https://doi.org/10.26219/heal.aueb.6628

ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

DEPARTMENT OF STATISTICS

POSTGRADUATE PROGRAM

CUSUM from a Bayesian Perspective

By

Maria Z. Andreou

A THESIS
Submitted to the Department of Statistics
of the Athens University of Economics and Business
in partial fulfilment of the requirements for
the degree of Master of Science in Statistics

Athens, Greece
July 2009

https://doi.org/10.26219/heal.aueb.6628

https://doi.org/10.26219/heal.aueb.6628

ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ

Μπευζιανή προσέγγιση του CUSUM

Μαρία Ζαχαρίου Ανδρέου

ΔΙΑΤΡΙΒΗ
Που υποβλήθηκε στο Τμήμα Στατιστικής
του Οικονομικού Πανεπιστημίου Αθηνών
ως μέρος των απαιτήσεων για την απόκτηση
Μεταπτυχιακού Διπλώματος Ειδίκευσης στη Στατιστική

Αθήνα
Ιούλιος 2009

https://doi.org/10.26219/heal.aueb.6628

https://doi.org/10.26219/heal.aueb.6628

DEDICATION

To my family

https://doi.org/10.26219/heal.aueb.6628

https://doi.org/10.26219/heal.aueb.6628

ACKNOWLEDGEMENTS

I would like to thank my mother and father who stand up to me during all this effort by giving me emotional and financial support. I would also like to thank my professor Dr. Panagiotis Tsiamyrtzis with whom I had an excellent collaboration. His directions and instructions made this work look easier. Finally I thank every single person that provided me with phycological or practical support.

https://doi.org/10.26219/heal.aueb.6628

VITA

I was born in Larnaca, Cyprus in 1985. I graduated Saint George's high school in July 2003 and in September of the same year I have entered my undergraduate studies at university of Patras, department of Mathematics. I have graduated from university of Patras in October 2007. At the same time, I was accepted in the postgraduate program of Athens University of Economics and Business, for master in Statistics where I have been studied the last two years.

https://doi.org/10.26219/heal.aueb.6628

https://doi.org/10.26219/heal.aueb.6628

ABSTRACT

Maria Andreou

CUSUM from a Bayesian Perspective

July 2009

The basic purpose of this thesis is to study the Classic Cumulative Sum method of Statistical Process Control (SPC), from a Bayesian prism. We consider the hypothesis testing problem of two possible models. Model under H_0 will be the model of the in control situation and model under H_1 will be the model of the out of control situation for a process. We wish to decide whether the model under H_1, has been more plausible than the model under H_0, or not. In other words, we want to decide whether a process has gone from an in control to an out of control situation.

The CUSUM charting is designed to detect small but persistent shifts of a process' characteristic as soon as possible, giving the optimal in control Average Run Length for a particular shift. From a Bayesian point of view, we consider the hypothesis testing problem as a decision theory problem, and using Bayes test theorem, we construct the Bayesian CUSUM. Due to the optimality properties of CUSUM (Moustakides (1986)), we aim to prove that the two forms are equivalent, thus, the Bayesian CUSUM we have constructed will be an optimal scheme for detecting a persistent shift in a process' characteristic.

https://doi.org/10.26219/heal.aueb.6628

https://doi.org/10.26219/heal.aueb.6628

ΠΕΡΙΛΗΨΗ

Μαρία Ανδρέου

Μπευζιανή προσέγγιση του CUSUM

Ιούλιος 2009

Βασικός σκοπός αυτής της διατριβής είναι η μελέτη του CUSUM Chart υπό το πρίσμα της Μπευζιανής στατιστικής. Θεωρούμε ένα πρόβλημα ελέγχου υποθέσεων το οποίο αποτελείται από δύο μοντέλα. Το μοντέλο υπό την H_0, όπου η παραγωγική διαδικασία βρίσκεται στην εντός ελέγχου κατάσταση και το μοντέλο υπό την H_1, που είναι το μοντέλο όπου η παραγωγική διαδικασία βρίσκεται στην εκτός ελέγχου κατάσταση. Στόχος μας είναι να αναγνωρίσουμε πότε το μοντέλο υπό την H_1, θα γίνει πιο πιθανό από το μοντέλο υπό την H_0, δηλαδή πότε η παραγωγική διαδικασία θα περάσει από την εντός ελέγχου κατάσταση, στην εκτός ελέγχου κατάσταση.

Το CUSUM είναι σχεδιασμένο έτσι ώστε να ανιχνεύει μικρές αλλά συνεχείς μεταβολές σε κάποιο χαρακτηριστικό μιας παραγωγικής διαδικασίας, το συντομότερο δυνατό, δίνοντας το βέλτιστο in control Average Run Length για την συγκεκριμένη μεταβολή για την οποία έχει σχεδιαστεί.

Στην Μπευζιανή προσέγγιση θεωρούμε το πρόβλημα ελέγχου υποθέσεων ως πρόβλημα αποφάσεων και χρησιμοποιώντας το θεώρημα του Bayes test, κατασκευάζουμε το Bayesian CUSUM. Επειδή έχει αποδειχθεί ότι το CUSUM δίνει το βέλτιστο in control Average Run Length για την ανίχνευση μεταβόλης του χαρακτηριστικού μιας διαδικασίας (Μουστακίδης (1986)), σκοπός μας είναι να δείξουμε ότι οι δύο μορφές του CUSUM είναι ισοδύναμες. Οπότε, θα αποδείξουμε ότι το Bayesian CUSUM που έχουμε κατασκευάσει είναι ένα αξιόλογο εργαλείο για την γρήγορη ανίχνευση της μεταβολής μιας παραγωγικής διαδικασίας στην εκτός ελέγχου κατάσταση.

https://doi.org/10.26219/heal.aueb.6628

https://doi.org/10.26219/heal.aueb.6628

TABLE OF CONTENTS

 https://doi.org/10.26219/heal.aueb.6628

https://doi.org/10.26219/heal.aueb.6628

LIST OF TABLES

https://doi.org/10.26219/heal.aueb.6628

 https://doi.org/10.26219/heal.aueb.6628

LIST OF FIGURES

https://doi.org/10.26219/heal.aueb.6628

https://doi.org/10.26219/heal.aueb.6628

Chapter 1

Introduction

1.1 Statistical Process Control

Statistical Process Control (SPC) is the area of statistics which studies the use of statistical methods and other problem-solving techniques to improve the quality of the products used by our society. Quality-improvement methods can be applied to any area within a company or organization, including manufacturing, process development, engineering design, finance and accounting, marketing, and field services of products. The question is how could we improve the quality of a product?

We consider a product coming out from a process for which a characteristic is of interest to us. A value that corresponds to the desired value for the quality characteristic is called target value for that characteristic. These target values are usually bounded by a range of values which do not impact the function or performance of the product if the quality of the characteristic is in that range. The largest allowable value for a quality characteristic is called the upper specification limit (USL), and the smallest allowable value is called the lower specification limit (LSL). For example, the diameter of a shaft used in an automobile transmission cannot be too large or it will not fit into the mating bearing, nor can it be too small, resulting in a loose fit, causing vibration, wear, and early failure of the assembly. In some processes we might have only one specification limit line. For instance, in the service industries, specifications are typically in terms of the maximum amount of time to process an order or to provide a particular service and there is no need to consider lower limit.

Generally we assume that the product comes from a process that is stable or repeatable. More precisely, the process must be capable of operating with little variability around the target or nominal dimensions of the product's quality characteristic. Statistical process control is a collection of problem-solving tools useful in achieving process stability and improving capability through the reduction of variability.

https://doi.org/10.26219/heal.aueb.6628

In any production process, regardless of how well designed or carefully maintained it is, a certain amount of inherent or natural variability will always exist. This natural variability is the cumulative effect of many small, essentially unavoidable causes. In the framework of statistical quality control, this natural variability is often called "a stable system of common causes". A process that is operating with only common causes of variation present is said to be in statistical control. Other kinds of variability may occasionally be present in the output of a process which usually arises from improperly adjusted or controlled machines, operator errors, or defective row material. Such variability usually represents an unacceptable level of process performance. We refer to these sources of variability that are not part of the common cause pattern as special causes. A process that is operating in the presence of special causes is said to be out of control. For further information see Deming (1986).

Often production processes will operate in the in control state, producing an acceptable product for relatively long periods of time. Eventually, however, special causes will occur, resulting in a shift to an out of control state, where a large proportion of the process output does not conform to requirements. When the process is in control, almost all of the production will fall between the lower and upper specification limits. When the process is out of control, a higher proportion of the process is expected to lie outside of these specifications. Nevertheless, it is possible for the process to shift out of control and all of the data still remain within specification limits.

A major objective of the SPC is to quickly detect the occurrence of special causes so that investigation of the process and corrective action may be undertaken before many nonconforming units are manufactured. The Control Chart is an on-line process-monitoring technique widely used for this purpose. After all, the eventual goal is the reduction of variability in the process. The Control Charts known to be an effective tool in reducing variability.

Through variability's control we are capable of improving the process and consequently the quality of the product. Thus we are interested in keeping the specific quality characteristic of the process inside some predefined boundaries, known as control limits: the upper control limit (UCL) and the lower control limit (LCL). Control limits are usually symmetric to a center line (CL), which represents the

 https://doi.org/10.26219/heal.aueb.6628

average value of the characteristic corresponding to the in control state. These control limits are chosen so that if the process is in control, nearly all of the sample points will fall between them. As long as the points lie between the control limits, the process is assumed to be in control and no action is necessary. However, a point that lies outside of the control limits is interpreted as evidence that the process is out of control, and investigation and corrective action are required to find and eliminate the special causes responsible for this behavior. This general procedure is summarized in control charts theory. A typical control chart is a graphical display of a quality characteristic that has been measured or computed from a sample versus the sample number of time. The chart contains the center line that represents the average value of the quality characteristic corresponding to the in control state, and two other horizontal lines representing the UCL and LCL respectively.

Figure 1.1 represents an in control situation, and Figure 1.2 an out of control situation example for the mean of a normal distribution.

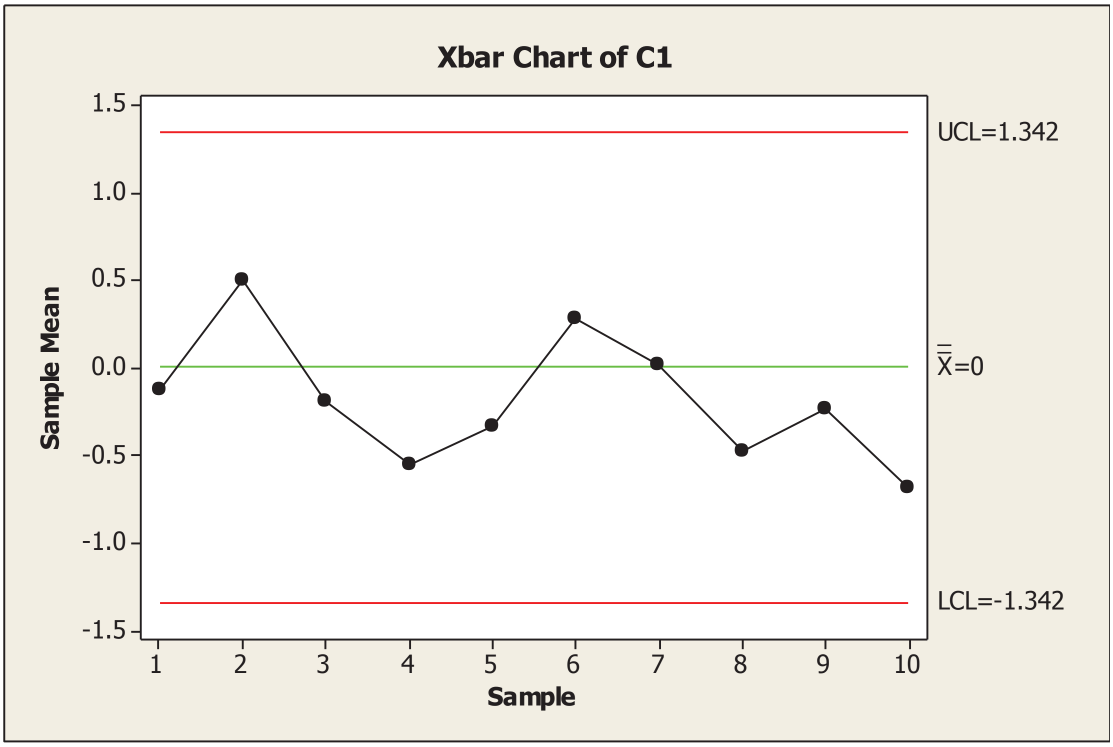


Figure 1.1: Control chart of the mean of a normal distribution for an in control process. In this control chart the in control situation is considered to be the $N(0,1)$, thus, the center line is CL=0.

https://doi.org/10.26219/heal.aueb.6628

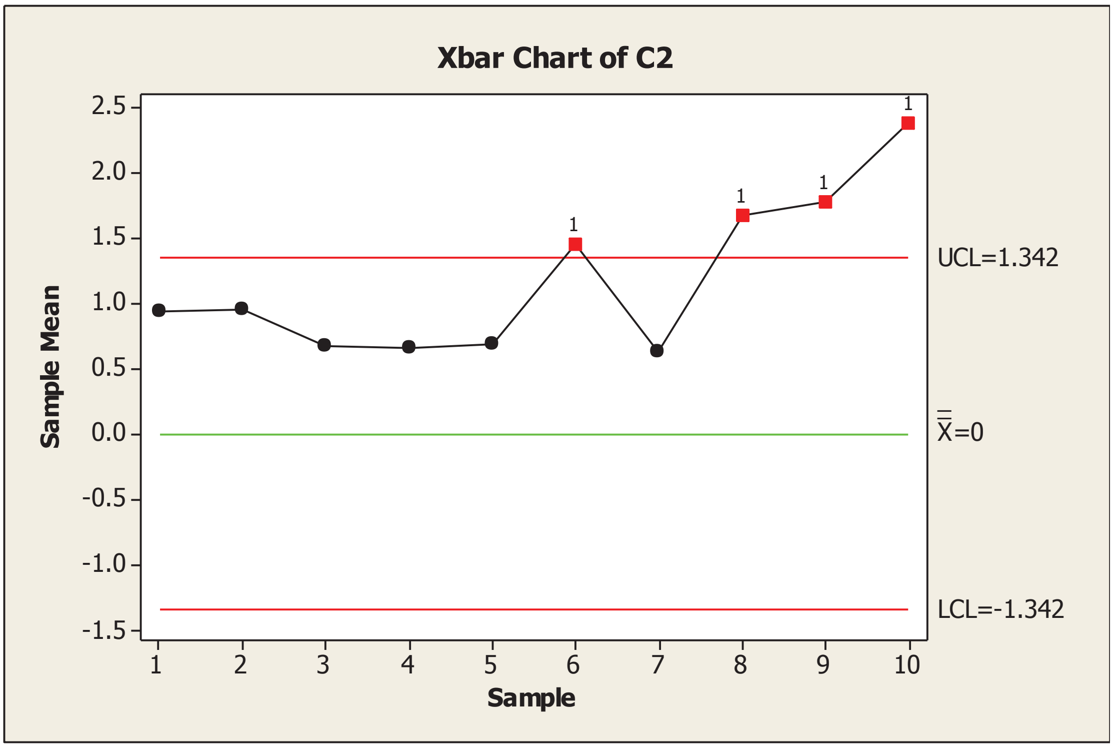


Figure 1.2: Control chart of the mean of a normal distribution for an out of control process. In this control chart the in control situation is considered to be the $N(0,1)$, thus, the center line is CL=0.

There is a close connection between control charts and hypothesis testing. To illustrate this connection, suppose that the vertical axis of a control chart is the sample mean $\bar{x}$. If the current value of $\bar{x}$ is between the control limits, we conclude that the process mean is in control, i.e. we cannot reject the null hypothesis that it is equal to the target value μ_0. On the other hand, if $\bar{x}$ exceed either control limit, we conclude that the process mean is out of control, that is, it is equal to some value $\mu_1 \neq \mu_0$. In a sense, then, the control chart is a test of the hypothesis that the process is in a state of statistical control. A point plotting within the control limits is equivalent to failing to reject the hypothesis of statistical control, and a point plotting outside the control limits is equivalent to rejecting the hypothesis of statistical control.

Two kinds of errors may be committed when testing hypotheses. If the hypothesis that the process is in a state of statistical control (null hypothesis), is rejected when it is true, then a type I error has occurred. If the null hypothesis is not

https://doi.org/10.26219/heal.aueb.6628

rejected when it is false then a type II error has occurred. The probabilities of these two types of error are denoted as:

$$\alpha = P\{type\ I\ error\} = P\{reject\ H_0 | H_0\ is\ true\}$$
$$\beta = P\{type\ II\ error\} = P\{fail\ to\ reject\ H_0 | H_0\ is\ false\}. \tag{1.1}$$

The general theory of control charts were introduced by Walter Shewhart (Shewhart (1931)), and they are known as Shewhart Control Charts.

To illustrate the idea of Shewhart Control Charts, we give an example. Suppose that every hour a random sample of n independent objects is taken from a population, the average of the sample say, $\bar{x}$ is computed, and $\bar{x}$ plotted on the chart. Because this control chart utilizes the sample average $\bar{x}$ to monitor the process mean, it is usually called an $\bar{x}$ control chart. In order to determine the control limits for this chart we use the central limit theorem to assume that $\bar{x}$ is approximately normally distributed. Thus, if the process is in control with a mean $\bar{x}$, we would expect $100(1-\alpha)\%$ of the sample mean to fall between $\bar{x} - Z_{\alpha/2}\sigma_{\bar{x}}$ and $\bar{x} + Z_{\alpha/2}\sigma_{\bar{x}}$ where α, is the probability of type I error given by equation (1.1). The quantities $\bar{x} - Z_{\alpha/2}\sigma_{\bar{x}}$ and $\bar{x} + Z_{\alpha/2}\sigma_{\bar{x}}$ represent the UCL and LCL of the control chart respectively. This is a general form of control limits for independent normal data using a multiple of standard deviation $\sigma_{\bar{x}}$. We consider $N(0,1)$ as the in control situation. It is known that $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ .Thus, for $\alpha = 0.0455, \left(Z_{\alpha/2} = 2\right)$, we fail to reject H_0 if the sample mean falls between the control limits:

$$\left[\bar{x} - Z_{\alpha/2}\sigma_{\bar{x}}, \bar{x} + Z_{\alpha/2}\sigma_{\bar{x}}\right] = \left[0 - 2 * \frac{\sigma}{\sqrt{n}}, 0 + 2 * \frac{\sigma}{\sqrt{n}}\right] = \left[-2 * \frac{1}{\sqrt{n}}, 2 * \frac{1}{\sqrt{n}}\right]$$

and we reject H_0, if the sample mean falls outside these limits. We deliberately derive 50 random variables from the in control distribution $N(0,1)$ and 50 from an out of control distribution $N(1,1)$. We compute the sample mean for subgroups of size 5 for each one of the two samples and then we plot it on the above control chart. Thus, control limits are computed to be:

$$\left[-2 * \frac{1}{\sqrt{n}}, 2 * \frac{1}{\sqrt{n}}\right] = \left[-2 * \frac{1}{\sqrt{20}}, 2 * \frac{1}{\sqrt{20}}\right] = [-0.447, 0.447]$$

 https://doi.org/10.26219/heal.aueb.6628

Figure 1.3 illustrates that particular $\bar{x}$ control chart for normal mean for a shift of one standard deviations. As we can notice, the last 10 points, respected to the out of control situation, are plotted above the upper control limit.

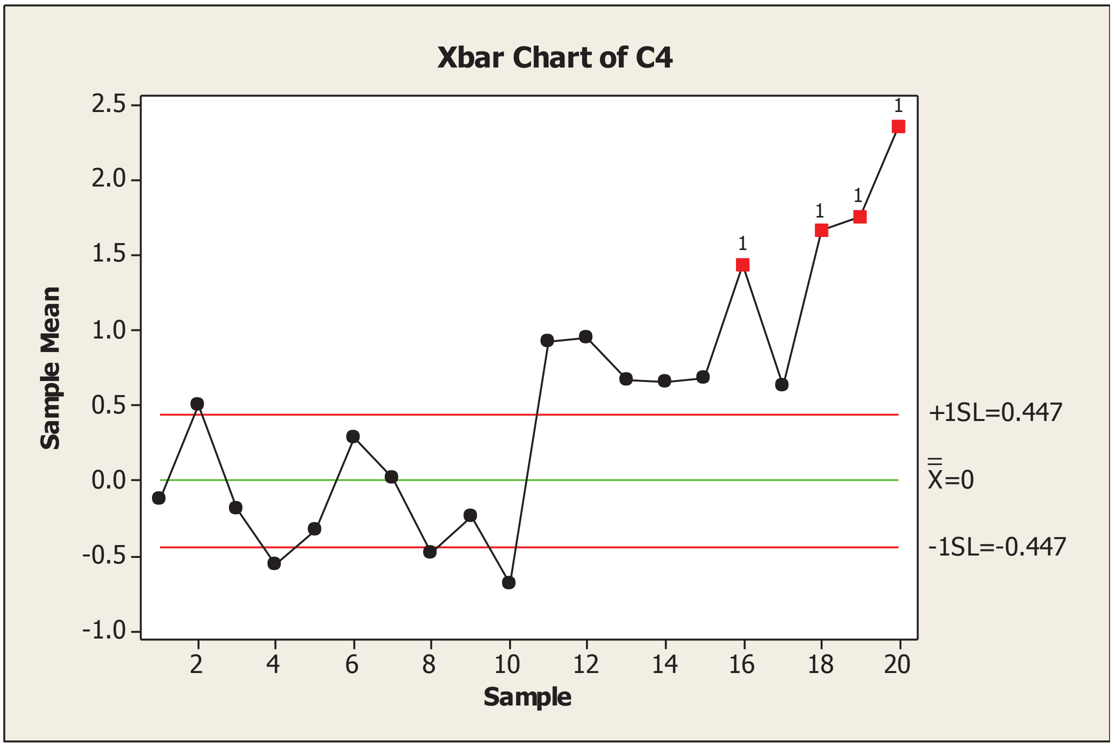


Figure 1.3: $\bar{x}$ control chart for normal mean for "two sigma" control limits.

The width of the control limits is inversely proportional to the sample size n for a given multiple of sigma, $\sigma_{\bar{x}}$, the standard deviation of the sample average $\bar{x}$. Note that choosing the control limits is equivalent to setting up the critical region for testing the hypothesis:

$$H_0: \mu = \mu_0$$

$$H_1: \mu \neq \mu_0$$

where $\sigma_{\bar{x}}$ is known. Essentially, the control chart tests this hypothesis repeatedly at different points in time.

We may give a general model for a control chart. Let w be a sample statistic that measures some quality characteristic of interest, and suppose that the mean of w

 https://doi.org/10.26219/heal.aueb.6628

is μ_w and the standard deviation of w is σ_w with μ_w, σ_w known. Then the center line, the upper control limit and the lower control limit become:

$$UCL = \mu_w + L\sigma_w$$

$$CL = \mu_w$$

$$LCL = \mu_w - L\sigma_w$$

where L is the distance of the control limit from the center line, expressed in standard deviation units. If the latest w value plots outside the control limits, then the control chart is considered to have given the signal that the process has gone out of control. Then we stop the process, and we reinstate the control limits. This is done by examining each of the out of control points, looking for a special cause. If a special cause is found, the point is discarded and the control limits are recalculated, using only the remaining points. Then these remaining points are reexamined for control. Note that points that were in control initially may now be out of control, because the new control limits will generally be tighter than the old ones. This process is continued until all points plot in control, at which point the control limits are adopted for current. As long as the points plot inside the control limits, no action is taken to alter the process. This rule can stop much unproductive repair that could take a process from a good state to a bad one. The control limits are placed sufficiently far from the center line so that very few samples should plot outside them if the process remains at its in control distribution. Thus, if the control chart does give a signal, it should be taken seriously. For further information see Montgomery (2001). Also see Hillier and Yang (1970) for control charts' limits based on small subgroups.

As we have already said, the construction of control charts, as we have defined it above, premises the independency of the data. For control charting of dependent data see Ermer (1980). Further more for correlated and autocorellated data see Harris and Ross (1991) and Montgomery and Mastrangelo (1991) respectively.

The control charts we have described, the Shwehart control charts, are member of a greater family of control charts known as Control Charts for variables. A quality characteristic that is measured on a numerical scale is called a variable.

https://doi.org/10.26219/heal.aueb.6628

Examples of variables include dimensions such as length or width. When dealing with a quality characteristic that is a variable, it is usually necessary to monitor both the mean value of the quality characteristic and its variability. Control of the mean is usually done with the $\bar{x}$ control chart. Process variability can be monitored with a control chart for standard deviation called the S chart or a control chart for the range, called the R chart.

Although Shewhart chart is very effective for detecting isolated special causes that lead to large shifts in the data, it is not very effective in detecting more moderate shifts, even if these more moderate shifts persist.

There have been various attempts to fix this inadequacy in the Sewhart chart by adding rules that will lead to a signal if the successive points have specified unusual patterns to them, even if they remain inside control limits (see Wester Electric Handbook (1956)).

Apart from variables, there are quality characteristics, which cannot be conveniently represented numerically. In such cases we usually classify each item inspected as either conforming or nonconforming to the specifications on the quality characteristic. Quality characteristics of this type are called attributes. In fact, control charts for attributes are control charts for categorical data contrary to control charts for variables, which are control charts for continuous data. For further information see Woodall (1997).

As far as the control charts for variables are concerned, despite the criteria which could be applied to Shewhart control charts in order to make them more sensitive into small shifts in process, the results are not always the desirable. The use of those rules, can dramatically reduce the average run length of the control chart when the process is in control (see Montgomery (2001)). Thus, two very effective alternatives to Shewhart control chart may be used when small shifts are of interest: the cumulative sum (CUSUM) control chart (see page (1954)) and the exponentially weighted moving average (EWMA) control chart (see Roberts (1959)).

The CUSUM chart directly incorporates all the information in a sequence of a sample values by plotting the cumulative sums of the deviations of the sample values from a target value. For example, suppose we want to test whether a process' mean is

https://doi.org/10.26219/heal.aueb.6628

equals to a target value μ or not. Suppose that samples of size $n \geq 1$ are collected and $\bar{x}_j$ is the average of the jth sample.The CUSUM control chart is formed by plotting the quantity

$$C_n = \sum_{i=1}^{n} (\bar{x}_i - \mu)$$

against n. C_n is called the cumulative sum up to and including the nth sample. Cumulative sum charts are more effective than Shewhart charts for detecting small process shifts. Furthermore, they are particularly effective with samples of size $n = 1$.

The exponentially weighted moving average (EWMA) control chart was introduced by Roberts (1959) and it has approximately equivalence performance to that of the CUSUM. It is typically used with individual observations. The EWMA is defined as:

$$z_i = \lambda * x_i + (1 - \lambda) * z_{i-1}$$

where $0 < \lambda \leq 1$ is a constant and the starting value (required with the first sample at $i = 1$) is the process target, so that:

$$z_0 = \mu\,.$$

The EWMA control chart is often used with individual measurements. However, if rational subgroups of size $n > 1$ are taken, then simply replace x_i with $\overline{x_i}$ and then the standard deviation of the random variable will be $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ (see Montgomery (2001)).

Therefore the EWMA control chart would be constructed by plotting z_i versus the sample number i. For further reading see Hunter (1986).

In this thesis we will focus in Cumulative Sum control chart. We will provide a brief introduction to classic CUSUM control chart in Chapter 2, giving various forms of CUSUM, ways to construct it and its optimality properties. We will also represent the relation of CUSUM with the sequential probability ratio test, especially for members of the exponential family. Afterwards, in Chapter 3, after a short introduction to Bayesian theory in subsection 3.1, we will provide a Bayesian approach of CUSUM using the Bayes test, and eventually we will prove the mathematical equivalence between the two forms of CUSUM in Chapter 4. Then, we will illustrate that equivalence of the two CUSUMs by simulation methods and finally

https://doi.org/10.26219/heal.aueb.6628

in Chapter 5 we will draw conclusions about the Bayesian CUSUM we will construct, and its optimality in detecting a shift in a process' characteristic.

https://doi.org/10.26219/heal.aueb.6628

Chapter 2

Cumulative Sum Control Chart

This Chapter constitutes a presentation about CUSUM control chart. We define cumulative sum chart and we introduce different ways to form it. We represent its statistical properties and its design. Based on the design of the CUSUM we illustrate its optimality properties in controlling the mean of a process, and particularly applying the relation between the Sequential Probability Ratio Test and the CUSUM we result to an algebraic form of CUSUM which we use in order to control the mean of any member of the exponential family.

2.1 Definition of the CUSUM

Cumulative sum charting is widely used for detecting small but persistent shifts in a process. Suppose we have to deal with the problem of whether the mean of a process, μ, has gone from an in control parameter value, μ_0, to an out of control value μ_1. Specifically, let's assume an independent and identically distributed (i.i.d) process for which the quality characteristic X follows the normal distribution with mean μ_0 and standard deviation σ. Thus, we intend to check whether the normal distribution $N(\mu_0, \sigma^2)$ of the characteristic X has shifted to $N(\mu_1, \sigma^2)$.We take a random sample of size n from the process and we write X_i for the i$^{\text{th}}$ observation. The idea of CUSUM is based on the use of quantities:

$$C_m = \sum_{i=1}^{m} (x_i - \mu). \tag{2.1.1}$$

The CUSUM plot consists of a graph of C_n against m with $m \leq n$.

Sometimes it is more useful to standardize the observations, so that they have zero mean and unit standard deviation. We define:

$$\begin{gathered} U_i = (x_i - \mu)/\sigma \\ S_n = \sum_{i=1}^{n} U_i. \end{gathered} \tag{2.1.2}$$

 https://doi.org/10.26219/heal.aueb.6628

It is obvious that the CUSUM C_n is the CUSUM S_n scaled by a factor of σ. Thus the two CUSUMs are identical except of the units of the vertical axis. The vertical axis of the S_n will be measured in multiples of the standard deviation σ of the data. Statistically the two CUSUMs contain the same information.

2.2 Statistical properties of CUSUM:

As we can notice C_n is the sum of independent normal $N(0, \sigma^2)$ quantities. Thus its distribution is:

$$C_n \sim N(0, n\sigma^2).$$

The standard deviation of C_n is proportional to the square root of n. This means that as n increases, C_n is increasingly likely to be far from zero.

The equation for C_n shows that it can be written in a recursive form where:

$$C_0 = 0$$

$$C_n = C_{n-1} + (X_n - \mu). \tag{2.2.1}$$

The recursive form implies that, while the process will be in control the CUSUM will be a random walk with no drift, (each point is the previous point plus the offset of the latest point from μ, an offset with zero mean). It will center on the horizontal axis and be subject to ever-widening excursions from the axis.

Suppose that at time m, the mean of process undergoes a persistent shift of size δ. Thus for any later instant $n > m$, CUSUM could be written as:

$$C_n = \sum_{i=1}^{n}(x_i - \mu) = \sum_{i=1}^{m}(x_i - \mu) + \sum_{i=m+1}^{n}(x_i - \mu). \tag{2.2.2}$$

Since, from instant m: $(x_i - \mu) \sim N(\delta, \sigma^2)$ we have

$$\sum_{i=m+1}^{n}(x_i - \mu) \sim N[(n-m)\delta, (n-m)\sigma^2]. \tag{2.2.3}$$

https://doi.org/10.26219/heal.aueb.6628

This means that at time $n > m$ the average value of the CUSUM is $(n-m)\delta$. Thus starting from the point (m, C_m), the CUSUM on average will trace out a path centered on a line of slope δ. While the process follows the in control $N(\mu, \sigma^2)$ distribution, the CUSUM follows a distribution centered on the horizontal axis. If the mean undergoes a step change, then the CUSUM develops a linear drift, and its distribution will be centered on a line whose slope δ equals the shift in mean. The idea of CUSUM is based on distinguishing the no-drift in control behavior from the linear drift behavior following a persistent mean shift.

There are both graphic and algebraic forms of CUSUM which could be used for testing for a shift in a process. Formula (2.2.1) represents the algebraic form and (2.2.3) gives the out of control distribution's algebraic form.

2.2.1 Graphic form of CUSUM

We will now illustrate the graphic form of CUSUM which is simply a plot of C_n versus n. For example, suppose we have 30 observations from $N(0,1)$ and 15 observations from $N(1,1)$. Thus, from instant m =31, the process has shifted from the in control distribution $N(0,1)$ to the out of control distribution $N(1,1)$. Figure 2.1 illustrates these results. As we can notice CUSUM detected the shift very quickly. The graph has shown an upward trend right after the shift occurred. The upward trend is a signal that the process has gone out of control. Thus, we are able to detect the shift, by the moment it occurred, since m =31.

 https://doi.org/10.26219/heal.aueb.6628

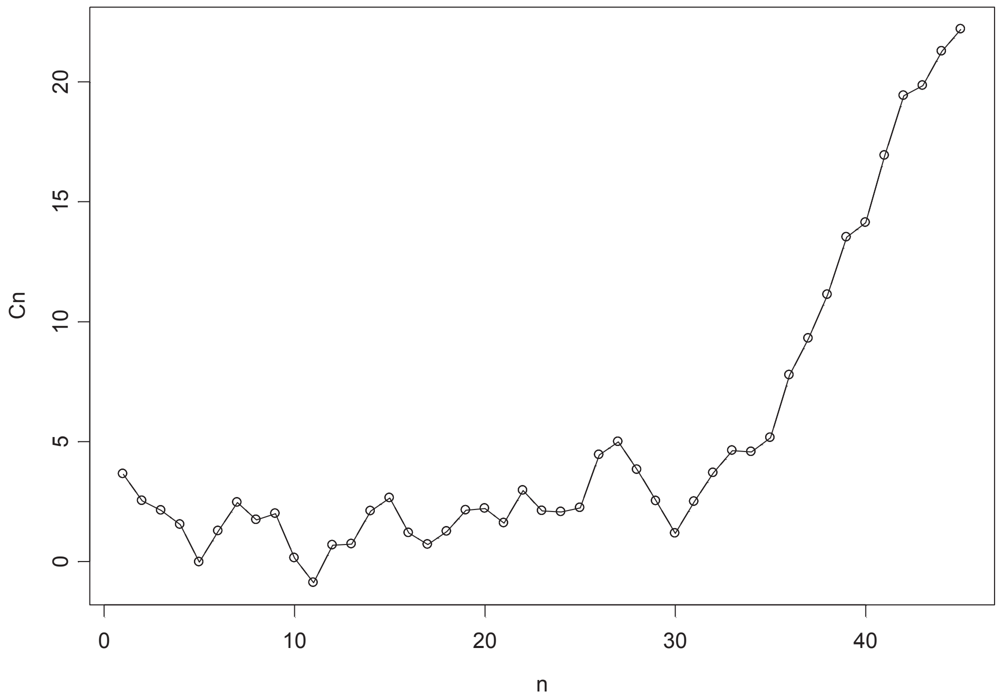


Figure 2.1*:* Illustration of the CUSUM of 30 observations from $N(0,1)$, the in control distribution, and 15 observations from $N(1,1)$ the out of control distribution

We compare CUSUM control chart with the corresponding Shewhart $\bar{x}$ control chart. We use the same 45 observations as we did for the construction of the CUSUM. Due to the definition of the Shewhart $\bar{x}$ control chart, we have to separate the sample into subgroups and compute the mean of every sample. Hence, we consider subgroups of sample size $n = 3$ and thus, we compute 10 sample means coming from the in control distribution $N(0,1)$ and 5 sample means coming from the out of control distribution $N(1,1)$. Figure 2.2 illustrates Shewhart $\bar{x}$ control chart for these results. As we can notice, Shewhart $\bar{x}$ control chart is unable to detect the shift in sample's mean since all the sample means of the subgroups have plotted inside the control limits.

