance, the outcome is usually expressed as 0 and 1, or as ok/nok, success/failure, profit/loss, in/out, or presence/absence of a particular characteristic. BS 5703-1, Clause 7 shows a practical example of how to construct and interpret a basic cusum chart for go/no-go test results on a particular attribute of a piston, the success or failure of a bond between the piston and a ring insert.

Data having two classes is termed "binomial" (binomial = "two names") data. A measure can be inherently binomial, e.g. was a profit or loss made, is someone in or out? Sometimes it is arrived at indirectly by categorizing some other numerical measure. Take, for instance, the case where telephone calls are classified on whether they last more than 10 min or, perhaps, whether they are answered within 6 rings.

The binomial distribution is used to obtain a standard error for cusum scaling and decision-making purposes. The binomial distribution is a statistical distribution giving the probability of obtaining a specified number of successes in a specified number of independent trials with a constant probability of success in each, where " $n$ " is the number of trials and " $p$ " the probability of success in each.

The conditions that should be satisfied for a binomial distribution are the following.

- There is a fixed number of trials,  $n$ .
- Only two possible outcomes are possible at each trial.
- The trials are independent.

d) There is a constant probability of success,  $p$ , in each trial.

e) The variable is the total number of successes in  $n$  trials.

The mean and standard deviation of a binomial distribution are given by:

$$\text{mean} = np; \quad (1)$$

$$\text{standard deviation} = \sqrt{np(1-p)}. \quad (2)$$

The binomial distribution is very cumbersome to calculate. Later, when dealing with ARLs and decision criteria, other distributions may be used as approximations to the binomial.

In standard statistical process control, binomial data is handled using:

- 1) “ $p$ ” charts for proportions, particularly when the sample size is variable; and
- 2) “ $np$ ” charts for numbers from samples of constant size.

#### 4.2.2 Method for setting up a basic cusum chart for classified data

The use of the binomial distribution is used here in the sense of scaling the vertical axis of the cusum chart. Later it will also be used in establishing the decision rules.

An estimate of the standard deviation [see equation (2)] will provide the basis for scaling the cusum chart thus:

**Step 1:** For a manual set-up, decide on a convenient interval for the horizontal scale then choose the vertical (cusum) scale such that  $2\sigma_e$  on that scale is roughly equal to 1 unit on the horizontal scale. This will ensure that a shift in the mean of  $2\sigma_e$  will give an angle of  $45^\circ$  to the horizontal. This enhances visual interpretation. Alternatively, if some specific shift is to be detected, the method of BS 5703-1, Clause 5, Step 5, may be used.

For a computerized set-up, the scales and aspect ratio of the chart can usually be readily adjusted for best effect.

**Step 2:** Decide on a target or reference value. If the chart is being used for retrospective analysis then the overall mean of the complete set of data will give the best discrimination of change in the mean.

**Step 3:** Calculate and plot the cusum using:

- a) for data consisting of individual test results, where,  $p$  = proportion of observations scoring 1 (an example of this is given in BS 5703-1, Clause 7):

$$\text{cusum} = \text{sum (observation value} - p) \quad (3)$$

- b) for data comprising samples of size,  $n$ , with proportion,  $p$ , of the attribute of interest (for example a nonconforming unit):

$$\text{cusum} = \text{sum (observation value} - np) \quad (4)$$

#### 4.2.3 Case study

**Application:** Housings are 100 % checked in production sequence for spot weld quality.

There are 20 spot welds per housing and the welds are classified ok or nok. In a production sequence of 40 housings the number of nonconforming welds per housing is as shown in Table 1.

Plot a cusum chart and compare it with a standard Shewhart control chart.

here:

sample size = the number of spot welds per housing = 20;

total nok welds = 20;

total welds inspected =  $40 \times 20 = 800$ ;

$p$  = overall proportion of welds faulty =  $20/800 = 0.025$ ;

$np$  = mean number of nok items per sample of 20 =  $0.025 \times 20 = 0.5$ ;

standard deviation =  $\sqrt{np(1-p)} = 0.7$ .

