Statistical Process Control
by
Quantile Approach

Osama Hasan Arif

A thesis submitted in partial fulfilment of the requirements of
Sheffield Hallam University
for the degree of Doctor of Philosophy

October 2000

Abstract

Most quality control and quality improvement procedures involve making assumptions about the distributional form of data it uses; usually that the data is normally distributed. It is common place to find processes that generate data which is non-normally distributed, e.g. Weibull, logistic or mixture data is increasingly encountered.

Any method that seeks to avoid the use of transformation for non-normal data requires techniques for identification of the appropriate distributions. In cases where the appropriate distributions are known it is often intractable to implement.

This research is concerned with statistical process control (SPC), where SPC can be apply for variable and attribute data. The objective of SPC is to control a process in an ideal situation with respect to a particular product specification. One of the several measurement tools of SPC is control chart. This research is mainly concerned with control chart which monitors process and quality improvement. We believe, it is a useful process monitoring technique when a source of variability is present. Here, control charts provides a signal that the process must be investigated.

In general, Shewhart control charts assume that the data follows normal distribution. Hence, most of SPC techniques have been derived and constructed using the concept of quality which depends on normal distribution. In reality, often the set of data such as, chemical process data and lifetimes data, etc. are not normal. So when a control chart is constructed for $\bar{x}$ or R, assuming that the data is normal, if in reality, the data is non-normal, then it will provide an inaccurate results.

Schilling and Nelson has (1976) investigated under the central limit theory, the effect of non-normality on charts and concluded that the non-normality is usually not a problem for subgroup sizes of four or more. However, for smaller subgroup sizes, and especially for individual measurements, non-normality can be serious problem.

The literature review indicates that there are real problems in dealing with statistical process control for non-normal distributions and mixture distributions. This thesis provides a quantile approach to deal with non-normal distributions, in order to construct median rankit control chart. Here, the quantile approach will also be used to calculate process capability index, average run length (ARL), multivariate control chart and control chart for mixture distribution for non-normal situations. This methodology can be easily adopted by the practitioner of statistical process control.

Acknowledgements

I wish to express sincere appreciation to my supervisor, Professor Gopal G. **Kanji** for his direction, assistance, encouragement and patience during the research's years. My appreciation also goes to the second supervisor, Professor Warren **Gilchrist** for his valuable comments, discussion and encouragement.

I would like to extend my special thanks to my mother, sisters and brothers for their love and constant encouragement. Finally my great appreciation goes to my wife **"Nisrin"** for taking care of our children "Faris and Deema" and for her love, support and patience while preparing this thesis.

Statistical Process Control by Quantile Approach

Table of Contents

Chapter 1 : Introduction

1.1 General Overview

The science of statistics itself goes back only to two or three centuries ago. Its greatest developments have been in the last 70 years. Early applications were not made until the 1920s, that is when theory of statistics began to be applied effectively to quality control. These statistical methods, which investigate the problems of quality control, were first suggested by Walter A. Shewhart of the Bell Telephone Laboratories. In a memorandum prepared on May 16, 1924, he made the first sketch of a modern "control chart", which he subsequently developed in various memoranda and articles. In 1931 Shewhart published a book on statistical quality control, titled "Economic Control of Quality of Manufactured Product".

Statistics has a very important role to play in the field of manufacturing, covering marketing plan, sales predictions, research and developments, and processes improvement. Statistics is also vital in manufacturing processes, such as incoming quality control, in-process quality control, outgoing quality control, quality assurance, etc. Therefore, statistical understanding plays a major role in product and service quality, care of customers through statistical process control (SPC), customer surveys, process capability and cost of quality etc.

In addition, experimental design using statistics is of an importance in distinguishing between special cause and common cause within Quality Improvement. The latter defined as the reduction of variability in processes and products. If we accept that all processes are variable and that there is a relationship between management action and quality, then statistical understanding becomes an essential aspect of quality improvement process. Here, quality improvement processes are about performance improvement of individuals, groups and organisations. "In order to improve performance, people need to know what to do, how to do it, to have the right tools to do it, to be able to measure performance and to receive feedback on current levels of achievement", (Kanji 1995).

Quality improvement is important and needed for achieving good quality product features with freedom from deficiencies. To maintain and increase sales revenue, companies must continually add new product features and introduce new improved processes to produce such features. Moreover, companies should realise that customers' needs are in a state of change, hence the need to be aware of meeting them. To keep cost competitive, companies must also continually aim at reducing the level of product and process deficiencies. Reduction in production costs besides improving the quality of a product, should be a prime concern to a company, as it would lead to customer satisfaction and consequently arise in sale and profit.

In reality, a company cannot survive in an open market or competitive economy for long if it is not achieving a reasonable level of profits. Such survival would require the company to look for improvement of product and cost reduction in its operations and carry out research and development. There are essential activities for an organisation to take in order to remain competitive and to maintain Business Excellence.

To manufacture a higher proportion of products within given specifications and to reduce the variability in quality of such products, it is necessary to increase the use of process control. In quality improvement, process control can be divided into two types: Statistical process control (SPC) and Engineering process control (EPC) or Automatic process control (APC). Statistical process control and Engineering process control, are two techniques relied on for quality improvement, which have developed independently. Box and Kramer (1992) provide an excellent comparison of SPC, which they refer to as statistical process monitoring and engineering process control, i.e. EPC. They explain the origin of statistical process monitoring as being in the parts industry, whereas APC is in the process industry. The aim of SPC and EPC techniques is the same, i.e., bringing all the process levels to their target with small variability.

Both techniques have the reduction of variability as their main objective, despite the fact that different methods have been employed to accomplish such an objective. SPC looks for signals representing assignable causes, which may be thought of as external disturbances that increase variability. It also assumes that the process data can be described in terms of statistically independent observations, which fluctuates around a constant mean. On the other hand, EPC actively reverses the effect of process disturbances by making regular adjustments to process variables. EPC is usually discussed in the framework of a process with a drifting mean, and the process adjustments to keep the output quality characteristics on target. EPC accomplishes this basically by transferring variability in the output variable to an input control variable.

The reason why EPC and SPC suggest different strategies for achieving the above mentioned goal, is because of the fact that traditionally they have different processes i.e. two different models. For many engineering systems, it is not only possible to describe them using control behaviour perspective, because they go out of control. This necessitates a form of intervention that will keep such systems in a state of equilibrium, with a small variance. On the other hand, in traditional applications of SPC, it is assumed that in normal conditions the process mean and variance are stable, but abrupt changes in the mean, variance or both, can occur at some unknown moments of time.

This research is concerned with the statistical process control (SPC). The objective of SPC is to control a process in an ideal situation with respect to a particular product specifications, (Chen, 1996). A widely used process indicator is its output distribution, characterised by the mean and variance. If the values of mean and variance are within prescribed limits, the process is operating in an in-control state. An assignable cause of variability may result in a shift in mean, variance or both, to an out-of-control state. Such shift leads to a defective product, down-time and costly corrective action. SPC uses the process information from samples to identify process shifts and to initiate timely remedial actions. SPC aims to maintain a process in its ideal status and to keep product quality loss at the minimum level during production. In addition, SPC's major aim within quality management is to decrease costs by improving process quality.

Usually, process quality can be improved by reducing output variability, the process failure rate or both.

Statistical process control can be divided into two types. These are on-line SPC and off-line SPC.

1.2 On-line SPC

On-line SPC methods are technical aid for quality and cost control in manufacturing. On-line SPC consists of preventative and screening processes. In preventative SPC, methods are always preferred, and in which the process itself is being inspected to avoid production of defective items. While, in screening SPC, the output of a process is checked by a system of sampling inspection. Screening helps to provide a basis for making decisions to investigate whether or not to accept the sample batch as satisfactory. This is always an expensive process because it takes more time and money to detect poor performance of the process. Taguchi (1978a) strongly believes that the main objective of an on-line SPC system should be prevention.

1.3 Off-line SPC

Off-line SPC methods are quality and cost control activities conducted at the product and process design stages, in order to improve product manufacturing and reliability, and reduce product development and lifetime costs. Design experiments are a major off-line SPC tool, because they are often used during activities and the early stages of manufacturing, rather than as a routine on-line procedure.

The Taguchi Method of experimental design (off-line quality control) has been promoted very strongly in the US and Europe; partly because it is thought to be a somewhat simpler and more defined approach to experimentation and partly because many successful applications have been attributed to it. However, the statistical content of Taguchi Method has been of an interest to statisticians and has been widely reviewed and criticised, (John 1990).

Taguchi, Elsayed and Hsiang (1989) discussed the robust design approach for determining the optimum configuration of design parameters for performance, quality and cost. The robust design method is an efficient, disciplined approach that can aid product delivery teams in designing for cost. Designing quality with product in mind would prove a cheaper process than trying to inspect and re-engineer such a product, after it hits the production floor, or worse, after it gets to the customer. The robust design method provides a systematic and efficient approach for finding the near optimum combination of design parameters, so that the product is functional, exhibits a high level of performance, and is robust to noise factors. Noise factors are those parameters that are uncontrollable or are too expensive to control.

However, introducing quality at the design stage to improve a process, requires the following overlapping factors:

- Inspection
- Quality control
- Quality improvement
- Quality by design

In order to minimise the effects of noise sources or error in the process, Taguchi suggests that certain counter measures have to be taken for the implementation of the following:

System design is the process of applying scientific and engineering knowledge to produce a basic functional prototype design, as in Kackar (1985). The prototype model defines the configuration and attributes of the product undergoing analysis or development. The initial design may be functional, but it may be far from optimum in terms of quality and cost.

Parameter design is an investigation conducted to identify the settings of design parameters that optimise the performance characteristic and reduce the sensitivity of engineering designs to the sources of variation (noise). Parameter design requires some form of experimentation for the evaluation of the effect of noise factors on the performance characteristic of the product, defined by a given set of values for the design parameters. This experimentation aims to select the optimum levels for the controllable design parameters.