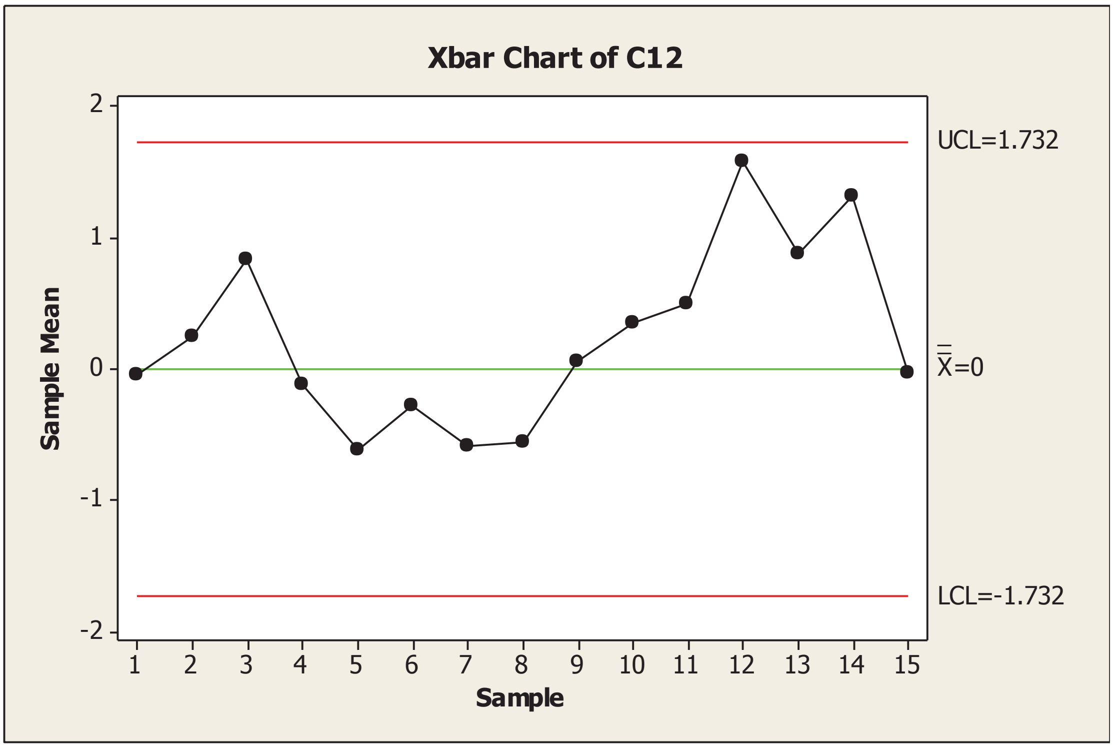


Figure 2:2: Illustration of the Shewhart $\bar{x}$ control chart, of 30 observations from $N(0,1)$, the in control distribution, and 15 observations from $N(1,1)$ the out of control distribution

2.3 Testing for a shift-the V mask (graphic)

The behavior of the CUSUM when the process is in and out of control shows what sort of behavior would constitute evidence of a shift in mean: a change from generally horizontal motion to a non horizontal linear drift. The historical formal tool for making this determination is the V-mask, so named because of its shape.

Suppose we are interested in seeing whether there has been a mean shift of size Δ standard deviations in the $U_i = (x_i - \mu)/\sigma$.

Suppose that we are at point m and we want to know whether it is the start of a non-horizontal linear drift. Using the plot of S_n against n, we fix upon a particular slope, let say Δ, for the non-horizontal drift we are looking for. This corresponds to a shift in mean of k standard deviations. So we want to decide between a slope of zero, and a slope of k for the segment leading out of the point $\{m, S_m\}$. A reasonable way to do this is to draw a line of slope midway between these two levels, that is of

 https://doi.org/10.26219/heal.aueb.6628

slope (½)Δ. If CUSUM remains close to or below this line, we will conclude that the mean has not shifted. But if the CUSUM goes substantially above this line, would be indicative that the drift of slope Δ is more believable than is slope zero. Thus, we say that if any future point is more than some threshold height h above this line, it will constitute the evidence of the shift.

Hence, we draw a line of slope $k = \frac{1}{2}\Delta$ starting from the point $\{n, S_n - h\}$ and going backward. If any preceding point lies below this line, then we signal an upward shift in mean.

If we are concerned about a possible downward shift of size Δ we could draw a line of slope $-k$, starting from $\{n, S_n + h\}$ and going backward. If any preceding point lies above this line, then we can conclude that the mean had shifted downward.

To check on shifts in mean for any subsequent point, we apply the same test every time a new point is added to the CUSUM by using a cutout mask. This mask has a flat front (right) end of height $2h$ and two legs of slope k and $-k$. As each new point is added to the CUSUM we align the center of the front edge with the point just plotted and we check whether all the previous points are contained within the mask. If they are then we conclude that the process is still in control. This means that the mean has not shifted. But if any preceding point projects out the mask, then we conclude that the mean has shifted.

If working with the un-standardized CUSUM (C_n), then both k and h need to be scaled up the factor σ.

For example we suppose we have the same 30 observations from normal $N(0,1)$ and 15 observations from $N(1,1)$ that we did for figures 2.1 and 2.2 above. Thus, from instant m =30, the process has shifted from the in control distribution $N(0,1)$ to the out of control distribution $N(1,1)$. So we construct the V-mask of CUSUM for h=4 and $k = 0.5$. Figure 2.3 illustrates those results. As we can notice, there are preceding points which are not contained within the mask. This is a sign that the process has gone out of control.

https://doi.org/10.26219/heal.aueb.6628

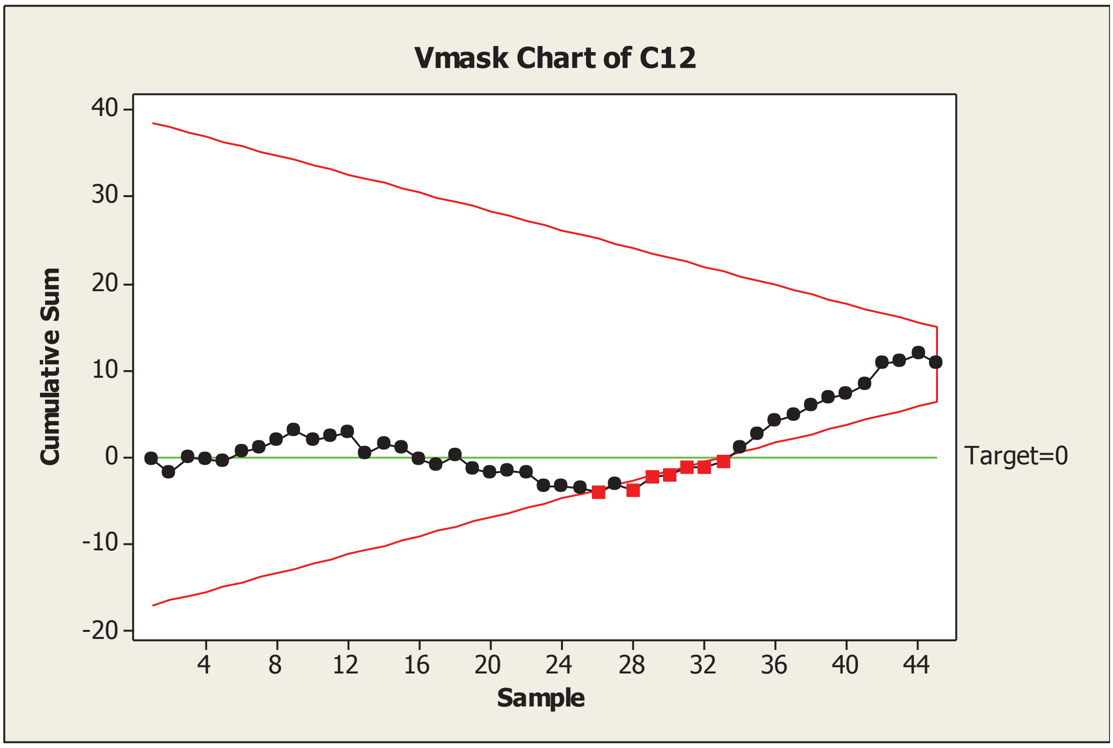


Figure 2.3: Illustrates V-mask form of the CUSUM for h=4 and $k = 0.5$, for 30 random variables of the in control distribution $N(0,1)$ and 15 of the out of control distribution $N(1,1)$.

2.4 Estimation

A CUSUM signal provides an estimate for both

- m: the reading following which the mean shift occurred
- δ: the magnitude of the shift

We estimate m as the observation number of the case that falls outside of the V-mask. If more than one of the previous points falls outside the V-mask, then we pick as the estimator of m the point lying farthest below the mask. Considering the example of section 2.2.1, we estimate that $m = 28$ because it seems to be the point lying farthest below the mask.

We estimate δ as the slope of the CUSUM over the out-of –control segment from the time m, the instant before the shift occurred, to the time n at which the shift is detected. The CUSUM slope estimator is

 https://doi.org/10.26219/heal.aueb.6628

$$\hat{\delta} = \frac{C_n - C_m}{n - m} \tag{2.4}$$

All the above values are not read from the plot but from the actual calculated values used to produce the plot.

2.5 The decision interval form of the CUSUM (algebraic)

The decision interval (DI) form of the CUSUM is an algebraic equivalent to setting up the V-mask form CUSUM C_n and diagnosing it, using a V-mask of a particular slope k with particular leg height h (see paragraph 2.3).

Monitoring X_n for an upward shift in mean is done by setting up the sequence:

$$C_0^+ = 0$$

$$C_n^+ = \max(0, C_{n-1}^+ + X_n - \mu - k). \tag{2.5.1}$$

Signaling an upward shift in mean if

$$C_n^+ > h. \tag{2.5.2}$$

If there is a signal, the estimate of the time of occurrence of the shift m, in the DI form is the most recent observation for which $C_m^+ = 0$.

DI form and V-mask forms are equivalent if we use the same k and h.

The DI CUSUM C_n^+ tests for upward mean shifts. Similarly to check whether there is a downward shift in mean, we set up the sequence:

$$C_0^- = 0$$

$$C_n^- = \min(0, C_{n-1}^- + X_n - \mu + k) \tag{2.5.3}$$

with a signal if

https://doi.org/10.26219/heal.aueb.6628

$$C_n^- < -h. \tag{2.5.4}$$

When there is an indication of a downward shift in mean, we use the last point m for which $C_m^- = 0$ as the estimate for the instant preceding the change in mean.

We can use the decision interval form of the CUSUM to estimate the magnitude of the shift in mean such as we did with the C_n form for an upward shift. The segment of the DI CUSUM leading to the signal starts at some case m for which $C_m = 0$ and then is positive up to the point at which it crosses the decision interval h. In this segment:

$$C_n^+ = C_m^+ + \sum_{i=m+1}^{n} (x_i - \mu - k). \tag{2.5.5}$$

Following the shift in the mean from μ to $\mu + \delta$, the summand is normally distributed with mean $(\delta - k)$. This means that we can estimate δ by adding k to the slope of the decision interval CUSUM from point m to point n , thus we get the estimate:

$$\hat{\delta} = k + \frac{C_n^+ - C_m^+}{n - m} \tag{2.5.6}$$

Because of $C_m^+ = 0$, the above estimator reduces to:

$$\hat{\delta} = k + \frac{C_n^+}{n - m}$$

which is the same estimator we got by looking at the slope of C_n form of the CUSUM in Equation (2.4).

Similarly, if the downward decision interval CUSUM signals a shift, then the magnitude of the shift can be estimated by

$$\hat{\delta} = -k + \frac{C_n^-}{n - m} \tag{2.5.7}$$

Generally, the introduction to the CUSUM and the demonstration of its ability to detect small but persistent shifts belongs to Page E.S (1954). His CUSUM had an equivalent form to the V-mask and the DI forms, but it was algebraically distinct. Page's CUSUM had the form:

 https://doi.org/10.26219/heal.aueb.6628

$$P_0 = 0$$
$$P_n = P_{n-1} + X_n - \mu - k.$$

As we can notice is the DI form without the step of resetting the CUSUM to zero if it should go negative. His rule for a signal was:

$$P_n - min_{0<m<n} P_m > h.$$

2.6 CUSUM design

The design of the CUSUM is based on the choice of the parameters k and h that define the decision interval scheme or the equivalent V-mask. The chart constant k is called the "reference value" and the chart constant h is called the "decision interval".

In the definition of the two equivalent forms of the CUSUM, in controlling the mean of $N(\mu, \sigma^2)$ quantities, it was suggested that $k = \frac{1}{2}\Delta$, where Δ is the anticipated shift in the process mean. This implies that we will pick different chart constants if we change our opinion about the likely size of the shift in process mean.

It will be proved later that if we work with standardized data for the shift in mean of a normal distribution, that intuitive choice for the reference value k is the right one (Hawkins and Olwell (1997). Choosing a particular value k implies that the CUSUM is being designed to detect a shift of $\Delta = 2k$ standard deviations.

One possible way of choosing Δ is that we choose the size of the shift for which we desire the quickest detection. This would be a shift large enough to have a meaningful impact on the process operation, but small enough not to be obvious by naked eye.

It is important to understand that k is chosen for optimal response to a shift of specified size Δ. However, it is not optimal for shifts of substantially different size (Moustakides (1986)).

 https://doi.org/10.26219/heal.aueb.6628

2.6.1 Average run length

Suppose we work with standardized data. The CUSUM chart starts out at its initial state S_0^+. From there, it may stay on the axis, or it may move into positive values. When the chart crosses the decision interval, h, this indicates that a shift has occurred and actions will be taken to diagnose and eliminate the cause of the shift. The CUSUM will then be restarted. The whole sequence going from the starting point to the CUSUM crossing the decision interval is called a run. The number of observations from the starting point up to the point at which the decision interval is crossed is called the run length.

Sometimes the CUSUM will signal when in fact no shift has occurred. This false alarm is analogous to a Type I error in classical hypothesis testing problem, and it is undesirable. Nevertheless, false alarms are inevitable. Thus, we would like the runs between false alarms to be as long as possible.

Analogous to Type II error in classical hypothesis testing, in control charting, is a chart remaining within its decision interval even though there has been a shift. If the shift is big enough to cause implications to the process we would like to detect it as soon as possible. Obviously we are interested in long runs for false alarms but short runs for signals of actual shifts. Of course these two objectives are conversely analogous, so it is necessary to make trade-offs between them.

The run length is a random variable, having distribution with some mean, and variance. Its mean is called the average run length denoted as ARL.

The in control behavior of the CUSUM is measured by its in control ARL. We want this to be sufficiently long. The out of control ARL is the ARL of the CUSUM following a shift in the process mean. We want this to be as short as possible.

As a result, the ARLs, in control and out of control, will be the performance measures for a CUSUM.

https://doi.org/10.26219/heal.aueb.6628

2.6.2 The choice of h (decision interval)

The in control ARL of the CUSUM depends on the values of k and h. Larger values of either of these parameters lead to larger in control ARLs. The decision interval h, is usually fixed by deciding on the minimum tolerable in control ARL.

When X is a continuous random variable, whatever value is given to k, it is possible to find some h such that the ARL will equal any given target greater than 1.

Thus for a given distribution of the process measurement X, choosing k and h implies an ARL, and choosing k and any ARL greater than 1 implies an h. This is true regardless of the distribution of X.

2.6.3 Calculating the k, h, ARL relationship

The CUSUM is designed by choosing k and h values that give some acceptable high target value for the in control ARL starting out from S_0^+. This can be done using tables, graphs or software. The reader is referred to Hawkins (1992) for further information. See also Vance (1986) and Brook and Evans (1972) for various techniques to calculate the ARL of a CUSUM.

A full CUSUM design involves two parts. First there is the problem of selecting a k, h pair that will provide acceptable in control ARL. The second involves assessing how quickly the CUSUM will respond to out of control situations when the parameter value has shifted.

2.7 Optimality of CUSUM based on the k, h, ARL relationship

Suppose we are interested in determining whether a process has shifted from one known distribution to a second known distribution. We consider the first known distribution to be the in control state and the second one to be a known out of control state. We also consider that the shift occurs at some unknown time t.

Our aim is to detect that change as soon as possible and there are various schemes that could spot this change. Some of them are the standard Shewhart type

https://doi.org/10.26219/heal.aueb.6628

control charts, the EWMA or the CUSUM. Among all procedure with the same in control ARL, the optimal procedure will have the smallest expected time until it signals a change, once the process shifts to the out of control state.

Moustakides (1986) proved that, among all tests with the same in control ARL, for any measure (statistic) of the quality characteristic, the CUSUM had the smallest expected out of control run length, thus he proved that the CUSUM scheme was optimal in the sense of the smallest expected time until it signals a change, once the process shifts to the out of control state indeed.

The optimality of the CUSUM is for detecting the shift to a single specific out of control distribution. The CUSUM that is optimal for one particular change is not optimal for detecting a different change. For example if a CUSUM is optimal for detecting a one standard deviation in mean then for detecting a two standard deviation in mean, a different CUSUM would be optimal. That means that the CUSUM is optimal for detecting the specific persistent shift for which it was designed only.

Usually in practice, there is no such thing as a single out of control distribution to which the process could shift. There are various possible shifts. In hypotheses testing terms, the alternative of real interest is not a point hypothesis but an interval hypothesis. We easily overcome that difficulty by selecting a particular out of control state for which we decide the CUSUM to be optimal. Thus we construct a CUSUM that will be the optimal diagnostic only if the process shifts to the particular alternative for which it is designed. Nevertheless due to the robust performance of CUSUM, it will nearly be as well as the optimal CUSUM for all shifts not too far from the particular shift point of the alternative hypothesis. Thus we are able to use the same CUSUM as the optimal scheme to detect a small area of shifts in mean, lying near to that particular value. For further reading in various optimality properties of CUSUM see Ritov (1990), Gan(1991) or Yaschin (1993).

2.8 Sequential Probability Ratio Tests (SPRT)

The SPRT tests a simple null hypothesis (H_0) against simple alternative hypothesis (H_1) with a probability density (or mass) function $f_0(x)$ and $f_1(x)$, respectively, for the measure x_n.

 https://doi.org/10.26219/heal.aueb.6628

Assume that we have a sequence of independent observations $\{x_i\}$ of length n and we wish to decide between H_0 and H_1. In the SPRT, we form the likelihood ratio Λ_n given by:

$$\Lambda_n = \prod_{i=1}^{n} \frac{f_1(x_i)}{f_0(x_i)}. \tag{2.8.1}$$

The test accepts H_0 if Λ_n is less than or equal to some cutoff value A and rejects H_0 in favor of H_1 if Λ_n is bigger than another cutoff constant $B > A$. If $A < \Lambda_n < B$, then the SPRT calls for another observation x_{n+1} and updates the likelihood ratio to incorporate it.

It is easier to work with logarithms of the likelihood ratio

$$\ln \Lambda_n = \sum_{i=1}^{n} \ln\left(\frac{f_1(x_i)}{f_0(x_i)}\right)$$

and we can write $\ln \Lambda_n$ as:

$$\ln \Lambda_n = \sum_{i=1}^{n} z_i\,, \quad z_i = \ln\left(\frac{f_1(x_i)}{f_0(x_i)}\right). \tag{2.8.2}$$

The score variable z_i is just a transformation of the random quantity x_i and it is also a random variable whose distribution can be calculated from that of the X.

Under the assumption that the original x_is are independent these z_is are also independent. The SPRT then:

- accepts H_0 if: $\ln \Lambda_n \leq \ln A$
- accepts H_1 if $\ln \Lambda_n \geq \ln B$
- calls for another observation if $\ln A \leq \ln \Lambda_n \leq \ln B$

The SPRT is optimal for deciding between two point hypotheses in the sense that the expected number of points sampled before a decision is made, is minimized with the SPRT.

In practice, SPRT works with the log-likelihood, or $\ln(\Lambda_n)$, so that the test is based on the cumulative sum of z_i. We decide to reject or not or to continue sampling based on the value of this cumulative sum. If the hypothesis H_1 is true, the z_is will

https://doi.org/10.26219/heal.aueb.6628

have a positive expected value and as points are being added the cumulative sum will tend to drift upward. If the hypothesis H_0 is true, the $z_i s$ will have negative expected value and the sum will tend to drift downward.

The constants A and B are related to Type I and Type II errors so deciding on these two error probabilities will provide the required values for A and B.

2.8.1 Relation of the SPRT to the CUSUM

The one sided CUSUM can be seen as a repeated SPRT. In the CUSUM the hypothesis of in control H_0, is never accepted. We are never content to decide that the process is in control and stop sampling. Instead we restart the test each time the evidence favors the hypothesis that we are in control.

The evidence favors the null (in control) hypothesis whenever the cumulative sum of Z_i is negative. Whenever that happens we start over by resetting the sum to zero, or in algebraic terms:

$$C_n^+ = \max\{0, C_{n-1}^+ + z_n\}$$

$$z_i = \ln\left(\frac{f_1(x_i)}{f_0(x_i)}\right) \Rightarrow C_n^+ = \sum_{i=1}^{n} z_i < 0 \tag{2.8.3}$$

$$\Rightarrow \ln\sum_{i=1}^{n}\left(\frac{f_1(x_i)}{f_0(x_i)}\right) \leq \ln A \Leftrightarrow \quad H_0\text{: accepted.} \tag{2.8.4}$$

In the CUSUM approach we sample until we reject the null hypothesis in favor of the alternative. The score function z_i turns out to be the same function seen in the SPRT, it is a function of x_i and depends on the parameters that the in control and the out of control distributions $f_0(x_i)$ and $f_1(x_i)$ respectively may have.

Next we will examine the relationship between the SPRT and the CUSUM for the distributions that belong to the exponential family starting with the basic definitions.

https://doi.org/10.26219/heal.aueb.6628

2.8.2 The general exponential family

Any member of the exponential family with a single parameter θ can be written as:

$$f(y|\theta) = exp\{a(y)b(\theta) + c(y) + d(\theta)\} \tag{2.8.5}$$

where,

θ: the parameter of the distribution

y: corresponding random variable

$a(y)$: minimal sufficient statistic

$b(\theta)$: natural parameter

$d(\theta)$: function of the parameter θ.

Suppose we want to test whether the process has gone from an in control parameter value θ_0 to an out of control value θ_1. Then the variables z_i simplify to:

$$z_i = \ln\left(\frac{f_1(y_i)}{f_0(y_i)}\right) = \ln\left(\frac{\exp\{a(y_i)b(\theta_1) + c(y_i) + d(\theta_1)\}}{\exp\{a(y_i)b(\theta_0) + c(y_i) + d(\theta_0)\}}\right) =$$

$$= a(y_i)\{b(\theta_1) - b(\theta_0)\} + \{d(\theta_1) - d(\theta_0)\}.$$

It has been shown that a generic decision interval CUSUM with its recursion is:

$$D_n = max(0, D_{n-1} + z_n)$$

which signals if $D_n > \ln A$.

For any member of the exponential family the decision interval CUSUM converts to

$$D_n = max(0, D_{n-1} + a(y_n)\{b(\theta_1) - b(\theta_0)\} + \{d(\theta_1) - d(\theta_0)\}). \tag{2.8.6}$$

We define:

$$X_n = a(y_n)$$

(2.8.7)

 https://doi.org/10.26219/heal.aueb.6628

$$k = -\frac{d(\theta_1) - d(\theta_0)}{b(\theta_1) - b(\theta_0)}.$$

Based on the sign of the difference $b(\theta_1) - b(\theta_0)$, we have the two possible scenarios:

- If $b(\theta_1) - b(\theta_0) > 0$ then we can rescale by diving the CUSUM through by $[b(\theta_1) - b(\theta_0)]$ getting the recursion:

$$\frac{D_n}{b(\theta_1) - b(\theta_0)} > \frac{A}{b(\theta_1) - b(\theta_0)} \leftrightarrow C_n^+ > h^+$$

where:

$$C_n^+ = max(0, C_{n-1}^+ + X_n - k) \quad \text{signals if } C_n^+ > h^+ \tag{2.8.8}$$

$$C_n^+ = \frac{D_n}{b(\theta_1) - b(\theta_0)}$$

$$h^+ = \frac{A}{b(\theta_1) - b(\theta_0)}$$

- If $b(\theta_1) - b(\theta_0) < 0$ then dividing CUSUM through $[b(\theta_1) - b(\theta_0)]$ reverses the direction CUSUM giving the recursion:

$$C_n^- = min(0, C_{n-1}^- + X_n - k) \quad \text{signals if } C_n^- < h^- \tag{2.8.9}$$

where:

$$C_n^- = \frac{D_n}{b(\theta_1) - b(\theta_0)}$$

$$h^- = \frac{A}{b(\theta_1) - b(\theta_0)}$$

the parameters h^+ and h^- are to be specified. It has been proved that CUSUM gives the smallest out of control ARL for all tests with the same in control ARL (Moustakides (1986)). Thus, we specify the desired in control ARL, by setting h.

https://doi.org/10.26219/heal.aueb.6628

2.8.3 Illustration for Normal Data (CUSUM for a normal mean shift)

Given the general formulation for the CUSUM expressed in Equation (2.8.6), we can easily derive the CUSUM for an upward or a downward shift in the mean of a normal distribution with a variance known.

Assume that $X \sim N(\theta, \sigma^2)$. Let σ be fixed and known. Then we can write the normal density in exponential family form as,

$$f(x|\theta) = exp\left(\frac{x\theta}{\sigma^2} - \frac{x^2}{2\sigma^2} - \ln\left(\sigma\sqrt{2\pi}\right) - \frac{\theta^2}{2\sigma^2}\right)$$

from which it follows that

$$a(x) = X_n = x$$

$$b(\theta) = \frac{\theta}{\sigma^2}$$

$$c(x) = \frac{x^2}{2\sigma^2} - \ln\left(\sigma\sqrt{2\pi}\right)$$

$$d(\theta) = -\frac{\theta^2}{2\sigma^2}$$

$$k = -\frac{d(\theta_1) - d(\theta_0)}{b(\theta_1) - b(\theta_0)} = -\frac{(-\theta_1^2 + \theta_0^2)/(2\sigma^2)}{(\theta_1 - \theta_0)/\sigma^2} = \frac{\theta_0 + \theta_1}{2}.$$

Thus, applying equations (2.8.7) and (2.8.8), the following CUSUM scheme for the mean θ_0 of a normal distribution, tuned for the alternative $\theta_1 > \theta_0$, results

$$C_0^+ = 0$$

$$C_n^+ = max(0, C_{n-1}^+ + X_n - k^+)$$

$$k^+ = \frac{\theta_0 + \theta_1}{2}. \tag{2.8.10}$$

This scheme signals if $C_n^+ > h^+$, where h^+ is chosen so the scheme has a specified ARL in control.

https://doi.org/10.26219/heal.aueb.6628

If $\theta_1 < \theta_0$, then because of the division of the standard form by the negative quantity $b(\theta_1) - b(\theta_0)$ the CUSUM is:

$$C_0^- = 0$$

$$C_n^- = min(0, C_{n-1}^- + X_n + k^-)$$

$$k^- = \frac{\theta_0 + \theta_1}{2} \tag{2.8.11}$$

signals if $C_n^- < -h^-$.

https://doi.org/10.26219/heal.aueb.6628

Chapter 3

Bayesian CUSUM

After we have introduced the CUSUM idea as a hypothesis testing problem, in Chapter 2, we will now represent an alternative approach. In Chapter 3 we regard hypothesis testing problem as a decision problem and we aim to construct a Bayesian form of CUSUM. Before that, we represent some basic principles of Bayesian statistics and then, using one of Bayesian hypothesis testing techniques, the Bayes factor, we compute the Bayesian CUSUM and then specify the form we have constructed for any member of the exponential family. Based on Bayes test, we construct Bayesian CUSUM for three different cases. We consider three kinds of hypothesis testing, the problem where both of hypotheses are simple hypotheses, problem where both of hypotheses are interval hypotheses and the one where the null hypothesis is a simple hypothesis but the alternative is an interval hypothesis. Classic CUSUM corresponds only to first case where both of the hypotheses are simple. The Bayes tests we use in order to define Bayesian CUSUM are based on Bayes factor, a useful and trivial tool for Bayesian statistics. West (1986) had purposed that kind of CUSUM providing cumulative Bayes factors. For further reading about various definitions about Bayesian CUSUM see Beibel (1996), Ritov (1990) and also Graves, Bisgaard and Kulachi (2005). Bayesian CUSUM is sufficiently widespread in applications of statistics especially in forecasting and dynamic models (see Harrison & West (1987) and (1999)).

3.1 Introduction to Bayesian approach

The Bayesian approach to statistics is fundamentally different from the known classical approach. In the classical approach, the parameter θ, is considered to be unknown, but fixed quantity. A random sample $x_1, x_2, \dots, x_n$ is drawn from a population indexed by θ, and based on the observed values in the sample knowledge about the value θ is obtained. In the Bayesian approach, θ is considered to be a random quantity whose variation can be described by a probability distribution (called the prior distribution). This is usually a subjective distribution, based on the experimenter's belief, and is formulated before the data are seen (hence the name

https://doi.org/10.26219/heal.aueb.6628

prior distribution). A sample is then taken from a population indexed by θ and the prior distribution is updated with this sample information. The updated prior is called the posterior distribution. This updating is done with the use of Bayes' Rule (3.1) written below, hence the name Bayesian statistics.

***Theorem 3.1*: ** Bayes' Rule

Let $A_1, A_2, \ldots$ be a partition of the sample space, and let B be any set. Then, for each $i = 1,2, \ldots,$

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{j=1}^{\infty} P(B|A_j)P(A_j)} \tag{3.1}$$

Thus, if we denote the prior distribution by $\pi(\theta)$ and the sampling distribution by $f(x|\theta)$, then the posterior distribution, i.e the conditional distribution of θ given the sample, x, is:

$$\pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{\int f(x|\theta)\pi(\theta)\, d\theta}$$

where $f(x|\theta)\pi(\theta) = f(x,\theta)$, and $f_x(x) = \int f(x|\theta)\pi(\theta)\, d\theta$ is the marginal distribution of X known as the normalizing constant of the posterior distribution. As a result Bayes' formula is often written as:

$$\pi(\theta|x) \propto f(x|\theta)\pi(\theta).$$

Notice that the posterior distribution is a conditional distribution, conditional upon observing the sample. The posterior distribution is now used to make statements about θ, which is still considered a random quantity. For instance, the mean of the posterior distribution can be used as a point estimate of θ, (known to be the Bayes estimator when squared error loss is used). We can also construct Credible Intervals for the random variable θ, or we can do hypotheses testing for the unknown parameter θ, based on the posterior distribution of the random variable.

 https://doi.org/10.26219/heal.aueb.6628

3.1.1 Decision Theory

Decision theory describes any situation in which choices are to be made among alternative courses of action with uncertain consequences. A decision problem is determined by three basic elements:

1. A set $A = \{a_i, i \in I\}$ of available actions, one of which is to be selected
2. for each action a_i, a set $\{E_j, j \in J\}$ of uncertain events, describing the uncertain outcomes of taking action a_i
3. corresponding to each set $\{E_j, j \in J\}$, a set of consequences $\{c_j, j \in J\}$.

The idea is as follows. Suppose we choose action a_i, then one and only one of the uncertain events $E_j, (j \in J)$, occurs and leads to the corresponding consequence $c_j, (j \in J)$. Each set of events $E_j, (j \in J)$ is a partition of total set of possibilities.

Decision theory represents one of the most important components of Bayesian inference. Thus, within the Bayesian approach, we consider the inference about a random variable θ as a decision theory problem.

We define a decision problem about an unknown parameter θ as follows:

i. Θ is the parameter space, which contains all possible values of the parameter θ, we are interested of.
ii. A is the action space, which contains all possible actions, which are available for a decision to be made about the parameter θ,
iii. $u(\theta, a): \Theta \times A \rightarrow \Re$ is the utility function, which reflects the utility we get if we choose action a while the real unknown parameter's value is θ.

Definition 3.1: Statistical Decision Problem

The triplet $(\Theta, A, u(\theta, a))$ accompanied with the data $x \sim f(x|\theta)$, where $\theta \in \Theta$, the value of the unknown parameter of interest, define a "statistical decision problem".