Table 1 — Nonconforming spot weld production figures and cusum table

| Housing no. | No. of nok welds<br>( $x$ ) | $x - np$ | Cusum of<br>( $x - np$ ) | Housing no. | No. of nok welds<br>( $x$ ) | $x - np$ | Cusum of<br>( $x - np$ ) |
|-------------|-----------------------------|----------|--------------------------|-------------|-----------------------------|----------|--------------------------|
| 1           | 2                           | 1.5      | 1.5                      | 21          | 0                           | -0.5     | 2.5                      |
| 2           | 0                           | -0.5     | 1.0                      | 22          | 1                           | 0.5      | 3.0                      |
| 3           | 0                           | -0.5     | 0.5                      | 23          | 0                           | -0.5     | 2.5                      |
| 4           | 1                           | 0.5      | 1.0                      | 24          | 1                           | 0.5      | 3.0                      |
| 5           | 1                           | 0.5      | 1.5                      | 25          | 0                           | -0.5     | 2.5                      |
| 6           | 1                           | 0.5      | 2.0                      | 26          | 1                           | 0.5      | 3                        |
| 7           | 0                           | -0.5     | 1.5                      | 27          | 0                           | -0.5     | 2.5                      |
| 8           | 1                           | 0.5      | 2.0                      | 28          | 0                           | -0.5     | 2.0                      |
| 9           | 0                           | -0.5     | 1.5                      | 29          | 0                           | -0.5     | 1.5                      |
| 10          | 1                           | 0.5      | 2.0                      | 30          | 1                           | 0.5      | 2.0                      |
| 11          | 1                           | 0.5      | 2.5                      | 31          | 0                           | -0.5     | 1.5                      |
| 12          | 0                           | -0.5     | 2.0                      | 32          | 0                           | -0.5     | 1.0                      |
| 13          | 1                           | 0.5      | 2.5                      | 33          | 1                           | 0.5      | 1.5                      |
| 14          | 0                           | -0.5     | 2.0                      | 34          | 0                           | -0.5     | 1.0                      |
| 15          | 0                           | -0.5     | 1.5                      | 35          | 0                           | -0.5     | 0.5                      |
| 16          | 0                           | -0.5     | 1.0                      | 36          | 1                           | 0.5      | 1.0                      |
| 17          | 0                           | -0.5     | 0.5                      | 37          | 1                           | 0.5      | 1.5                      |
| 18          | 3                           | 2.5      | 3.0                      | 38          | 0                           | -0.5     | 1.0                      |
| 19          | 1                           | 0.5      | 3.5                      | 39          | 0                           | -0.5     | 0.5                      |
| 20          | 0                           | -0.5     | 3.0                      | 40          | 0                           | -0.5     | 0                        |