Tolerance design is the process of determining tolerances around the nominal settings identified in the parameter design process. Tolerance design is required if a robust design cannot produce the required performance without costly special components or high process accuracy.

In this thesis, off-line methods will not be discussed, partly because the aim of the research is to develop and improve quality of product or process through statistical process control using quantile approach. Therefore, the focus of this thesis is on the on-line preventative SPC on process variable.

1.4 Outlines of Thesis

This thesis is divided into ten chapters. Chapter one presents a general review of quality control and process control. Process control is divided into two kinds of SPC i.e. on-line SPC and off-line SPC.

Chapter two introduces SPC methodologies, techniques and strategies. It defines control chart under the assumption of normality and discusses the effects of non-normality on control chart, the source of process variation i.e. common cause and assignable (special) cause, Average Run Length (ARL) and the hypothesis test used in SPC. In addition, it reviews the capability index, multivariate control chart and mixture distribution, in normal situation.

Chapter three introduces control chart methodology for non-normal situation and the effect of non-normality on control chart. Some techniques dealing with non-normal situation e.g. Q-chart, Box-Cox transformation are considered. Finally, quantile approach is introduced to deal with the non-normal situation of quality control chart.

Chapter four provides the theoretical development of quantile approach for continuous and discrete distributions. For continuous distributions, Uniform, Extreme-value, Exponential, Logistic, Weibull, Power and Pareto are considered. For discrete distributions Geometric is discussed.

Chapter five is dedicated to developing the theoretical aspects of quantile approach, which have been discussed in chapter four, in order to construct quality control charts for non-normal situation.

Chapter six discusses the capability index for non-normal situation, using quantile approach. It also discusses the performance of control charts using average run length (ARL) in chapter seven. Chapter eight, extends the quantile approach to dealing with the multivariate control chart and its applications.

Chapter nine, provides the quantile control chart for mixture distribution and its application. A conclusion of the thesis and future work in this area, are presented in chapter ten.

Chapter 2: Literature Review for Statistical Process Control

2.1 Introduction

The idea of using statistical methods for quality improvement easily extends to the general problems of process improvement. A good way to approach any of these problems is to define performance, measure it, determine the special causes of poor performance, and monitor it, which would result in continuous improvement in quality. Such approaches are generally known as Statistical Process Control (SPC), Carlyle, *et al*. (2000); Montgomery and Woodall (1997).

Control charts and other related techniques for statistical process control monitoring are in widespread use. The last 20 years have seen increasing emphasis on statistical process control, as practical approach for reducing variability in industrial processes. Control charts and other related methods for process monitoring are discussed, as Multivariate quality control in Kourti and MacGregor (1996), Sullivan and Woodall (1996), Mason *et al*. (1997) and Tracy, *et al*. (1992). Autocorrelated data has considered by Faltin, *et al*. (1997) and Zhang (1998). For Shewart control chart, various contributions can be seen in Woodall and Montgomery (1999), Amin and Ethridge (1998), Palm, *et al*. (1997), Chen (1996), Montgomery (1997), Wood (1995), Patel (1993) and Rigdon, *et al.* (1994). Economic design and related issues are discussed in Keats, *et al.* (1997). The relationship of Statistical process monitoring and direct process adjustment through engineering control, Integration and comparison of SPC and Engineering Process Control (EPC), are also discussed in Montgomery, *et al.* (1994), Box, *et al.* (1997) and Box and Kramer (1992). In addition, Capability Process Index is discussed in Palar and Wesolowsky (1999), Kotz and Lovelace (1998), Rodriguez (1992) and Kane (1986).

Statistical Process Control (SPC) have several major tools which can be applied to any process. They are, histogram or stem-and-leaf display, check sheet, pareto chart, cause and effect diagram, defect concentration diagram, scatter diagram and control chart. This research is mainly concerned with control chart which monitors process and improvement. We believe it is a useful process monitoring technique when an unusual source of variability is present, i.e. when the sample average values lie outside the control limits. This provides a signal that the process must be investigated to undertake corrective action

2.2 Statistical Process Control

Statistical process control (SPC) is part of a statistical quality control (SQC), which provides a system of quality control used in place of industrial or other operations.

The purpose of SPC is to control a process in an ideal status with respect to a particular product specification (Chen, 1996). A widely used process indicator is its output distribution characterised by the mean and variance. If the values of mean and variance are within prescribed limits, the process is operating in-control state. An assignable cause of variability may result in a shift in mean or variance, or both to an out-of-control state, and thus lead to a defective product, downtime, costly corrective resulting in action. SPC uses the process information from samples to identify process shifts, and to initiate timely remedial actions. SPC aims to maintain a process in its ideal status and keep product quality loss at a minimum during production. Furthermore, the objective of SPC is to monitor the performance of a process over time in order to detect any unusual events that may occur. By finding assignable causes for these events, improvements in the process and in the product quality can be achieved, by eliminating the causes, improving the process or its operating procedures, (Kourti and MacGregor, 1996). The purpose of statistical process control (SPC) is to find as many sources of variation as possible and then eliminate them. When stable process with small variation is achieved, the target is to maintain or, if possible, improve the process even further. In these cases, it is often not possible to make improvement by eliminating sources of variation. Instead, a creative change in the process structure is need.

In addition, one of SPC's major concerns relates to quality management, which is to decrease costs by improving process quality. Usually, process quality can be improved by reducing output variability or the process failure rate, or both. This is in order to quickly detect the occurrence of assignable causes or possible shift, so that investigation of the process and corrective action may be undertaken, before many nonconforming units are manufactured.

Usually, statistical process control uses control charts for monitoring the evolution of a manufacturing process: upper and lower control limits are computed, and if the process operates outside these limits, it is declared out of control and a search for an explanation of this abnormal behavior is initiated. An important tool in statistical process control for finding assignable causes and for monitoring a manufacturing process is the use of the control chart.

2.3 Control Chart

A control chart is a graph of a quality measurement, plotted against time with control lines superimposed to show statistically significant deviations from the normal level of performance. Any significant deviations are assumed to correspond to assignable or special causes, which deserve investigation. A large number of different control charts are discussed in the literature. Each of these charts has the same underlying format but embodies a different statistical model. Control charts can be used for two main purposes. Firstly, it gives an indication of how the level of performance varies with time. Secondly, it monitors improvement, (Wood 1995). Control charts are the basic statistical tools used to monitor and control processes. They can be easily constructed, visualised and interpreted.

The basic Shewhart $\bar{x}$ or x-Chart for monitoring the mean of a process, consists of a centre line at the historical process level, upper and lower control limits. Sample means are plotted over time. An out-of-control signal is given when a sample mean falls beyond the control limits. The control limits are most often set at $\pm 3\sigma$ from the centrelines, where sigma is estimated standard error of the sample means. Other methods have been proposed to improve sensitivity to small and moderate sized shifts in the mean. In particular, runs rules have been used to signal for other unusual patterns on the chart, such as having eight sample means in a row either all above or all below the centreline. Runs rules improve the sensitivity, but also increase the number of false alarms. Some of these run rules, which are useful with an $\bar{x}$ chart in detecting a small sustained shift in the mean, such are rule 1-of-1, rule 2-of-3, rule 4-of-5, rule 9-of-9 an so on, For more details, see Nelson (1984) and Lucas and Saccucci (1990).A typical Shewart control chart is shown in figure 2.1.

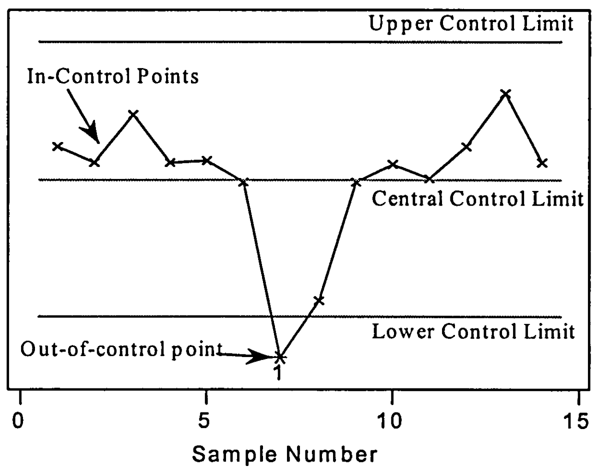


Figure 2.1 Shewhart Control Chart

In practice, Shewhart charts have been widely used for process monitoring because of an interest in involving production operators in quality improvement and the feeling that they cannot be trained to use other charting methods. Lucas (1976), Crowder (1987) and Lucas and Saccucci (1990) have shown that CUSUM and EWMA charts provide faster detection of small step changes than a non-modified Shewhart chart without an increase in the false-alarm rate.

Control charts for individual measurements are often used when production volume is too low to justify subgrouping or when automated inspection equipment is used to measure every unit produce. Montgomery (1997) considered control charts for individuals measurements, and Rigdon, *et al.* (1994) suggested that the Individual control limits should be based on a short-term estimate of the process variability, such as the moving average, rather than a long-term estimate, such as sample standard deviation of the process.

The primary purpose of a control chart is thus to quickly detect whenever a change has occurred in a process resulting in an alteration in the mean value or in the dispersion. Control charts may be used to estimate the parameters of a production process and process capability through this estimate. The control chart may also provide useful information for improvement of the process. The eventual goal of statistics process control is the elimination of variability in the process. It may not be possible to completely eliminate variability, but the control chart is an effective tool in reducing variability as much as possible.

In application of statistical method to quality engineering, it is very important to classify data on quality characteristics as either variable or attribute data. Attributes data are usually discrete measurement, often taken the form of counts. On the other hand, variable data are usually continuous measurement, such as length of stay. Most of the work in this thesis will dealing with variable data.