It has been proved that if for a problem there is an optimal decision, then Bayesian approach can lead to that optimal solution (DeGroot (1970)).

https://doi.org/10.26219/heal.aueb.6628

The basic hypothesis testing problem usually considered by statisticians may be described as a decision problem with elements:

$$\Omega = \{\omega_0 = [H_0 : \theta \in \Theta_0],\ \omega_1 = [H_1 : \theta \in \Theta_1]\}$$

where $\theta \in \Theta = \Theta_0 \cup \Theta_1$, is the parameter labeling a parametric model, $f(x|\theta)$, and Θ_0, Θ_1 is a partition of the parameter space, $A = \{\alpha_0, \alpha_1\}$ is the action space with α_0 corresponding to rejecting hypothesis H_1, and α_1 corresponding to rejecting H_0. Finally we have the loss function $l(a_i, \omega_j) = -u(a_i, \omega_j) = -u_{ij},\ i, j \in \{0,1\}$ with $-u_{ij}$ reflecting the relative seriousness of the four possible consequences:

$$l(a_0, \omega) = \begin{cases} 0, & \omega = \omega_0 \\ -u_{01} = l_{01}, & \omega = \omega_1 \end{cases} \quad \text{and} \quad l(a_1, \omega) = \begin{cases} -u_{10} = l_{10} & \omega = \omega_0 \\ 0, & \omega = \omega_1 \end{cases}$$

We will have further conversation about Bayesian hypothesis testing and its relevance with decision problem theory later. Before that we have to define some standard hypothesis testing methods in Bayesian inference. For further reading see Jeffreys (1948), Good (1965), Smith (1986).

A basic tool in Bayesian statistical inference overall, is Bayes factor, introduced by Jeffreys (1948).

Definition 3.2: Bayes Factor

Assume that we are interested in testing:

$$\begin{Bmatrix} H_0 : \theta \in \ \Theta_0 \\ H_1 : \theta \in \ \Theta_0^c \end{Bmatrix}$$

If we denote $P(H_i|\underline{x})$, i=0,1 the posterior probabilities of the hypothesis H_i, $i = 0{,}1$ when the data $\underline{x}$ were observed and $P(H_i)$ denote the prior probabilities of H_i, $i = 0{,}1$ then the Bayes factor, B, is defined as the ratio of the posterior odds of H_0 to the prior odds of H_0:

https://doi.org/10.26219/heal.aueb.6628

$$B_{01}(x) = \frac{\frac{P(H_0|x)}{P(H_1|x)}}{\frac{P(H_0)}{P(H_1)}} = \frac{\frac{P(H_0|x)}{1-P(H_0|x)}}{\frac{P(H_0)}{1-P(H_0)}}. \tag{3.1.1}$$

The Bayes factor provides a measure of whether the data $\{x_i\}$ have increased or decreased the odds on H_0 relatively to H_1.

Following Carlin and Louis (1996) we can write:

$$B_{01}(x_i) = \frac{\frac{P(H_0|x_i)}{P(H_1|x_i)}}{\frac{P(H_0)}{P(H_1)}} = \frac{\left[\frac{P(x_i|H_0)P(H_0)}{P(x_i)}\right] \Big/ \left[\frac{P(x_i|H_1)P(H_1)}{P(x_i)}\right]}{\frac{P(H_0)}{P(H_1)}} = \frac{P(x_i|H_0)}{P(x_i|H_1)}, \tag{3.1.2}$$

where

$$P(x_i|H_j) = \int f(x_i|\theta_j)\pi(\theta_j)d\theta_j \quad (\theta_j \text{ indicates } H_j \text{ , j= 0,1}) \tag{3.1.3}$$

$\pi(\theta_j)$= a-priori probability density for the parameter θ under H_j.

Given data $x_1, x_2, \dots, x_n \sim f(x_i|\theta_j)$ i.i.d we calculate the likelihood function

$$p(\underline{x}|\theta_j) = \prod_{i=1}^{n} f(x_i|\theta_j). \tag{3.1.4}$$

Thus, applying Equations (3.1.4) to (3.1.3),

$$P(x_1, .. x_n|H_j) = \int p(\underline{x}|\theta_j)\pi(\theta_i)d\theta_i \quad (\theta_i \text{ indicates } H_i \text{ , } i = 0{,}1). \tag{3.1.5}$$

We consider $P(\underline{x}|H_i) = P_i(\underline{x})$ as the marginal distribution of the model under H_i hypothesis. The set of the alternative models is $H = \{H_0, H_1\}$. Obviously we are dealing with an H-closed case, meaning that one of the two models is true without the explicit knowledge of which of them is the true model.

From this point of view, the overall model specified beliefs for $\underline{x}$ is of the form:

$$P(x) = \sum_{i=0}^{1} P(H_i)\, P(x|H_i)$$

with $P(H_i)$ denoting prior weights on the component models $\{H_i, i = 0,1\}$.

Jeffreys (1948) provided the following table of cutoff values for B, to be used when we decide about rejecting H_0 or not (Table 3.1):

Table 3.1: Jeffreys cutoff values for Bayes factors

$1 < B$	Null hypothesis supported
$10^{-1/2} < B < 1$	Evidence against H_0,but not worth more than a bare mention
$10^{-1} < B < 10^{-1/2}$	Evidence against H_0 substantial
$10^{-3/2} < B < 10^{-1}$	Evidence against H_0 strong
$10^{-2} < B < 10^{-3/2}$	Evidence against H_0 very strong
$B < 10^{-2}$	Evidence against H_0 decisive

(When $B = 1$ we cannot draw any decision for or against H_0)

Thus a common Bayesian hypothesis testing method is to calculate Bayes factor and compare it to the Jeffreys cutoff values to decide whether to reject a hypothesis or not.

3.2 Model comparison as a model decision problem

A hypothesis testing decision problem involves only the choice of a model H_i, without any subsequent action, so that the utility function has the form $u(a_i, w)$ where, w is some unknown hypothesis of interest, and the action a_i corresponds to choosing mode H_i.

From an H-closed perspective, an example of an obvious w of interest might be the H_i for which imagining a large future sample of observations, $x = (x_1 \dots x_s)$, $P(H_i|x) \to 1\ as\ s \to +\infty$. w in this case labels the "true model" and the utility of choosing a particular model then, depends on whether a correct choice has been made.

 https://doi.org/10.26219/heal.aueb.6628

Our main intention is to maximize the expected utility. Maximizing expected utility implies that the optimal model choice a^* is given by:

$$\overline{u}(a^*|x) = \sup_{\mathrm{i}=0,1}\left(\overline{u}(a_i|x)\right) \tag{3.2.1}$$

where:

$$\overline{u}(a_i|x) = \sum_{i\,\in I=\{0,1\}} u(a_i, w)p(w|x) \tag{3.2.2}$$

with $p(w|x)$ representing actual beliefs about w having observed x.

In the H-closed case:

$$p(w|x) = \sum_{i\in I=\{0,1\}} p_i(w|x)P(H_i|x) \tag{3.2.3}$$

where:

$$P(H_i|x) = \frac{P(H_i)P(x|H_i)}{\sum_{j\,\in J=\{0,1\}} P\left(H_j\right)P\left(x\middle|H_j\right)}$$

is the posterior probability given x, of model H_i, being true and

$$p_i(w|x) = p(w|H_i, x)$$

is given by standard (posterior or predictive) calculations conditional on model under H_i, $i \in I = \{0,1\}$.

3.2.1 Zero-one utility function

Consider the point of choosing a model from H, without any subsequent decision, when the "state of the world" of interest is defined to be the true model, H_i, so that assuming a future sample $x = (x_1, \dots, x_s)$, $P(H_i|x) \to 1 \; as \; s \to +\infty$

In this case a usual form of utility function is the 0-1 utility function:

https://doi.org/10.26219/heal.aueb.6628

$$u(a_i, w) = \begin{cases} 1, if\ w = H_i \\ 0, if\ w \neq H_i \end{cases} .$$

The expected utility of the decision a_i (choosing H_i) given x, is given by Equation (3.2.2) We consider:

$$p_i(w|x) = \begin{cases} 1, if\ w = H_i \\ 0, if\ w \neq H_i \end{cases} . \tag{3.2.4}$$

Hence, applying Equations (3.2.3) and (3.2.4) to (3.2.2) we conclude that

$$\overline{u}(a_i|x) = \sum_{i \in I=\{0,1\}} u(a_i, w)p(w|x) = P(H_i|x)\,, \qquad i = 0,1. \tag{3.2.5}$$

The optimal decision is therefore to "choose the model which has the highest posterior probability".

3.2.2 Hypothesis Testing

Two alternative models, H_0 and H_1 are under consideration and both are special cases of the predictive model:

$$p(x) = \int p(x|\theta)dQ(\theta)$$

with the same assumed parametric form $p(x|\theta)\,, \theta \in \Theta$, but with different choices of Q. This corresponds to an assumption of a random sample from the unknown distribution $p(x|\theta)$, together with the prior distribution Q, for $p(x|\theta)$, defined over the space of all distribution functions on $\Re$.

If for model H_i, Q_i assigns probability one to a specific value θ_i, the model is said to reduce to a simple hypothesis for θ (recalling that the form $p(x|\theta)$ is assumed throughout). If for model H_j, Q_j defines a non-degenerate density $P_j(\theta)$ over $\theta_j \subseteq \Theta$, the model is said to reduce to a composite hypothesis for θ.

https://doi.org/10.26219/heal.aueb.6628

If a simple hypothesis is being compared with a composite hypothesis so that $\Theta_j = \Theta - \{\theta_i\}$, the latter is called a general alternative hypothesis.

In the situation where the "state of the world" of interest w, is defined to be the true model H_i, we can generalize slightly the zero-one utility structure used earlier by assuming that:

$$u(a_i|w) = -l_{ij}, w = H_j, i, j = 0,1$$

with $l_{00} = l_{11} = 0$ and $l_{10,}\, l_{01} > 0$, known as generalized zero-one utility function (Berger (1985)).

In this utility function $l_{10,}$ is the cost of type I error (the error of falsely rejecting H_0) and l_{01} is the cost of type II error (the error of falsely accepting H_0). As we will see below, in comparing tests all that really matters is the ratio $l_{01}/\, l_{10}$ and not the two individual numbers (Casella and Berger (1990)).

Intuitively there is a loss in choosing the wrong model (possibly asymmetric), and there is no loss in choosing the correct model.

Given data x, and using again:

$$p_i(w|x) = \begin{cases} 1, if\ w = H_i \\ 0, if\ w \neq H_i \end{cases}$$

the expected utility of a_i, it is then easily proved to be:

$$\overline{u}(a_i|x) = -[l_{i0}P(H_0|x) + l_{i1}P(H_1|x)]$$

and then we have:

$$\overline{u}(a_0|x) < \overline{u}(a_1|x) \Leftrightarrow l_{01}P(H_1|x) > l_{10}P(H_0|x).$$

Thus we prefer H_1 to H_0 if and only if:

$$\frac{P(H_0|x)}{P(H_1|x)} < \frac{l_{01}}{l_{10}} \qquad (3.2.6)$$

https://doi.org/10.26219/heal.aueb.6628

revealing a balance of the posterior odds against the relative seriousness of the two possible ways of selecting the wrong model.

In the symmetric case $l_{01} = l_{10}$ the choice reduces to choosing the a-posteriori most likely model as shown earlier for the zero-one case.

Applying Equations (3.1.1) and (3.2.6) we conclude that:

$$B_{01}(\underline{x}) < \frac{l_{01}}{l_{10}} \frac{P(H_1)}{P(H_0)} \ .$$

Thus, from Bayes theory we have the following proposition with the use of Bayes factor (Bernardo and Smith (1994))

Proposition 3.1: *Bayes Tests*

Assume that we are interested in comparing two models which are defined by parametric hypotheses for $P(\underline{x}|\theta)$ with utility structure:

$u(m_i|w)$, where $w = H_j$, m_i =model choice $\quad i, j = 0,1$

with $l_{00} = l_{11} = 0$ and $l_{10,}\ l_{01} > 0$. Thus, model under H_1 is preferred to the model under $H_{0,}$ if and only if:

$$B_{01}(\underline{x}) < \frac{l_{01}}{l_{10}} \frac{P(H_1)}{P(H_0)}$$

where for:

- Simple versus simple test

$$\begin{Bmatrix} H_0 \colon \theta = \theta_0 \\ H_1 \colon \theta = \theta_1 \end{Bmatrix}$$

$$B_{01}(\underline{x}) = \frac{P(\underline{x}|\theta_0)}{P(\underline{x}|\theta_1)} < \frac{l_{01}}{l_{10}} \frac{P(H_1)}{P(H_0)} \tag{3.2.7}$$

 https://doi.org/10.26219/heal.aueb.6628

- Simple versus composite test

$$\begin{cases} H_0: \theta = \theta_0, \Theta_0 = \{\theta_0\} & \text{with a-priori probability density } \pi_0(\theta) \\ H_1: \theta \neq \theta_0, \theta \in \Theta_1 = \Theta_0^c & \text{with a-priori probability density } \pi_1(\theta) \end{cases}$$

$$B_{01}(x) = \frac{P(\underline{x}|\theta_0)}{\int_{\theta \neq \theta_0} P(\underline{x}|\theta)\pi_0(\theta)\, d\theta} < \frac{l_{01}}{l_{10}} \frac{P(H_1)}{P(H_0)} \tag{3.2.8}$$

- Composite versus composite test

$$\begin{cases} H_0: \theta \in \Theta_0 & \text{with a-priori probability density } \pi_0(\theta) \\ H_1: \theta \in \Theta_1 = \Theta_0^c & \text{with a-priori probability density } \pi_1(\theta) \end{cases}$$

$$B_{01}(\underline{x}) = \frac{\int_{\Theta_0} P(\underline{x}|\theta)\pi_0(\theta)\, d\theta}{\int_{\Theta_1} P(\underline{x}|\theta)\pi_1(\theta)\, d\theta} < \frac{l_{01}}{l_{10}} \frac{P(H_1)}{P(H_0)} \tag{3.2.9}$$

3.3 Bayesian CUSUM

3.3.1 Simple versus Simple

Suppose we have to deal with the decision problem whether a process has two possible values θ_0 or $\theta_1$with the former referring to the in control situation and the latter to the out of control situation.

We are interested in testing:

$$\begin{Bmatrix} H_0: \theta = \theta_0 \\ H_1: \theta = \theta_1 \end{Bmatrix}$$

where the parameter space Θ is partitioned into $\Theta_0 = \{\theta_0\}$, the in control situation and $\Theta_0^c = \theta_1$ the out of control situation with an a-priori probability density:

https://doi.org/10.26219/heal.aueb.6628

$$\pi(\theta) = \begin{cases} P(H_0) = \pi(\theta_0) = \pi_0 \\ P(H_1) = \pi(\theta_1) = \pi_1 = 1 - \pi_0 \end{cases}.$$

The action space has two elements: $A = \{\alpha_0 = \text{accept } H_0, \alpha_1 = \text{reject } H_0\}$

We define a utility function as:

$$u(\alpha_0, \theta) = \begin{cases} 0\,, & \theta \in \Theta_0 \\ -l_{01}, & \theta \in \Theta_1 \end{cases} \quad and \quad u(\alpha_1, \theta) = \begin{cases} -l_{10}, & \theta \in \Theta_0 \\ 0\,, & \theta \in \Theta_1 \end{cases}$$

According to Bayes test the model under H_1 is preferable to model under H_0 if and only if:

$$B_{01}(\underline{x}) = \frac{P(\underline{x}|\theta_0)}{P(\underline{x}|\theta_1)} = \prod_{i=1}^{n} \frac{f(x_i|\theta_0)}{f(x_i|\theta_1)} < \frac{l_{01}}{l_{10}} \frac{P(H_1)}{P(H_0)}\ .$$

It is easier to work with logarithms of the Bayes factor

$$\ln B_{01}(\underline{x}) = \ln\left(\prod_{i=1}^{n} \frac{f(x_i|\theta_0)}{f(x_i|\theta_1)}\right) = \sum_{i=1}^{n} \ln\left(\frac{f(x_i|\theta_0)}{f(x_i|\theta_1)}\right). \tag{3.3.1}$$

We set:

$$L_i = \ln\left(\frac{f(x_i|\theta_0)}{f(x_i|\theta_1)}\right).$$

Then (3.3.1) becomes:

$$\ln B_{01}(\underline{x}) = \sum_{i=1}^{n} \ln\left(\frac{f(x_i|\theta_0)}{f(x_i|\theta_1)}\right) = \sum_{i=1}^{n} L_i.$$

Thus if

$$\sum_{i=1}^{n} L_i = \ln\left(\frac{f(x_1|\theta_0)}{f(x_1|\theta_1)}\right) + \cdots + \ln\left(\frac{f(x_n|\theta_0)}{f(x_n|\theta_1)}\right) < \ln\left(\frac{l_{01}}{l_{10}} \frac{P(H_1)}{P(H_0)}\right) \tag{3.3.2}$$

it is evidence in favor of H_1.

 https://doi.org/10.26219/heal.aueb.6628

We can construct a "Decision Interval" form, based on the cumulative sum of the quantities L_i, by using the logarithm of the Bayes factor.

According to the Jerrreys cutoff values for Bayes factor shown on Table 3.1, if:

$$1 < B \Leftrightarrow \ln(1) < \ln B \Leftrightarrow 0 < \ln B \tag{3.3.3}$$

the null hypothesis is definitely supported. The null hypothesis may be supported for smaller values of Bayes factor as well such as $10^{-1/2} < B < 1$, because the evidence against H_0 are not enough to reject it. Nevertheless, we cannot be confident that for any $10^{-1/2} < B < 1$ the null hypothesis is not supported .

Due to the fact that in the CUSUM, the null hypothesis is never accepted (we either reject or fail to reject H_0), we turn out to restart the test each time the evidence favors the in control hypothesis.

As we can notice from Equation (3.3.3), whenever the cumulative sum of L_i is positive it is a sign that we certainly are in control. Thus, every time that happens, we start over by resetting the CUSUM to zero.

Thus, based on (3.3.2) and (3.3.3) we define the quantities:

$$W_n = \min(0, W_{n-1} + L_n)$$

where:

$$W_0 = 0$$

$$W_1 = \min(0, W_0 + L_1) = \min\left(0, \ln\left(\frac{f(x_1|\theta_0)}{f(x_1|\theta_1)}\right)\right)$$

$$\vdots$$

$$W_n = \min(0, W_{n-1} + L_n) = \min\left(0, W_{n-1} + \ln\left(\frac{f(x_n|\theta_0)}{f(x_n|\theta_1)}\right)\right). \tag{3.3.4}$$

Obviously the Decision Interval form of the CUSUM signals if and only if:

$$W_n < \text{cutoff value}$$

Applying Proposition 3.1 for logarithm of Bayes factor, we get:

https://doi.org/10.26219/heal.aueb.6628

$$\ln B_{01}(\underline{x}) < \ln\left(\frac{l_{01}}{l_{10}}\frac{P(H_1)}{P(H_0)}\right).$$

As we can notice from Equation (3.3.4),

$$W_n \leq \ln B_{01}(\underline{x}).$$

Thus, CUSUM signals if and only if:

$$W_n < \text{cutoff value} \leq \ln\left(\frac{l_{01}}{l_{10}}\frac{P(H_1)}{P(H_0)}\right).$$

Cutoff point values will be decided by simulations and they will depend on the in control and out-of control ARLs. They will also depend on the distance between the in control parameter value θ_0 and the out of control parameter value θ_1, $(\theta_1 - \theta_0)$, just like the h does in the classic CUSUM. Given the general formulation of the CUSUM expressed in Equation (3.3.4), we will construct the Bayesian CUSUM for any member of the exponential family.

3.3.1.1 The general exponential family

The probability density for any member of the exponential family with a single parameter θ can be written as:

$$f(x|\theta) = \exp\{a(x)b(\theta) + c(x) + d(\theta)\}.$$

Then the quantities $L_i = \ln\left(\frac{f(x_i|\theta_0)}{f(x_i|\theta_1)}\right)$ simplify to

$$L_i = \ln\left(\frac{f(x_i|\theta_0)}{f(x_i|\theta_1)}\right) = \ln\left(\frac{\exp\{a(x_i)b(\theta_0) + c(x_i) + d(\theta_0)\}}{\exp\{a(x_i)b(\theta_1) + c(x_i) + d(\theta_1)\}}\right)$$

$$= \alpha(x_i)\{b(\theta_0) - b(\theta_1)\} + \{d(\theta_0) - d(\theta_1)\}. \quad (3.3.5)$$

Thus the decision interval CUSUM becomes:

$$W_n = \min(0, W_{n-1} + L_n) = \min\left(0, W_{n-1} + \ln\left(\frac{f(x_n|\theta_0)}{f(x_n|\theta_1)}\right)\right)$$

https://doi.org/10.26219/heal.aueb.6628

which signals if $W_n <$ cutoff value and for the general exponential family will take the form:

$$W_n = \min(0, W_{n-1} + \alpha(x_i)\{b(\theta_0) - b(\theta_1)\} + \{d(\theta_0) - d(\theta_1)\}). \tag{3.3.6}$$

As we can notice W_n depends on the quantities $\{b(\theta_0) - b(\theta_1)\}$ and $\{d(\theta_0) - d(\theta_1)\}$ thus, the cutoff value, for which the DI CUSUM signals, will depend on the distance between the two parameter values θ_0 and θ_1 and the functions $b(\theta)$ and $d(\theta)$. For further information see Harrison and Veerapen (1994))

3.3.2 Composite versus Composite

Suppose that we want to test whether the parameter θ has gone from an in control parameter value θ_0 to an out of control parameter value θ_1.

$$\begin{Bmatrix} H_0: \theta = \theta_0 \in \Theta_0 = [a, b], \theta \sim \pi_0(\theta) \\ H_1: \theta = \theta_1 \in \Theta_1 = [c, d], \theta \sim \pi_1(\theta) \end{Bmatrix}$$

where the parameter space Θ is partitioned into $\Theta_0 = [a, b]$, the in control situation and $\Theta_1 = [c, d]$ the out of control situation. Note that Θ_1 is not necessarily equals to Θ_0^c. Assume that prior distributions are the same, but they are defined for different intervals for the two hypotheses, and they are conjugate to the likelihood function $f(\underline{x}|\theta)$, such that $\pi_0(\theta) = \pi_1(\theta) = \pi(\theta)$, (they are the same distributions defined for different intervals).

The action space has two elements: $A = \{\alpha_0 = \text{accept } H_0, \alpha_1 = \text{reject } H_0\}$

We define the utility function as:

$$u(\alpha_0, \theta) = \begin{Bmatrix} 0, & \theta \in \Theta_0 \\ -l_{01}, & \theta \in \Theta_1 \end{Bmatrix} \quad and \quad u(\alpha_1, \theta) = \begin{Bmatrix} -l_{10}, & \theta \in \Theta_0 \\ 0, & \theta \in \Theta_1 \end{Bmatrix}.$$

According to Bayes Test Proposition and Equation (3.2.9) the Bayes factor for composite versus composite case is:

https://doi.org/10.26219/heal.aueb.6628

$$B_{01}(\underline{x}) = \frac{\int_a^b f(\underline{x}|\theta)\pi_0(\theta)d\theta}{\int_c^d f(\underline{x}|\theta)\pi_1(\theta)d\theta} = \frac{\int_a^b f(\underline{x}|\theta)\pi(\theta)d\theta}{\int_c^d f(\underline{x}|\theta)\pi(\theta)d\theta} .$$

It is known that the posterior probability of $\theta|\underline{x} \sim p(\theta|\underline{x})$ is given by the formula:

$$p(\theta|\underline{x}) = \frac{f(\underline{x}|\theta)\pi(\theta)}{\int_{-\infty}^{+\infty} f(\underline{x}|\theta)\pi(\theta)} = \frac{f(\underline{x}|\theta)\pi(\theta)}{f_x(\underline{x})} . \tag{3.3.7}$$

As we can notice,

$$f(\underline{x}|\theta)\pi(\theta) = p(\theta|\underline{x})f_x(\underline{x}).$$

Thus,

$$B_{01}(\underline{x}) = \frac{\int_a^b f(\underline{x}|\theta)\pi(\theta)d\theta}{\int_c^d f(\underline{x}|\theta)\pi(\theta)d\theta} = \frac{\int_a^b p(\theta|\underline{x})f_x(\underline{x})d\theta}{\int_c^d p(\theta|\underline{x})f_x(\underline{x})d\theta} =$$

$$= \frac{f_x(\underline{x}) \int_a^b p(\theta|\underline{x})d\theta}{f_x(\underline{x}) \int_c^d p(\theta|\underline{x})d\theta} = \frac{\int_a^b p(\theta|\underline{x})d\theta}{\int_c^d p(\theta|\underline{x})d\theta} . \tag{3.3.8}$$

It is easier to work with the $\ln\left(B_{01}(\underline{x})\right)$ thus,

$$\ln\left(B_{01}(\underline{x})\right) = \ln\left(\frac{\int_a^b p(\theta|\underline{x})d\theta}{\int_c^d p(\theta|\underline{x})d\theta}\right) \Rightarrow$$

$$\ln\left(B_{01}(\underline{x})\right) = \ln\left(\int_a^b p(\theta|\underline{x})d\theta\right) - \ln\left(\int_c^d p(\theta|\underline{x})d\theta\right). \tag{3.3.9}$$

According to Jeffreys' cutoff values, if $1 < B$ then the null hypothesis is supported. Thus if $0 < \ln B$ the process is in control and the parameter value: $\theta \in \Theta_0 = [a, b]$.

Hence, if

https://doi.org/10.26219/heal.aueb.6628

$$\ln\left(B_{01}(\underline{x})\right) = \ln\left(\int_a^b p(\theta|\underline{x})d\theta\right) - \ln\left(\int_c^d p(\theta|\underline{x})d\theta\right) > 0 \tag{3.3.10}$$

the process is in control.

Applying equations (3.3.9) and (3.3.10) we construct Bayesian CUSUM for Composite versus Composite case

$$W_0 = 0$$

$$W_1 = min\left[0, \ln\left(\int_a^b p(\theta|x_1)d\theta\right) - \ln\left(\int_c^d p(\theta|x_1)d\theta\right)\right]$$

$$\vdots$$

$$W_n = min\left[0, \ln\left(\int_a^b p(\theta|x_1, x_2, \ldots, x_n)d\theta\right) - \ln\left(\int_c^d p(\theta|x_1, x_2, \ldots, x_n)d\theta\right)\right] \tag{3.3.11}$$

signals if and only if:

$$W_n < \text{cutoff value} .$$

Note that in the Composite versus Composite case we are not able to define the best in control ARL based on a specified (h, k). Hence we will use the same cutoff values which have been defined by simulation for simple versus simple hypothesis.

3.3.3 Simple versus Composite

Simple versus Composite case could be considered as a special case of Composite versus Composite case, with the in control distribution being a single distribution for a constant known in control parameter value $\theta = \theta_0$ (and the Simple versus Simple is also a special case of the Composite versus Composite).

Suppose we want to test whether the parameter θ has gone from an in control parameter value θ_0 to an out of control value.

https://doi.org/10.26219/heal.aueb.6628

$$\begin{Bmatrix} H_0: \theta = \theta_0 , P(H_0) = p \\ H_1: \theta \neq \theta_0, P(H_1) = 1 - p \end{Bmatrix}$$

where the parameter space Θ is partitioned into $\Theta_0 = \{\theta_0\}$, the in control situation and $\Theta_0^c = \Theta_1 = \{\theta : \theta \neq \theta_0\}$, the out of control situation with an a-priori probability density:

$$\pi(\theta) = \begin{cases} \pi_0(\theta) = \delta_{\theta_0}(\theta), & w.p\ \ p \\ \pi_1(\theta) = \pi_1 = \dfrac{\pi(\theta)}{\pi(\theta \neq \theta_0)}, & w.p\ \ (1-p) \end{cases}$$

$\delta_{\theta_0}(\theta)$ is point mass at θ_0 and π_1 is truncated at θ_0 with $\pi(\theta)$ conjugate to $f(\underline{x}|\theta)$.

The above prior distribution of θ could be written as:

$$\pi(\theta) = p\delta_{\theta_0}(\theta) + (1-p)\pi(\theta)I_{\theta \neq \theta_0}(\theta) \tag{3.3.12}$$

where:

$$I_{\theta \neq \theta_0}(\theta) = \begin{cases} 1, \theta \neq \theta_0 \\ 0, \theta = \theta_0 \end{cases}.$$

Using the above prior (3.3.12) and applying the formula (3.3.7) for the posterior distribution, we get:

$$p(\theta|\underline{x}) = \frac{pf(\underline{x}|\theta)\delta_{\theta_0}(\theta) + (1-p)f(\underline{x}|\theta)\pi(\theta)I_{\theta \neq \theta_0}(\theta)}{\int [pf(\underline{x}|\theta)\delta_{\theta_0}(\theta) + (1-p)f(\underline{x}|\theta)\pi(\theta)I_{\theta \neq \theta_0}(\theta)]d\theta}$$

Thus, we get the respective posterior distributions for each one of the two models under the two alternative hypotheses:

If $\theta = \theta_0$, $$p(\theta_0|\underline{x}) = \frac{pf(\underline{x}|\theta_0)}{pf(\underline{x}|\theta_0) + (1-p)\int_{\theta \neq \theta_0} f(\underline{x}|\theta)\pi(\theta)\,d\theta} \tag{3.3.13}$$

If $\theta \neq \theta_0$, $$p(\theta|\underline{x}) = \frac{(1-p)f(\underline{x}|\theta)\pi(\theta)}{pf(\underline{x}|\theta_0) + (1-p)\int_{\theta \neq \theta_0} f(\underline{x}|\theta)\pi(\theta)\,d\theta} \tag{3.3.14}$$

 https://doi.org/10.26219/heal.aueb.6628

Applying Equations (3.3.13) and (3.3.14) to definition 3.2 for Bayes factor we get:

$$B_{01}(x)=\frac{\frac{P(H_0|x)}{P(H_1|x)}}{\frac{P(H_0)}{P(H_1)}}=\frac{\frac{p(\theta_0|\underline{x})}{\int_{\theta\neq\theta_0}p(\theta|\underline{x})}}{\frac{p}{(1-p)}}=$$

$$=\frac{\frac{pf(\underline{x}|\theta_0)}{pf(\underline{x}|\theta_0)+(1-p)\int_{\theta\neq\theta_0}f(\underline{x}|\theta)\pi(\theta)\,d\theta}}{\frac{\int_{\theta\neq\theta_0}\left(\frac{(1-p)f(\underline{x}|\theta)\pi(\theta)}{pf(\underline{x}|\theta_0)+(1-p)\int_{\theta\neq\theta_0}f(\underline{x}|\theta)\pi(\theta)\,d\theta}\right)d\theta}{\frac{p}{(1-p)}}}=$$

$$=\frac{\frac{pf(\underline{x}|\theta_0)}{(1-p)\int_{\theta\neq\theta_0}f(\underline{x}|\theta)\pi(\theta)d\theta}}{\frac{p}{(1-p)}}=\frac{f(\underline{x}|\theta_0)}{\int_{\theta\neq\theta_0}f(\underline{x}|\theta)\pi(\theta)d\theta}.$$

As we can notice we have resulted to the known Bayes test for Simple versus Composite case (Proposition 3.1).

We have already set the action space A in previous hypothesis testing. A has two elements: $A=\{\alpha_0 = \text{accept } H_0, \alpha_1 = \text{reject } H_0\}$ We have also defined the utility function as:

$$u(\alpha_0,\theta)=\begin{Bmatrix}0\,, & \theta\in\Theta_0\\ -l_{01}, & \theta\in\Theta_1\end{Bmatrix}\ \ and\ \ u(\alpha_1,\theta)=\begin{Bmatrix}-l_{10}, & \theta\in\Theta_0\\ 0\,, & \theta\in\Theta_1\end{Bmatrix}.$$

According to Bayes tests and Equation (3.2.8), the model under H_1 is preferable to model under H_0 if and only if:

$$B_{01}(x)=\frac{f(\underline{x}|\theta_0)}{\int_{\theta\neq\theta_0}f(\underline{x}|\theta)\pi(\theta)\,d\theta}<\frac{l_{01}}{l_{10}}\frac{P(H_1)}{P(H_0)}.$$

As we have already shown, the posterior probability of $(\theta\neq\theta_0)|\underline{x}\sim p(\theta\neq\theta_0|\underline{x})$ for the simple versus composite case is given by the formula:

$$p(\theta|\underline{x})=\frac{(1-p)f(\underline{x}|\theta)\pi(\theta)}{pf(\underline{x}|\theta_0)+(1-p)\int_{\theta\neq\theta_0}f(\underline{x}|\theta)\pi(\theta)\,d\theta}.$$

From now on we refer to $\theta\neq\theta_0$.

https://doi.org/10.26219/heal.aueb.6628

We set:

$$f_x(\underline{x}) = pf(\underline{x}|\theta_0) + (1-p)\int_{\theta \neq \theta_0} f(\underline{x}|\theta)\pi(\theta)\, d\theta.$$

Thus,

$$p(\theta|\underline{x}) = \frac{(1-p)f(\underline{x}|\theta)\pi(\theta)}{f_x(\underline{x})} \Rightarrow$$

$$f(\underline{x}|\theta)\pi(\theta) = \frac{p(\theta|\underline{x})f_x(\underline{x})}{(1-p)}.$$

Thus,

$$B_{01}(\underline{x}) = \frac{f(\underline{x}|\theta_0)}{\int_{\theta \neq \theta_0} f(\underline{x}|\theta)\pi(\theta)\, d\theta} = \frac{f(\underline{x}|\theta_0)}{\int_{\theta \neq \theta_0} \frac{p(\theta|\underline{x})f_x(\underline{x})}{(1-p)}\, d\theta} =$$

$$= \frac{f(\underline{x}|\theta_0)}{\frac{f_x(\underline{x})}{(1-p)}\int_{\theta \neq \theta_0} p(\theta|\underline{x})\, d\theta}. \qquad (3.3.15)$$

It is easier to work with the $\ln\left(B_{01}(\underline{x})\right)$ thus,

$$\ln\left(B_{01}(\underline{x})\right) = \ln\left(\frac{f(\underline{x}|\theta_0)}{\frac{f_x(\underline{x})}{(1-p)}\int_{\theta \neq \theta_0} p(\theta|\underline{x})\, d\theta}\right) \Rightarrow$$

$$\ln\left(B_{01}(\underline{x})\right) = \ln\left(f(\underline{x}|\theta_0)\right) - \ln\left(\frac{f_x(\underline{x})}{(1-p)}\int_{\theta \neq \theta_0} p(\theta|\underline{x})\, d\theta\right) \Rightarrow$$

$$\ln\left(B_{01}(\underline{x})\right) = \ln\left(f(\underline{x}|\theta_0)\right) - \left[\ln\left(\frac{f_x(\underline{x})}{(1-p)}\right) + \ln\left(\int_{\theta \neq \theta_0} p(\theta|\underline{x})\, d\theta\right)\right]. \qquad (3.3.16)$$

According to Jeffreys' cutoff values, if $1 < B$ then the null hypothesis is supported. Thus if $0 < \ln B$ the process is in control and the parameter value: $\theta \in \Theta_1 = \{\theta : \theta \neq \theta_0\}$.

https://doi.org/10.26219/heal.aueb.6628

Hence, if

$$\ln\left(B_{01}\left(\underline{x}\right)\right) = \ln\left(f\left(\underline{x}\middle|\theta_0\right)\right) - \left[\ln\left(\frac{f_x\left(\underline{x}\right)}{(1-p)}\right) + \ln\left(\int_{\theta\neq\theta_0} p\left(\theta\middle|\underline{x}\right) d\theta\right)\right] > 0 \quad (3.3.17)$$

the process is in control.