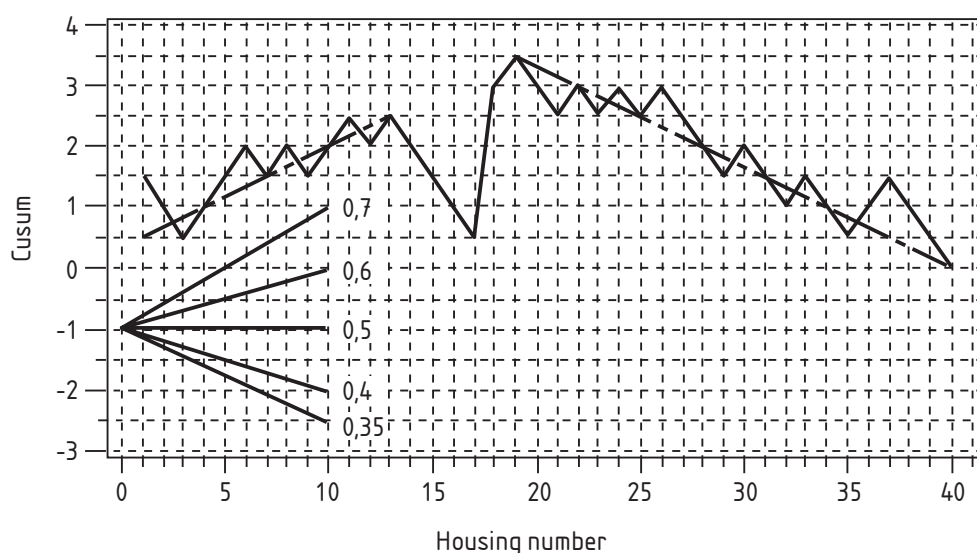


Figure 1 — Basic cusum plot of spot weld data

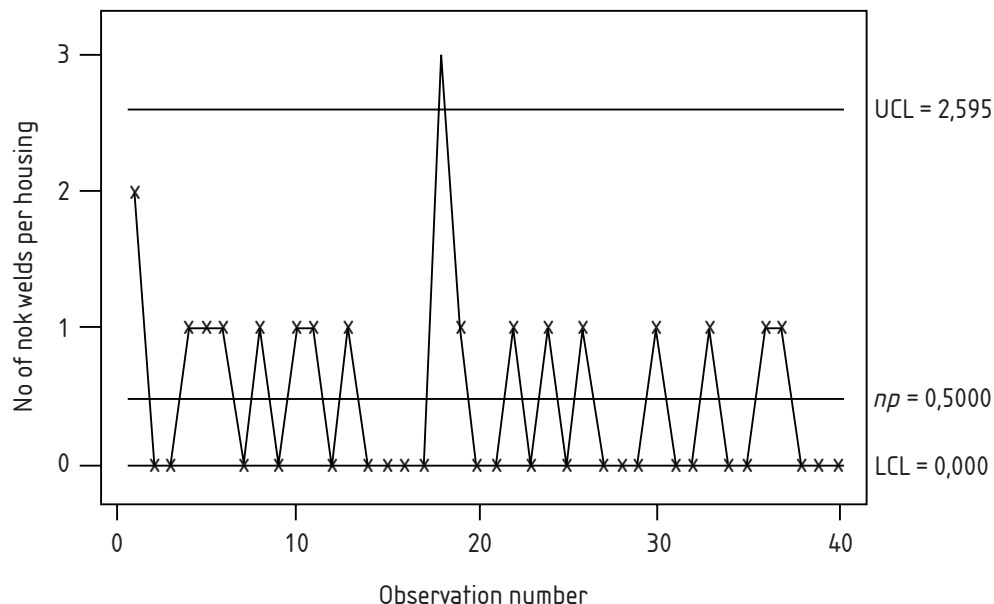


Figure 2 — Standard shewhart  $np$  control chart for spot weld data

## Conclusions

a) The cusum chart of Figure 1 indicates a mean number of 0.65 nonconforming items per 20 from sample 1 to sample 13. A period of 5 samples followed with erratic performance due to a transient problem with the progressive up-keep of inspection/test records. From sample 19 on the performance improved to some 0.35 nonconforming items per 20. This confirmed the success of a continual improvement programme being applied to the housing.

b) A comparison of Figure 1 and Figure 2 indicates that the cusum chart is far superior to the standard Shewhart chart in detecting sustained changes in the mean. The standard Shewhart chart of Figure 2 is, however, better at signalling transient changes in the mean.

## 4.3 Countable data

### 4.3.1 Event count: Poisson data

Countable data relates to counts of events where each item of data is the count of the number of particular events per given time period or quantity of product. Instances are: number of accidents or absentees per month, number of operations or sorties per day, number of incoming telephone calls per minute, or number of nonconformities per unit or batch. BS 5703-1, Clause 8 illustrates the application of basic cusum techniques to countable data, in the form of injuries per month, in an engineering organization.

The Poisson distribution has two principal parts to play in cusum analysis:

a) as an approximation to the more cumbersome binomial when:

$n$  is large and  $p$  is small, say,  $n > 20$  and  $p \leq 0.1$  (i.e.  $\leq 10\%$ );

b) as a distribution in its own right, when events occur randomly in time or space and the observation is made of the number of events in a given interval.

The validity of the Poisson model hinges on the independence of events and their occurrence at an average rate that is assumed to be stable (in the absence of special causes).

Two fundamental cusum related features of the Poisson distribution are: *one*, the connection between its standard deviation and mean, and: *two*, the results of summation of two independent Poisson variables.