The essential idea of a statistical control chart is that a reference distribution of 'usual background noise' may be obtained by pooling experiences from groups of observations, called rational subgroup taken over short periods within which the process is judged to be stable. Continuous comparison of current with control limits based on this reference distribution can lead to the detection of unusual and undesirable distributions. Moreover, the idea of a control chart is to take a number of units produced by the process at regular intervals and check one or more characteristics of them. This information is then weighed together in a suitable manner, for instance to an arithmetic mean or to a standard deviation, and plotted in a diagram. Not only is the process variation illustrated as a function of the time, but process changes are indicated too. Another way to increase sensitivity, is to use more information from the collected data, for instance by also using information from earlier plotted points in the chart. Control charts are widely used in manufacturing to distinguish between variation that is inherent (common) to the process and variation that signals a special (assignable) event or problem.

2.4 Source of Process Variation

A control chart is a statistical tool used to study and control repetitive processes in industrial setting. Shewhart control charts developed to help distinguish between variation in manufacturing that is intrinsic to the production system and variation which is due to external factors. In many production processes, there are many small sources of variation that are inherent in the system itself, which are summarised under the name chance (common) variation. In addition, there is variation that is relatively large and can be assigned to a particular cause, and this is called assignable (special) variation.

A system that only exhibits chance variation is said to be in statistical control; otherwise, it is out of control. There are many types of control charts for different situations, such as individual control charts, x-bar control charts etc. Control charts have upper and lower control limits, often placed three standard deviations from the average. If an observation falls outside these limits, it is considered to be a signal that the process is not in control. These upper and lower control limits are based on estimates of the mean and variance of the process when it is in statistical control.

The ability to separate special/common cause of variations within a process, has enabled management to analyse data and take the necessary actions to improve quality and productivity, at economical cost levels. The basis of such improvement, however, is in the selection, application and interpretation of statistical data generated through the use of the correct type of control charts, (Patel, 1993).

A widely used process indicator is its output distribution characterised by the mean and variance. If the values of means and variance are within prescribed limits, the process is operating in an in-control state. An assignable cause of variability may result in a shift in mean or variance or both, to an out of control state, and thus lead to a defective product, (Chen, 1996).

Variation remaining in a stable process reflects common causes, which cannot be removed easily from the process without fundamental changes in the process itself. If the underlying probability distribution of the quality characteristic is stable over time, the process is said to be in statistical control. One purpose of a control chart is to detect unusual variation due to assignable causes. When the control chart signals the possible presence of an assignable cause, an effort is made to find and remove it from the process, if this action is to reduce variability or improve quality. It is also important to detect improvements in process performance, (Woodall and Montgomery, 1999).

2.5 Hypothesis Testing in SPC

There is a connection between hypothesis testing and control charts. Suppose that the vertical axis in figure 2.1 is the sample average (process, say). If the process points lie between the control limits, we conclude that the process mean is in-control and the processes have the same mean and average over time. On the other hand, if the process points exceed control limits, then we conclude that the process mean is out-of-control. It indicates that the process being monitored by SPC control chart, does not have the same mean and variance over time. There is significant evidence that the process is not in-statistical control. Two kinds of error can be occur in testing hypotheses, the first is commonly called a type I error (α), which occurs, if the null hypotheses rejected when it is true. The second error called a type II error (β), it takes place, if the null hypotheses is not rejected when it is false. In quality control studies, α is called the producer's risk and β is called the consumer's risk.

2.6 Capability Index

The concept of process capability was introduced by Juran *et al.* (1974), but did not gain considerable acceptance until the early 1980s. The concept enhances the idea of achieving a process output with minimal variation centred at a target value. Juran realised that there was a need in industry for the development of a single ratio or index, in order to compare the specification interval with the actual process variation. Therefore, Juran defined the first process capability index C_p as $C_p = \dfrac{USL - LSL}{6\sigma}$, where USL and LSL are the Upper and Lower specification limits, respectively, and σ is the standard deviation of the process. The general idea of C_p is to understand what the process is actually doing, in order to reflect the usability of the product, by controlling the process.

Juran & Gryna (1993) and Montgomery (1997) suggested that the purposes of Process Capability are to:

- Meet or exceed the customer need.
- Predict how well the process will hold the tolerances.
- Assist product developers/designers in selecting or modifying a process.
- Assist in establishing an interval between sample for process monitoring.
- Specify performance requirements for new equipment.
- Select between competing vendors.
- Plan the sequence of production processes when there is an interactive effect of processes on tolerance.
- Reduce the variability in a manufacturing process.

The process capability indices are appropriate only when measurements of the process data are independent, normally distributed and statistically process control. For various development of rules, confidence limits for C_p, C_{pk}, C_{pm}, C_{pmk} and various assumption, see Kane (1986), Bissell (1990), Chou *et al.* (1990), Boyles (1991) and Rodriguez (1992) and Gilchrist (2000).

Process capability indices are numerical values capable of demonstrating the relationship between the customer specification and the process variation. If the process follows normal distribution, then C_p, C_{pk}, C_{pm} and C_{pmk} can be obtained as follows:

The C_p index

The C_p index measures potential capability of the process, assuming that the process average is equal to the midpoint of the specification limits and the process is operating under statistical control. Here C_p only provides the process variability σ without indicating any sensitivity of the process departure.

The process capability index C_p relates the allowable (tolerance) process speed to the actual (natural) process spread in the form of a ratio

$$C_p = \frac{\text{allowable process variability}}{\text{actual process variability}} = \frac{USL - LSL}{6\sigma}$$

Supposing the process follows normal variation and the process is exactly capable i.e. $C_p = 1$, then the process target is at the midpoint of specification limits

$$\text{Target} = \frac{USL + LSL}{2}$$

The probability of obtaining a value outside the specification limits is $2\Phi(-3C_p)$=0.0027, where $\Phi(.)$ denotes the standard normal cumulative distribution function. When C_p=1, the Upper Specification Limit (USL) and Lower Specification Limit (LSL) equal the Upper Control Limit (UCL) and Lower Control Limit (LCL), which means that the variability of the distribution is exactly the width of the specification interval (for instance, Kane (1986)).

The actual process spread is taken to be six-sigma, which is represented in normal distribution, i.e. the width of the interval contains 99.73% of the population. The difference in the specification limits is used to indicate allowable process spread. The allowable process spread is considered fixed, while the actual process spread must be estimated.

C_p was considered as a measure of non-conforming product. If C_p is one, which represents 2700 parts per million (ppm) non-conforming, while 1.33 represents 63 ppm, 1.5 represents 7 ppm, 1.66 represents 0.6 ppm and 2 represents 0.0018 ppm. These results are correct if the process measurement arises from a normal distribution (see chapter 6 for non-normal situation). A minimum value of C_p=1.33 is generally used for an ongoing process, (see Juran, Gryna and Bingham 1979, pp. 9-22). If the value of six-sigma is less than the tolerance, the process is capable of meeting the specification, and if not, then process is incapable of meeting the specification.

The C_{pk} index

In the previous section, C_p assumes that the process has both upper and lower specification limits. It does not take into account the possibility that the process mean μ may differ from the centre (midpoint) M. If $\mu \neq M$, then the value of C_p=1 will correspond to an expected non-conforming proportion, greater than the nominal 0.27%. To avoid this situation(i.e. C_p), a C_{pk} index is more suitable to use. It is better to work with C_{pk}, because it represents both the spread and location of the process. Kane (1986) used the terms of process potential and process performance indices for C_p and C_{pk} respectively.

$$C_{pk} = \min(C_{pu}, C_{pl}) = (1-k)C_p$$

where

$$C_{pu} = \frac{USL-\mu}{3\sigma} = \frac{USL-T}{3\sigma}\left(1-\frac{|T-\mu|}{USL-T}\right) \quad ,$$

$$C_{pl} = \frac{\mu-LSL}{3\sigma} = \frac{T-LSL}{3\sigma}\left(1-\frac{|T-\mu|}{T-LSL}\right)$$

$$k = \frac{2\,|\,T-\mu\,|}{USL-LSL} \quad , \qquad 0 \le k \le 1$$

has been suggested for symmetric tolerance i.e. $T=\mu$. If the process is on-target then k=0 (i.e. $T=\mu$).

The C_{pk} is one side of the C_p specification limit nearest to the process mean. The value of C_{pk} does not determine the probability of non-conformance. It does, however, provide its limits, and in fact, the probability of non-conformance is never more than $2\phi(-3C_p)$. C_{pk} is yield-based and is independent of the target T. This fails to account for process centring with symmetric tolerance, and presents an even greater problem with asymmetric tolerance (see, Pearn and Chen (1998)).

The goal of C_{pk} is impossible to meet when source of variability in a measurement error is large (Herman (1989)). However, C_{pk} provides a meaningful measure of process quality, when a process is not in statistical control. C_{pk} should not be computed by either method if the process is unstable, because without statistical control, a process is unpredictable (Gunter (1989)).

The C_{pm} index

A capability index can also be calculated around a target value rather than the actual average. This index called C_{pm} or the Taguchi index, focuses on reduction of variation from a target value rather than reduction to meet specifications. See Chan *et al.* (1988), Pearn *et al.* (1992), Boyles (1991), Spiring (1991) and Kane (1986) for more dissuasion.

Chan, Cheng and Spiring (1988), proposed the index

$$C_{pm} = \frac{USL - LSL}{6\sigma'} = \frac{USL - LSL}{6\sqrt{E[(X-T)^2]}} = \frac{USL - LSL}{6\sqrt{\sigma^2 + (\mu - T)^2}}$$

$$C_{pm} = \frac{C_p}{\sqrt{1 + \left(\frac{(\mu - T)}{\sigma}\right)^2}} = \frac{C_{pk}}{\left(1 - \frac{|\mu - M|}{d}\right)\sqrt{1 + \frac{(\mu - T)^2}{\sigma^2}}}$$

According to C_{pm} above, if the process variance increase or decrease, the denominator of C_{pm} increase or decrease too, and C_{pm} will decrease or increase. If the process drifts from its target value, the denominator of C_{pm} will again increase, causing C_{pm} to decline. When the process mean and process variance change, the C_{pm} index changes as well. Note that the quadratic term $(\mu - T)^2$ reduces the value of the index, as a penalty for lack of co-ordination between the process and the desired results.