Applying Equations (3.3.16) and (3.3.17) we construct Bayesian CUSUM for Simple versus Composite case

$$W_0 = 0$$

$$W_1 = min\left[0, \ln\left(f(x_1|\theta_0)\right) - \left[\ln\left(\frac{f_x(x_1)}{(1-p)}\right) + \ln\left(\int_{\theta\neq\theta_0} p(\theta|x_1)\, d\theta\right)\right]\right]$$

$$\vdots$$

$$W_n = min\left[0, \ln\left(f\left(\underline{x}\middle|\theta_0\right)\right) - \left[\ln\left(\frac{f_x\left(\underline{x}\right)}{(1-p)}\right) + \ln\left(\int_{\theta\neq\theta_0} p\left(\theta\middle|\underline{x}\right) d\theta\right)\right]\right] \quad (3.3.18)$$

signals if and only if:

$$W_n < \text{cutoff value.}$$

Based on Equation (3.3.18), we will provide Bayesian CUSUM for any member of the exponential family.

3.3.3.1 The general exponential family

For any member of the exponential family, $f\left(\underline{x}\middle|\theta\right)$, the conjugate prior π(θ) is also a member of the exponential family.

$$f(x_i|\theta) = exp\{a(x_i)b(\theta) + c(x_i) + d(\theta)\}$$

$$f\left(\underline{x}\middle|\theta\right) = \prod_{i=1}^{n} exp\{a(x_i)b(\theta) + c(x_i) + d(\theta)\}$$

https://doi.org/10.26219/heal.aueb.6628

$$f(\underline{x}|\theta) = \exp\left(b(\theta)\sum_{i=1}^{n} a(x_i) + \sum_{i=1}^{n} c(x_i) + nd(\theta)\right). \tag{3.3.19}$$

The a-priori distribution of the parameter θ is also a member of the exponential family and it is conjugated with the likelihood function of the random variables x_i.

$$\pi(\theta) = \exp(a'(\theta)b'(q) + c'(\theta) + d'(q))$$

$$f(\underline{x}|\theta)\pi(\theta) = exp\left(\left[b(\theta)\sum_{i=1}^{n} a(x_i) + a'(\theta)b'(q)\right] + nd(\theta) + c'(\theta) + [\sum_{i=1}^{n} c(x_i) + d'(q)]\right).$$

As we have defined in the previous section:

$$f_x(\underline{x}) = pf(\underline{x}|\theta_0) + (1-p)\ \int_{\theta\neq\theta_0} f(\underline{x}|\theta)\pi(\theta)\, d\theta. \tag{3.3.20}$$

Because of the fact that $\theta = \theta_0$ has a point mass, for continuous θ_0 , we consider the integration region $\theta \neq= \Theta - \{\theta_0\}$, as Θ, thus we integrate in the entire domain of the parameter θ.

Thus, we consider

$$\int_{\theta\neq\theta_0} f(\underline{x}|\theta)\pi(\theta)\, d\theta \ \text{ as } \ \int_{\Theta} f(\underline{x}|\theta)\pi(\theta)\, d\theta.$$

Hence,

$$f_x(\underline{x}) = pf(\underline{x}|\theta_0) + (1-p)\ \int_{\Theta} f(\underline{x}|\theta)\pi(\theta)\, d\theta \tag{3.3.21}$$

where:

$$\int_{\Theta} f(\underline{x}|\theta)\pi(\theta)\, d\theta = \int_{-\infty}^{+\infty} f(\underline{x}|\theta)\pi(\theta)\, d\theta =$$

$$\int exp\left(\left[b(\theta)\sum_{i=1}^{n} a(x_i) + a'(\theta)b'(q)\right] + nd(\theta) + c'(\theta) + [\sum_{i=1}^{n} c(x_i) + d'(q)]\right) d\theta =$$

$$\exp\left(\sum_{i=1}^{n} c(x_i) + d'(q)\right)\int exp\left(\left[b(\theta)\sum_{i=1}^{n} a(x_i) + a'(\theta)b'(q)\right] + \left[nd(\theta) + c'(\theta)\right]\right) d\theta$$

https://doi.org/10.26219/heal.aueb.6628

where the quantity $\int exp([b(\theta)\sum_{i=1}^{n} a(x_i) + a^{'}(\theta)b^{'}(q)] + [nd(\theta) + c^{'}(\theta)])\, d\theta$ is the normalizing constant of the posterior probability function.

We write the normalizing constant as:

$$\int exp\left(\left[b(\theta)\sum_{i=1}^{n} a(x_i) + a^{'}(\theta)b^{'}(q)\right] + [nd(\theta) + c^{'}(\theta)]\right) d\theta = C. \qquad (3.3.22)$$

Thus, applying Equation (3.3.22)

$$\int_{\Theta} f(\underline{x}|\theta)\pi(\theta)\, d\theta = C\exp\left(\sum_{i=1}^{n} c(x_i) + d^{'}(q)\right). \qquad (3.3.23)$$

Hence, applying (3.3.19) for $\theta = \theta_0$ and (3.3.23) to (3.3.21) we have:

$$f_x(\underline{x}) = pexp\left\{b(\theta_0)\sum_{i=1}^{n} a(x_i) + \sum_{i=1}^{n} c(x_i) + nd(\theta_0)\right\} +$$

$$+(1-p)Cexp\left\{\sum_{i=1}^{n} c(x_i) + d^{'}(q)\right\}.$$

If we work with the $\log B_{01}(x)$, using Equation (3.3.17) we get:

$$\ln\left(B_{01}(\underline{x})\right) = \ln\left(f(\underline{x}|\theta_0)\right) - \left[\ln\left(\frac{f_x(\underline{x})}{(1-p)}\right) + \ln\left(\int_{\theta\neq\theta_0} p(\theta|\underline{x})\, d\theta\right)\right] =$$

$$\ln\left[\prod_{i=1}^{n} f(x_i|\theta_0)\right] - \left[\ln\left(\frac{f_x(\underline{x})}{(1-p)}\right) + \ln\left(\int_{\theta\neq\theta_0} p(\theta|\underline{x})\, d\theta\right)\right] \Rightarrow$$

$$\ln\left(B_{01}(\underline{x})\right) = \sum_{i=1}^{n} \ln(f(x_i|\theta_0)) - \left[\ln\left(\frac{f_x(\underline{x})}{(1-p)}\right) + \ln\left(\int_{\theta\neq\theta_0} p(\theta|\underline{x})\, d\theta\right)\right]. \qquad (3.3.24)$$

Applying Equations (3.3.19) for $\theta = \theta_0$, and (3.3.23) to (3.3.24) we get:

https://doi.org/10.26219/heal.aueb.6628

$$\ln\left(B_{01}\left(\underline{x}\right)\right) = b(\theta_0)\sum_{i=1}^{n} a(x_i) + \sum_{i=1}^{n} c(x_i) + nd(\theta_0)$$

$$- \left[\ln\left(p \times exp\left\{b(\theta_0)\sum_{i=1}^{n} a(x_i) + \sum_{i=1}^{n} c(x_i) + nd(\theta_0)\right\} + (1-p)\ C\right.\right.$$

$$\left.\left. \times\ exp\left\{\sum_{i=1}^{n} c(x_i) + d^{'}(q)\right\}\right) - \ln(1-p) - \ln\left(\int_{\theta\neq\theta_0} p\left(\theta\middle|\underline{x}\right) d\theta\right)\right].$$

Thus,

$$\ln\left(B_{01}\left(\underline{x}\right)\right) = b(\theta_0)\sum_{i=1}^{n} a(x_i) + \sum_{i=1}^{n} c(x_i) + nd(\theta_0) -$$

$$-\ln\left(\frac{f_x\left(\underline{x}\right)}{1-p}\right) - \ln\left[\int_{\theta\neq\theta_0} p\left(\theta\middle|\underline{x}\right) d\theta\right].$$

Applying to Equation 3.3.18 we construct Bayesian CUSUM for Simple versus Composite case:

$$W_0 = 0$$

$$W_1 = min\left[0, b(\theta_0)a(x_1) + c(x_1) + d(\theta_0) - \ln\left(\frac{f_x\left(\underline{x_1}\right)}{1-p}\right) - \ln\left[\int_{\theta\neq\theta_0} p(\theta|x_1)\, d\theta\right]\right]$$

$$\vdots$$

$$W_n = min\left[0, b(\theta_0)\sum_{i=1}^{n} a(x_i) + \sum_{i=1}^{n} c(x_i) + nd(\theta_0) - \ln\left(\frac{f_x\left(\underline{x}\right)}{1-p}\right) - \ln\left[\int_{\theta\neq\theta_0} p\left(\theta\middle|\underline{x}\right) d\theta\right]\right]$$

signals if and only if:

$$W_n < \text{cutoff value.}$$

Similarly with the Composite versus Composite case we are not able to define cutoff point values given the best in control ARL based on a specified k. Hence we will use the same cutoff values which have been defined by simulation for simple versus simple hypothesis.

https://doi.org/10.26219/heal.aueb.6628

Chapter 4

Illustration for Normal Data

As we have already provided in Chapter 2, Classic CUSUM has the optimal in control ARL for detecting a particular shift in a process' mean. Thus, our main purpose after we construct the Bayesian CUSUM is to prove a mathematical equivalence between the two forms of CUSUM, so we will prove that the Bayesian CUSUM is optimal as well as the Classic CUSUM for the Simple versus Simple case. Then we will illustrate this equivalence by simulations for normal data. Simulations will not only used for Simple versus Simple case but for all the other cases as well.

4.1 Simple versus Simple Case:

Suppose we are interested to test whether a process has gone from an in control parameter value θ_0 to an out of control parameter value θ_1. Thus, we test:

$$\begin{Bmatrix} H_0\colon \theta = \Theta_0 \\ H_1\colon \theta = \theta_1 \end{Bmatrix}$$

Connaturally to classic CUSUM, the Bayesian CUSUM design involves two parts too. Firstly we have to select an appropriate cutoff point that will give an acceptable in control ARL for each different value of the distance between the two parameters θ_0 and θ_1 respectively. Then we will check out how quickly the Bayesian CUSUM will respond to out of control situations when the parameter value has shifted.

Suppose that: $x_1, x_2, \dots, x_n \sim N(\theta, \sigma^2)$, with σ^2: known. Then:

$$f(x_i|\theta) = \frac{1}{\sigma\sqrt{2\pi}} exp\left(-\frac{(x_i-\theta)^2}{2\sigma^2}\right) =$$

$$= \frac{1}{\sigma\sqrt{2\pi}} exp\left(-\frac{x_i^2}{2\sigma^2} + \frac{x_i\theta}{\sigma^2} - \frac{\theta^2}{2\sigma^2}\right) =$$

https://doi.org/10.26219/heal.aueb.6628

$$= f(x_i|\theta) = exp\left(-\frac{x_i^2}{2\sigma^2} + \frac{x_i\theta}{\sigma^2} - \frac{\theta^2}{2\sigma^2} - \frac{1}{2}\ln(2\pi\sigma^2)\right).$$

Where, according to the general exponential family form:

$$a(x_i) = x_i$$

$$b(\theta) = \frac{\theta}{\sigma^2}$$

$$c(x_i) = -\frac{x_i^2}{2\sigma^2} - \frac{1}{2}\ln(2\pi\sigma^2)$$

$$d(\theta) = -\frac{\theta^2}{2\sigma^2} \tag{4.1.1}$$

Thus, applying (4.1.1) to Equation (3.3.6), the decision interval form of the Bayesian CUSUM is:

$$W_n = \min(0, W_{n-1} + \alpha(x_i)\{b(\theta_0) - b(\theta_1)\} + \{d(\theta_0) - d(\theta_1)\}) =$$

$$W_n = \min\left[0, W_{n-1} + x_i\left(\frac{\theta_0}{\sigma^2} - \frac{\theta_1}{\sigma^2}\right) + \left(-\frac{\theta_0^2}{2\sigma^2} - \left(-\frac{\theta_1^2}{2\sigma^2}\right)\right)\right].$$

Thus,

$$W_n = \min\left[0, W_{n-1} + x_i\left(\frac{\theta_0}{\sigma^2} - \frac{\theta_1}{\sigma^2}\right) + \left(\frac{\theta_1^2}{2\sigma^2} - \frac{\theta_0^2}{2\sigma^2}\right)\right] \tag{4.1.2}$$

signaling if W_n < cutoff.

4.1.1 Assessment of the In Control ARL

In order to define in control ARLs, we use software and specifically an S-Plus routine (see appendix). Without loss of generality we consider in control parameter value $\theta = \Theta_0$ and known variance $\sigma^2 = 1$. We define a vector of forty different

https://doi.org/10.26219/heal.aueb.6628

possible cutoff point values, which lies between points -2 and -5.9. For each one of these forty possible cutoff point values we compute the in control ARL. We work out with three different out of control parameter values θ_1=2, θ_1=3 and θ_1=4. For each one of these three different values, ARL comes up as the mean of a vector of $N = 1000$ different run lengths per cutoff value. Simulation results for the corresponding three out of control parameter values, per cutoff value are presented in Table 4.1 below.

https://doi.org/10.26219/heal.aueb.6628

Table 4.1:The in control ARL for various cutoff points for each one of three different out of control parameter values for Bayesian form of CUSUM.

Aa	Cutoff point	$\theta_1=2$ In control ARL	$\theta_1=3$ In control ARL	$\theta_1=4$ In control ARL
1	-2	35.969	66.043	169.978
2	-2.1	38.229	66.962	173.781
3	-2.2	43.789	75.519	181.356
4	-2.3	47.587	78.032	196.436
5	-2.4	51.556	86.44	205.752
6	-2.5	58.494	97.212	238.933
7	-2.6	61.883	103.002	251.521
8	-2.7	74.542	114.759	278.3017
9	-2.8	75.223	127.066	295.967
10	-2.9	87.908	125.778	309.459
11	-3	96.225	140.144	305.607
12	-3.1	102.349	154.838	340.137
13	-3.2	115.535	171.43	396.903
14	-3.3	127.199	182.849	398.613
15	-3.4	138.618	199.641	459.568
16	-3.5	152.033	225.334	482.555
17	-3.6	164.87	254.867	488.435
18	-3.7	194.056	275.336	565.903
19	-3.8	204.911	281.842	615.367
20	-3.9	232.779	346.451	659.171
21	-4	260.459	346.506	733.814
22	-4.1	289.172	370.884	754.933
23	-4.2	312.062	423.86	880.526
24	-4.3	340.609	471.94	937.495
25	-4.4	412.188	506.553	1003.998
26	-4.5	421.291	563.217	1017.102
27	-4.6	475.058	586.387	1152.883
28	-4.7	545.316	624.627	1225.355
29	-4.8	605.587	759.666	1372.017
30	-4.9	624.071	770.792	1411.107
31	-5	729.153	872.447	1629.927
32	-5.1	790.75	979.666	1652.114
33	-5.2	908.59	1045.781	1970.969
34	-5.3	1002.213	1198.944	2163.876
35	-5.4	1079.039	1379.622	2293.918
36	-5.5	1272.989	1381.925	2455.999
37	-5.6	1273.36	1579.191	2687.45
38	-5.7	1409.141	1746.98	2988.419
39	-5.8	1635.378	1867.566	3181.985
40	-5.9	1823.771	2109.508	3389.649

Generally, entries in which the in control ARL becomes very large are not of interest and so are omitted (Hawkins and Olwell (1997)).

https://doi.org/10.26219/heal.aueb.6628

In order to decide whether Bayesian CUSUM is an acceptable scheme for checking whether a process has gone from an in control parameter value to an out of control value, we compare the above in control ARLs with the already known in control ARLs of the classic CUSUM, for a specific pair(h, k) for which the classic CUSUM was designed.

In classic CUSUM there are tables showing the ARL for selected values of k and h and they may be scanned to find the h that gives an acceptable ARL. These tables were computed using simulation methods. For more information see Hawkins and Olwell (1997).Table 4.2 gives ARL as a function of k and h for classic CUSUM for mean shift of standardized normal data. Case $k = 1$ corresponds to $\theta_1 = 2$ and $k = 1.5$ to θ_1 =3 respectively.

Table 4.2: ARL as a function of k and h for classic CUSUM for mean shift of standardized normal data for $\theta_1 = 2$ and θ_1 =3

H	k=1	k=1.5
1.000	35.3	142.2
1.125	44.8	196.8
1.250	57.2	274.9
1.375	73.1	387.2
1.500	93.8	549.7
1.625	120.7	786.0
1.750	155.5	1130.8
1.875	200.5	1635.8
2.000	258.7	2376.8
2.125	333.8	3465.4
2.250	430.7	5065.1
2.375	555.5	7414.5
2.500	716.0	10861.4
2.625	922.2	15910.5
2.750	1187.0	23294.0
2.875	1526.8	34071.6
2.125	1962.8	49777.5

https://doi.org/10.26219/heal.aueb.6628

Before comparing the results of the in Control ARL for classic and Bayesian CUSUM we will show that they are equivalent and give the relationship among the various parameters. We will work with the normal family when for $\theta_1 > \theta_0$, the two forms of the CUSUM will be:

$$C_n = \frac{D_n}{b(\theta_1) - b(\theta_0)} = \frac{max(0, D_{n-1} + a(x_n)\{b(\theta_1) - b(\theta_0)\} + \{d(\theta_1) - d(\theta_0)\})}{b(\theta_1) - b(\theta_0)}$$

$$C_n = max(0, C_{n-1} + X_n - k)$$

$$k = \frac{\theta_0 + \theta_1}{2}$$

$$C_n^+ = max\left(0, C_{n-1}^+ + X_n - \frac{\theta_0 + \theta_1}{2}\right)$$

$$\text{signaling if } C_n = \frac{D_n}{b(\theta_1) - b(\theta_0)} > \frac{\ln A}{b(\theta_1) - b(\theta_0)} = h$$

representing the Classic scheme of CUSUM.

According to Equation (3.3.6)

$$W_n = \min(0, W_{n-1} + \alpha(x_n)\{b(\theta_0) - b(\theta_1)\} + \{d(\theta_0) - d(\theta_1)\}) \Rightarrow$$

$$W_n = \min_x \left[0, W_{n-1} + x_n\left(\frac{\theta_0}{\sigma^2} - \frac{\theta_1}{\sigma^2}\right) + \left(\frac{\theta_1^2}{2\sigma^2} - \frac{\theta_0^2}{2\sigma^2}\right)\right]$$

signaling if $W_n <$ cutoff, represents the Bayesian scheme of CUSUM.

Dividing D_n by $-\{b(\theta_1) - b(\theta_0)\} = b(\theta_0) - b(\theta_1)$ maximum turns to minimum and the inequality is being reversed. Thus,

$$\frac{D_n}{b(\theta_0) - b(\theta_1)} = \frac{max(0, D_{n-1} + a(x_n)\{b(\theta_1) - b(\theta_0)\} + \{d(\theta_1) - d(\theta_0)\})}{b(\theta_0) - b(\theta_1)} =$$

https://doi.org/10.26219/heal.aueb.6628

$$= \frac{max(0, D_{n-1} + x_n(\theta_1 - \theta_0)/\sigma^2 + (-\theta_1^2 + \theta_0^2)/(2\sigma^2))}{(\theta_1 - \theta_0)/\sigma^2}. \qquad (4.1.3)$$

Thus,

$$C_n = \max(0, C_{n-1} - a(x_n) + k) =$$

$$C_n^- = \min(0, C_{n-1} + a(x_n) - k)$$

signals if $C_n < -h$

where:

$$k = -\frac{(-\theta_1^2 + \theta_0^2)/(2\sigma^2)}{(\theta_1 - \theta_0)/\sigma^2} = \frac{\theta_0 + \theta_1}{2} .$$

Furthermore dividing W_n by $\{b(\theta_0) - b(\theta_1)\}$ we get:

$$\frac{W_n}{b(\theta_0) - b(\theta_1)} = \frac{min(0, W_{n-1} + a(x_n)\{b(\theta_0) - b(\theta_1)\} + \{d(\theta_0) - d(\theta_1)\})}{b(\theta_0) - b(\theta_1)} =$$

$$\frac{\min\left[0, W_{n-1} + x_i\left(\frac{\theta_0}{\sigma^2} - \frac{\theta_1}{\sigma^2}\right) + \left(\frac{\theta_1^2}{2\sigma^2} - \frac{\theta_0^2}{2\sigma^2}\right)\right]}{\left(\frac{\theta_0}{\sigma^2} - \frac{\theta_1}{\sigma^2}\right)}$$

$$= \min\left(0, \frac{W_{n-1}}{b(\theta_0) - b(\theta_1)} + x_i - k\right) = C_n^-. \qquad (4.1.4)$$

From the above we conclude that:

$$-h = \frac{cutoff}{b(\theta_0) - b(\theta_1)} . \qquad (4.1.5)$$

Thus, it has been proved that for simple versus simple case Bayesian CUSUM is equivalent to classic CUSUM as far as the in control ARL is concerned.

Table 4.3 shows the comparison between the two schemes of CUSUM for some values of parameter h and cutoff point. We will only compare values of cutoff point that are multiples of h, with the respective value of h, hence we compare points

https://doi.org/10.26219/heal.aueb.6628

that satisfy the formula (4.1.5) in order to show the equivalence between the two schemes of CUSUM. That equivalence cannot be displayed for values of cutoff point and h that are not proportional. (For cases that cutoff point is not a multiply of h the respective cells of the Table 4.3 will remain blank). As we can notice in control ARL values are relatively close for the two forms of CUSUM for values of h and cutoff point which satisfy the formula (4.1.5). Their differences are due to Monte Carlo error.

Table 4.3: Comparison for the in control ARL of classic and Bayesian CUSUM for specific values of parameter h and cutoff point.

$\theta_1-\theta_0$	cutoff	h=1		h=1.250		h=1.5		h=1.750	
CUSUM form		B.C	C.C	B.C	C.C	B.C	C.C	B.C	C.C
2	-2	35.969	35.3						
	-2.5			58.494	57.2				
	-3					96.225	93.8		
	-3.5							152.033	155.5
3	-3	140.144	142.2						
	-4.5					563.217	549.7		

As far as the in control ARL is concerned, it is known by definition that is the mean of a geometric distribution, the distribution of run lengths. Thus, the in control ARL is given by the equation:

$$ARL = \frac{1}{p}, 0 < p < 1$$

where p is the probability that any point exceeds the control limits while the process is in control (see paragraph 2.6.1). In other words p is the probability of a false alarm. Obviously ARL and p are inversely related. Taking the logarithm of the above equation, the relationship between $\log(ARL)$ and $\log(p)$ becomes linear with negative slope:

$$\log(ARL) = \log\left(\frac{1}{p}\right) \Rightarrow$$

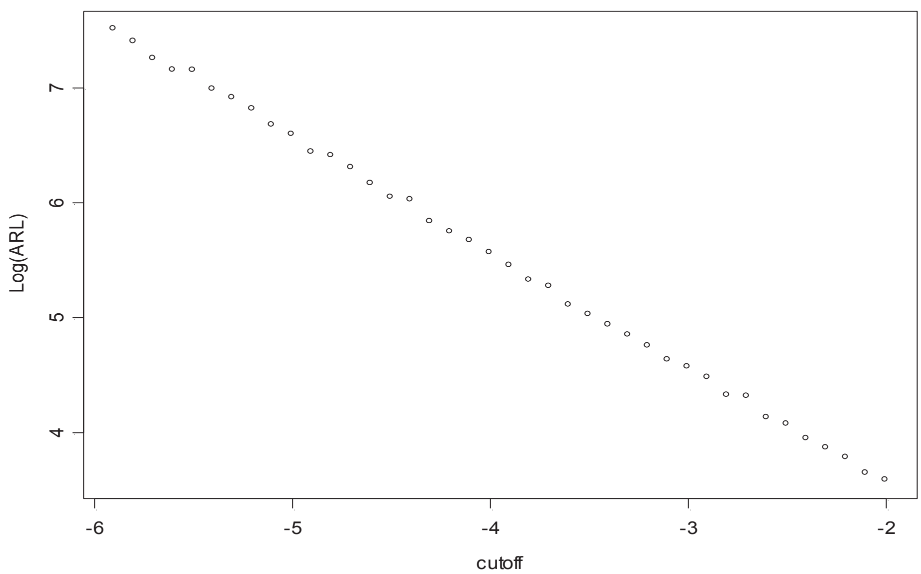
https://doi.org/10.26219/heal.aueb.6628

$$\log(ARL) = \log 1 - \log(p)\,, 0 < p < 1 \;\Rightarrow$$

$$\log(ARL) = 0 - \log(p),\ \log(p) < 0.$$

Using simulation's data of Table 4.1 it comes out that ARL increases as the value of cutoff point decreases. It is reasonable that as the cutoff point value decreases the probability of a false alarm decreases too. Thus, the ARL and cutoff are inversely related too. So we expect that relationship between ARL and cutoff will be similar as for ARL and p. That relation is shown in Figure 4.1.1, 4.1.2 and 4.1.3 below:

Figure 4.1: The logarithm of the in control ARL, Log(ARL) for the out of control parameter value $\theta_1 = 2$, against the cutoff point values.



https://doi.org/10.26219/heal.aueb.6628

Figure 4.2: The logarithm of the in control ARL, Log(ARL) for the out of control parameter value $\theta_1 = 3$, against the cutoff point values.

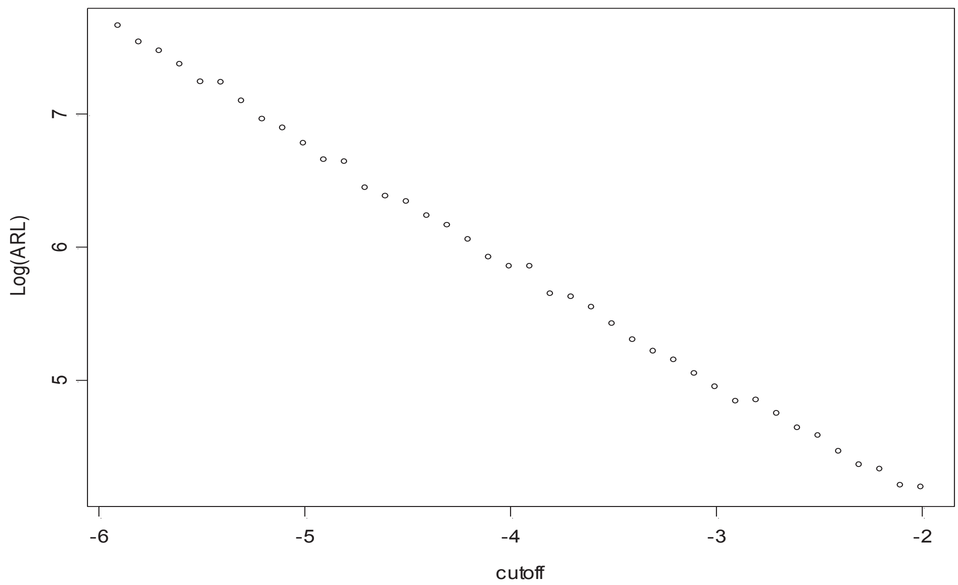
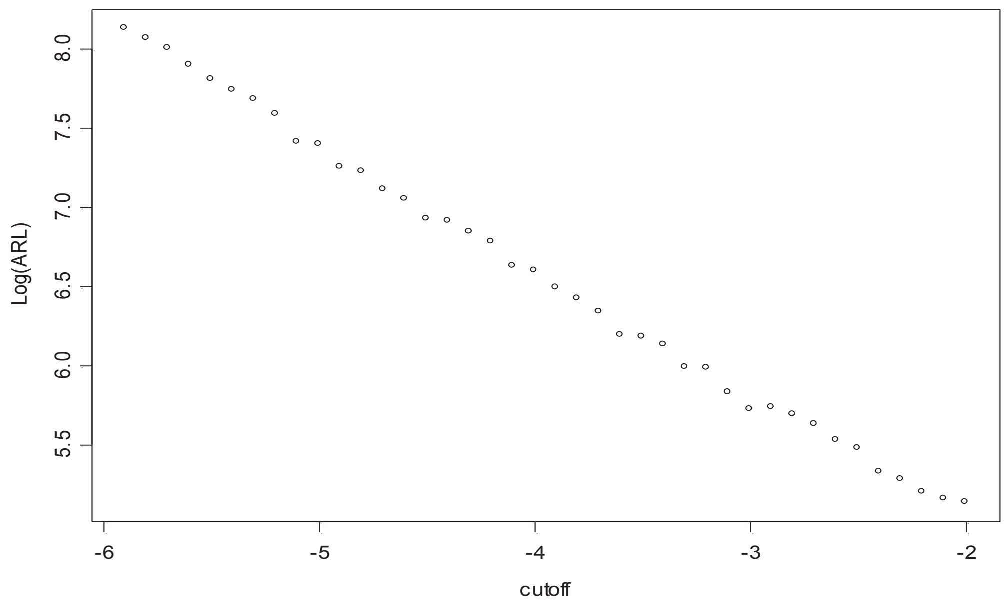


Figure 4.3: The logarithm of the in control ARL, Log(ARL) for the out of control parameter value $\theta_1 = 4$, against the cutoff point values.



 https://doi.org/10.26219/heal.aueb.6628

4.1.2 Assessment of the Out of Control ARL

Simulation techniques will be used to check out how quickly the Bayesian CUSUM will respond to out of control situations when the parameter value has shifted. We consider the in control parameter value $\theta_0 = 0$ and known variance $\sigma^2 = 1$ for normal data. We use the same vector of different possible cutoff point values, which lies between points -2 and -5.9. We work out with three different out of control parameter values $\theta_1 = 2$, $\theta_1 = 3$ and $\theta_1 = 4$. For each one of these three values we deliberately start to derive normally distributed random variables from the out of control distribution respectively. We compute the out of control ARL for each one of the three cases and the last point for which Equation 4.1.4 was equal to zero. Thus, we compute n=instant at which the shift is detected (out of control ARL) and m=last time W_n was equal to zero before the shift was detect. Then we calculate the distance between those two instants, for which we would like to be very small. The above ARLs once again come out as the Average Run Lengths for each one of the forty different cutoff values. Each one of those ARLs is the mean of a vector of $N = 10000$ different run lengths per cutoff value. Thus the quantity $(n - m)$ is the distance of each one of those out of control ARLs from the respective average of the last points for which the W_n was equal to zero and which also comes up as an average of $N = 10000$ different iterations. The simulation results are shown at Table 4.4 below:

https://doi.org/10.26219/heal.aueb.6628

Table 4.4: Out of Control ARL and the distance $(n - m)$ for various cutoff points for three different out-of control parameter values for the Bayesian CUSUM

b(θ$_1$)-b(θ$_0$)	**θ$_1$-θ$_0$=2-0=2**		**θ$_1$-θ$_0$=3-0=3**		**θ$_1$-θ$_0$=4-0=4**	
Cutoff	**ARL**	**Average n-m**	**ARL**	**Average n-m**	**ARL**	**Average n-m**
-2	2.7791	1.9865	2.2408	2.0001	2.0698	2.0008
-2.1	2.8267	1.9916	2.2577	2.0107	2.076	2.0024
-2.2	2.8658	1.9913	2.2638	2.0086	2.081	1.9987
-2.3	2.9121	1.9902	2.2765	1.9961	2.0848	1.9964
-2.4	2.9334	1.9764	2.2951	2.0002	2.0802	1.995
-2.5	3.0025	2.034	2.3001	2.0025	2.0884	2.003
-2.6	3.0328	1.9865	2.3198	2.0052	2.0865	1.9958
-2.7	3.0996	1.9972	2.336	2.0026	2.095	2.0011
-2.8	3.1308	1.9952	2.3424	1.9932	2.1015	2.0098
-2.9	3.1654	1.9644	2.3687	2.0024	2.1122	1.9894
-3	3.2396	1.9987	2.3836	2.0081	2.116	1.99
-3.1	3.2829	1.9908	2.3946	2.002	2.1181	1.9987
-3.2	3.3312	1.9801	2.4245	2.0211	2.1272	1.9974
-3.3	3.393	1.9946	2.4198	1.9911	2.1295	1.9958
-3.4	3.4297	1.9826	2.4445	1.99	2.1303	1.993
-3.5	3.4914	2.0205	2.4533	1.9937	2.1409	2.0039
-3.6	3.5324	1.9894	2.4768	2.0005	2.1448	2.0042
-3.7	3.5824	2.0039	2.5005	2.0039	2.1523	2.0065
-3.8	3.6401	1.9953	2.511	1.9925	2.1578	1.9948
-3.9	3.6872	2.0027	2.531	1.9908	2.1653	1.9944
-4	3.7543	1.9852	2.5633	1.9982	2.1728	1.9933
-4.1	3.7962	2.0106	2.566	1.9929	2.1735	2.0061
-4.2	3.8624	2.0386	2.6042	2.0057	2.1815	1.9998
-4.3	3.8993	2.0096	2.6148	1.9945	2.1924	1.9946
-4.4	3.9204	1.9958	2.6287	2.0029	2.1977	1.9838
-4.5	3.9802	1.9846	2.6596	2.0022	2.206	1.9966
-4.6	4.041	2.008	2.6751	1.9988	2.2167	1.9874
-4.7	4.0866	1.9533	2.6864	1.9992	2.2203	2.0019
-4.8	4.1609	2.0355	2.7163	1.9892	2.2375	1.9966
-4.9	4.2189	2.0128	2.7476	2.0115	2.2347	2.0055
-5	4.2606	2.0231	2.7561	1.9943	2.2487	1.9972
-5.1	4.3018	1.9862	2.7896	1.9912	2.2541	1.9997
-5.2	4.324	1.9927	2.8205	2.0109	2.2625	1.9927
-5.3	4.3818	1.9968	2.8444	2.0155	2.2783	2.0138
-5.4	4.4946	2.0603	2.8419	1.9991	2.2852	1.996
-5.5	4.4526	1.9611	2.8681	1.9941	2.3012	1.9968
-5.6	4.5398	1.9923	2.9043	1.9956	2.3008	1.9994
-5.7	4.5916	1.9919	2.9166	1.9757	2.3158	1.9923
-5.8	4.6506	1.9781	2.9594	2.0154	2.3155	2.0035
-5.9	4.7326	2.044	2.9756	2.0014	2.3442	2.0018

https://doi.org/10.26219/heal.aueb.6628

As we can notice from Table 4.4, the out of control ARL seems to be quite good for Simple versus Simple case. Bayesian scheme of CUSUM has a quick respond to out of control situations. The greater values for ARL are reported for $\theta_1 = 2$. That delay in detection happens because of the very small shift of the parameter value, and it is noticed only for large values of the cutoff point parameter. Generally out of control ARL does not surpass 5, and for bigger shifts than 2σ, it lies between 2 and 3. For smaller cutoff points the detection occurs as quickly for 2σ's shift, as it does for larger shifts. Furthermore, as it was expected, the distance between n=the instant at which the shift was detected (out of control ARL) and m=last time W_n was equal to zero before the shift was detect (last point for which Equation (4.1.4) was equal to zero), is relatively small and does not exceed 2.1. That means that the detection is being noted fairly fast after the instant the shift occurs. Thus, Bayesian CUSUM represents a power full tool for Statistical Process Control.