Regarding point *one*: the mean and standard deviation of a Poisson distribution are given by:

$$\text{mean} = m; \quad (5)$$

$$\text{standard deviation} = \sigma_e = \sqrt{m}. \quad (6)$$

This relationship is useful for establishing the validity of a Poisson model for a particular set of data. While many real life situations might suggest that Poisson is a reasonable underlying model it would be unwise to take this for granted. Sometimes a degree of clustering of events, e.g. accidents and faults, is present. This has the effect of increasing the spread beyond that of the Poisson distribution. This wider spread, termed over-dispersion, gives rise to shorter ARLs than Poisson, namely, a higher false alarm rate.

Of course, this effect is not peculiar to the cusum chart. It applies also to conventional count charts, such as the application of the Shewhart “c” chart, where it is generally ignored.

To reduce the risk of assuming an incorrect cusum model a simple “dispersion” test is recommended. This involves comparing the sample variance of the data with the sample mean.

Equation (6) indicates that the variance (the square of the standard deviation) is equal to the mean for a Poisson distribution. Hence if:

- 1) sample variance equals the sample mean, the Poisson model is plausible;
- 2) sample variance is “much greater than”<sup>1)</sup> the sample mean, the Poisson model is implausible.

Regarding *point two*: The distribution of the sum of independent Poisson variables is still Poisson. For example: if two sets of Poisson data are independent with separate means  $m_1$  and  $m_2$ , their sum is also Poisson with mean  $(m_1 + m_2)$ .

This effect has considerable impact on counts where multiple characteristics are involved, e.g. adding together different types of flaws, absenteeism for different reasons and injuries of different kinds. Another application is where events of a particular kind have a very small mean so that counts are frequently zero. This enables one to increase the number of sampled items (e.g. flaws per 10 or 100 panels rather than per single panel), or unit time intervals (e.g. counts per month rather than per day or week).

#### 4.3.2 Method for setting up a basic cusum chart for countable data

The Poisson distribution is used here in the sense of scaling the vertical axis of the cusum chart. Later it will also be used in establishing decision rules.

An estimate of the standard deviation will provide the basis for scaling the cusum chart thus:

**Step 1:** For a manual set-up, decide on a convenient interval for the horizontal scale then choose the vertical (cusum) scale such that  $2\sigma_e$  on that scale is roughly equal to 1 unit on the horizontal scale. This will ensure that a shift in the mean of  $2\sigma_e$  will give an angle of  $45^\circ$  to the horizontal. This enhances visual interpretation. Alternatively if some specific shift is to be detected the method of BS 5703-1, Clause 5, Step 5, may be used.

With a computerized set-up for retrospective analysis, the scales and aspect ratio of the chart can usually be readily set-up subjectively to obtain the best effect.

**Step 2:** Decide on a target or reference value. If the chart is being used for retrospective analysis then the overall mean,  $m$ , of the complete set of data will give the best discrimination of changes in the mean.

**Step 3:** Calculate and plot the cusum using:

$$\text{cusum} = \text{sum (observation value count - target value)} \quad (7)$$

<sup>1)</sup> Annex A quantifies what is meant by “much greater than” and provides a specification for the dispersion test.

### 4.3.3 Case study

**Application:** printing errors in 24 successive issues of a magazine are given in Table 2.

Plot a cusum plot and compare it with a standard Shewhart control chart.

**Table 2 — Printing errors and basic cusum tabulation for magazine issues**

| Magazine issue no. | No. of printing errors<br>( $x$ ) | $x - 4$ | Cusum of<br>( $x - 4$ ) | Magazine issue no. | No. of printing errors<br>( $x$ ) | $x - 4$ | Cusum of<br>( $x - 4$ ) |
|--------------------|-----------------------------------|---------|-------------------------|--------------------|-----------------------------------|---------|-------------------------|
| 1                  | 4                                 | 0       | 0                       | 13                 | 7                                 | 3       | 1                       |
| 2                  | 4                                 | 0       | 0                       | 14                 | 8                                 | 4       | 5                       |
| 3                  | 3                                 | -1      | -1                      | 15                 | 2                                 | -2      | 3                       |
| 4                  | 4                                 | 0       | -1                      | 16                 | 4                                 | 0       | 3                       |
| 5                  | 5                                 | 1       | 0                       | 17                 | 1                                 | -3      | 0                       |
| 6                  | 2                                 | -2      | -2                      | 18                 | 3                                 | -1      | -1                      |
| 7                  | 5                                 | 1       | -1                      | 19                 | 6                                 | 2       | 1                       |
| 8                  | 1                                 | -3      | -4                      | 20                 | 3                                 | -1      | 0                       |
| 9                  | 6                                 | 2       | -2                      | 21                 | 4                                 | 0       | 0                       |
| 10                 | 5                                 | 1       | -1                      | 22                 | 2                                 | -2      | -2                      |
| 11                 | 4                                 | 0       | -1                      | 23                 | 4                                 | 0       | -2                      |
| 12                 | 3                                 | -1      | -2                      | 24                 | 6                                 | 2       | 0                       |