Parlar and Wesolowsky (1998) have noticed that C_p, C_{pk}, and C_{pm} are related by the formula

$$C_{pk} = C_p - \frac{1}{3}\sqrt{\left(\frac{C_p}{C_{pm}}\right)^2 - 1}$$

then

$$C_{pm} = \frac{C_p}{\sqrt{1 + 9(C_p - C_{pk})^2}}$$

Accordingly,

$$C_p \geq \max(C_{pk}, C_{pm})$$

The C_{pmk} Index

The third generation index C_{pmk} was introduce by Pearn *et al.* (1992). C_{pmk} is constructed by combining the modification of C_p that produces C_{pk} and C_{pm}. C_{pk} is obtained from C_p by modifying the numerator; C_{pm} is obtained by modifying the denominator of C_p . If the C_{pk} and C_{pm} are combined then C_{pmk} is produced as follows:

$$\begin{aligned} C_{pmk} &= \frac{\min(USL - \mu, \mu - LSL)}{3\sqrt{\sigma^2 + (\mu - T)^2}} \\ &= \frac{C_{pk}}{\sqrt{1 + \left(\frac{\mu - T}{\sigma}\right)^2}} = \left(1 - \frac{|\mu - M|}{d}\right)C_{pm} = \frac{d - |\mu - M|}{3\sqrt{\sigma^2 + (\mu - T)^2}} \end{aligned}$$

The concept of variation about the target provided by Hsiang and Taguchi (1985) as $\tau^2 = \sigma^2 + (\mu - T)^2$, which illustrate that τ^2 incorporates two variance component, variance about the process mean and variance of the process mean about the target. The term $(\mu - T)^2$ in the denominator may be viewed as an additional penalty to lack of process quality, i.e. the departure of process mean from target. This penalty ensures that C_{pmk} will be more sensitive to departure than C_{pk} and therefore C_{pmk} is better for distinguishing between off-target and on-target processes. Wallgren (1996) found that the advantage of C_{pmk} is having more sensitivity to deviations from target than C_{pk} or C_{pm}. Vannman (1995) compared C_{pmk} index to C_p, C_{pk}, C_{pm} and found that C_{pmk} is more restrictive, with regard to process means deviation from the target value, than the other indices.

In most statistical literature and quality assurance, distribution of properties of indices discussed above, are investigated under the assumption that the process measurement arise from normal distribution. However, in the real situation, most of the process data is non-normal distributed, (Clement, 1989) and (Gunter, 1989). The process capability indices for non-normal situation will be discussed in chapter 6.

2.7 Average Run Length (ARL)

The run length of a control chart is defined as the sample number until a signal is issued by the chart and the expectation of run length is commonly defined as the average run length (ARL). ARL will be large when the process is in-control and small when the process is out-of-control.(Gan, 1996)

ARL is the average number of points that must be plotted before a point indicates an out of control, where the run length is the number of samples required to obtain a signal. For normal situation or Shewhart control chart, the run length of the basic $\bar{x}$ chart is geometric random variable with expected value

$$ARL = \frac{1}{p}$$

where p is the probability of any point plot out-of-control chart, i.e. the probability of a signal at a given time period when the process is in control, (see Quesenberry 1995c).

For $\bar{x}$-chart or individual chart with 3σ limits, p=0.0027 "3.09σ limit, p=0.002 in British Standard" is the probability that a single point falls outside the limits when the process is in control. The number of observations until an observation falls outside of the control limits is geometrically distributed since the sample statistics are independent. So, the ARL of the control chart when the process is in control, i.e. normal situation, is

370 samples "500 samples for British Standard", that means, on the average, if the process remains in control, an out of control signal will be generated every 370 samples. These run length properties are calculated under the conditions of normality.

Optimum design criteria for EWMA control chart can be found from Crowder (1987) and Lucus and Saccucci (1990) who derive theoretical properties for the chart in order to ARL. The latter compare EWMA chart to the CUSUM chart, concluding that there is little difference between them.

2.8 Multivariate Control Chart

Control charts play a very important role in industrial situations for monitoring processes. Multivariate control chart is necessary when monitoring of several correlated quality characteristics simultaneously is desired. Traditional multivariate control chart based on T^2 statistics, which are very effective for detecting events, when the multivariate space is not very large, (Kourti and MacGregor, 1996).

Many of the concepts of multivariate quality control are associated with Hotelling (1947). Several approaches to multivariate control chart have been discussed in the literature such as economic design, can be found in Alt (1985), chart based on principle components can be found in Jackson (1980,1981a, 1981b, 1985), Ryan (1989) and Montgomery (1997). Jackson (1985) proposed using principle component analysis (PCA) for selecting the problem variables. The PCA technique decomposes the T^2 statistic into a sum of independent squared principal components, which are linear combinations of the original variables. These principle components must be examined to see why the process is statistical out of control.

Kourti and McGregor (1996) provide a newer approach based on PCA. T^2 is expressed in terms of the normalised principal component scores of the multinormal variables. When an out of control signal is received, the normalised scores with high values are detected, and contribution plots are used to find the variables responsible for the signal.

Alloway (1994) have considered the accuracy of multivariate control charts. The latter can be improved through a three step graphic process: identify and remove outliers, examine the distribution of the data relative to assumptions and use alternative approaches if the assumption of normality is not justified.

The values plotted on multivariate control charts are usually statistical based on his well-known Hotelling's T^2 distribution. This distribution is the multivariate counterpart to student's t distribution. The multivariate T^2 chart is particularly appropriate when the characteristics of interest are correlated.

In constructing the multivariate control charts, it is assumed that the covariance matrix is constant over time. One of the visual method for checking this assumption is to monitor the process variability.

An obvious idea is to consult the corresponding univariate control charts when a multivariate control chart signals that the process is out of control. Two aspects must be considered. Firstly, the overall significance level of the simultaneous use of p univariate control charts is difficult to determine. Secondly, it is not necessarily one quality characteristic that causes an out of control situation.

Development of Multivariate Control Chart

The SPC approach for process monitoring, currently in practice in several industries, is to chart a small number of variables, usually the final product quality variables, and examine them one at a time. However, when the quality of a product is defined by more than one property, all the properties should be studied collectively. Multivariate SPC charts developed for this purpose have been based on the χ^2 statistics or on Hotelling T^2 statistic.

Assume that the p-quality characteristics are jointly distributed as a p-variate normal and that a random sample of size n is available from the process. The likelihood ratio test of $H_0 : \mu = \mu_0$ vs. $H_1 : \mu \neq \mu_0$ specifies that the null hypothesis be reject if

$$\chi^2 = (\bar{x} - \mu_0)' \sum_0^{-1} (\bar{x} - \mu_0) \succ \chi^2_{\alpha,p}$$

Where $\bar{x}$ denoted the $(P \times 1)$ vector of sample mean and $\chi^2_{\alpha,p}$ is corresponding χ^2-percentile. Plotting the value of χ^2 versus time with an upper control limit (UCL) given by $\chi^2_{\alpha,p}$, where α is an appropriate significance level for performing the test (e.g. α= 0.05 or 0.01). The χ^2 statistic represents the direct or weighted distance (Mahalanobis distance) of any point from μ. If χ^2 statistic plot above the upper control limit, the process mean is out-of-control, and assignable causes of variation are sought. For the two quality characteristics, an elliptical control region, centred at μ_0, can be used in place for χ^2-chart.

When the in-control covariance matrix Σ is not known and must be estimated from a limited amount of data, it is suitable to plot Hotelling T^2 statistic given by

$$T^2 = (\bar{x} - \bar{\bar{x}})' s^{-1} (\bar{x} - \bar{\bar{x}})$$

Where s is an estimate of covariance matrix Σ. An upper control limit T^2_{UCL} is then obtained based on the F distribution and will depend upon the degree of freedom available for the estimate s, (Wierda, 1994a).

There are two distinct phases in constructing control charts, Alt (1982, 1985). The first phase, which offers a retrospective view, involves testing whether the processes were in-control, when the initial individual or subgroup data were collected on the process. A subgroup represents a sample of observations taken at some point in the process, such as a sample taken during a specified time period. This phase is often termed the start-up stage of the process for the purpose of obtaining a set of data to establish the control limits for monitoring purposes. The goal of this stage is to establish statistical control and find accurate control limits for stage two. The second phase consists of using the control chart to maintain control, that is, detecting any departure from the process standards as future subgroups are drawn. The multivariate T^2 statistic is often utilised as the charting statistic for both phases of control chart construction.

The phase 1 control limits for the T^2 is given by

$$UCL = \frac{p(m-1)(n-1)}{mn-m-p+1} F_{\alpha,p,mn-m-p+1}$$
$$LCL = 0$$

In phase 2, when the chart is used for monitoring future production, the control limits are as follows:

$$UCL = \frac{p(m+1)(n-1)}{mn-m-p+1} F_{\alpha,p,mn-m-p+1}$$
$$LCL = 0$$

where F_{v_1,v_2} is Snedecor's F with v_1 and v_2 degree freedom, p is the number of quality characteristics , m is number of preliminary sample and n is size of preliminary sample.

When μ and Σ are estimated from a large number of preliminary samples, it is customary to use UCL= $\chi^2_{\alpha,p}$ as the upper control limit in both phase 1 and phase 2. Retrospective analysis of the preliminary samples to test for statistical control and establish control limits also occurs in the univariate control chart setting. For the $\bar{x}$ - chart, it is well known that if we use $m \geq 20 \text{ or } 25$ preliminary samples, the distinction between phase 1 and phase 2 limits will nearly coincide, (Montgomery, 1997, pp: 366-367).