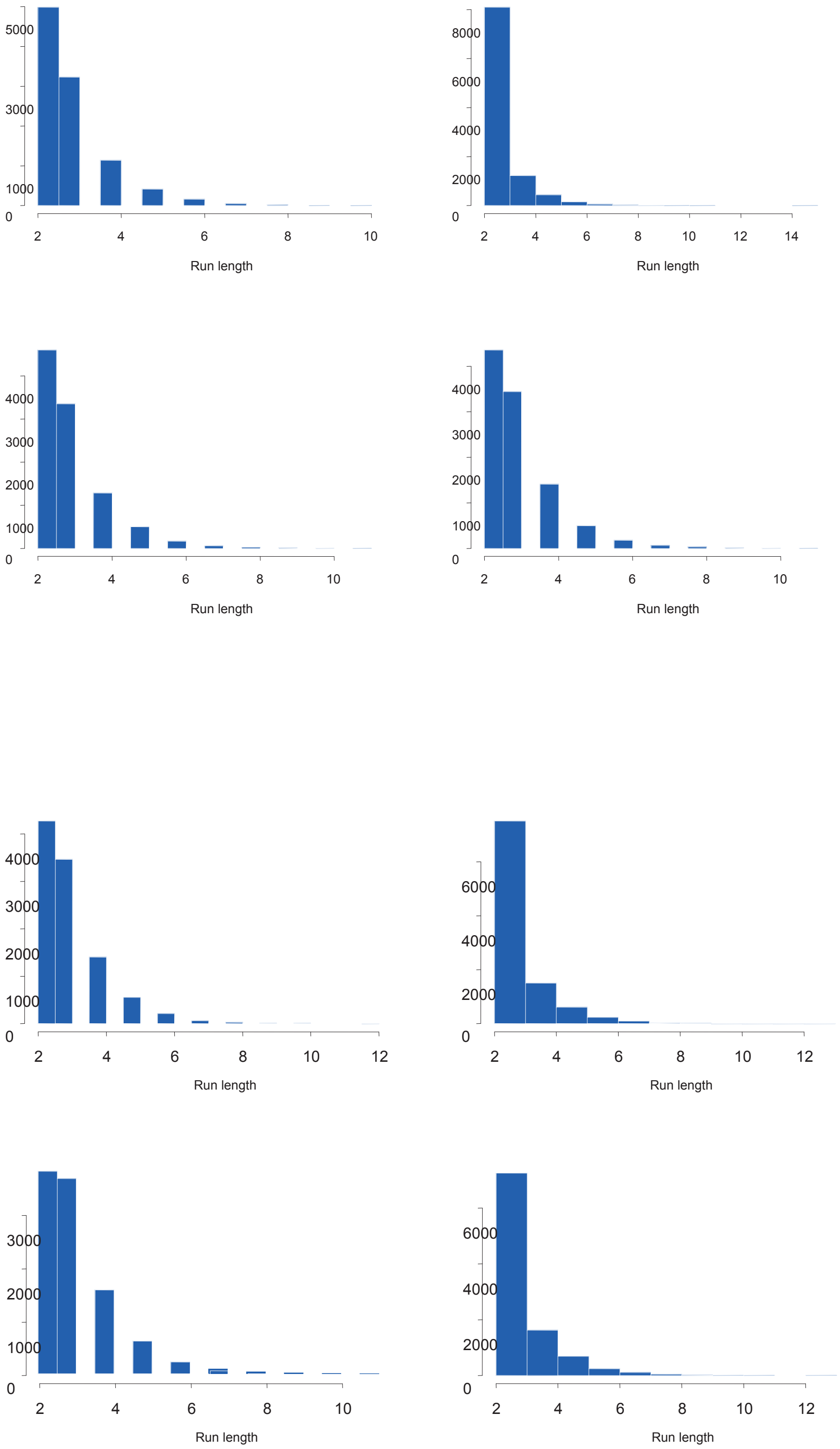
In order to draw better conclusions about Bayesian scheme of CUSUM we study the distribution of run length, for each one of the three different out of control parameter values per cutoff values. Thus we array forty different histograms corresponding to the forty cutoff values for each one of the three different out of control parameter values to show the way run length is distributed.

It is known that the distribution of run length for a control chart has a geometric distribution with the in control Average Run Length representing the mean of that distribution. Therefore, we expect to view a particular shape drawing out of these histograms, corresponding to the geometric distribution (see Montgomery (2001)).

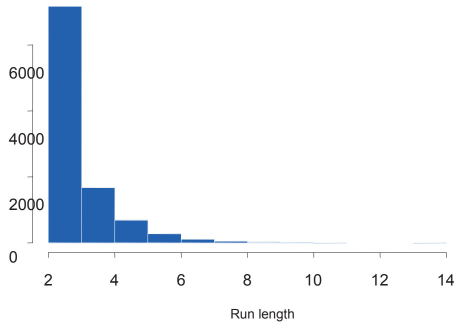
We group histograms into three different categories, with respect to the out of control parameter value θ_1. For each one of these three categories we have forty different histograms corresponding to the forty different plausible values of cutoff point parameter. For every one of these histograms the axis of xx' represents the different values of run length and the axis of yy' represents the frequency of a particular value of run length that was noted for the specific out of control parameter value per cutoff point value. The results are shown in Figures 4.2.1, 4.2.2 and 4.2.3 below:

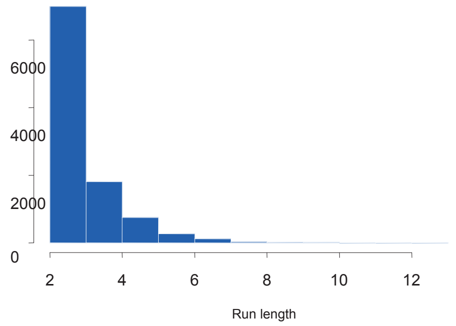
https://doi.org/10.26219/heal.aueb.6628

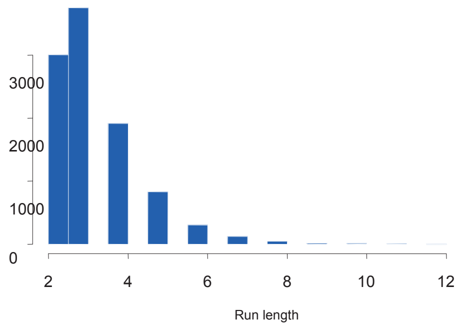
Figure 4.2.1: histograms of the quantities [run length] for $\theta_1 - \theta_0 = 2 - 0 = 2$, for the forty different cutoff point values of Bayesian CUSUM.

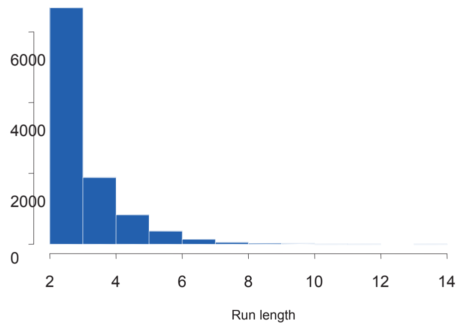


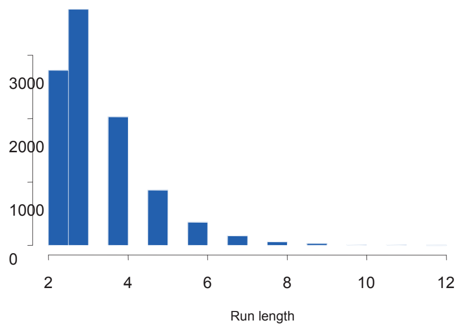
https://doi.org/10.26219/heal.aueb.6628

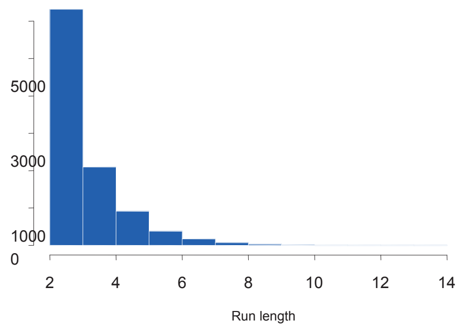


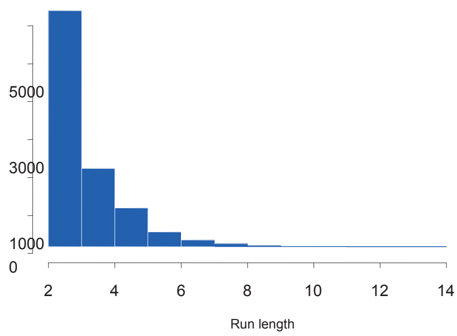


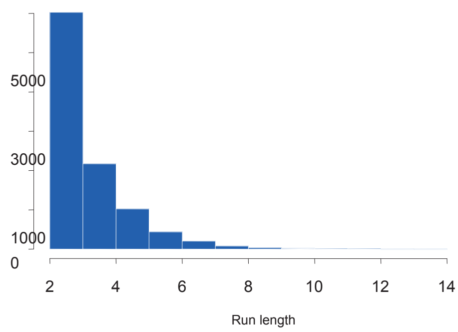




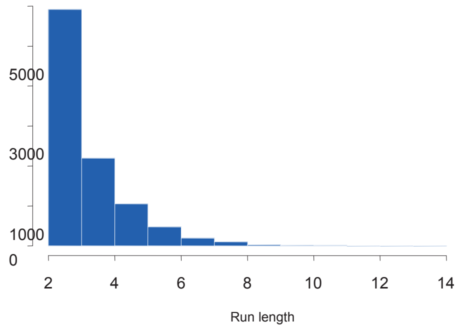


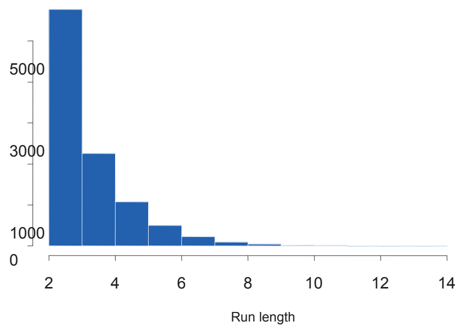


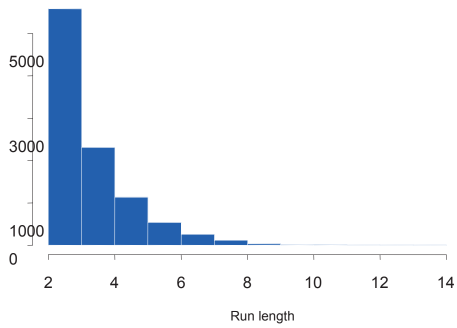


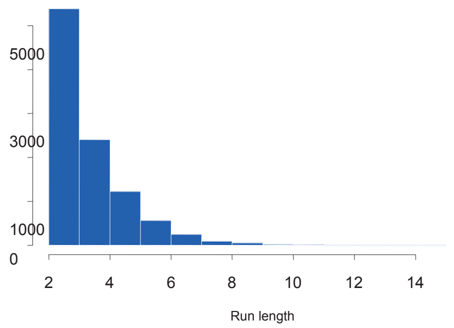


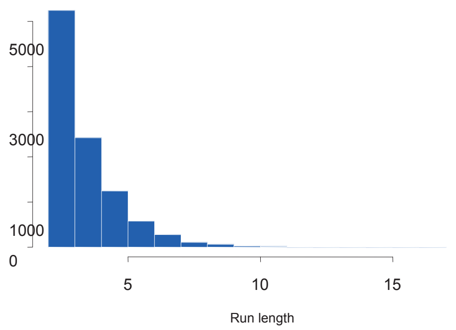
https://doi.org/10.26219/heal.aueb.6628

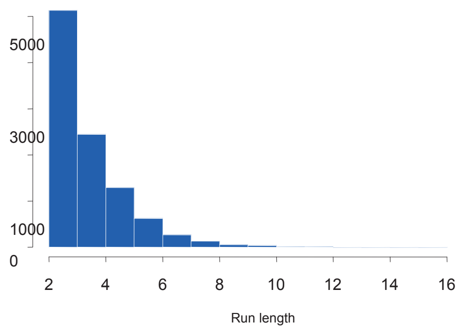


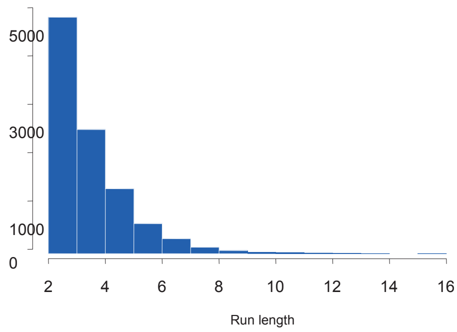


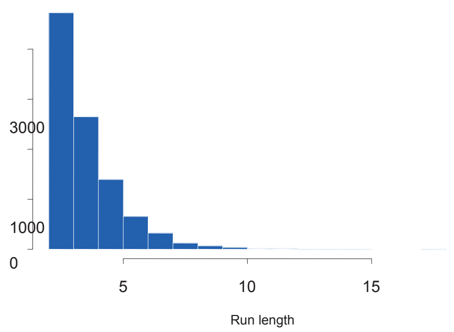




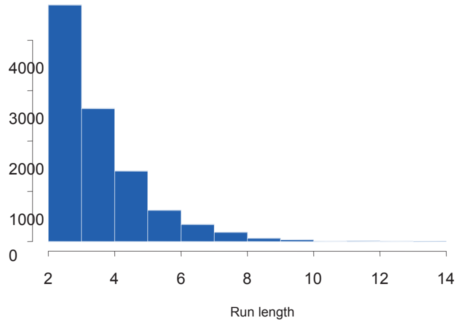


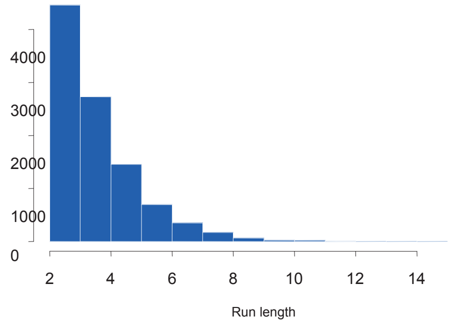


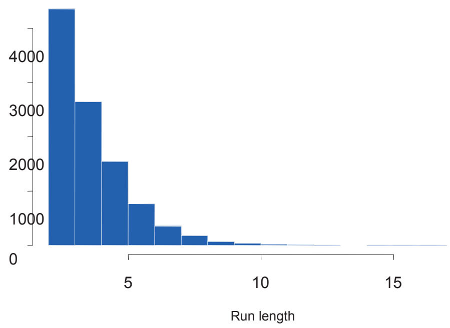


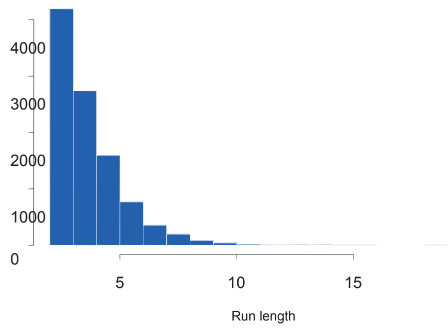


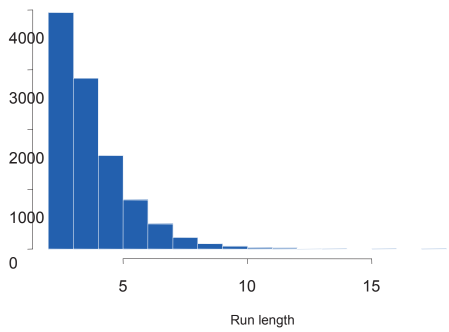
https://doi.org/10.26219/heal.aueb.6628

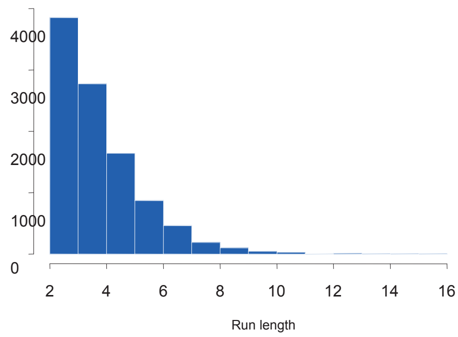


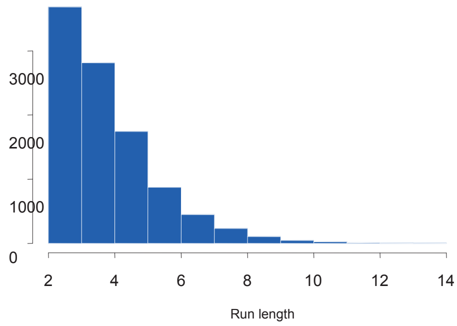


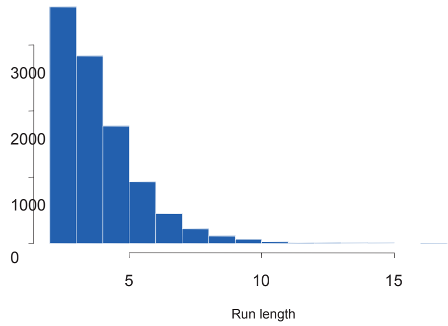




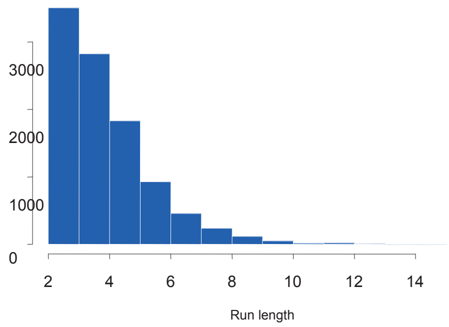


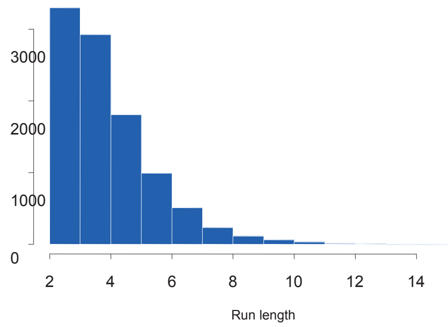


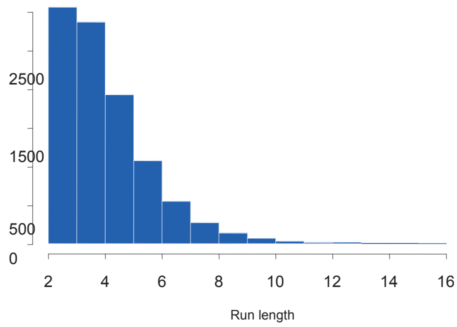


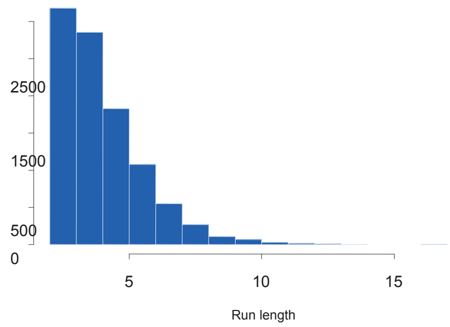


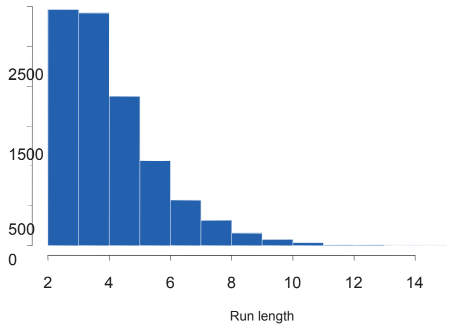
https://doi.org/10.26219/heal.aueb.6628

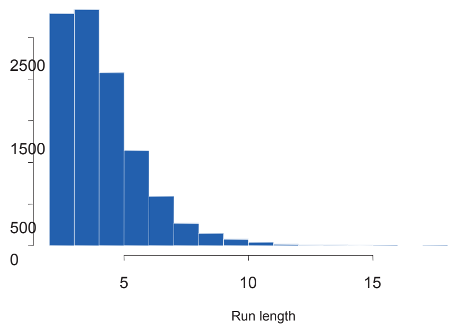


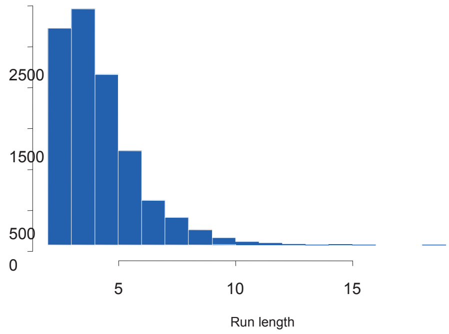


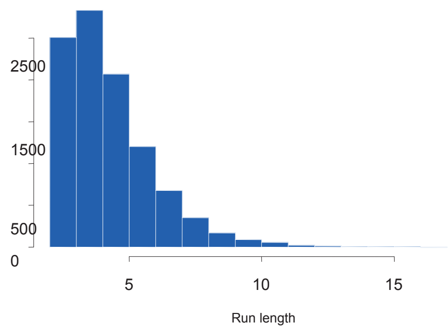






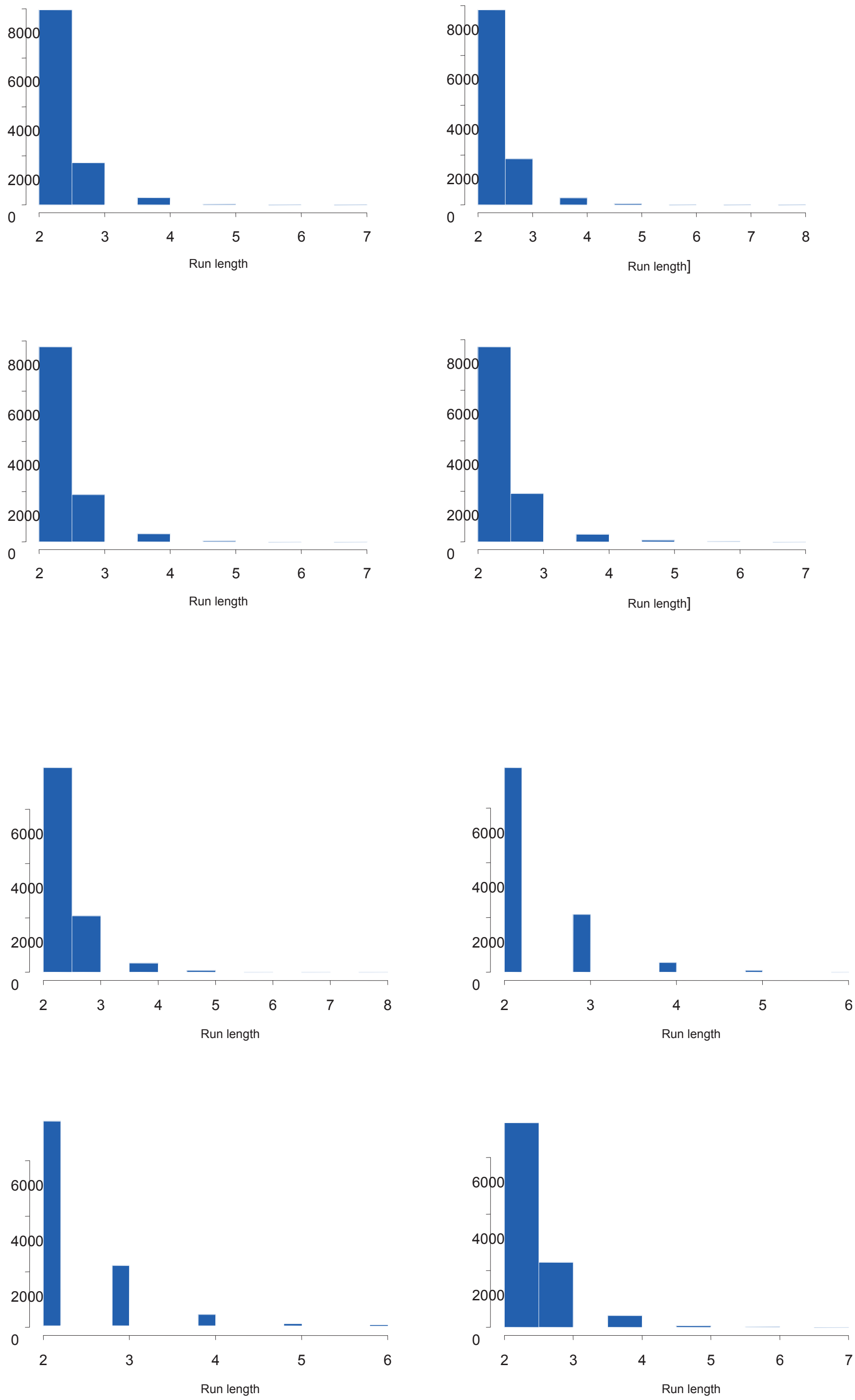




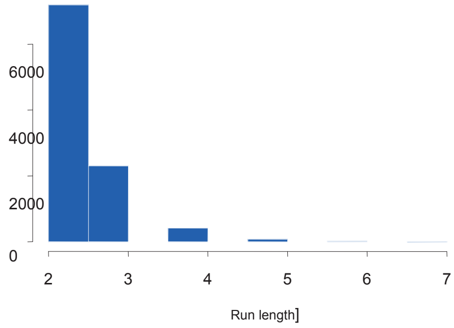


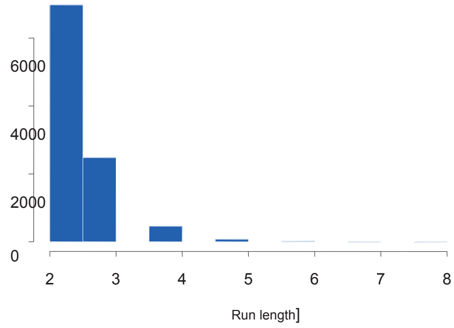
https://doi.org/10.26219/heal.aueb.6628

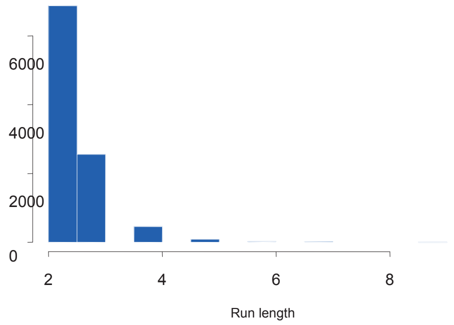
Figure 4.2.2: histograms of the quantities [run length] for $\theta_1 - \theta_0 = 3 - 0 = 3$, for the forty different cutoff point values of Bayesian CUSUM.