here:

the sample size = one magazine;

total printing errors over 24 issues = 96;

mean =  $m = 96/24 = 4$ ;

standard deviation =  $\sigma_e = \sqrt{4} = 2$ .

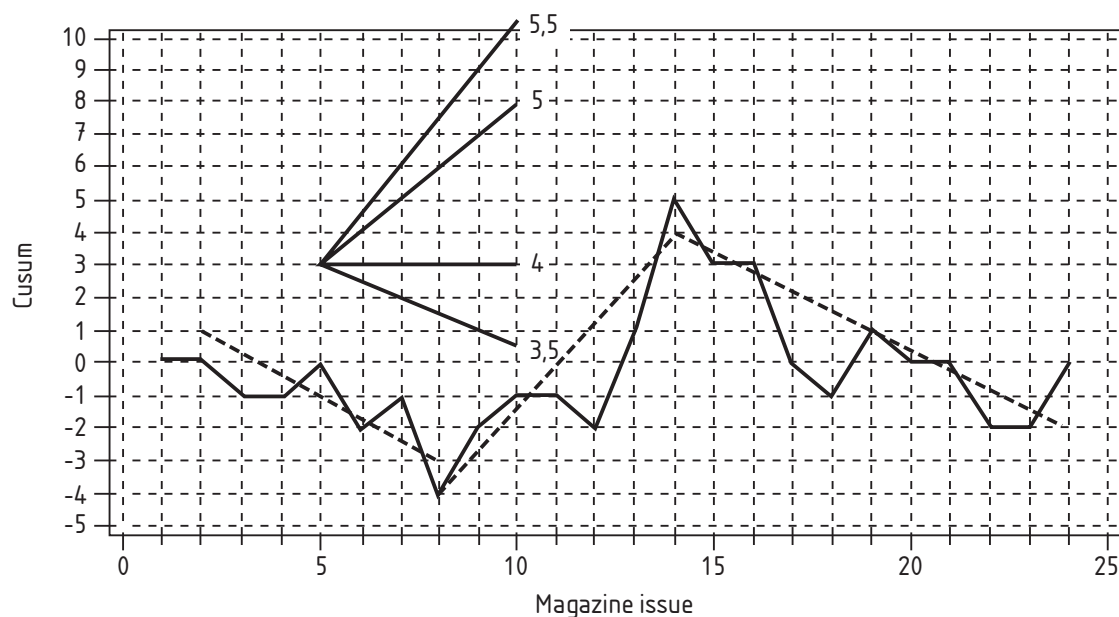


**Figure 3 — Cusum plot of magazine printing errors by issue number**

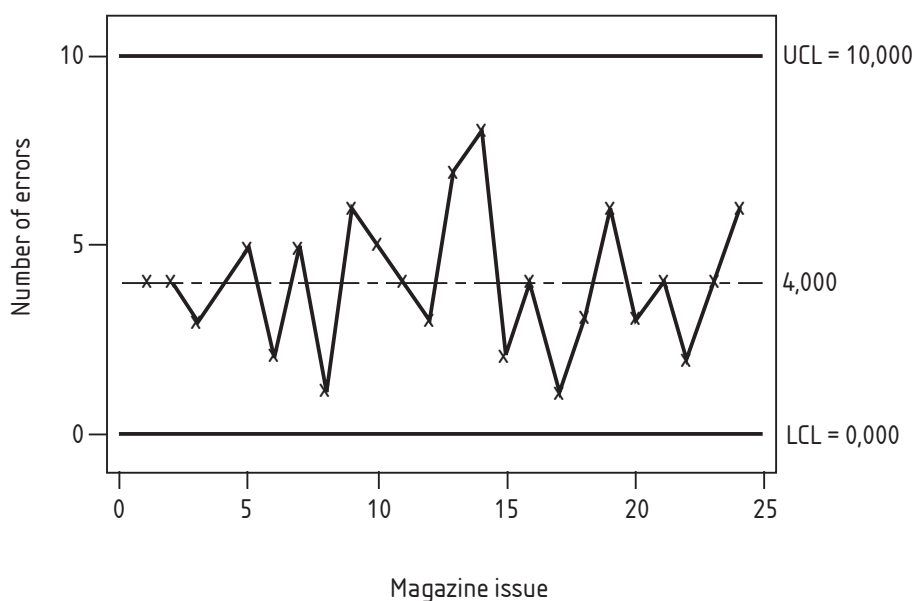


**Figure 4 — Standard Shewhart  $c$  chart for magazine printing errors**

## Conclusions

- a) The standard Shewhart chart of Figure 4 indicates a process in a state of statistical control throughout about an average value of 4 printing errors per magazine.
- b) The cusum plot of Figure 3 reveals quite a different picture. It indicates:
  - 1) an average of 3.5 errors per magazine up to issue number 8 and from issue 15 onwards;
  - 2) an average of 5.25 errors per magazine in issues 9 to 14. This coincided with the employment of a different proofreader during this period.

## 5 Cusum decision schemes for discrete data

### 5.1 General

#### 5.1.1 Introduction

Clause 4 indicates that it is simple, straightforward, and highly rewarding, to construct basic cusum charts for discrete (count and classified) data and interpret them in a purely visual, subjective sense. Setting up formal statistical decision rules and calculating appropriate average run lengths (ARLs) is, however, quite a different matter.

In BS 5703-2 the underlying distribution model on which to base the decision rules was taken to be “normal”. A “normal” distribution is symmetrical and bell shaped. The term “normal” is not used here in its usual context of “usual”, or perhaps, more to the point, to indicate that all other distributions are in some way abnormal. The usual underlying distribution for discrete data is asymmetrical, for attribute data it is binomial and for countable data it is Poisson. Decision rules, mask characteristics and ARLs are critically dependent on the various shapes of these distribution models. Hence it is standard practice to perform statistical tests to check the validity of a particular distribution model.

Due to the general lack of symmetry associated with the binomial and Poisson models, different decision rules should be used for evaluating shifts in the upwards and the downwards directions. Therefore, if a truncated V-mask is used the mask will not be symmetrical as before. There will be different values for the slope and decision interval for the upper and lower halves.