Multivariate control chart for individual case i.e. n=1

This case always occurs in the chemical and process industries. Since these industries frequently have multiple quality characteristics that must be monitored, multivariate control charts with n=1.

Suppose that m sample (preliminary sample), each of size n=1 are available, and that p is the number of quality characteristics observed in each sample. Let $\bar{x}$ and s be the sample mean vector and covariance matrix, respectively, of these observations. The Hotelling T^2 statistic in the above becomes

$$T^2 = (x-\bar{x})' s^{-1} (x-\bar{x})$$

T^2 test statistic is distributed as

$$T^2 \sim \frac{(m-1)^2}{m}\beta(p/2,(m-p-1)/2)$$

see Sullivan and Woodall (1996) and Gnanadesikan and Kettenring (1972).

Then the control limits for this statistic are suggested by Tracy *et al.* (1992) as follows

$$UCL = \frac{(m-1)^2}{m} * \beta(\alpha/2; p/2,(m-p-1)/2)$$
$$= \frac{(m-1)^2}{m} * \frac{(p/(m-p-1)\,F(\alpha/2; p, m-p-1)}{1+(p/(m-p-1))\,F(\alpha/2; p, m-p-1)}$$

and

$$LCL = \frac{(m-1)^2}{m} * \beta(1-\alpha/2; p/2,(m-p-1)/2)$$
$$= \frac{(m-1)^2}{m} * \frac{(p/(m-p-1)\,F(1-\alpha/2; p, m-p-1)}{1+(p/(m-p-1))\,F(1-\alpha/2; p, m-p-1)}$$

where $\beta(\alpha/2; p/2,(m-p-1)/2)$ and $\beta(1-\alpha/2; p/2,(m-p-1)/2)$ are the $1-\alpha$ percentile of the beta distribution.

As well as control limits for a single $px1$ multivariate observation vector and an estimate s based on m past multivariate are

$$UCL = \frac{p(m+1)(m-1)}{m^2-mp}F_{(\alpha,p,m-p)}$$
$$LCL = 0$$

and when the number of preliminary sample is large, i.e. m>100, many practitioners use an approximate control limit

$$UCL = \frac{p(m-1)}{(m-p)}F_{(\alpha;\,p,m-p)} \qquad \text{or} \qquad UCL = \chi^2_{\alpha,p}$$

2.9 Mixture distribution

Mixture distribution needs when the data represented by two or more kinds of distribution, for example, Laplace and Normal distribution. In this thesis the author assumes that the data from the mixture distributions are statistically independent from each other.

Statistical analysis of mixture data has proved not to be straightforward, for two main reason. Firstly, explicitly formulae generally do not exist for estimators of the various parameters, so the numerical methods are required. Secondly, theoretical difficulties which arise in certain aspects of the statistical analysis reveal some common mixture problems to be non-standard .

As a result, detailed investigation of the analysis of finite mixture problems offers more than just a catalogue of straightforward applications of standard methods to a particular class of statistical methods.

In this thesis will dealing with quantile approach for mixture distribution in order to develop quality control chart.

2.10 Effect of Non-normality on control chart

One of the underlying assumptions of SPC is the use of the normal distribution. Such assumptions are implicit in the construction of control charts and process capability studies. It has long been realised that the variability associated with many engineering processes does not have a normal distribution. In continuous batch manufacture the normality assumption is often justified, but the distribution of the process variation is more critical when considering the sample sizes associated with small batch manufacture.

An unstable process can lead to a seemingly non-normal distribution. If the process shifted upward after two-thirds of the data were collected, then the histogram would be skewed to the right. A mixture of two processes could lead to the same problem. In

these cases a transformation would be inappropriate. It is thus that the data be taken from a stable process.
Schilling and Nelson (1976) investigated the effect of non-normality on charts and concluded that the non-normality is usually not a problem for subgroup sizes of four or more. For smaller subgroup sizes, and especially for individual measurements, non-normality can be serious problem.

Control charts and process capability calculations remain fundamental techniques for statistical process control. However, it has long been realised that the accuracy of these calculations can be significantly affected when sampling from a non-normal population. Many quality practitioners are conscious of these problems but are not aware of the effects; such problems might have on the integrity of their results. Use is made of the Johnson system of distributions as a simulation technique to investigate the effects of non-normality of control charts and process control calculations. An alternative technique is suggested for process capability calculations which alleviates the problems of non-normality while retaining computational efficiency, (Spedding, 1994).

In general, there is the need for widespread realisation that non-normality can be a major problem for a wide variety of control chart procedures. For sample sizes, less than five, the central limit theorem does not apply. This has been demonstrated for an $\bar{X}$ chart by Yourstone and Zimmer (1992), Ryan and Howley (1999), Janacek and Meikle (1997) , Moore (1957) and for attributes charts by Ryan and Schwertman (1997) and Ryan (1989).

For positively skewed data, simple transformations such as the logarithmic, cube root or square are often useful. If we are dealing with proportions and if binomial variations is found, the inverse sine of the square root may remedy the problem.

Shewhart control charts assume that the variable of interest is normally distributed. Often, in practice, this assumption is violated (Montgomery, 1997). The distribution of the variable in question may be strongly skewed, as for example when measuring the eccentricity of a part or hole-drilling errors in a manufactured part, (Gunter, 1991). Further, a test on the variable may reject the normality assumption, but use of a transformed variable is generally not desired, due to resulting difficulties in interpretation of control charts. In such circumstances, the standard method of assuming a normal distribution may perform poorly, especially for very skewed process distribution, (Burr, 1967) and (Schilling & Nelson, 1976).

The above literature review indicates that there are real problems in dealing with statistical process control for non-normal distributions and mixture distributions. The main purpose of this thesis is to develop quality control charts and capability index for non-normal distribution and mixture distribution which can be easily adopted by the practitioner of statistical process control.

Many techniques can be used to deal with the data violating the assumption of normality, e.g. Quesenberry technique, Box-Cox transformation, Quantile technique, etc. These techniques will be discussed in the following chapters. This chapter has identified the limitation of the small sample sizes and the transformation as well as the inability of traditional SPC chart to cope with mixture distribution situation. In order to address this problem, we will develop and use the quantile method which offers relatively new and generally powerful techniques for non-normal and mixture situation in the area of SPC.

Chapter 3: Control Chart Methodology for Non-Normal Situation

3.1 Introduction

In general, Shewhart control charts assume that the set of data comes out from the process, following normal distribution, and the probabilities of points falling outside control limits, when the process is in control is 0.0027. Hence, most of SPC techniques have been derived and constructed from the concept of quality characteristics which depends on normal distribution, (see the reference of the central limit theorem in chapter 2). In reality, often the set of data such as, chemical process data, lifetimes data and cutting tool wear processes are not normal. So when constructed, a control chart of $\bar{x}$ or R, supposes that the data is normal and the actual sets of data are not normal. Therefore, it will give inaccurate results of quality characteristics.

Measuring quality characteristics often involve non-normal distribution. The point which arises from that is the effect of non-normality on the accuracy of control limits. Schiling and Nelson (1976) investigated the effect of non-normality on quality control charts. They concluded that the effect of non-normality on quality control is not a serious problem, when subgroup size is four or more. There is a serious problem of non-normality effect, faced, when the sample size of subgroup is less than four, especially for individual measurements. Moving range and individual measurement charts provide non suitable control limits for non-normal data, (Montgomery, 1997). By using the fourth root of the set of data, the positive skewed exponential distribution can be made into almost symmetric distribution, then plotted on individual measurement and EWMA and CUSUM for SPC, (Kittlitz, 1999). It has been discovered that the capability indices gives the false or inaccurate process fallout rates for non-normal data. Therefore, this issue for non-normality situation will be discussed in chapter 6.

There are some useful and validity techniques for transforming the non-normal data to normality situation. Therefore, it is possible to perform SPC technique on non-normal data. Rigdon, *et al.* (1994) suggest two remedies for dealing with non-normality, using a suitable non-normal distribution for a particular data, by physical consideration of the process; and seeking a transformation of the original data, which leads to an approximate normal distribution. From the literature search, it was found that there are many techniques used for such procedures. This chapter deals with some of these techniques, such as Quesenberry transformation or Q-Chart, Box-Cox transformation (1964) and Quantile Approach. In addition, there are other techniques, such as Johnson transformation, Pearson System and others, which will not be discussed here.

3.2 Quesenberry Technique (Q-Chart)

Statistical Transformation

In classical mathematics e.g. Laplace transformation when the original data is transform and a solution is found we perform an iverse transformation on the situation. Thus the solution refers to the original data. However, in statistics when we perform a transformation we model the relationships and solution in the transformation space only and by inference we claim that the same relationship exist in the original data. Quantile technique overcome this deficiency of refered to the original data at all times.

Quesenberry (1991) has suggested a new technique for short-run SPC using a transformation. This technique plays a role in monitoring a process mean or variance for a normally distributed quality variable. He refers to this technique as Q-Chart and defines it as being distributed approximately as standard normal statistics and is also approximately independent. The technique is plotted on standard normal scale, when the parameters are known and unknown. He notes that, the technique can be used for short-run and for long-run production

Q-chart concept converts independent identical distribution x_r into independent identical distribution standard normal (0,1) observation Q(x_r) called Q statistics . Q statistics are plotted on a Shewhart chart with control limits at ± 3 and centreline at 0.

Quesenberry uses the Q-chart for variable data $\bar{x}, s$ or R and for individual measurement of the process mean and the process variance. Both processes are discussed for the four cases below, which are, (μ known, σ known); (μ unknown, σ known); (μ known, σ unknown) and (μ unknown, σ unknown). Table 1 provides the Q statistics for individual measurement and Q statistics for sample mean, and Table 2 provides the Q statistics variance process for individual measurement and Q statistics for sample variance.