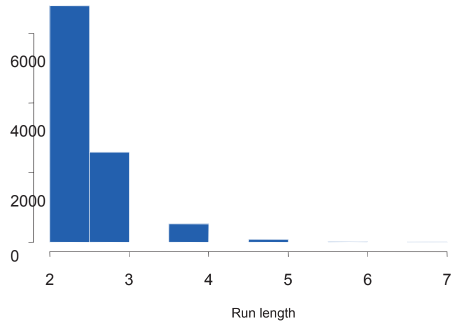


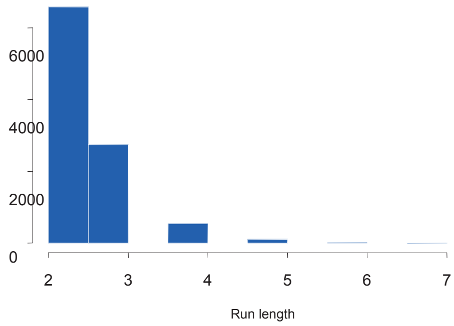
https://doi.org/10.26219/heal.aueb.6628

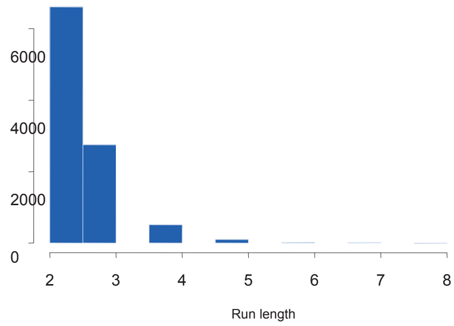


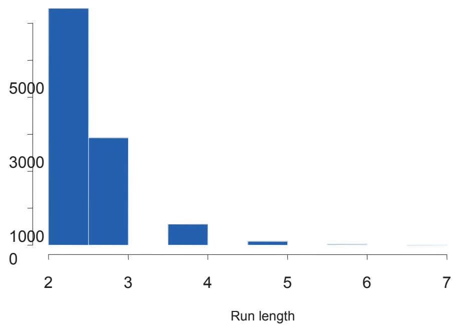


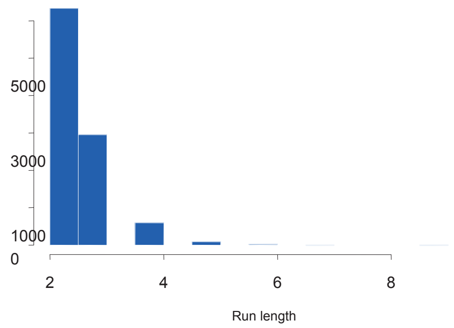




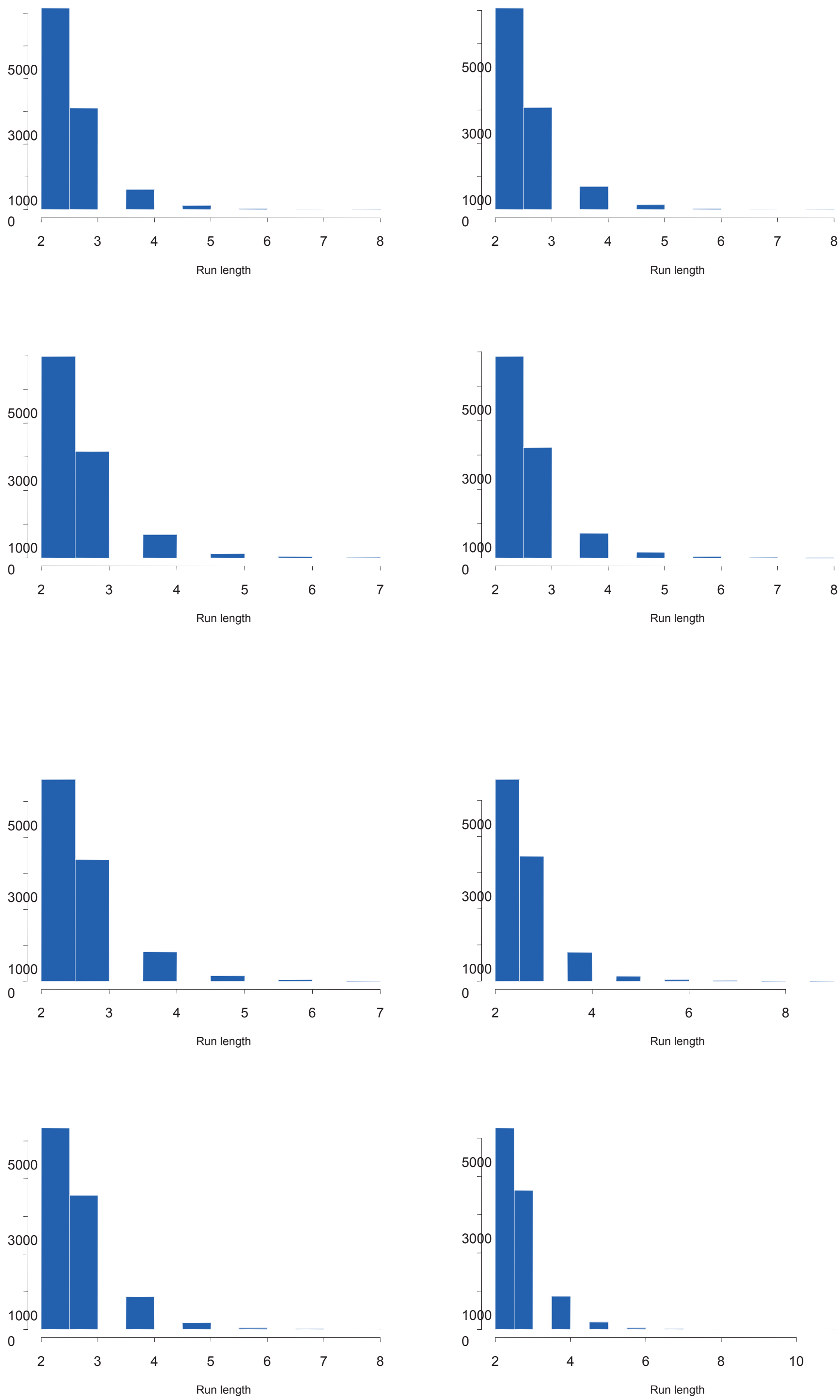




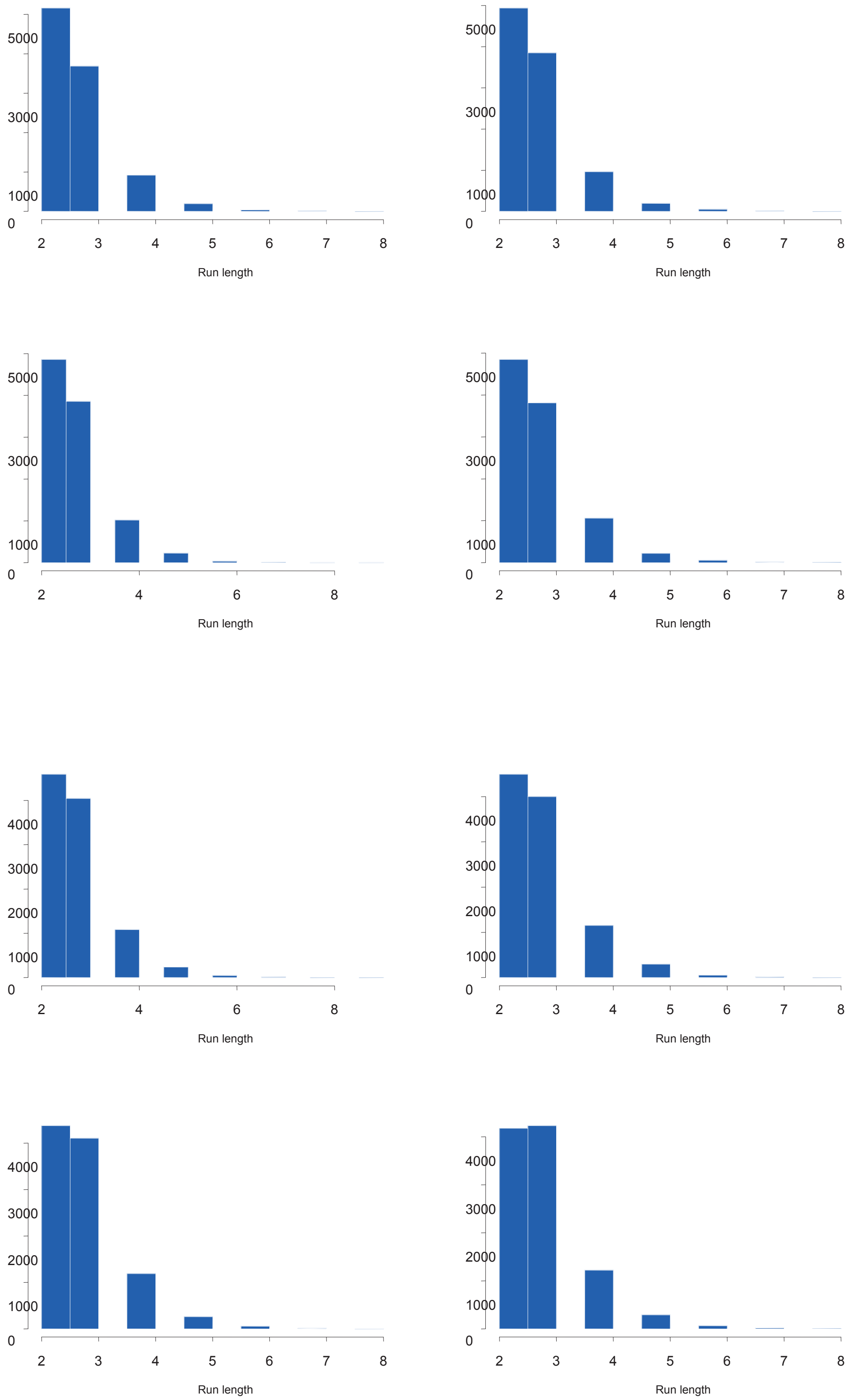




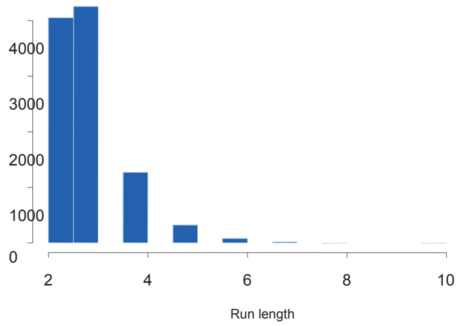
https://doi.org/10.26219/heal.aueb.6628

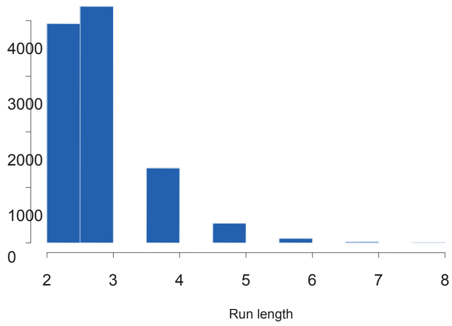


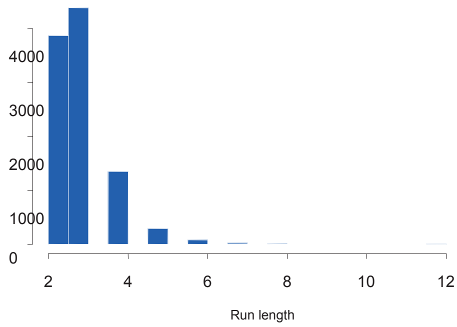
https://doi.org/10.26219/heal.aueb.6628

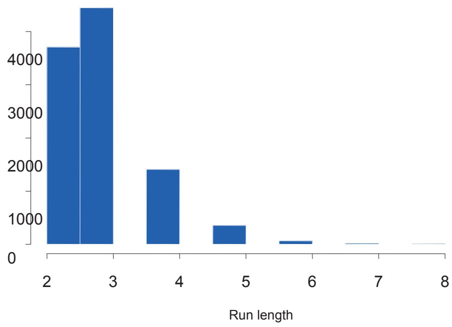


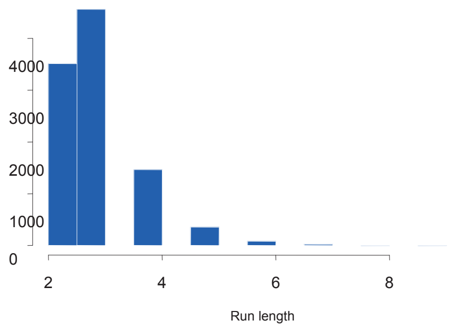
https://doi.org/10.26219/heal.aueb.6628

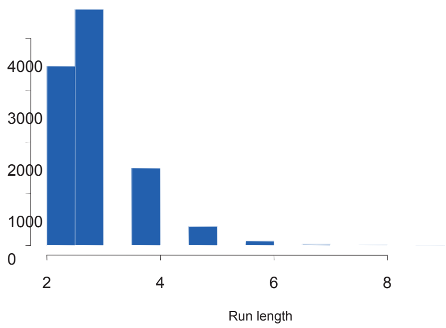


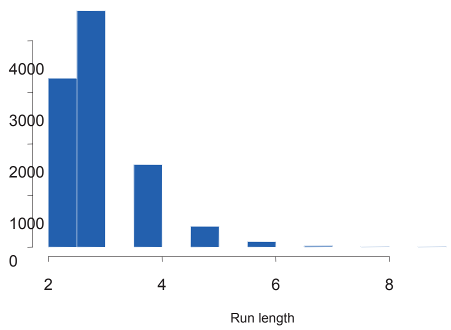


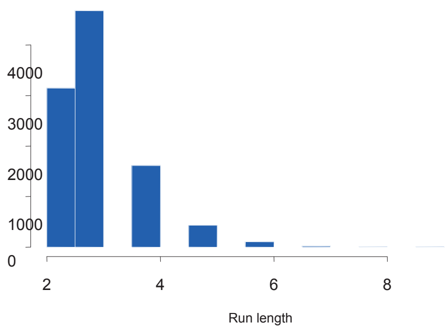






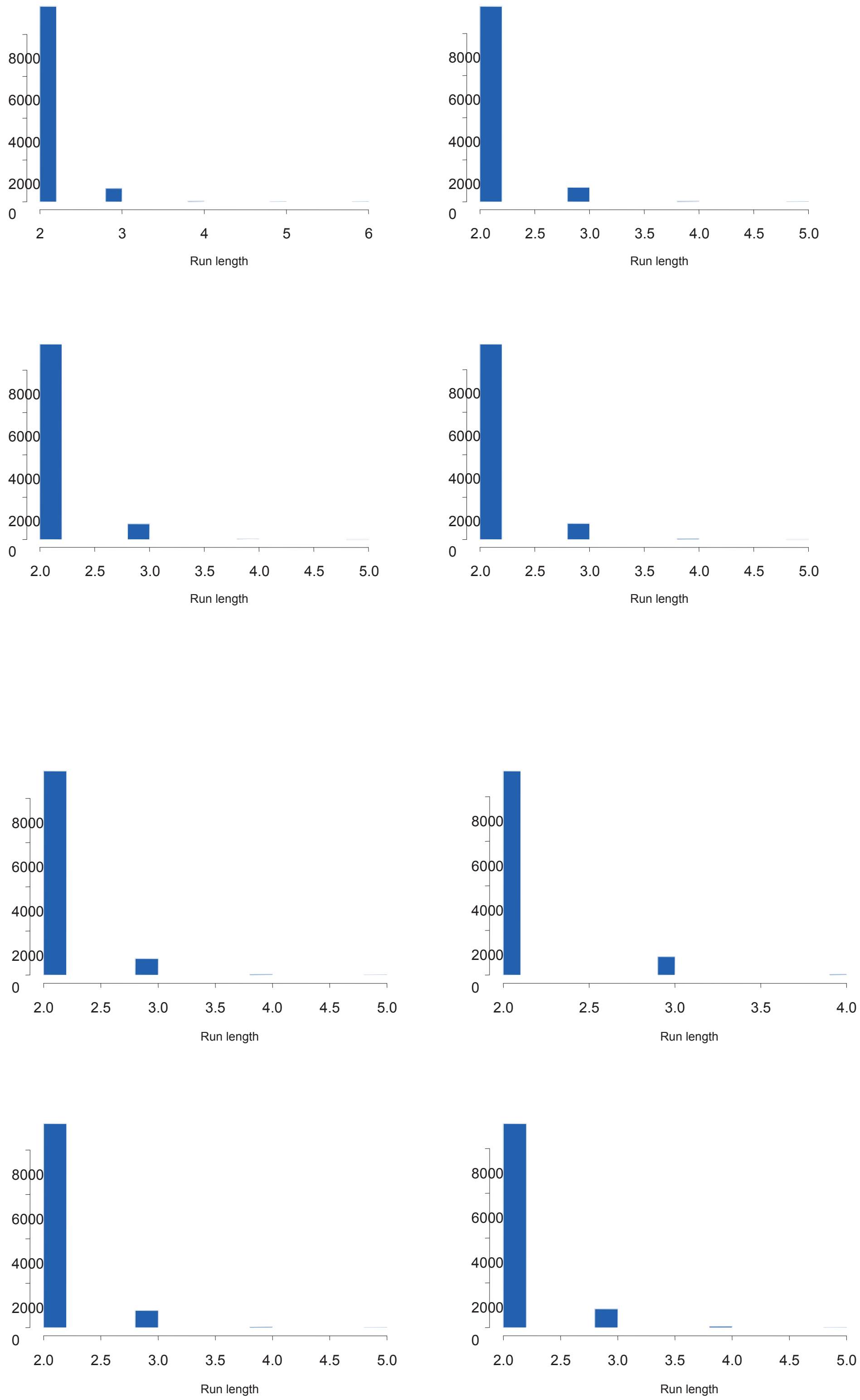




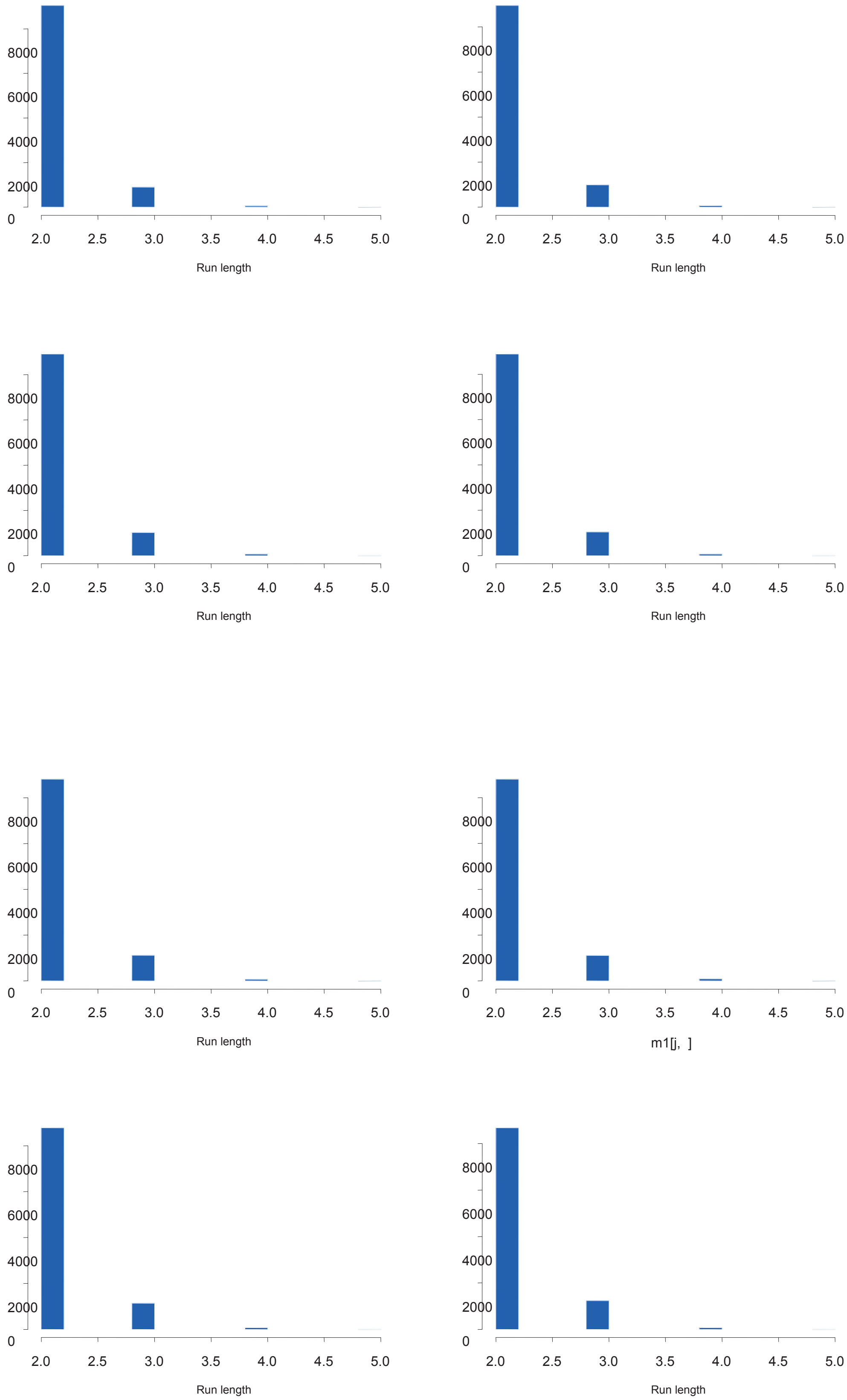


https://doi.org/10.26219/heal.aueb.6628

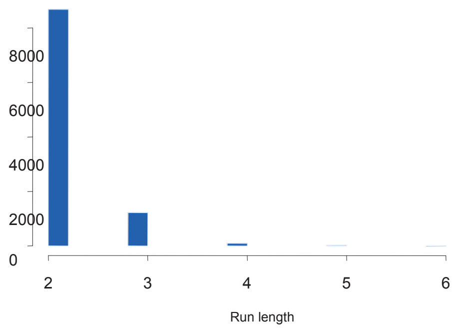
Figure 4.2.3: histograms of the quantities [run length] for $\theta_1 - \theta_0 = 4 - 0 = 4$, for the forty different cutoff point values of Bayesian CUSUM.

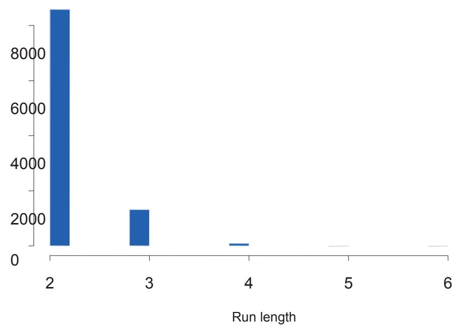


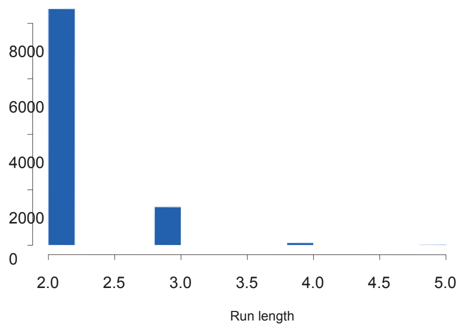
https://doi.org/10.26219/heal.aueb.6628

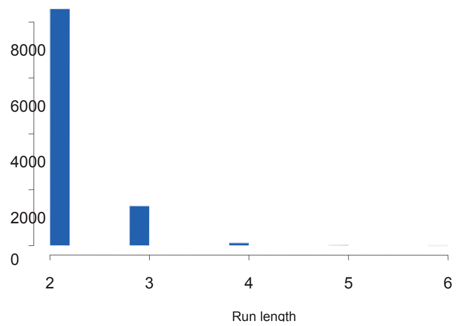


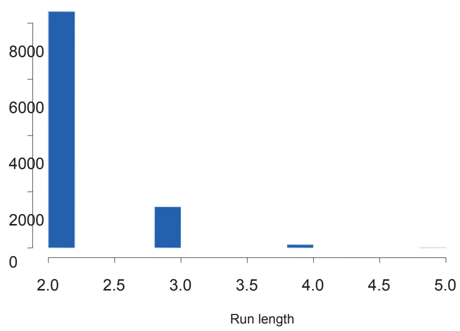
https://doi.org/10.26219/heal.aueb.6628

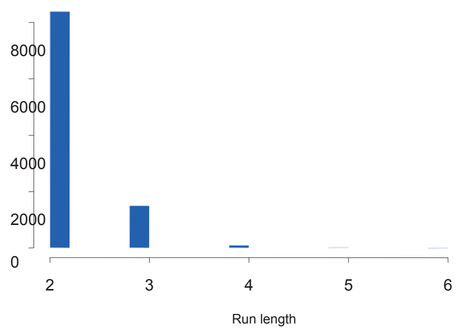


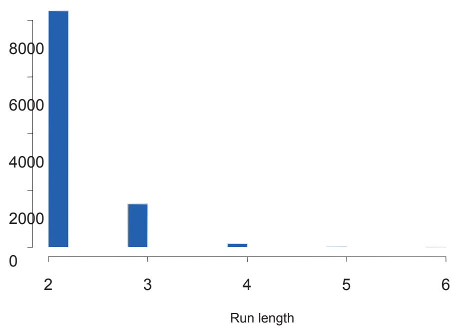


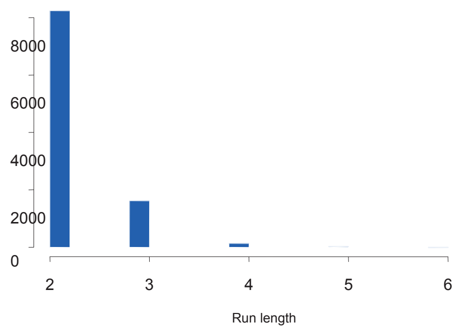




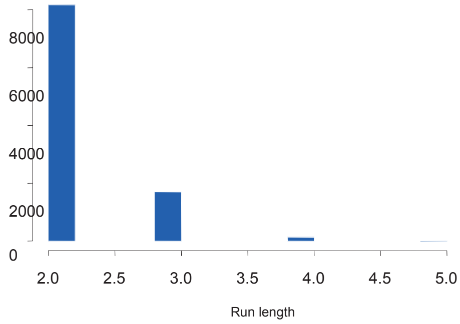


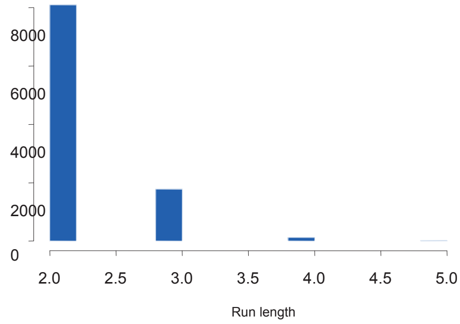


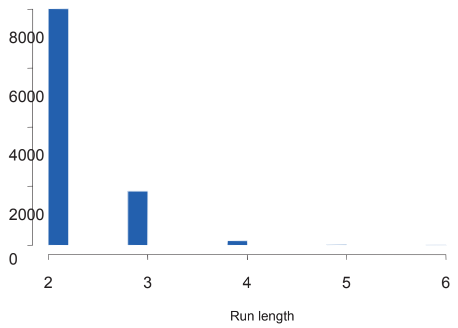


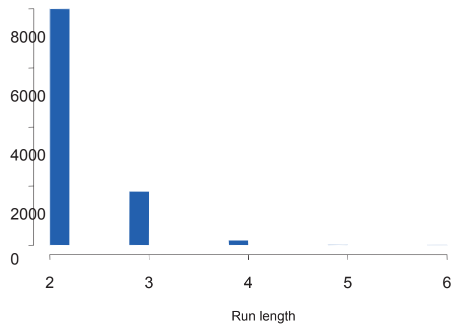


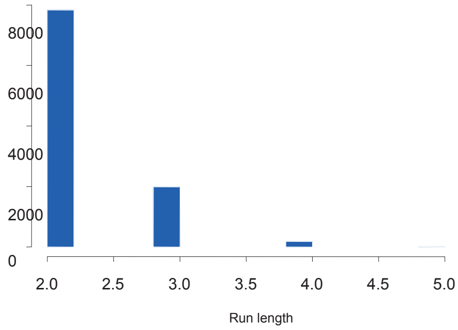
https://doi.org/10.26219/heal.aueb.6628

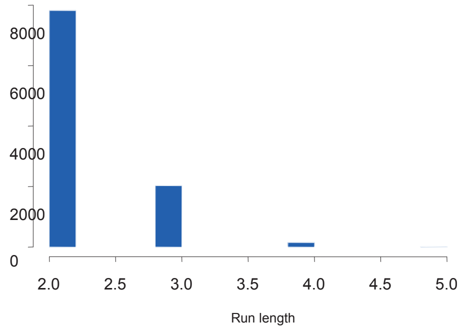


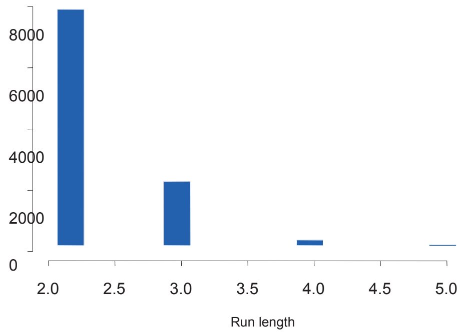


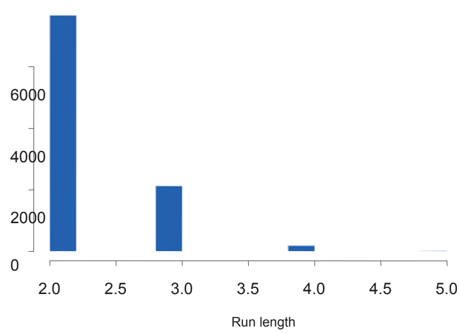




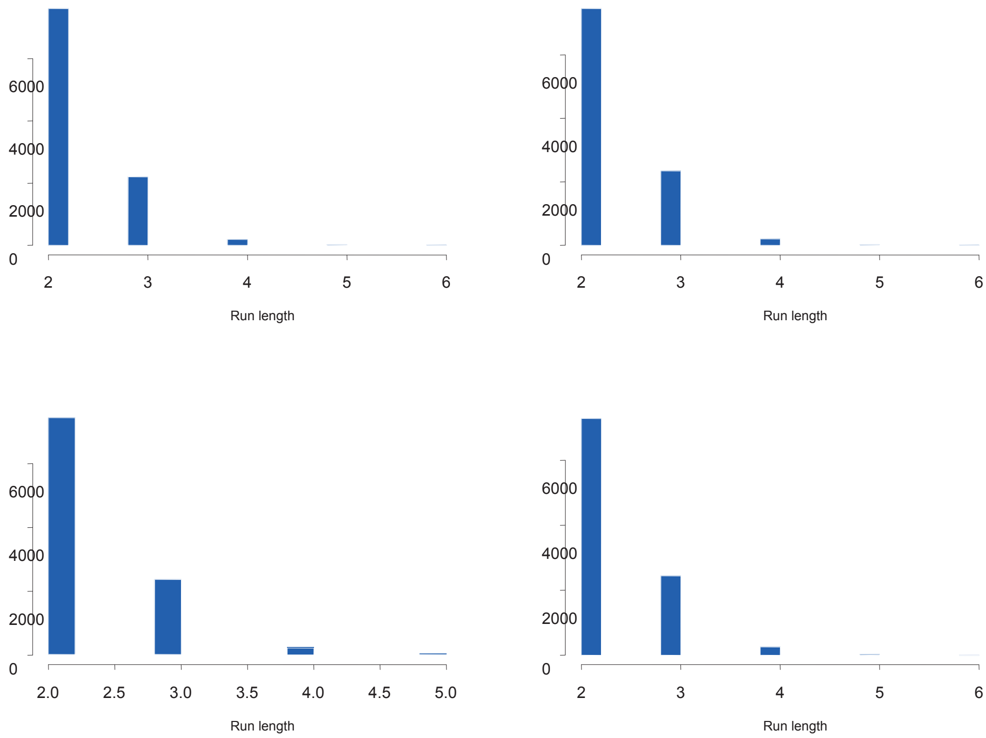


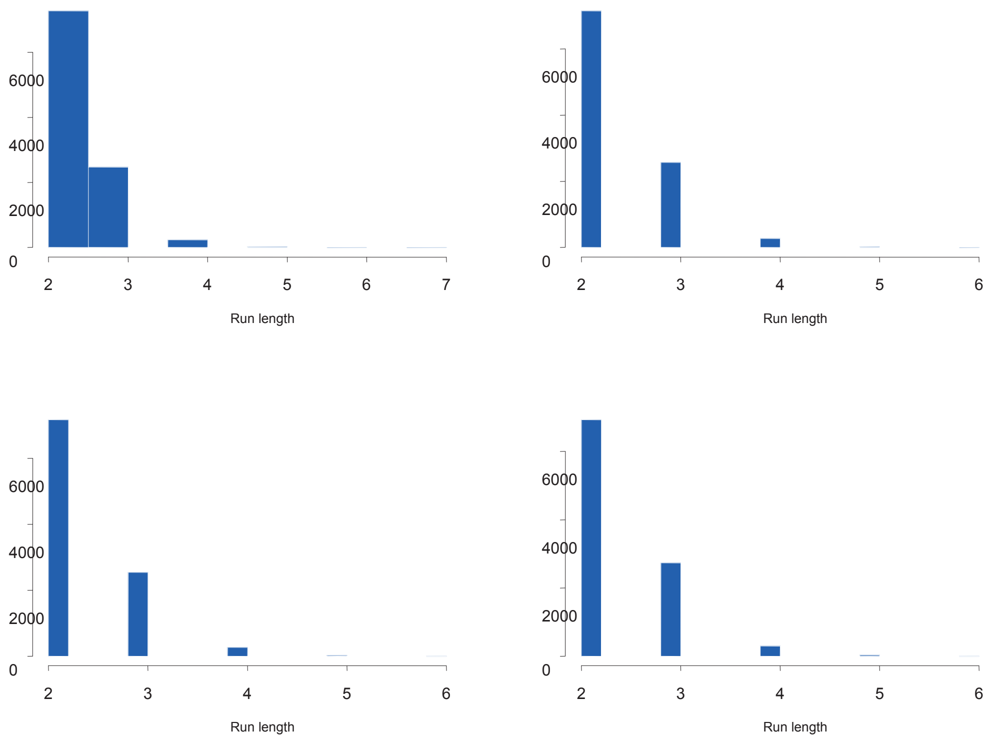






https://doi.org/10.26219/heal.aueb.6628





https://doi.org/10.26219/heal.aueb.6628

We will compare the way the Run Length for Bayesian CUSUM is distributed, with the corresponding histograms for the Classic CUSUM. But before that, we will make comparison between the out of control ARLs represented in Table 4.4 and the out of control ARLs for the Classic CUSUM, defined by Equation (2.8.10), for a specific pair (h, k) for which the classic CUSUM was designed.

We will obtain the out of control ARLs via simulations (S-Plus routine, see appendix). As we already know in the Classic CUSUM, the cutoff point h depends on the in control ARL and on the parameter k. Thus, we compute out of control ARLs for twenty different values of the parameter h. We will work again with the same three out of control parameter values $\theta_1 = 2$, $\theta_1 = 3$ and $\theta_1 = 4$. For each one of these values we derive a sample of $n_0 = 100$ normally distributed random variables. As we did for Bayesian CUSUM, we compute the out of control ARL for each one of the three cases, for the twenty different cutoff values. Each one of those ARLs came up as the mean of a vector of $N = 10000$ different run lengths per cutoff value. We also compute the last point for which Equation (2.8.10) was equal to zero. Thus, we compute n=instant at which the shift is detected (out of control ARL) and m= last time W_n was equal to zero before the shift was detect. Then we calculate the shift between those two instants, for which we would like to be very small. The simulation results are shown at Table 4.5 below:

https://doi.org/10.26219/heal.aueb.6628

Table 4.5: Out of Control ARL and distance $(n-m)$ for various cutoff points for three different out-of control parameter values for the Classic CUSUM

$k = \frac{\theta_0 + \theta_1}{2}$	$k = 1$		$k = 1.5$		$k = 2$	
H	**ARL**	**Average n-m**	**ARL**	**Average n-m**	**ARL**	**Average n-m**
1	2.7898	2.0167	2.365	1.986	2.1708	2.0013
1.125	2.8736	1.9797	2.4432	2.0155	2.2119	2.005
1.25	2.9894	1.9746	2.5132	2.0117	2.2533	2.0062
1.375	3.0999	2.0012	2.5825	1.9969	2.3001	2.0046
1.5	3.2318	2.0061	2.6551	2.0046	2.3406	2.0032
1.625	3.3677	2.0028	2.7218	1.9871	2.3913	1.9953
1.75	3.467	1.9863	2.842	2.0263	2.4538	2.0007
1.875	3.635	2.0308	2.9138	2.0152	2.5165	1.9861
2	3.7489	2.0168	2.9863	1.9978	2.5874	1.9969
2.125	3.8793	2.0207	3.0894	2.0094	2.6538	2.0034
2.25	3.9866	1.9849	3.1634	2.0028	2.7165	1.9935
2.375	4.1361	1.9968	3.2732	1.9999	2.783	2.0069
2.5	4.2624	2.02	3.3375	1.9961	2.8511	1.9992
2.625	4.3824	2.0037	3.4258	2.0057	2.9231	1.9967
2.75	4.4959	2.0048	3.5066	1.9975	3.001	2.0078
2.875	4.6259	2.0174	3.5881	1.9979	3.05	2.002
3	4.7804	2.0641	3.679	2.0196	3.1238	1.996
3.125	4.851	1.975	3.7664	1.9853	3.1861	2.004
3.25	5.0068	2.015	3.8594	2.0215	3.2344	1.9889
3.375	5.1245	2.0108	3.9296	1.9999	3.317	2.0098

The results for Classic CUSUM are similar to those of the Bayesian CUSUM. It is known that the Classic CUSUM is optimal for detecting small specific shifts of a parameter value, for which it was design. Thus, we aim to prove equivalence between the Classic CUSUM and the constructed Bayesian CUSUM. We have already shown that as far as the in control ARL is concerned, the two schemes are equivalent and

 https://doi.org/10.26219/heal.aueb.6628

their cutoff points satisfy Equation (4.1.5). Now we prove the equivalence for the out of control ARL.

We illustrate comparison between the two schemes of CUSUM for some values of parameter h and cutoff point in Table 4.6. As we can notice out of control ARL values are almost the same for the two forms of CUSUM for values of h and cutoff which satisfy the formula (4.1.5).

Table 4.6: Comparison for the out of control ARL of classic and Bayesian CUSUM for specific values of parameter h and cutoff point.