As a further apparent complication, introduced for ease of calculation, certain distributions are sometimes approximated to by others. For example, under certain conditions the Poisson or normal distribution serves as an approximation to the binomial, and in others the normal for the Poisson.

In BS 5703-2 and BS 5703-3 the ARLs for normally distributed data were determined simply from the ARLs of a standardized normal distribution having mean = 0 and standard deviation = 1. Discrete distributions do not possess this feature. Each parameter should be individually calculated. Hence *tabulations* for discrete cusum design purposes have been, of necessity, restricted to selected combinations for movement in the upward direction only. More recently the ready availability of *software routines* has increased significantly the choice of cusum designs for discrete data. These two approaches (tabular and software), to the design of cusum schemes are initially discussed in general in 5.1.3 and 5.1.4.

### 5.1.2 General cusum decision rule design considerations

A cusum scheme for discrete data is uniquely specified in terms of the type of distribution of the data and two parameters, “*K*” the datum value, and “*H*” the decision interval. Key mental markers in the choice of the parameters are the following.

- a) The design of a decision scheme cusum is essentially a two stage process:
  - 1) selection of a *K* and *H* combination to give the desired in-control ARL;
  - 2) determination of the swiftness of the signal response at various appropriate shifts in the mean.
- b) The datum value *K* should be chosen on the basis of the specified shift in mean for which a response is to be signalled. A convenient *K* is at a value between the in-control mean and the out-of-control mean for which the cusum should have maximum sensitivity. The particular value of *K* is dependent on the type of distribution of the data and the definition of what is an acceptable value for the mean. For a:
  - 1) normal distribution, *K* lies midway between the in-control and out-of-control means. Hence a cusum scheme, for normally distributed data, with *k* = 0.5, would be used to optimize the detection of a shift in the mean of 2*k* standard deviations. The in-control mean is sometimes taken to be the largest acceptable mean not necessarily the nominal or target mean;
  - 2) for a Poisson distribution with upward and downward shift, theoretical *K* is positioned between the out-of-control mean, *m*<sub>1</sub>, and the in-control mean, *m*<sub>0</sub>, thus:

$$\text{For upward shift, } K^+ = \frac{m_1 - m_0}{1n \ m_1 - 1n \ m_0} \quad (8)$$

where 1n = logarithm to base e

$$\text{For downward shift, } K^- = \frac{m_0 - m_1}{1n \ m_0 - 1n \ m_1} \quad (9)$$

Example: Calculate *K*<sup>+</sup> and *K*<sup>−</sup> for a Poisson process with a mean of 2. It is required to detect an upward movement to 4 and a downward movement to 1.

$$K^+ = \frac{4 - 2}{1n \ 4 - 1n \ 2} = \frac{2}{1.39 - 0.69} = 2.86$$

$$K^- = \frac{2 - 1}{1n \ 2 - 1n \ 1} = \frac{1}{0.69 - 0} = 1.45$$

- 3) for a binomial distribution, theoretical *K*<sup>+</sup> and *K*<sup>−</sup> are given by:

$$K^+ = K^- = -\text{sample size} \times \frac{1n \left( \frac{1-p_1}{1-p_0} \right)}{1n \left\{ \frac{p_1(1-p_0)}{p_0(1-p_1)} \right\}} \quad (10)$$

where

*p*<sub>1</sub> = in-control proportion;

*p*<sub>0</sub> = out-of-control proportion.

Example: Calculate *K*<sup>+</sup> and *K*<sup>−</sup> for a binomial process with an in-control proportion of 0.01 It is required to detect an upward shift to 0.02 and a downward shift to 0.006.

$$K^+ = -n \left[ \frac{\ln\left(\frac{1-0.02}{1-0.01}\right)}{\ln\left\{\frac{0.02(1-0.01)}{0.01(1-0.02)}\right\}} \right] = -n \left\{ \frac{\ln\left(\frac{0.98}{0.99}\right)}{\ln\left(\frac{0.01998}{0.0098}\right)} \right\} = -n \left( \frac{-0.0102}{0.703} \right) = n0.0145^{2)}$$

$$K^+ = -n \left[ \frac{\ln\left(\frac{1-0.006}{1-0.01}\right)}{\ln\left\{\frac{0.006(1-0.01)}{0.01(1-0.006)}\right\}} \right] = -n \left\{ \frac{\ln\left(\frac{0.994}{0.99}\right)}{\ln\left(\frac{0.00594}{0.00994}\right)} \right\} = -n \left( \frac{-0.00403}{-0.5149} \right) = n0.0078^{2)}$$

c) The actual value of  $K$  used in a cusum scheme is usually rounded off. This is done to limit the number of possible combinations that rises significantly with each increase in decimal place and hence the extent of the calculations required to determine ARLs. It also simplifies the arithmetic needed in operational use of the cusum. For example, a theoretical  $K = 4.853$  could well be rounded off to 4.85, 4.9 or even 5.

d) Having determined  $K$ , a value of  $H$  is now found that achieves a minimum in-control ARL. This predetermined ARL might not be precisely achievable due to the step nature of discrete distributions.

e) The in-control ARL increases with increases in  $K$  or  $H$ . For  $K = H = 0$ ,  $ARL = 1$ .

f) The relationships between  $K$ ,  $H$  and ARL are critically dependent on the underlying distribution of the data.

These considerations provide the basis for the selective combination tables and software approaches to cusum decision scheme design now discussed.

### 5.1.3 Designing and using selective combination tables

A very wide selection of decision rules can be devised having approximately the same ARLs at the target level. They would often differ in their speed of reaction to shifts from the target conditions in mean rate or proportion of occurrences under consideration. A short list of preferred sets of decision rules, termed C1 and C2 schemes, have already been introduced in BS 5703-2. There are four principles adopted in selecting parameters for the C1 and C2 procedures.