Table 1: Quesenberry statistics from sample mean

Cases	Q statistics for individual measurement	Q statistics for sample mean
μ known σ known	$Q_r(x_r) = \frac{x_r - \mu_0}{\sigma_0}$; $r = 1,2,..$	$Q_r(\bar{x}_r) = \frac{\sqrt{n_r}(\bar{x}_r - \mu_0)}{\sigma_0}$; $r = 1,2,...$
μ unknown σ known	$Q_r(x_r) = \left(\frac{r-1}{r}\right)^{1/2} \frac{(x_r - \bar{x}_{r-1})}{\sigma_0}$; $r = 2,3,4...$	$Q_r(\bar{x}_r) = \sqrt{\frac{n_r(n_1 + ... + n_{r-1})}{n_1 + ... + n_r}}\left(\frac{\bar{x}_r - \bar{\bar{x}}_{r-1}}{\sigma_0}\right)$ $r = 2,3,...$
μ known σ unknown	$Q_r(x_r) = \Phi^{-1}\left\{t_{r-1}\left(\frac{x_r - \mu_0}{s_{0,r-1}}\right)\right\}$; $r = 2,3,4,...$ $S_{0,r}^2 = \frac{1}{r}\sum_{j=1}^{r}(x_j - \mu_0)^2$	$Q_r(\bar{x}_r) = \Phi^{-1}\left\{t_{n_1+...+n_r}\left(\frac{\sqrt{n_r}(\bar{x}_r - \mu_0)}{s_{0,r}}\right)\right\}$ $r = 2,3,...$ $s_{0,r}^2 = \frac{\sum_{\alpha=1}^{r}\sum_{j=1}^{n_\alpha}(x_{\alpha j} - \mu_0)^2}{n_1 + ... + n_r}$
μ unknown σ unknown	$Q_r(x_r) = \Phi^{-1}\left\{t_{r-2}\left[\left(\frac{r-1}{r}\right)^{1/2}\left(\frac{x_r - \bar{x}_{r-1}}{s_{r-1}}\right)\right]\right\}$; $r = 3,4,5,...$	$Q_r(\bar{x}_r) = \Phi^{-1}\left[t_{n_1+...+n_{r-1}}(w_r)\right]$ $r = 2,3,...$ $w_r = \sqrt{\frac{n_r(n_1 + ... + n_{r-1})}{n_1 + ... + n_{r-1}}}\left(\frac{\bar{x}_r - \bar{\bar{x}}_{r-1}}{s_{p,r}}\right)$

Table 2: Quesenberry statistics from sample variance

Case	Q statistics variance process for individual measurement	Q statistics for sample variance
σ known	$Q_r = \Phi^{-1}\left\{\chi_1^2\left(\frac{R_r^2}{2\sigma_0^2}\right)\right\}$; $r = 2,4,6,...$ $R_r = x_r - x_{r-1}$	$Q_r(S_r^2) = \Phi^{-1}\left\{\chi_{n_r-1}^2\left[\frac{(n_r-1)S_r^2}{\sigma_0^2}\right]\right\}$ $r = 1,2,...$
σ unknown	$Q_r = \Phi^{-1} = \left\{F_{1,v}\left(\frac{vR_r^2}{R_2^2 + R_4^2 + ... + R_{r-2}^2}\right)\right\}$ $r = 4,6,...\quad v = (r/2) - 1$	$Q_r(S_r^2) = \Phi^{-1}\left[F_{n_r-1,n_1+..+n_{r-1}-r+1}(w_r)\right]$ $w_r = \frac{(n_1 + ... + n_{r-1} - r + 1)S_r^2}{(n_1-1)S_1^2 + ... + (n_{r-1}-1)S_{r-1}^2}$ $r = 2,3,..$

Quesenberry also applies the Q-chart for attributes of Binomial, Poisson and Geometric distributions. Q-chart can be applied for common distribution, which are used to describe variable and attribute data. Table 3 gives the summary of Q-chart for attributes of such distributions. The transform observations from such distributions are given in table 3, for attribute case, values plotted on standardised normal Q charts, for the two cases when the parameter is known and unknown before charting is begun.

By transforming the observations through such distributions function in the table 3, the u_i's are approximately uniform on (0,1), and the Q_i's are approximately standard normal distribution, $N(0,1)$. The values of $Q_1, Q_2, Q_3,...$ can be plotted on a chart with control limits at LCL=-3, CL=0 and UCL=3. The distributions function for unknown parameters for the results in column 3, table 3, are derived by using the uniform minimum variance unbiased (UMVU) estimating distribution function, where the UMVU of Binomial is Hypergeometric distribution, the UMVA of Poisson is Binomial distribution and the UMVU of Geometric distribution is Geometric distribution.

Table 3: Quesenberry statistics for attribute

Distribution	Parameters	
	Known	Unknown
Binomial	$u_i = B(x_i; n_i, p)$ $Q_i = \Phi^{-1}(u_i)$, i=1,2,...	$u_i = H(x_i; t_i, n_i, N_{i-1})$ $Q_i = \Phi^{-1}(u_i)$, $i = 2,3,..$ where $N_i = \sum_{j=1}^{i} n_j$ and $t_i = \sum_{j=1}^{i} x_j$
Poisson	$u_i = F(y_i; n_i\lambda)$ $Q_i = \Phi^{-1}(u_i)$; $i = 1,2,..$	$u_i = B(y_i; t_i, n_i / N_i)$ $Q_i = \Phi^{-1}(u_i)$; $i = 2,3,...$ where $N_i = \sum_{j=1}^{i} n_j$ and $t_i = \sum_{j=1}^{i} y_j$
Geometric	$u_i = G(x_i; p_0) = 1-(1-p_0)^{x_i}$ $Q_i = -\Phi^{-1}(u_i)$, $i = 1,2,...$	$u_i = \tilde{G}(x_i t, n)$ $Q_i = -\Phi^{-1}(u_i)$, $i = 2,3,...$ where $t = \sum_{i=1}^{n} x_i$ $\tilde{G}(x_i; t, n)$ defines in Q95, p308

Quesenberry concludes from the distributions above, that Q-chart can be applied for these two cases. The interpretation of the Q-charts for the two cases, are nearly the same, but basic differences must be borne in mind. Q-chart for unknown parameters are plotted from the second sample, but no points are plotted for the first sample because, the parameter must be estimated from the set of data. Meanwhile, Q-chart for known parameters are plotted from the first sample. Quesenberry discussed some examples where to apply Q-chart on the distributions above. The observations plotted on these charts were very similar for both cases, when parameters are known and unknown.

Furthermore, Quesenberry (1995 a,b,c) discussed the properties of Q-chart for variable and attribute, such as the sensitivity of four test on Shewhart Q-chart and EWMA and CUSUM Q charts to detect one-step permanent shift of a Binomial, Poisson and Geometric. He found that the classic test of one point outside 3-sigma control limits i.e.

the 1-of-1 test, have poor sensitivity. Whereas, the test consisting of four out of five points beyond one sigma control limits i.e. the 4-of-5 test, is found to be a good test. The EWMA and CUSUM Q charts are most sensitive and are nearly comparable in overall performance.

Del Castilo and Montgomery (1994) have investigated the average run length performance of the Q-chart for variables and show that in some cases the ARL performance is inadequate. They suggested some modifications to the Q-chart procedures and some alternative methods based on the EWMA and a related technique called the Kalman filter which have better ARL performance than the Q-chart.

3.3 Box-Cox Transformation

Most statistical methods were created under the assumption of normality. Shewhart (1931) mentioned that most industrial measurements violate this assumption. Quality characteristics are always required to be normally distributed. If quality characteristics are not normally distributed, but the techniques are based on normality, then we will have inaccurate results. So, it is important to transform the data to normal situation. In most cases, the choice of the transformation is not obvious. For positive measurements, i.e. skew to the right, a family of power distribution was introduced by Tukey (1957). It is convenient to transform the data to normality using the formula below

$$
\begin{aligned}
y_i &= x_i^{\lambda} \quad &;\lambda \neq 0 \\
\mathrm{y_i} &= \log x_i \quad &;\lambda = 0
\end{aligned}
$$

One of the best techniques for choosing a transformation which could simultaneously achieve:

1- normality of distribution

2- constancy of error variance, i.e. independence between cell means and cell variance

3- simplicity (linearity) of the model structure is Box and Cox's (1964). In addition, such transformation chosen to achieve independence between cell mean and cell variance often has the effect of improving the closeness to normality.

Box and Cox suggested a useful modification for family of power transformation, which is defined only for positive values, using a maximum likelihood estimate of λ. However, this technique is not a restrictive one, because a single constant can be added to the data if there are some negative values.

$$y^{(\lambda)} = \frac{y^{\lambda} - 1}{\lambda \dot{y}^{\lambda-1}} \quad ;\lambda \neq 0$$
$$y^{(\lambda)} = \dot{y} \log y \quad ;\lambda = 0$$

Where $\dot{y} = (\Pi_{i=1}^{n} y_i)^{1/n} = \exp(\frac{1}{n}\Sigma \ln y_i)$;$y_i > 0$ is the geometric mean.

The family of power transformations are chosen, where each value is replaced by x^{λ} at $\lambda \neq 0$, where λ is always one of the value below:

λ	-2	-1	-0.5	0	0.5	1	2
$y_i = x_i^{\lambda}$	$\frac{1}{x_i^2}$	$\frac{1}{x_i}$	$\frac{1}{x_i^{0.5}}$	$\log x_i$	$x_i^{0.5}$	x_i	x_i^2

For λ =-0.5 ,0 ,0.5, the data values must all be positive. To use these transformations when there are negative and positive values, a constant can be added to all the data values, which must be greater than 0. If all the data values are negative, the data instead should be multiplied -1. However, in this situation, data suggesting skewness to the right would now become data suggesting skewness to the left.

Box-Cox transformation has found more practical utility in the empirical determination of functional relationship in a variety of fields, (Sakia, 1992). This method is almost applicable for most positive skewness data. The disadvantage of Box-Cox transformation is that it works only with non-negative and non-discrete distribution.