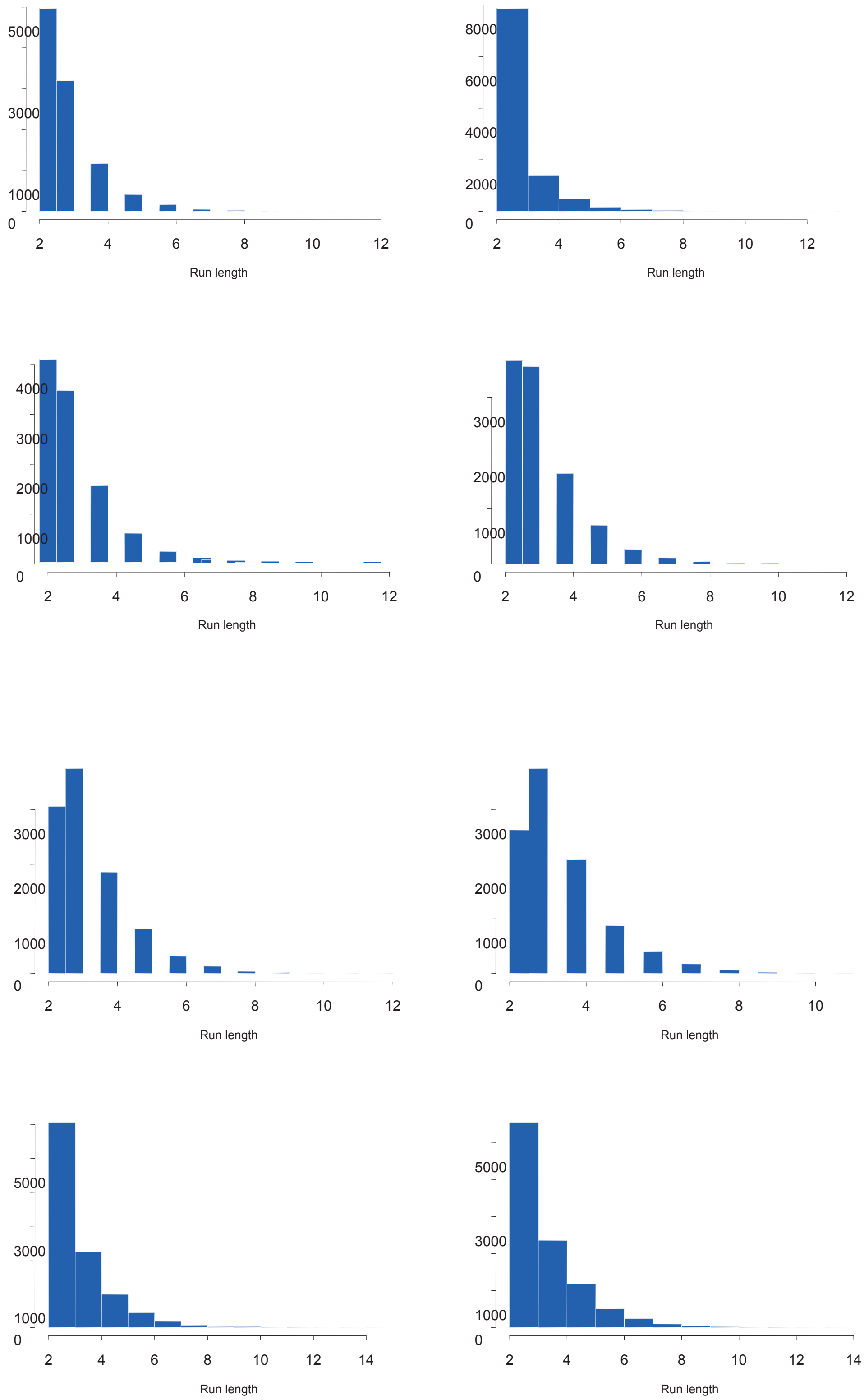
		h=1		h=1.250		h=1.5		h=1.750	
θ_1-θ_0	Cutoff	B.C	C.C	B.C	C.C	B.C	C.C	B.C	C.C
2	-2	2.7791	2.7898						
	-2.5			3.0025	2.9894				
	-3					3.2396	3.2318		
	-3.5							3.4914	3.467
3	-3	2.3836	2.365						
	-4.5					2.6596	2.6551		
4	-4	2.1728	2.1708						
	-5			2.2484	2.2533				

Finally we arrange histograms corresponding to the twenty different values of h for each one of the three different out of control parameter values, to show the way run length in Classic CUSUM is distributed.

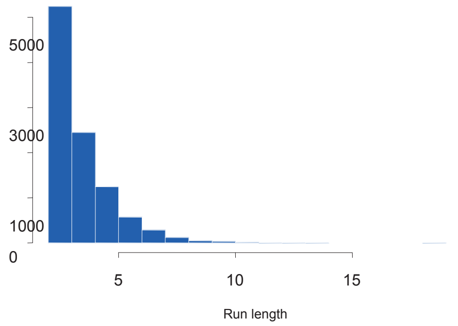
Following the same technique as we did for Bayesian CUSUM, we group histograms into three different categories, with respect to the out of control parameter value θ_1. Our intention is to compare the way the Run Length is distributed while it shifts amongst different values of cutoff point per parameter value, for the two schemes of CUSUM. The results are shown in Figures 4.2.4, 4.2.5 and 4.2.6:

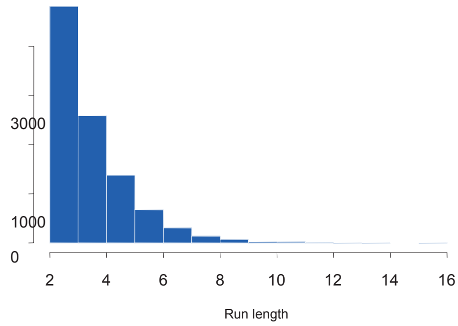
https://doi.org/10.26219/heal.aueb.6628

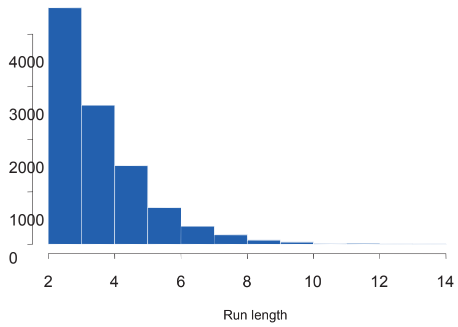
Figure 4.2.4: histograms of the quantities [run length] for $\theta_1 - \theta_0 = 2 - 0 = 2$, for the twenty different cutoff point values of Classic CUSUM.

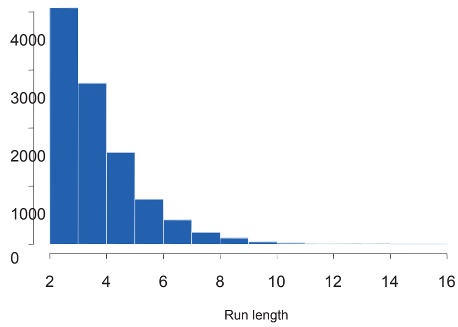


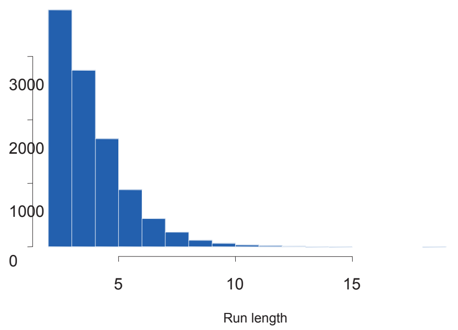
https://doi.org/10.26219/heal.aueb.6628

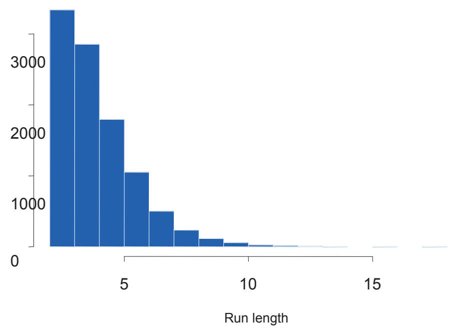


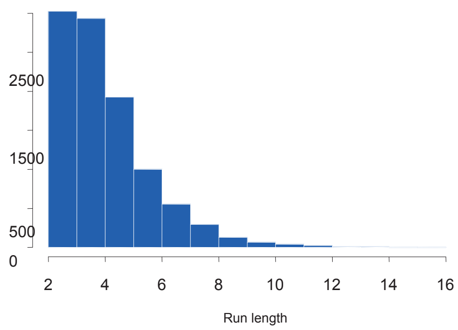


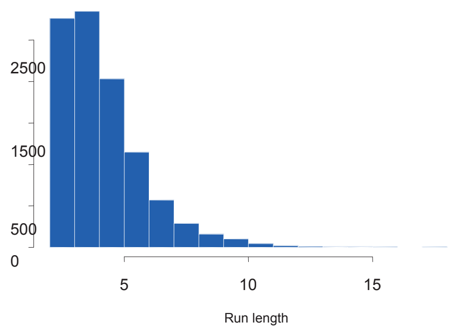




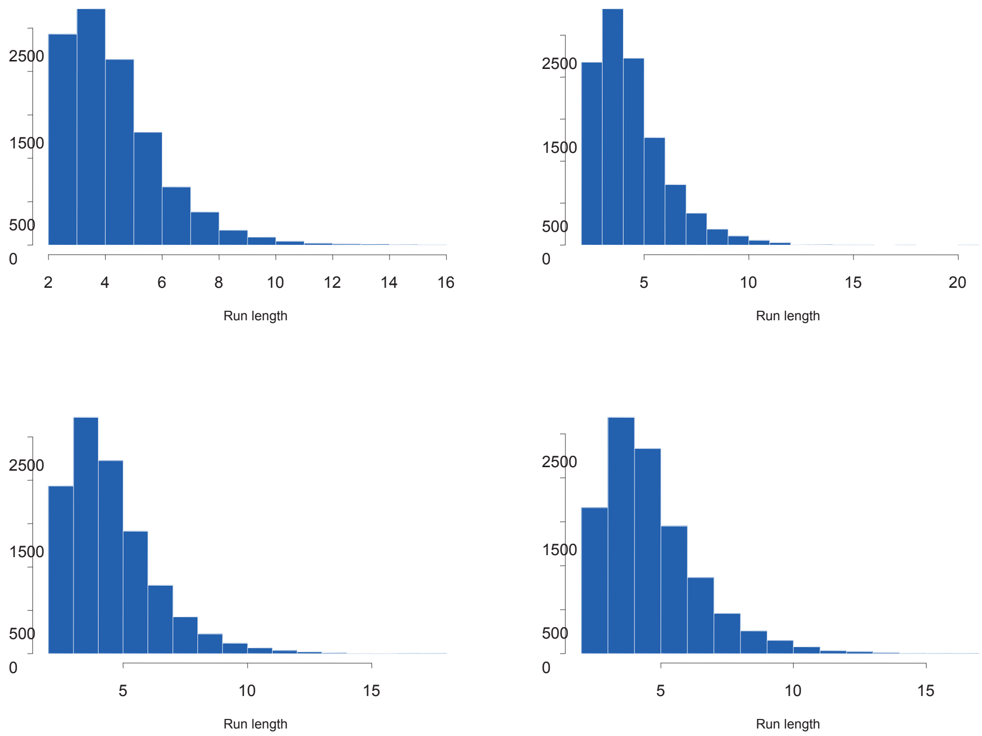






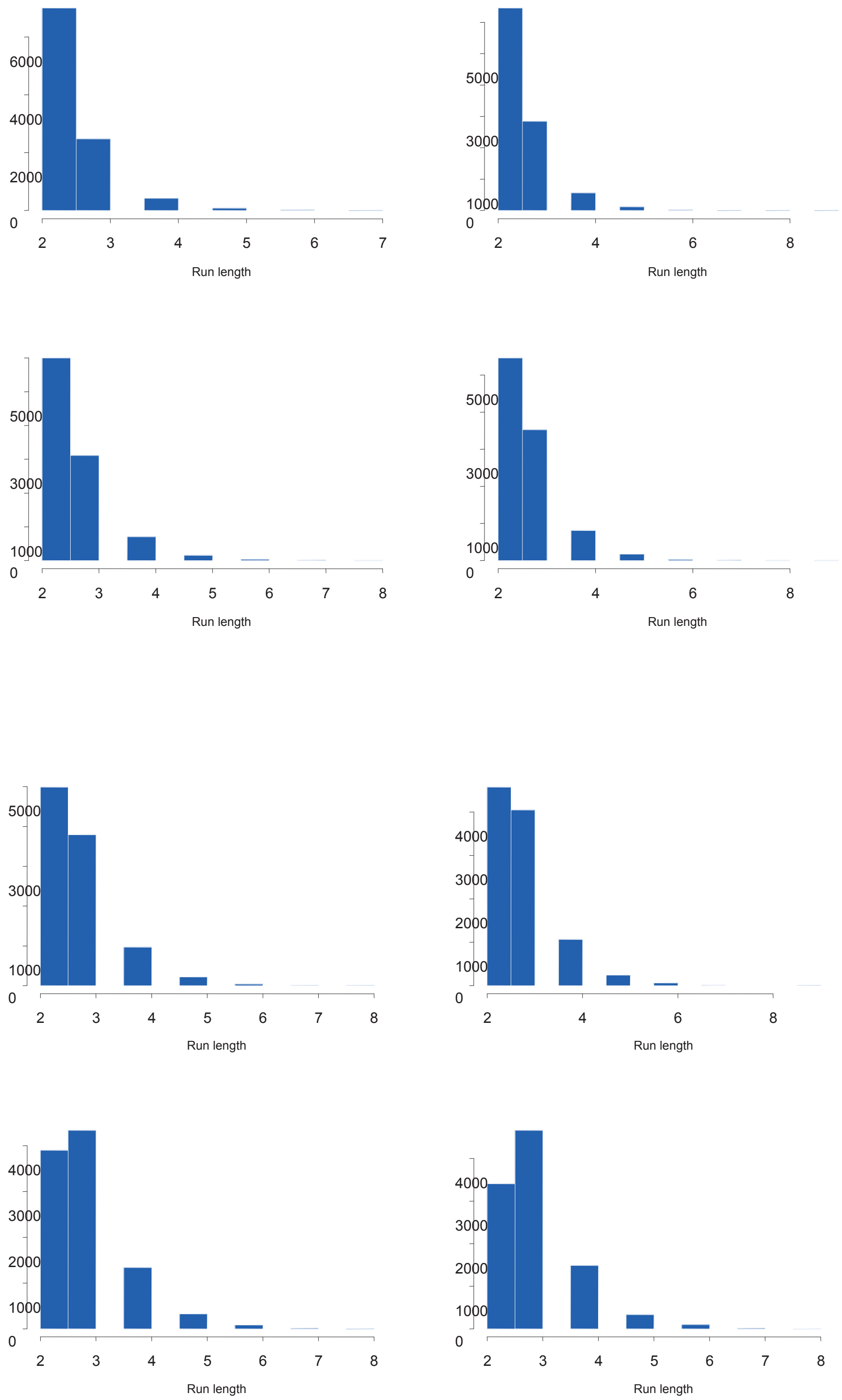


https://doi.org/10.26219/heal.aueb.6628

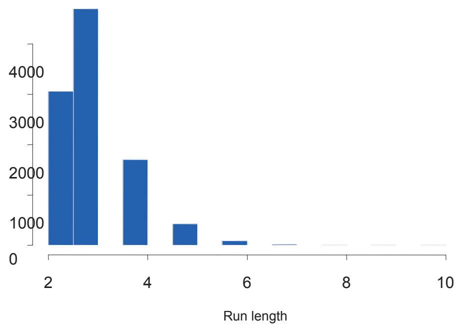


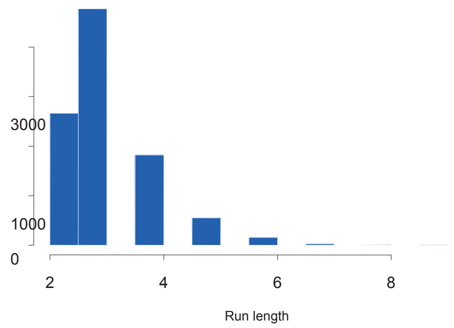
https://doi.org/10.26219/heal.aueb.6628

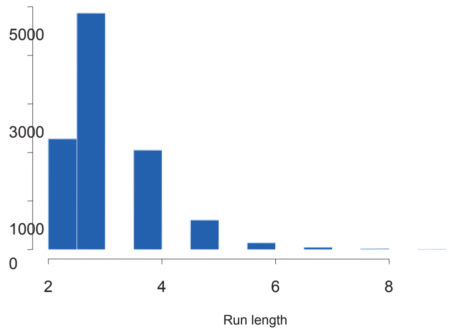
Figure 4.2.5: histograms of the quantities [run length] for $\theta_1 - \theta_0 = 3 - 0 = 3$, for the twenty different cutoff point values of Classic CUSUM.

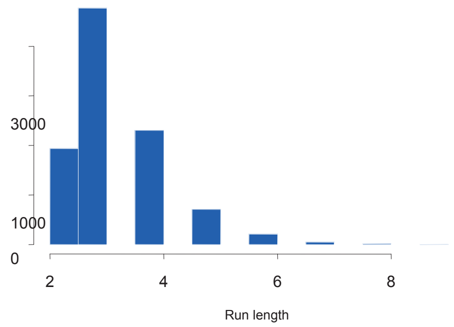


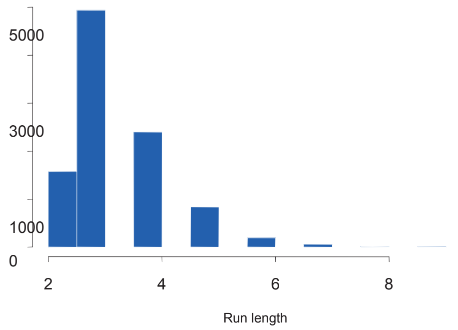
https://doi.org/10.26219/heal.aueb.6628

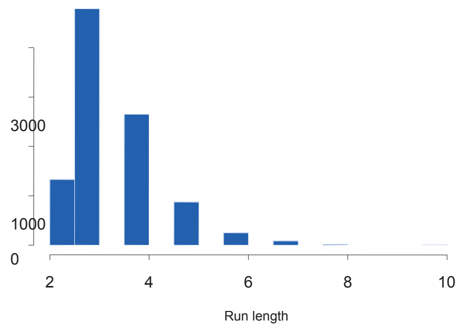


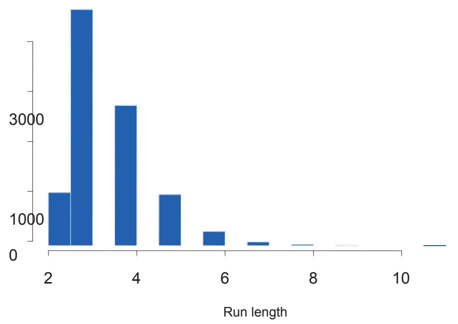


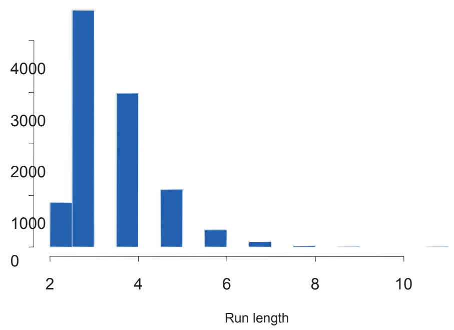




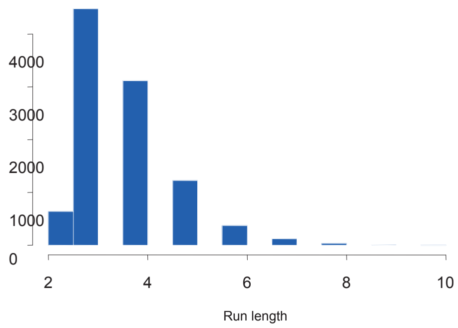


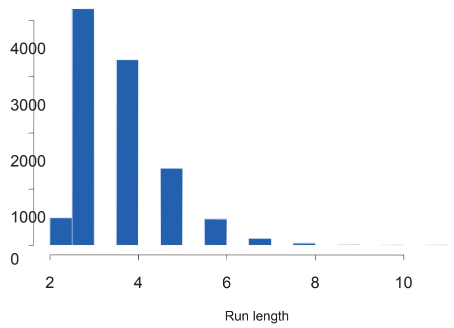


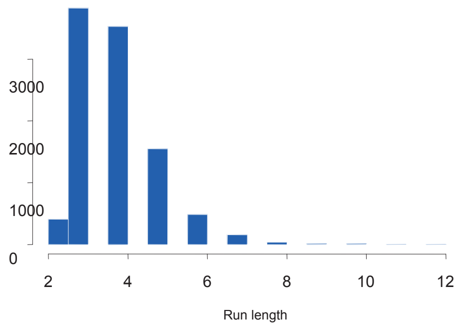


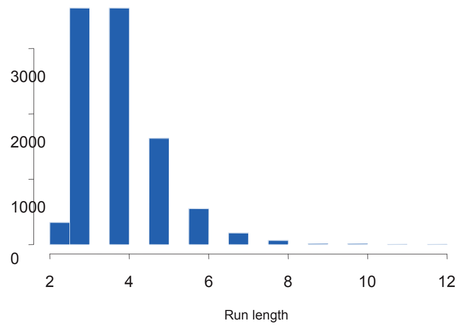


https://doi.org/10.26219/heal.aueb.6628



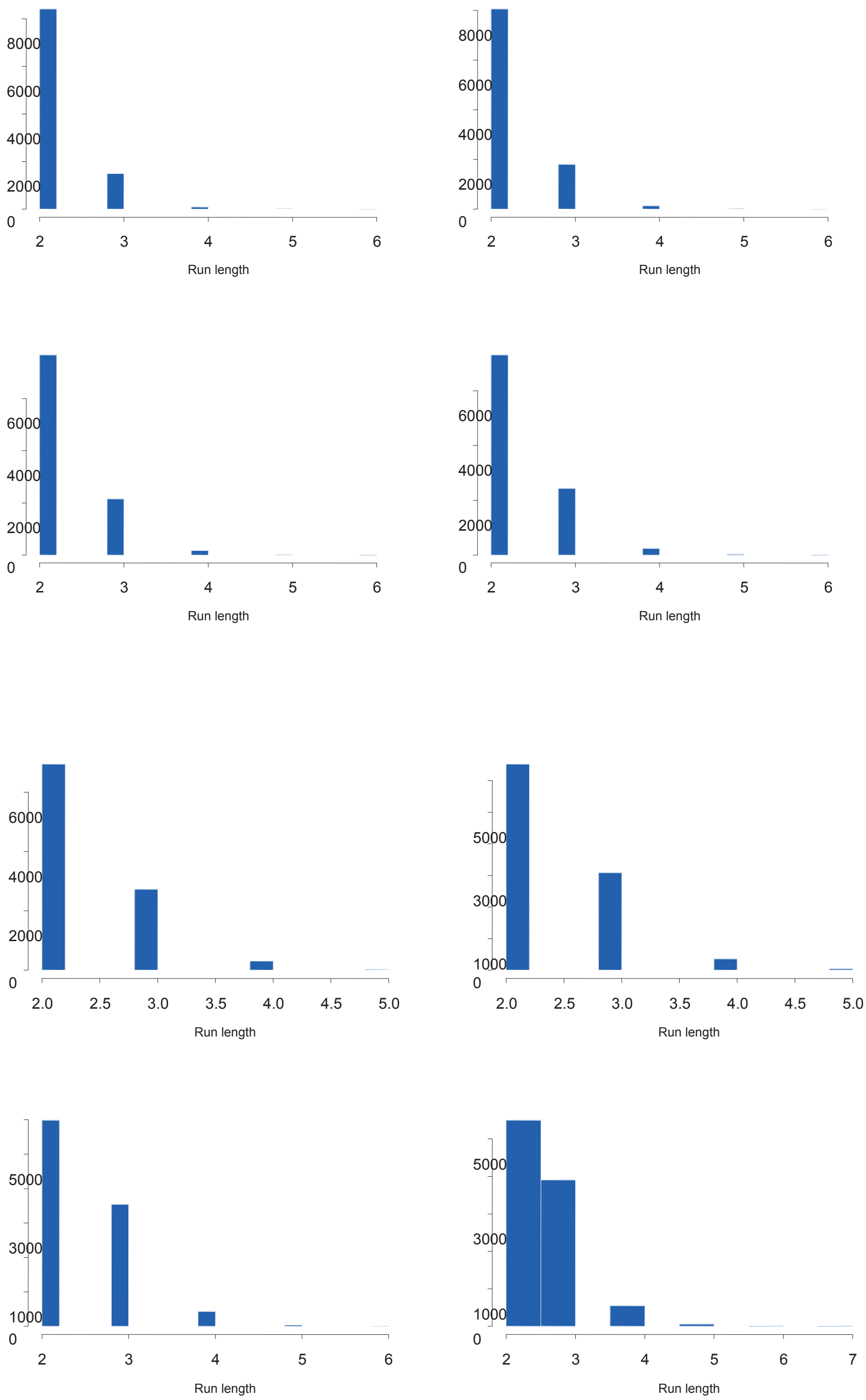




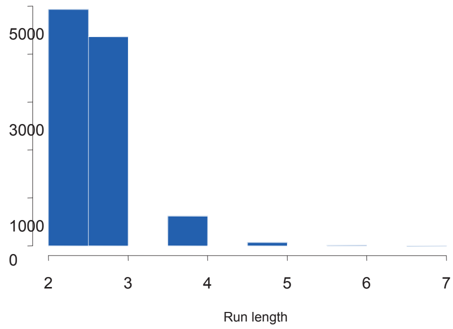


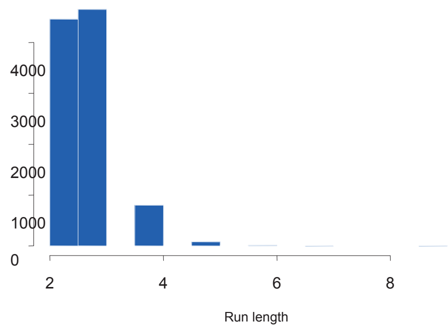
https://doi.org/10.26219/heal.aueb.6628

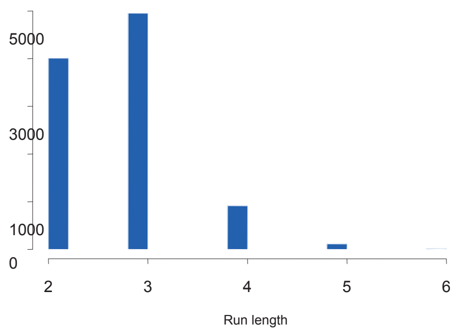
Figure 4.2.6: histograms of the quantities [run length-n_0] for $\theta_1 - \theta_0 = 4 - 0 = 4$, for the twenty different cutoff point values of Classic CUSUM.

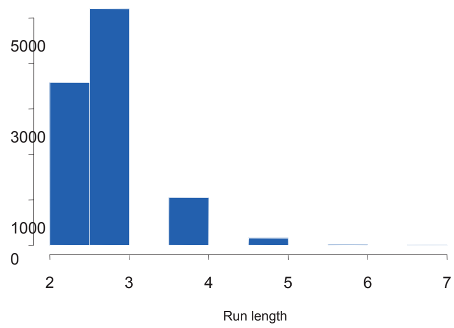


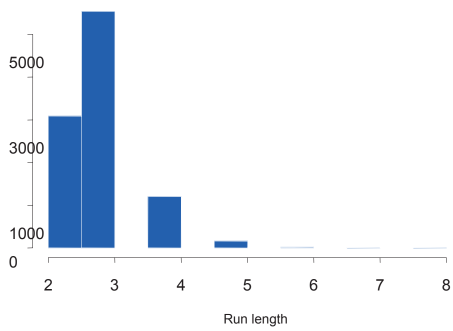
https://doi.org/10.26219/heal.aueb.6628

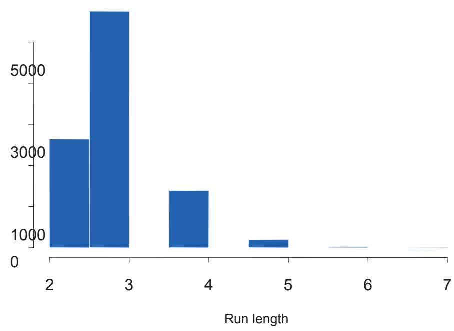


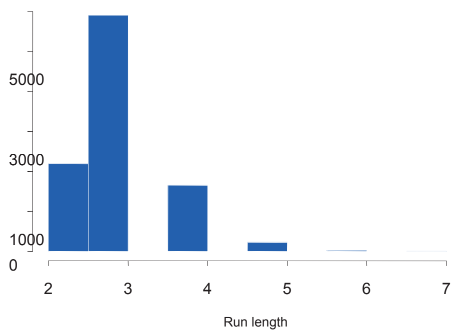


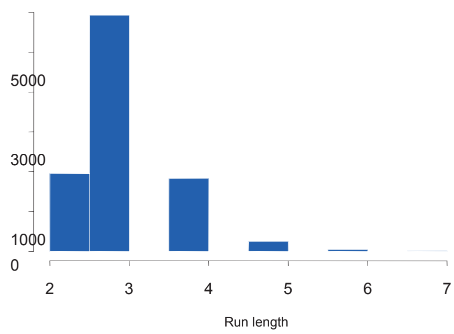




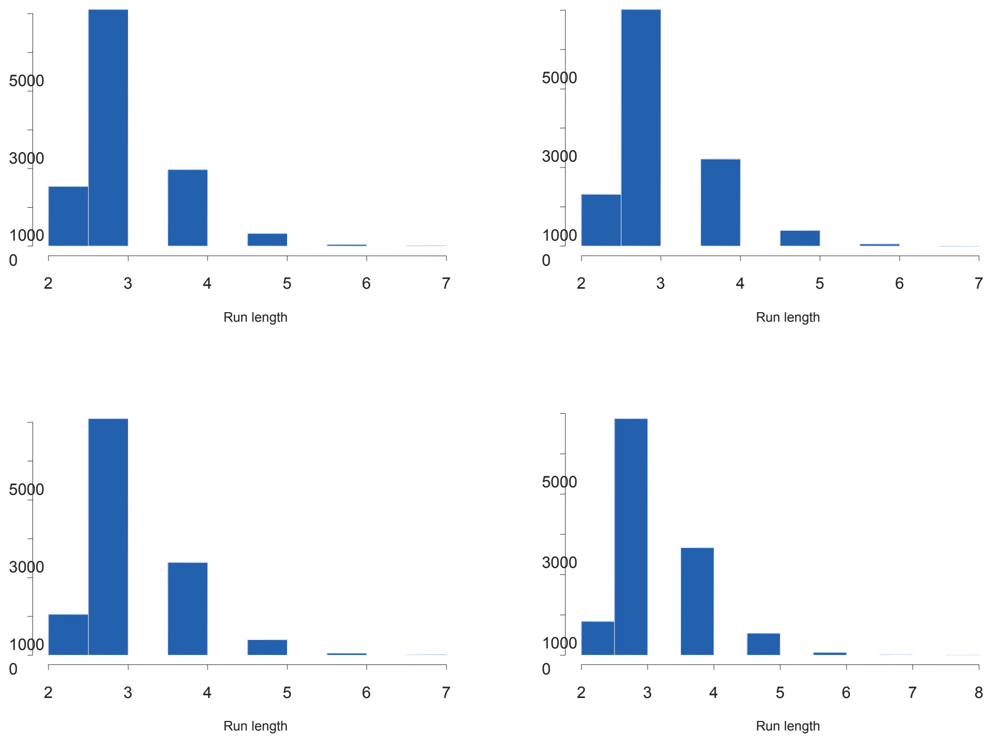








https://doi.org/10.26219/heal.aueb.6628



As we can notice from the above histograms, the way the Run Length is distributed seems to be similar for the respective cut point values that satisfy Equation (4.1.5), for each one of the three cases for the two schemes of CUSUM respectively. Furthermore Run Length has the geometric distribution, indeed for both of the forms of CUSUM (classic and Bayesian).

4.2 Composite versus Composite Case:

Working with the same approach as we did for the simple versus simple case, we will firstly select a cutoff value that gives an appropriate in control ARL and then we will test how quickly the Bayesian CUSUM will respond to an out of control situation. Contrary to the simple versus simple case, the Classic CUSUM for Composite versus Composite case will not give the optimal in control ARL, because, the optimality of the CUSUM is for detecting the shift to a particular out of control parameter value only. In Composite versus Composite case the null and alternative hypotheses are interval hypotheses, and CUSUM is not designed to detect a specific shift as quickly as possible in order to give the optimal in control ARL. Hence, we do

https://doi.org/10.26219/heal.aueb.6628

not expect that the cutoff values will be relative to the in control ARLs, computed for a specific distance between the in control parameter value and the out of control parameter value. We also do not expect such precise results as the ones we had for the Simple versus Simple case.

Suppose that: $x_1, x_2, .., x_n \sim N(\theta, \sigma^2)$, σ^2: known

$$f(x_i|\theta) = \frac{1}{\sigma\sqrt{2\pi}} exp\left(-\frac{(x_i-\theta)^2}{2\sigma^2}\right)$$

$$f(\underline{x}|\theta) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} exp\left(-\frac{(x_i-\theta)^2}{2\sigma^2}\right) =$$

$$\frac{1}{\left(\sigma\sqrt{2\pi}\right)^n} exp\left(-\frac{\sum_{i=1}^{n}(x_i-\theta)^2}{2\sigma^2}\right). \tag{4.2.1}$$

The a-priori distribution of the parameter θ is:

$\theta \sim N(q, p^2)$, q, p^2: known

$$\pi(\theta) = \frac{1}{p\sqrt{2\pi}} exp\left(-\frac{(\theta-q)^2}{2p^2}\right). \tag{4.2.2}$$

Applying (4.2.1) and (4.2.2) to theorem 3.1 (Bayes' Rule) we compute the posterior distribution for normal data.

Posterior Distribution:

The posterior distribution of normal data $x_1, x_2, .., x_n \sim N(\theta, \sigma^2)$, σ^2: known, with a-priori distribution $\theta \sim N(q, p^2)$, q, p^2: known, is given by:

$$f(\theta|x) \propto f(x|\theta)\pi(\theta)$$

where:

$$\pi(\theta) = \frac{1}{p\sqrt{2\pi}} exp\left(-\frac{(\theta-q)^2}{2p^2}\right)$$ Is the prior distribution, and

 https://doi.org/10.26219/heal.aueb.6628

$$f(\underline{x}|\theta) = \frac{1}{(\sigma\sqrt{2\pi})^n} exp\left(-\frac{\sum_{i=1}^{n}(x_i-\theta)^2}{2\sigma^2}\right)$$ Is the likelihood function.