- One scheme, C1, should be used to give an ARL at the target level of at least 1 000. The other scheme, C2, should be used to give an ARL of at least 200, at the same target level.
- Greater sensitivity should be given to shifts yielding an ARL in the region of 5 to 10 observations.
- The parameters should be kept simple, generally as integers, but fractional steps should be adopted of not less than  $\frac{1}{4}$  for small rates (or proportions) of occurrences.

NOTE Some special cases arise with very small rates or proportions, as detailed in Clause 8.

- The tables contained in this standard, to avoid undue complication, relate to *upward* shifts in the mean only. The alternative software based design method described in 5.1.3 can also be used for *downward* shifts or for constructing appropriate associated tables.

The target conditions are specified by the mean rates of occurrences of events for Poisson variables ( $m_a$ ) and the proportion of values having some particular attribute for binomial variables ( $p_a$ ). Tables are provided in Annex B for count (Poisson) data, and Annex C for binomial data, appropriate to target levels  $m_a$  and  $p_a$ . In the latter case, the expected number of binomial events per sample of an item is also represented, for convenience, by  $m_a = np_a$ .

<sup>2)</sup> It is seen that these values are very close to half way between the in-control and out-of-control levels in this case. When rounded off, as discussed in c) they do, indeed, coincide.

The general procedure for use of these tables is to choose a target value ( $m_a$  or  $p_a$ ) from the tables, and to use the appropriate pair of cusum parameters  $H$  and  $K$  in one of two ways:

- 1) for DI tabulation: form sum  $(x - K)$  and test against  $H$ ;
- 2) for charting: plot sum  $(x - \text{target value})$  and use a truncated V-mask with decision interval,  $H$ , and slope,  $F$ , where  $F = K - \text{target value}$ .

#### 5.1.4 Designing using software routines

Hawkins and Olwell have developed software programs [4], as a companion to their book on cumulative sum charting [1], that facilitate the design of cusum schemes for both measured and discrete data. These programs can be freely accessed through the cusum website of the School of Statistics, University of Minnesota at <http://www.stat.umn.edu>. Two programs used in this part of BS 5703 are:

- ANYGETH.EXE a program that finds the decision interval ( $H$ ) upon specification of the distribution and the reference value “ $K$ ”;
- ANYARL.EXE a program that finds the ARL for a cusum scheme when the distribution, process mean, reference value “ $K$ ” and decision interval “ $H$ ” are specified.

#### 5.2 Truncated V-mask scheme

The features of the cusum chart that indicate the level of the monitored statistic are the steepness of the cusum slope and the number of observations over which this slope persists. Truncated V-masks of the type indicated in BS 5703-2 take these two features into account. For count data the decision interval ( $H$ ) and slope of the decision lines ( $F$ ) of these V-masks require some modification. The principal reason for this is that usually distributions relating to attribute and count data are skewed to varying degrees. In such cases the symmetrical V-mask used in BS 5703-2 is not appropriate. One-sided masks are then designed to detect a shift in the mean in a single direction at a time. Three construction parameters of such a mask are particularly relevant. These are:

$H$  = decision interval;

$F$  = slope per unit of the decision line;

$K$  = datum value =  $T + F$ .

where

$T$  = target, or nominal, mean value from which departures are required to be detected.

The outline constructional features of such a mask, for detecting upward shifts in the level of the monitored variable, are shown in Figure 5. The detection of downwards shifts entails similar constructional features but with different values due to the general lack of symmetry of discrete distributions.

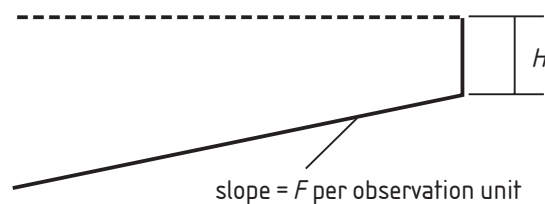


**Figure 5 — Outline features of one-sided V-mask for discrete data**

For cusum charts:

$$\text{cusum} = \text{sum } (X - T).$$

where

$X$  = observed value;

$T$  = target or nominal control value.

A signal occurs if the cusum path cuts the decision line, based on  $H$  and  $F$ , when the mask is overlaid on the cusum plot such that the plotting point of interest (usually the most recent one) is at the intersection of the vertical and horizontal lines.