3.4 New Approach

Quantile function Q(p) can be used to provide non-parametric measures of location, scale etc. Q(p) can be applied for continuous and discrete random variable. Unfortunately, Q(p) does not exist for all points of the p^{th} quantile in case of discrete random variable, but it gives general indication on the attitude of set of data. Q(p) is defined as the inverse of distribution function of the random variable. So quantile function is defined as $Q(p) = F^{-1}(p),\ 0 \le p \le 1$, and the sample quantile function define as

$$\tilde{Q}(p) = F^{-1}(\tilde{p}) = x_i \qquad ; \frac{i-1}{n} \le p \le \frac{i}{n}.$$

The density quantile function f(Q(p)) can be obtained by deriving the quantile distribution function

$$p = F(Q(p)) \qquad 3.1$$

where $F(.)\ and\ Q(.)$ are the inverse function of each other.

Differentiation (3.1) in respect to p

$$1 = f(x)q(p) \quad ; \quad x = Q(p)$$

then

$$f(x) = 1/q(p)$$

is the density quantile function. So a plot $f(x)$ against $x = Q(p)$, will give the desired density plot. For more details see Parzen 1979.

Assessment of the suitability of the normal situation for a set of data is provided by quantile-quantile plot, (theoretical quantile $q_{(i)}$ vs. empirical data $x_{(i)}$). If the distributions are nearly the same, then quantiles will be nearly the same.

Quantile population $q_{(i)}$ for standard normal distribution N(0,1) are defined by

$$P(X \le q_{(i)}) = P_{(i)} = \frac{i}{n+1}$$

If the data follows a normal distribution, the plotting of theoretical quantile against observed quantile will be approximately linearly related. If the plotting of data does not give linear, then the derivation from this line will reveal how the distribution differs. So, the quantile approach for non-normal situation is discussed here.

For non-normal distribution, data can be transformed to normality, by using the square root for all random variable, Somerville and Montgomery (1996) or by taking the fourth root of the data, Kittlitz (1999). Moreover, some authors have recommended the use of distribution of power family or its extension, as done by Box-Cox (1964). On the other hand, Quesenberry technique (1991) can be used for common distribution, in order to deal with non-normal data.

The advantage of the quantile method is that it is very simple and fully applicable and can be easily used by a practitioner. Quantile approach also plays a very important role in continuous random variable.

So in the following two chapters, we will concentrate on the development of quantile approach for non-normal situation. In addition, we will discuss the theory of quantile approach for non-normality for various distributions, and then create the quantile control chart for some of these distributions.

Chapter 4: Theoretical Development of Quantile Approach

4.1 Introduction

Statistical process control techniques are widely used in industry for process monitoring and quality improvement. Various statistical control charts have been developed to monitor the process mean and variance. Traditional SPC methodology is based on the fundamental assumption that the process data are statistical normal distributed. Although, the process data are always non-normal distributed, (see Box and Luceno, 1997, p.6). For example, Chemical reactions follow Logistic; Bulb life follow Weibull, Power, Lognormal; Mechanical properties of material follow Extreme-value, etc.

As discuss before, the effects of non-normality on quality control charts have been suggested by Schiling and Nelson (1976), and concluded that the non-normality is usually not a problem for subgroup size of four or more. But for small subgroup size and especially for individual measurements, non-normality can be a serious problem. There are two ways of dealing with non-normality: firstly, using an appropriate non-normal distribution for the particular data suggested by the physical considerations of the process charts for the Weibull distribution (see Nelson, 1979) and secondly, seeking a transformation of the original data that results in an approximate normal data, such as the Box-Cox transformation, SPC Q chart proposed by Quesenberry, (1995) and the use of distribution families, e.g. Pearson, Johnson.

In recent years customers have exerted enormous pressure on organisations to improve the quality of their products and services. As a result, many organisations have implemented various quality improvement processes, as part of their everyday business activities. Some of these improvements are due to the application of statistical process control and process improvement methods (Blache *et al.* 1988). Due to complexity of data, it is sometimes difficult for people to interpret the various approaches of statistical process control, especially when they are modified with the help of various transformations (DuBois *et al.* 1991).

It is desirable that the data for statistical control charts be normally distributed. However, if the data is not normal, then a transformation can be used to produce a suitable control chart. A control chart is proposed which monitors the conformance of a sample using the quantile or inverse cumulative distribution function. This method also helps to detect changes in the distributional shape, which may be undetected in control charts that are based on summary statistics.

A successful quality improvement process must be based on proper interpretation of statistical data and quality improvement methods. In this chapter we will be discussing quality improvement process through quantile distribution. In doing so we will first discuss the quantile process of monitoring and control, then develop a quality control chart for this purpose using the median rankit.

4.2 Quantile Approach

Tukey (1960) has introduced a family of random variables defined by the transformation

$$x_p = [p^{\lambda} - (1-p)^{\lambda}]/\lambda \qquad 4.1$$

where p is a uniformly distributed random variable on (0,1) and $-\infty \prec \lambda \prec \infty$. It can be shown that the rectangular and logistic distributions are also members of the above family. For example a limiting form of (4.1) when $\lambda \to 0$ is given by

$$x_p = \ln p - \ln(1-p) \qquad 4.2$$

where x_p is known as the quantile function of logistic distribution. Location and scale parameters could be introduced to obtain the Generalised Lambda Distribution (GLD), which is also true for all quantile distribution function. One of the important aspects of the lambda family is that the percentage points are available directly for use (Joiner & Rosenblatt, 1971).

Various distributions about Generalised Lambda Distributions (GLD) can be found in Shapiro & Gross (1981), Ramberg and Schmeiser (1972) and Ramberg *et al.* (1979). However, a new quantile distribution can be obtained by using the inverse function of the generalised lambda distribution. For example, in GLD, a new quantile distribution, which is an extension of Tukey lambda distribution, Ramberg and Schmeiser (1974), can be obtained as follows

$$x_p = \lambda_1 + \{p^{\lambda_3} - (1-p)^{\lambda_4}\} / \lambda_2 \qquad ; 0 \le p \le 1 \qquad 4.3$$

where the range of x_p can be determined by setting p=0 and p=1. Range of λ values are discussed in Ramberg (1974), e.g. if λ_2, λ_3 and λ_4 are all-negative and $\lambda_2 \to 0$ then the range is $(-\infty, \infty)$.

In equation (4.3), if p is a uniform random variable, then x_p will have a GLD. The skewness and peakedness of the GLD can be determined by λ_3 and λ_4 and the scale by λ_2. The location of GLD can then be given any value using appropriate choice of λ_1. However, if the GLD is asymmetric ($\lambda_3 \neq \lambda_4$), then its expected value will not be equal to λ_1 as is the case with the symmetric situation. Furthermore, if $\lambda_3 = \lambda_4$, the original lambda distribution will be given, i.e. symmetric random variable.

Tukey's lambda distribution in a generalised form provides an algorithm for generating unimodel asymmetric random variables. This can also be generalised by using three or four parameters in the unimodel asymmetric distribution.

Suppose x be a random variable with a distribution function F. The root of equation

$$F(x_p) = p = prob(X \le x_p)$$

is called the p-th quantile of the distribution $F(x_p)$. The p-th quantile is also called the 100pth percentile. The p_i-th percentile of the population described by the distribution Q(p) is simply $Q(p_i)$, where $100p_i$-th is a suitable percentage. The root of the above equation, for p=0.5, is corresponds to the median of F, and for p=0.25 and p=0.75 which correspond to the lower and upper quantiles of F.

Here, the inverse cumulative distribution or quantile distribution Q (p) can be expressed as follows

$$x_p = Q(p) = F^{-1}(p) = [x : F(x) = p], p \in (0,1)$$

Here, F (x), f (x) and Q (p) (i.e. cumulative distribution, density function and quantile distribution, respectively) can be used as alternative starting points for defining distributions (Parzen, 1979). Quantile density function is defined as

$$f(Q(p)) = 1/q(p)$$

Kanji & Arif (2000) have shown that the quantile approach can be used to develop the quantile distribution, which can be used to develop a control chart. For example, if we consider a distribution with parameters λ, η, θ where θ represents one or more parameters, e.g. Weibull, Pareto, Power, then Q(p) is

$$Q(p) = \lambda + \eta R(p;\theta) \qquad 4.4$$

can be defined as a quantile distribution. A standard quantile distribution can be expressed as

$$z_p = \frac{x_p - \lambda}{\eta} = R(p,\theta)$$

where λ and η as location and scale parameters, and $R(p,\theta)$ depends on the parameters (e.g. skewness, shape).

Furthermore, a quantile distribution, which requires only two parameters (i.e. location and scale parameters), can be expressed as

$$Q(p) = \lambda + \eta R(p) \tag{4.5}$$

where $R(p)$ does not depend on the parameter (θ). Distributions such as Exponential, Extreme value and Uniform belong to this category.

Probability rules for Quantile Approach

Various properties of the quantile distribution function (QDF) can be described as follows:

- If $Q_1(p)$ and $Q_2(p)$ are QDF then $Q_1(p)+Q_2(p)$ is also QDF. This follows from the simple fact that all we require of a QDF is that it is a non-decreasing function of p.
- If the product of the two is inherently non-decreasing QDF then the $Q_1(p) * Q_2(p) = Q(p)$ is also QDF.
- The distribution $x_p = -Q(1-p)$ is the reverse of the distribution $x_p = Q(p)$.

In some situations QDF is heavily biased or weighted towards a specific tail area. For such situations, it is necessary to look at the tails separately in order to apply a suitable one-tailed model. However, for some of the standard distributions, a simple transformation gives a linear QDF.

For example, let us consider some continuous distributions, such as Logistic, Uniform, Exponential, Extreme-value, Weibull, Power and Pareto, to construct the quantile function for each of them.