Thus:

$$f(\theta|\underline{x}) \propto \; exp\left(-\frac{(\theta-q)^2}{2p^2}\right) exp\left(-\frac{\sum_{i=1}^{n}(x_i-\theta)^2}{2\sigma^2}\right) =$$

$$exp\left\{-\frac{(\theta-q)^2}{2p^2} - \frac{\sum_{i=1}^{n}(x_i-\theta)^2}{\sigma^2}\right\} =$$

$$exp\left\{-\frac{1}{2}\left[\frac{(\theta-q)^2}{p^2} + \frac{\sum_{i=1}^{n}(x_i-\theta)^2}{2\sigma^2}\right]\right\} =$$

$$exp\left\{-\frac{1}{2}\left[\frac{(\theta-q)^2}{p^2} - \frac{\sum_{i=1}^{n}(\theta-x_i)^2}{\sigma^2}\right]\right\}. \qquad (4.2.3)$$

It is known that:

$$\sum A_i(x-a_i)^2 = \left[\sum A_i\right]\left[x - \frac{\sum a_i A_i}{\sum A_i}\right]^2 + C.$$

The expression:

$$\frac{(\theta-q)^2}{p^2} - \frac{\sum_{i=1}^{n}(\theta-x_i)^2}{\sigma^2}$$

could be written as:

$$\sum_{i=0}^{n} A_i(\theta-a_i)^2 = \left[\sum_{i=0}^{n} A_i\right]\left[\theta - \frac{\sum_{i=0}^{n} a_i A_i}{\sum_{i=0}^{n} A_i}\right]^2 \qquad (4.2.4)$$

where:

$$A_0 = \frac{1}{p^2}, \qquad A_i = \frac{1}{\sigma^2}, i = 1, \dots, n$$

$$a_0 = q, \qquad a_i = x_i, i = 1, \dots, n.$$

https://doi.org/10.26219/heal.aueb.6628

Thus,

$$\frac{(\theta-q)^2}{p^2}-\frac{\sum_{i=1}^{n}(\theta-x_i)^2}{\sigma^2}=$$

$$=\left[\frac{1}{p^2}+\sum_{i=1}^{n}\frac{1}{\sigma^2}\right]\left[\theta-\frac{\frac{q}{p^2}+\frac{x_1}{\sigma^2}+\cdots+\frac{x_n}{\sigma^2}}{\frac{1}{p^2}+\sum_{i=1}^{n}\frac{1}{\sigma^2}}\right]^2=$$

$$\left[\frac{1}{p^2}+\frac{n}{\sigma^2}\right]\left[\theta-\frac{\frac{q}{p^2}+\frac{\overline{x}n}{\sigma^2}}{\frac{1}{p^2}+\frac{n}{\sigma^2}}\right]^2. \tag{4.2.5}$$

Thus, according to Equation (4.2.5), the posterior distribution of $\theta|x$ for normal data is proportional to:

$$f(\theta|\underline{x})\propto exp\left\{-\frac{1}{2}\left[\frac{1}{p^2}+\frac{n}{\sigma^2}\right]\left[\theta-\frac{\frac{q}{p^2}+\frac{\overline{x}n}{\sigma^2}}{\frac{1}{p^2}+\frac{n}{\sigma^2}}\right]^2\right\}.$$

Thus,

$$\theta|\overline{x}\sim N\left(\frac{\frac{q}{p^2}+\frac{n\overline{x}}{\sigma^2}}{\frac{1}{p^2}+\frac{n}{\sigma^2}},\frac{1}{\frac{1}{p^2}+\frac{n}{\sigma^2}}\right). \tag{4.2.6}$$

Hence, applying (4.2.6) to (3.3.8), we compute Bayes factor for Composite vs Composite case for normal data:

$$B_{01}(\underline{x})=\frac{\int_a^b N\left(\frac{\frac{q}{p^2}+\frac{n\overline{x}}{\sigma^2}}{\frac{1}{p^2}+\frac{n}{\sigma^2}},\frac{1}{\frac{1}{p^2}+\frac{n}{\sigma^2}}\right)d\theta}{\int_c^d N\left(\frac{\frac{q}{p^2}+\frac{n\overline{x}}{\sigma^2}}{\frac{1}{p^2}+\frac{n}{\sigma^2}},\frac{1}{\frac{1}{p^2}+\frac{n}{\sigma^2}}\right)d\theta}. \tag{4.2.7}$$

 https://doi.org/10.26219/heal.aueb.6628

In order to compute Bayes factor we have to standardize normal distribution. Thus if:

$$\theta|\overline{x} \sim N\left(\frac{\frac{q}{p^2}+\frac{n\overline{x}}{\sigma^2}}{\frac{1}{p^2}+\frac{n}{\sigma^2}}, \frac{1}{\frac{1}{p^2}+\frac{n}{\sigma^2}}\right).$$

Then,

$$z|\overline{x} = \frac{\theta - \left(\frac{\frac{q}{p^2}+\frac{n\overline{x}}{\sigma^2}}{\frac{1}{p^2}+\frac{n}{\sigma^2}}\right)}{\frac{1}{\frac{1}{p^2}+\frac{n}{\sigma^2}}} \sim N(0,1). \tag{4.2.8}$$

In addition to this, if $\theta \in \Theta_0 = [a,b]$, so that $a < \theta < b$, then $z \in [a_1, b_1]$

$$a_1 = \frac{a - \left(\frac{\frac{q}{p^2}+\frac{n\overline{x}}{\sigma^2}}{\frac{1}{p^2}+\frac{n}{\sigma^2}}\right)}{\frac{1}{\frac{1}{p^2}+\frac{n}{\sigma^2}}} < \frac{\theta - \left(\frac{\frac{q}{p^2}+\frac{n\overline{x}}{\sigma^2}}{\frac{1}{p^2}+\frac{n}{\sigma^2}}\right)}{\frac{1}{\frac{1}{p^2}+\frac{n}{\sigma^2}}} < \frac{b - \left(\frac{\frac{q}{p^2}+\frac{n\overline{x}}{\sigma^2}}{\frac{1}{p^2}+\frac{n}{\sigma^2}}\right)}{\frac{1}{\frac{1}{p^2}+\frac{n}{\sigma^2}}} = b_1. \tag{4.2.9}$$

Similarly, if $\theta \in \Theta_1 = [c,d]$, so that c$< \theta < d$, then $z \in [c_1, d_1]$

$$c_1 = \frac{c - \left(\frac{\frac{q}{p^2}+\frac{n\overline{x}}{\sigma^2}}{\frac{1}{p^2}+\frac{n}{\sigma^2}}\right)}{\frac{1}{\frac{1}{p^2}+\frac{n}{\sigma^2}}} < \frac{\theta - \left(\frac{\frac{q}{p^2}+\frac{n\overline{x}}{\sigma^2}}{\frac{1}{p^2}+\frac{n}{\sigma^2}}\right)}{\frac{1}{\frac{1}{p^2}+\frac{n}{\sigma^2}}} < \frac{d - \left(\frac{\frac{q}{p^2}+\frac{n\overline{x}}{\sigma^2}}{\frac{1}{p^2}+\frac{n}{\sigma^2}}\right)}{\frac{1}{\frac{1}{p^2}+\frac{n}{\sigma^2}}} = d_1. \tag{4.2.10}$$

Thus, applying (4.2.8),(4.2.9) and (4.2.10) to (4.2.7) we get:

$$B_{01}(\underline{x}) = \frac{\int_{a_1}^{b_1} N(0,1)d\theta}{\int_{c_1}^{d_1} N(0,1)d\theta} = \frac{\Phi(b_1) - \Phi(a_1)}{\Phi(d_1) - \Phi(c_1)}. \tag{4.2.11}$$

Working with $\ln\left(B_{01}(\underline{x})\right)$ we have that:

https://doi.org/10.26219/heal.aueb.6628

$$\ln\left(B_{01}(\underline{x})\right) = \ln\big(\Phi(b_1) - \Phi(a_1)\big) - \ln\big(\Phi(d_1) - \Phi(c_1)\big). \tag{4.2.12}$$

Hence,

$$W_n = \ln\left(\int_a^b p(\theta|x_1, x_2, \dots, x_n)d\theta\right) - \ln\left(\int_c^d p(\theta|x_1, x_2, \dots, x_n)d\theta\right) =$$

$$= \ln\big(\Phi(b_1) - \Phi(a_1)\big) - \ln\big(\Phi(d_1) - \Phi(c_1)\big) \tag{4.2.13}$$

signals if and only if:

$$W_n < \text{cutoff}.$$

4.2.1 Assessment of the in Control ARL

In order to define in control ARLs, we use software and specifically an S-Plus routine. We consider the in control situation $\theta_0 \in [\alpha, b]$ and known variance $\sigma^2 = 1$ and the out of control situation $\theta_1 \in [c, d]$. We compute the in control ARL for various values of the parameters α, b, c and d. We work with prior distribution $N(0,1)$ for the parameter θ.

As it comes out from simulations, for interval hypotheses, the in control ARL is extremely high, (most of the times gets greater than 10000). And that is for any value of cutoff point we have tested for. This probably happens because of the fact that the quantities a_1, b_1, c_1 and d_1, become very informative while data $\{x_i\}$ are updated for $i = 1, .., n$. As i grows the parameters a_1, b_1, c_1 and d_1 obtain either very big positive or very big negative values (usually a_1 has a large negative value and the other have large positives). That is because as long as $i = 1, .., n$ is getting bigger the denominator of the parameters a_1, b_1, c_1 and d_1, is getting very small. Applying prior distribution $N(0,1)$ to the quantities a_1, b_1, c_1 and d_1 and setting $\sigma^2 = 1$ we have:

While: $n \uparrow$ then the denominator is $\dfrac{1}{\frac{1}{p^2} + \frac{n}{\sigma^2}} = \dfrac{1}{1+n} \downarrow$

https://doi.org/10.26219/heal.aueb.6628

Thus,

$$a_1 = \frac{a - \frac{n\overline{x}}{\frac{1}{1+n}}}{\frac{1}{1+n}}, b_1 = \frac{b - \frac{n\overline{x}}{\frac{1}{1+n}}}{\frac{1}{1+n}}, \; c_1 = \frac{c - \frac{n\overline{x}}{\frac{1}{1+n}}}{\frac{1}{1+n}}, \; d_1 = \frac{d - \frac{n\overline{x}}{\frac{1}{1+n}}}{\frac{1}{1+n}}.$$

The parameters: a_1, b_1, c_1 and d_1 obtain large absolute values. It is reasonable for a_1 to become negative because $\overline{x}$ is certainly greater than a. Data $\{x_i\}$are being derived from $N(\theta_0, 1)$, $\theta_0 \in [a, b]$. Thus, most of the times $\Phi(\, b_1), \Phi(\, c_1), \Phi(\, d_1) = 1$ and $\Phi(\, a_1) = 0$.

Hence, we usually have to deal the problem that,

$$\Phi(\, b_1) - \Phi(\, a_1) = 1 \text{ and } \Phi(\, d_1) - \Phi(\, c_1) = 0.$$

Consequently,

$$\ln\big(\Phi(b_1) - \Phi(a_1)\big) - \ln\big(\Phi(d_1) - \Phi(c_1)\big) = 0 - (-\infty) = \infty.$$

Thus,

$$\ln\left(B_{01}\left(\underline{x}\right)\right) = \ln\big(\Phi(b_1) - \Phi(a_1)\big) - \ln\big(\Phi(d_1) - \Phi(c_1)\big)$$

never becomes smaller than cutoff value for normal data. We end up to enormous in control ARLs which are meaningless to be studied for CUSUM scheme. Table 4.7 provides suggestively results for the values of a_1, b_1, c_1 and d_1, and for the in control ARL of Composite versus Composite case. Simulations were planned to consider for run lengths greater than 1000, as equal to 1000. We consider as the in control situation, the interval hypothesis $\theta \in [-3.5, -3]$ and as the out of control situation $\theta \in [1.5,2]$

https://doi.org/10.26219/heal.aueb.6628

Table 4.7: Suggestive results for the values of the parameters a_1, b_1, c_1, d_1 and the in control ARL respectively for Composite vs Composite case

	cutoff=-3			cutoff=-4		
a_1	-279.983	-6.14083	-302.963	-249.682	-296.524	-306.055
b_1	220.0167	-5.14083	197.0375	250.3179	203.4756	193.9454
c_1	4720.017	3.859169	4697.037	4750.318	4703.476	4693.945
d_1	5220.017	4.859169	5197.037	5250.318	5203.476	5193.945
ARL	1000	1000	1000	1000	1000	1000

As we have already seen, in the Classic CUSUM, the problem of an interval hypothesis for the out of control situation is being overcome by selecting a particular out of control state, thus, the construction of the CUSUM turns to be a Simple versus Simple case. As a matter of fact we select the particular shift for which we would like the CUSUM to be optimal. Because of the robustness of CUSUM, it will nearly be as well as the optimal CUSUM for all shifts close to the particular shift point of the alternative hypothesis (Hawkins and Olwell (1997)). Hence, we can use the same CUSUM as the optimal scheme to detect a small area of shifts in mean, lying near to that particular value we have selected.

Following the same technique for the Bayesian CUSUM, we could select particular representative points from the two interval hypotheses and turn the problem to be the construction of a Simple versus Simple case Bayesian CUSUM.

4.3 Simple versus Composite Case:

Suppose once again that we have to deal with normally distributed data: $x_1, x_2, .., x_n \sim N(\theta, \sigma^2)$, σ^2: known. The likelihood function $f(\underline{x}|\theta) \sim N(\theta, \sigma^2)$ is given by Equation (4.2.1):

$$f(\underline{x}|\theta) = exp\left(-\frac{\sum_{i=1}^{n} x_i^2}{2\sigma^2} + \frac{\theta \sum_{i=1}^{n} x_i}{\sigma^2} - \frac{n\theta^2}{2\sigma^2} - n\ln\left(\sigma\sqrt{2\pi}\right)\right)$$

where:

https://doi.org/10.26219/heal.aueb.6628

$$b(\theta) = \theta$$

$$\sum_{i=1}^{n} a(x_i) = \frac{\sum_{i=1}^{n} x_i}{\sigma^2}$$

$$\sum_{i=1}^{n} c(x_i) = -\left(\frac{\sum_{i=1}^{n} x_i^2}{2\sigma^2} + n\ln\left(\sigma\sqrt{2\pi}\right)\right)$$

$$nd(\theta) = -\frac{n\theta^2}{2\sigma^2}$$

the prior distribution of the parameter θ, $\pi(\theta) \sim N(q, v^2)$ is given by (4.2.2):

$$\pi(\theta) = \exp\left(-\frac{\theta^2}{2v^2} + \frac{\mathrm{q}\theta}{v^2} - \ln\left(\mathrm{p}\sqrt{2\pi}\right) - \frac{q^2}{2v^2}\right)$$

where:

$$a^{'}(\theta) = \frac{\theta}{v^2}$$

$$b^{'}(q) = q$$

$$c^{'}(\theta) = -\frac{\theta^2}{2p^2} - \ln\left(v\sqrt{2\pi}\right)$$

$$d^{'}(q) = -\frac{q^2}{2v^2}$$

and the posterior distribution of the unknown parameter $\theta|\overline{x}$, is then given by (4.2.6):

$$\theta|\overline{x} \sim N\left(\frac{\frac{q}{v^2} + \frac{n\overline{x}}{\sigma^2}}{\frac{1}{v^2} + \frac{n}{\sigma^2}}, \frac{1}{\frac{1}{v^2} + \frac{n}{\sigma^2}}\right).$$

As we have shown in chapter 3, the Bayesian CUSUM for any member of the exponential family, for Simple versus Composite case is given by:

$$W_n = min\left[0, b(\theta_0)\sum_{i=1}^{n} a(x_i) + \sum_{i=1}^{n} c(x_i) + nd(\theta_0) - \ln\left(\frac{f_x(\underline{x})}{1-p}\right) - \ln\left[\int_{\theta \neq \theta_0} p(\theta|\underline{x})\, d\theta\right]\right]$$

where:

$$f_x(\underline{x}) = pexp\left\{b(\theta_0)\sum_{i=1}^{n} a(x_i) + \sum_{i=1}^{n} c(x_i) + nd(\theta_0)\right\} +$$

$$+(1-p)\ \ Cexp\left\{\sum_{i=1}^{n} c(x_i) + d^{'}(q)\right\}.$$

We have also shown that C is the normalizing constant of the posterior distribution. Thus, for normally distributed data with $\pi(\theta) \sim N(q, v^2)$, prior distribution:

$$C = \frac{1}{\sqrt{\frac{1}{v^2} + \frac{n}{\sigma^2}}\sqrt{2\pi}}. \tag{4.3.1}$$

For the computation of the quantity $\int_{\theta \neq \theta_0} p(\theta|\underline{x})\, d\theta$ we integrate the posterior distribution for the entire parameter space Θ, excluding a very small area around the out of control parameter value θ_0, say $[\theta_0 - \varepsilon, \theta_0 + \varepsilon]$. Thus, for $\varepsilon \to 0$,

$$\int_{\theta_0-\varepsilon}^{\theta_0+\varepsilon} p(\theta|\underline{x})\, d\theta = 1 - \Phi(\theta_0 + \varepsilon) + \Phi(\theta_0 - \varepsilon) \xrightarrow{\varepsilon \downarrow 0} 1. \tag{4.3.2}$$

Finally we set:

$$b(\theta_0)\sum_{i=1}^{n} a(x_i) + \sum_{i=1}^{n} c(x_i) + nd(\theta_0) =$$

$$= \frac{\theta_0 \sum_{i=1}^{n} x_i}{\sigma^2} - \frac{\sum_{i=1}^{n} x_i^2}{2\sigma^2} - n\ln\left(\sigma\sqrt{2\pi}\right) - \frac{n{\theta_0}^2}{2\sigma^2} = A^*. \tag{4.3.3}$$

Applying (4.3.1), (4.3.2) and (4.3.3) to Bayesian CUSUM scheme W_n, for Simple versus Composite case we get:

$$W_n = A^* - \ln\left(\frac{pexp\{A^*\} + (1-p)\dfrac{1}{\sqrt{\frac{1}{v^2} + \frac{n}{\sigma^2}}\sqrt{2\pi}}\, exp\left\{-\frac{\sum_{i=1}^{n} x_i^2}{2\sigma^2} - n\ln\left(\sigma\sqrt{2\pi}\right) - \frac{q^2}{2v^2}\right\}}{1-p}\right) - \ln 1$$

signaling if $W_n <$ cutoff.

4.3.1 Assessment of the in Control ARL

The Simple versus Composite case, as a special case of interval hypotheses testing, faces the same problem of enormous in control ARLs (most of the times greater than 1000), as the Composite versus Composite case does. This is probably happening for the same reason as it does for Composite versus Composite case too. After a large number of observations $\{x_i\}$, the logarithm of the Bayes factor becomes very informative. Thus, the type I error (the error of falsely rejecting H_0) becomes very small and as a consequence we get huge in control ARLs.

To define the in control ARL we consider the hypothesis testing:

$$\begin{Bmatrix} H_0 : \theta = 0 \\ H_1 : \theta \neq 0 \end{Bmatrix}$$

Thus, without loss of generality we derive random variables $x_i \sim N(0,1)$ and we regard the prior distribution of parameter θ, as $\pi(\theta) = N(q, v^2) \equiv N(0,1)$ too.

We apply the Bayesian CUSUM test:

$$W_n = A^* - \ln\left(\frac{p exp\{A^*\} + (1-p)\dfrac{1}{\sqrt{\dfrac{1}{v^2} + \dfrac{n}{\sigma^2}}\sqrt{2\pi}} exp\left\{-\dfrac{\sum_{i=1}^{n} x_i^2}{2\sigma^2} - n\ln\left(\sigma\sqrt{2\pi}\right) - \dfrac{q^2}{2v^2}\right\}}{1-p} \right) - \ln 1$$

signaling if $W_n <$ cutoff.

Obviously, from Equation (4.3.3), as n grows, A^* gets large negative values. Thus,

$$\exp(A^*) \to 0$$

and also,

$$\exp\left(-\frac{\sum_{i=1}^{n} x_i^2}{2\sigma^2} - n\ln\left(\sigma\sqrt{2\pi}\right) - \frac{q^2}{2v^2}\right) \overset{n\uparrow}{\to} 0.$$

Thus,

$$\frac{p exp\{A^*\} + (1-p)\dfrac{1}{\sqrt{\dfrac{1}{v^2} + \dfrac{n}{\sigma^2}}\sqrt{2\pi}} exp\left\{-\dfrac{\sum_{i=1}^{n} x_i^2}{2\sigma^2} - n\ln\left(\sigma\sqrt{2\pi}\right) - \dfrac{q^2}{2v^2}\right\}}{1-p} \longrightarrow 0$$

https://doi.org/10.26219/heal.aueb.6628

Hence,

$$\ln\left(\frac{pexp\{A^*\}+(1-p)\dfrac{1}{\sqrt{\dfrac{1}{v^2}+\dfrac{n}{\sigma^2}}\sqrt{2\pi}}exp\left\{-\dfrac{\sum_{i=1}^n x_i^2}{2\sigma^2}-n\ln\left(\sigma\sqrt{2\pi}\right)-\dfrac{q^2}{2v^2}\right\}}{1-p}\right)\to \ln(0).$$

Consequently,

$$W_n = A^* - (-\infty) - 0 = +\infty$$

never becomes smaller than cutoff value for normal data. That is the reason why we obtain large ARLs.

https://doi.org/10.26219/heal.aueb.6628

https://doi.org/10.26219/heal.aueb.6628

Chapter 5

Conclusions

The main intention of this thesis was to provide an alternative form of Classic CUSUM scheme, following the Bayesian approach. Generally CUSUM's theory considers the in control situation of a process as the null hypothesis and the out of control situation as the alternative hypothesis of a hypothesis testing problem respectively. Considering hypothesis testing problem as a decision problem and using the Bayes factor as the test tool, via Bayes test, we have obtained the Bayesian form of CUSUM.

Taking into consideration the well known optimality properties of classic CUSUM in the Simple versus Simple case, we interested in proving the equivalence between the two forms of CUSUM.

Following the Bayes test theorem, we splot the hypothesis test into three different cases:

- Simple versus simple test

$$\begin{Bmatrix} H_0: \theta = \theta_0 \\ H_1: \theta = \theta_1 \end{Bmatrix}$$

- Simple versus composite test

$$\begin{cases} H_0: \theta = \theta_0, \theta \in \Theta_0 & \text{with a-priori probability density } \pi_0(\theta) \\ H_1: \theta \neq \theta_0, \theta \in \Theta_1 = \Theta_0^c & \text{with a-priori probability density } \pi_1(\theta) \end{cases}$$

- Composite versus composite test

$$\begin{cases} H_0: \theta \in \Theta_0 & \text{with a-priori probability density } \pi_0(\theta) \\ H_1: \theta \in \Theta_1 = \Theta_0^c & \text{with a-priori probability density } \pi_1(\theta) \end{cases}$$

Based on the above, we constructed three different forms of Bayesian CUSUM, one for each distinct case of test respectively.

Then, using simulations for normal data, we tested that the Bayesian CUSUM we have constructed with the Classic CUSUM for the Simple versus Simple case. The two forms of CUSUM resulted similar in and out of control ARLs for appropriate cut point values. In other words we have proved that the Bayesian CUSUM for Simple versus Simple case gives the optimal in control ARL for a particular predefined mean shift and for a specific cutoff value, as long as the appropriate cutoff value is selected. We have also proved that for that particular cutoff value, Bayesian CUSUM has a quick detection for specific out of control shifts.

Classic CUSUM does not deal with hypothesis problem for which either one or both hypotheses are interval hypotheses. Instead of that, the theory of Classic CUSUM solves that kind of problems by selecting a particular in control state and a particular out of control state, and reducing the problem into a Simple versus Simple case. Thus, there is no Classic form of CUSUM, dealing with interval hypotheses to compare with the Bayesian form of CUSUM, we have constructed.

As far as the other two cases of hypothesis testing are concerned, for large number of observations $\{x_i\}$, the logarithm of Bayes factor turned out to be very informative. Hence, we ended up to huge in control ARLs.

For future work in order to deal with the problem of large in control ARLs, we could use an appropriate window of data such as to avoid large values of n, greater than a threshold. Thus, we will be able to surpass the fact of a very small variance, and consequently we will avoid enormous in control ARLs.

https://doi.org/10.26219/heal.aueb.6628

Appendix

S-Plus Routines used in simulations for the three cases of hypothesis testing.

1. Simple versus Simple case

1.1 S-Plus routine for the in control ARL

```
f<-function(d0,d1,s,n,cut) {

rl<-NULL

for (k in 1:n) {

w1<-0

w<-0

i<-1

while (w >=cut ) {

x<-rnorm(1,d0,s)

l<-((x*(d0-d1))/(s^2)+((d1^2)-(d0^2))/(2*(s^2)))

w1<-w+l

w<-min(0,w1)

i<-i+1

}

rl<-c(rl,i)

arl<-mean(rl)}

list(rl=rl,arl=arl)

}

t<-seq(-2,-5.9,-0.1)

bayesian.in.control.arl<-function(d0,d1,s,n) {
```

https://doi.org/10.26219/heal.aueb.6628

```
v<-rep(NA,n*length(t))

m<-matrix(v,ncol=n,nrow=length(t))

arl<-NULL

for (j in 1:length(t)) {

m[j,]<-f(d0,d1,s,n,t[j])$rl

arl<-c(arl,mean(m[j,]))

}

data.frame(cut=t,ARL=arl)

}

for (l in 2:4) {

print(bayesian.in.control.arl(0,l,1,10000))

}
```

https://doi.org/10.26219/heal.aueb.6628

1.2 S-Plus routine for the out of control ARL

```
f1<-function(d0,d1,n0,s,n,cut) {

time<-NULL

rl<-NULL

for (k in 1:n) {

x<-rnorm(n0,d1,s)

w1<-0

w<-0

i<-1

u<-NULL

while (w >=cut) {

l<-((x[i]*(d0-d1))/(s^2)+((d1^2)-(d0^2))/(2*(s^2)))

w1<-w+l

w<-min(0,w1)

u<-c(u,w)

i<-i+1

}

z<-u

d<-i-1

if (z[d]!=0) {

d<-d-1

}

time<-c(time,d)
```

```
rl<-c(rl,i)

arl<-mean(rl)

}

list(rl=rl,arl=arl,time=time)}

t<-seq(-2,-5.9,-0.1)

bayesian.cusum<-function(d0,d1,n0,s,n) {

v<-rep(NA,n*length(t))

m1<-matrix(v,ncol=n,nrow=length(t))

m2<-matrix(v,ncol=n,nrow=length(t))

par(mfrow=c(5,4))

arl<-NULL

average.time<-NULL

for (j in 1:length(t)) {

m1[j,]<-f1(d0,d1,n0,s,n,t[j])$rl

m2[j,]<-f1(d0,d1,n0,s,n,t[j])$time

arl<-c(arl,mean(m1[j,]))

average.time<-c(average.time,mean(m2[j,]-m1[j,]))

hist(m1[j,])

}

data.frame(cut=t,ARL=arl,average.time=average.time)

}

for (l in 2:4){

print(bayesian.cusum(0,l,100,1,10000)) }
```

https://doi.org/10.26219/heal.aueb.6628

2. Composite versus Composite case

2.1 S-Plus routine for the in control ARL

```
f.cVSc<-function(a,b,c,d,s,q,p,n,cut) {

rl<-NULL

for (k in 1:n) {

w<-0

x<-NULL

i<-1

while (w >=cut ) {

d0<-runif(1,a,b)

x<-c(x,rnorm(1,d0,s))

a1<-(a-((q/p^2)+(i*mean(x)/s^2))/((1/p^2)+(i/s^2)))/(1/((1/p^2)+(i/s^2)))

b1<-(b-((q/p^2)+(i*mean(x)/s^2))/((1/p^2)+(i/s^2)))/(1/((1/p^2)+(i/s^2)))

c1<-(c-((q/p^2)+(i*mean(x)/s^2))/((1/p^2)+(i/s^2)))/(1/((1/p^2)+(i/s^2)))

d1<-(d-((q/p^2)+(i*mean(x)/s^2))/((1/p^2)+(i/s^2)))/(1/((1/p^2)+(i/s^2)))

w<-min(0,log(pnorm(b1)-pnorm(a1))-log(pnorm(d1)-pnorm(c1)))

i<-i+1

if (i >= 1000) {
w<-cut-1 } }

rl<-c(rl,i)

arl<-mean(rl) }

list(rl=rl,arl=arl)

 }
```

https://doi.org/10.26219/heal.aueb.6628

3. Simple versus Composite case

3.1 S-Plus routine for the in control ARL

```
h<-function (d0,p,s,q,v,n,cut) {

rl<-NULL

for (k in 1:n) {

w<-0

i<-1

x<-NULL

while (w >= cut) {

x<-c(x,rnorm(1,d0,s))

A<-((d0*sum(x))/s^2)-(sum(x^2)/(2*s^2))-(i*log(s*sqrt(2*pi)))-((i*d0^2)/(2*s^2))

C<-1/(sqrt((1/v^2)+(i/s^2))*sqrt(2*pi))

O<-p*exp(A)+(1-p)*C*exp(-(sum(x^2)/(2*s^2))-(i*log(s*sqrt(2*pi)))-(q^2/(2*v^2)))

w<-min(0,A-log(O/(1-p)))

i<-i+1

if (i > 1000) {

w<-cut-1}

}

rl<-c(rl,i-1)

arl<-mean(rl)}

list(rl=rl,arl=arl)

}
```

https://doi.org/10.26219/heal.aueb.6628

References

Beibel, M. (1996). A note on Ritov's Bayes approach to the minimax property of CUSUM procedure, *The Annals of Statistics*, Vol. 24, No 4, 1804-1812

Berger, J.O (1985). *Statistical Decision Theory and Bayesian Analysis*, Berlin: Springer

Berbardo, J. and Smith, A. (1994). *Bayesian Theory*, West Sussex: Wiley

Brook, D. and Evans, D.A. (1972). An Approach to the Distribution of Run Lengths, *Biometrika* , Vol.59, 539-549

Casella, G. and Berger, R.L. (1990). *Statistical Inference*, Pacific Drove, CA: Wadsworth

Carlin, P. and Louis, T. (1996). *Bayes and Empirical Bayes Methods for Data Analysis*, Chapman & Hall, London

DeGroot, M.H. (1970). *Optimal Statistical Decisions*, New York: McGrow-Hill

Deming, W.E. (1986). *Out of Crisis*, Cambridge: Massachusetts Institute of Technology, Center for advanced Engineering study.

Ermer, DC. (1980). A control chart for Dependent Data, *Annual Technical Conference Transactions of the American Society for Quality Control*, 34, 121-128

Gan, F.F. (1991). An optimal Design of CUSUM Quality Control Charts, *Journal of Quality Technology*, Vol.23, No.4, 279-286

Good, I.J. (1965). The *elimination of Probabilities, An essay on Modern Bayesian Methods*, Cambridge, Mass: The MIT Press.

Graves, S. , Bisgaard, S. and Kulachi, M. (2005). A Bayes-Adjusted Cumulative Sum, *Bayes-Adjusted CUSUM 4b.doc*

Harris, T.J. and Ross, W.H. (1991). Statistical Process Control Producers of correlated observations, *Canadian Journal of chemical Engineering*, Vol.69

 https://doi.org/10.26219/heal.aueb.6628

Harrison, P.J. and West, M. (1987). Practical Bayesian Forecasting, *Journal of the Royal Statistical Society*, *series D (The Statistician)*, Vol.36, No.2/3, special issue: Practical Bayesian Statistics, 115-125

Harrison, P.J. and West, M. (1999). *Bayesian Forecasting and Dynamic Models,* 2nd edition, 2nd corrected printing, NY: Springer

Harrison, P.J. and Veenapen, P.P (1994). A Bayesian Decision Approach to Model Monitoring and CUSUMs, *Journal of forecasting*, Vol.13, 29-36

Hawkins, D.M. (1992). A fast Accurate Approximation of Average Run Lengths of CUSUM Control Charts, *Journal of Quality Technology*, Vol.24, No.1, 37-43

Hawkins, D.M and Olwell, D. (1997). *Cumulative Sum Charts and Charting for Quality Improvement*, New York: Springer-Verlag

Hilliers, F.S. and Yang, C. (1970). $\bar{X}$ and R Charts Control Limits Based on a Small Number of Subgroups, *Journal of Quality Technology*, Vol.2, No.1, 9-16

Hunter, J.S. (1986). The Exponentially Weighted Moving Average, *Journal of Quality Technology*, Vol.18, No.4, 203-210

Jeffreys, H. (1948). *Theory of Probability*, Clarendon Press Oxford, second edition

Jeffreys, H. (1939/1961). *Theory of Probability,* Oxford: University Press, third edition in 1961, Oxford: University Press

Montgomery, D. (2001). *Introduction to Statistical Process Control*, 4th edition, New York: Wiley

Montgomery, D and Mastrangelo (1991). Some statistical Process Control Methods for autocorrelated Data (with discussion), *Journal of Quality Technology*, Vol.23

Moustakides, G. (1986). Optimal Stopping Times for Detecting Changes in Distributions, *Annals of Statistics*, Vol.14, 1379-1387

Page, E.S. (1954). Continuous inspection schemes, *Biometrika*, Vol.41, 100-115

Roberts, S.W. (1959). Control Chart tests Based on Geometric Moving Averages, *Technometrics*, Vol.1, No.3, 239-250

https://doi.org/10.26219/heal.aueb.6628

Ritov, Ya'cov (1990). Decision Theoretic Optimality of CUSUM Procedure, *Annals of Statistics*, Vol.18, No3, 1464-1469

Shewhart, W.A. (1931). *Economic Control of Quality of Manufactured Product*, New York: Van Nostrand

Smith, A. (1986). Some Bayesian thoughts on modeling and model choice, *The statistician*, Vol.35, 97-102

Vance, L.C. (1986). Average Run Lengths of Cumulative Sum Control Charts and Controlling Normal means, *Journal of Quality Technology*, Vol.18, No.3, 189-193

West, M. (1986). Bayesian Model Monitoring, *Journal of the Royal Statistical Society, series B (Methodological)*, Vol.48, No.1, 70-78

Western Electric Handbook (1956). Statistical Quality Control Handbook, *Western Electric Corporation*, Indianapolis, IN

Woodall, W.H. (1997). Conflicts Between Deming's Philosophy and Economic Design of Control Charts, *in Frontiers in Statistical Quality Control* edited by H.J Lenz, G.B. Wetherill and P.T Wilvich, Physica-Verlag, Vienna

Yaschin, E. (1993). Statistical Control Schemes: Methods, Applications and Generalizations, *International Statistical Review*, Vol.61, No1, 41-66