4.3 Quantile Function for Logistic Distribution.

The density function of logistic distribution is define by

$$f(x) = \frac{\exp(x)}{(1+\exp(x))^2} \quad ; -\infty \le x \le \infty \qquad \textbf{4.6}$$

and the Cumulative Distribution Function (CDF) in general is

$$F(x) = 1 - \frac{1}{1+\exp(x)}$$

Then

$$F(x_{(p)}) = p = prob(X \le x_p) = prob(x \prec Q(p))$$

where $x_p = Q(p)$ is the Quantile Distribution Function (QDF), i.e. the QDF is the inverse of the CDF

$$F(x_p) = p = 1 - \frac{1}{1+\exp(x_p)} \quad ; p \in (0,1)$$

Reversing this to get the QDF simply gives

$$1 - p = \frac{1}{1+\exp(x_p)}$$

$$x_p = \ln \frac{p}{1-p}$$

Then the basic quantile distribution function for the logistic is

$$Q(p) = x_p = \ln \frac{p}{1-p}$$

The range of the quantile distribution from p=0 to p=1 is $(-\infty, \infty)$

Hence, the general quantile distribution for one tail gives us the exponential quantile distribution from a left tail or a right tail as follows

$$Q(p) = \lambda + \eta \ln(p) \quad \text{or} \quad Q(p) = \lambda + \eta(-\ln(1-p))$$

Hence, by using the property of the quantile distribution, and adding the two tails given above, we can obtain the logistic quantile distribution.

The logistic quantile distribution is

$$Q(p) = \lambda + \eta(\ln(p) - \ln(1-p)) \tag{4.7}$$

Hence the range of the quantile distribution from p=0 to p=1 is $(-\infty, \infty)$, and λ, η are location and scale parameters respectively.

Here, we have a left and a right tailed distribution, which if combined, then a potential model for the data will be given. A convenient form of weighting brings in position and scale parameters is

$$Q(p) = \lambda + \frac{\eta}{2}((1-\delta)\ln(p) - (1+\delta)\ln(1-p)) \tag{4.8}$$

defining the quantile logistic function for logistic distribution (Gilchrist, 1997), where δ represents the skewness of Q(p).

Properties of Logistic Quantile Distribution

Some properties of the logistic quantile distribution can be derived from e.g. 4.8 as follows

- Median

$$M = Q(0.5) = \lambda + \delta\eta \ln 2 \tag{4.9}$$

- Inter Percentile Range

$$R = Q((1-p)\lambda, \eta, \delta) - Q(p, \lambda, \eta, \delta)$$

$$R = -\eta \ln \frac{p}{1-p} \qquad ; p \prec \frac{1}{2} \qquad 4.10$$

- Difference between the Upper Tails and Lower Tails

$$D = Q((1-p), \lambda, \eta, \delta) + Q(p, \lambda, \eta, \delta) - 2m$$

i.e.

$$D = -\delta\eta(\ln p(1-p) + 2\ln 2) = -\delta\eta \ln 4p(1-p) \qquad 4.11$$

- Inter Percentile Range (R) > Difference

Estimation of Parameters

It is natural to describe Q (p) in term of percentiles/quantile rather than the method of moments. Therefore, we will look at a method of percentiles, as described by Dudewicz, Ramberg and Tadikamalla, 1974. The method of percentiles is the simplest method of estimation, which uses the natural percentile properties of distributions. However, the interest is often in the skewness and shape of distributions and in the limits that are exceeded with only low probability.

Parameters estimation of Logistic quantile distribution

Estimated parameters of logistic quantile distribution can be expressed as follows, using method of percentile (see eq. 4.9, 4.10, 4.11)

- Location

$$\hat{\lambda} = m + (\frac{d}{\ln 4p(1-p)}) \ln 2 \qquad 4.12$$

- Scale

$$\hat{\eta} = -\frac{r}{\ln \frac{p}{1-p}} \qquad 4.13$$

- Skewness

$$\hat{\delta} = \frac{d \ln \frac{p}{1-p}}{r \ln 4p(1-p)} \qquad 4.14$$

where m, d, r represents the sample median, difference between upper and lower tails and inter percentile range of sample population, respectively.

Measurement of the distance of tails from the median, the Right Percentile Range (RPR_q) and Left Percentile Range (LPR_q) for logistic quantile distribution can be expressed as follows

$$RPR_q = x_{1-p} - m = \eta\{(\frac{1}{2}(\ln(1-p) - \delta \ln(1-p) - \ln p - \delta \ln p)) - (\delta \ln 2)\} \quad ; p \prec \frac{1}{2}$$

$$LPR_q = m - x_p = \eta\{(\delta \ln 2) - \frac{1}{2}(\ln p - \delta \ln p - \ln(1-p) - \delta \ln(1-p))\} \quad ; p \prec \frac{1}{2}$$

The Method Least Absolute

The aim of the distribution of least absolute is based on choosing parameters $\underline{\theta}$, to minimise the sum of absolute deviation of the order observation values, which is

$$\xi = \Sigma | x_r - M_r |$$

For more details see Bloomfield and Steiger (1983) and Dodge (1987).

The median of the distribution of the rth order observation is called the median rankit, M_r, and the median rankit is defined as

$$M_r = Q(BETAINV(0.5, r, n-r+1)) = Q(p_r)$$

where p_r is define as

$$p_r = BETAINV(0.5, r, n-r+1)$$

When the distributions are not symmetric. The method of distribution of least absolute is more robust than the method of distribution of least square. Therefore, the distribution of least absolute is based on median rankit. It is also an advantage that the least absolute method of estimation can be implemented using Solver in Excel. However, the median in same cases is not unique. It is also not a sufficient statistics and most importantly, it is a biased estimator of the mean and can never be classified as UMVUE, (see page 38). For skew data and data from mixture distribution the quantile approach based on median is statistically and mathematically (untransformed) a superior statistic and for that reason alone quantile approach will be the basis for the research of the remaining chapters.

Residual plot for best estimate

Most of the models we have considered (Gilchrist 1997), are of the form $Q(p) = \lambda + \eta R(p)$ where R(p) contains two parameters e.g. exponential and extreme-value, and $Q(p) = \lambda + \eta R(p,\theta)$ contains more than two parameters e.g. logistic and pareto distribution. If we have ordered data, $x_{(r)}$, a fitted $\hat{R}(p)$ or $\hat{R}(p,\theta)$ and values of the median percentiles $p_{(r)}$, then for a correct model, a plot of $x_{(r)}$ versus $\hat{R}(p_{(r)})$ will be linear. For a fully fitted model $\hat{Q}(p_{(r)})$, we should get approximately a 45^0 line through the origin; such diagrams are a natural approach for identifying the appropriate QDF. A useful supplementary plot of the residual can be shown, using

$$\hat{e}_{(r)} = | x_{(r)} - \hat{Q}(p_{(r)}) |$$

in order to indicate the suitability of the model. It will use the same criteria which was used here, to the remaining distributions.

4.4 Quantile Function for Exponential Distribution

The density function of exponential distribution is

$$f(x) = e^{-x} \qquad ; \quad x \succ 0$$

and the Cumulative Distribution Function (CDF) is

$$F(x) = 1 - e^{-x}$$

Then

$$F(x_{(p)}) = p = prob(X \leq x_p) = prob(X \prec Q(p))$$

where $x_p = Q(p)$ is Quantile distribution function (QDF), i.e. the QDF is the inverse of the CDF

$$F(x_p) = p = 1 - e^{-x_p} \quad ; x \succ 0$$

Reversing this to get the QDF simply gives

$$x_p = -\ln(1-p)$$

Then the basic quantile distribution function for exponential is

$$Q(p) = x_p = -\ln(1-p)$$

The range of the distribution from p=0 to p=1 is $(0, \infty)$

Hence, the exponential quantile distribution is

$$Q(p) = \lambda + \eta(-\ln(1-p))$$

The range of the distribution from p=0 and p=1 is (λ, ∞).

Properties of exponential Quantile Distribution

Some properties of exponential quantile distribution can be seen as follows

- Median

$$M = Q(0.5) = \lambda + \eta \ln 2 \qquad 4.15$$

- Inter p - Range

$$R = Q((1-p),\lambda,\eta) - Q(p,\lambda,\eta)$$

$$R = \eta\{(-\ln p) + \ln(1-p)\} \quad ; p \prec \frac{1}{2} \qquad 4.16$$

- Difference

$$D = Q((1-p),\lambda,\eta) + Q(p,\lambda,\eta) - 2m$$

$$D = -\eta\{\ln 4p(1-p)\} \qquad 4.17$$

Estimation of Parameters

- Location

$$\hat{\lambda} = m - \ln 2 * \left(\frac{r}{-\ln p + \ln(1-p)}\right) \qquad 4.18$$

- Scale

$$\hat{\eta} = \left(\frac{r}{-\ln p + \ln(1-p)}\right) \qquad 4.19$$

- Skewness (Galton p-Skewness)

δ =Quantile p-Difference/ Quantile inter p-Range

$$\delta = \frac{-\{\ln 4p(1-p)\}}{\{(-\ln p) + \ln(1-p)\}} \qquad \textbf{4.20}$$

Measurement of the distance of tails from the median, the Right Percentile Range (RPR_q) and Left Percentile Range (LPR_q) of exponential quantile distribution, can be expressed as follows

$$RPR_q = x_{1-p} - m = -\eta\{\ln 2 + (1/\ln p)\}$$

$$LPR_q = m - x_p = \eta\{\ln 2 + (1/\ln(1-p)\}$$

4.5 Quantile Function for Uniform Distribution

The density function of Uniform distribution is

$$f(x) = 1; \qquad 0 \le x \le 1$$

and the Cumulative Distribution Function (CDF) is

$$F(x) = x$$

Then

$$F(x_{(p)}) = p = prob(X \le x_p) = prob(X \prec Q(p))$$

where $x_p = Q(p)$ is Quantile distribution function (QDF), i.e. the QDF is the inverse of the CDF

$$F(x_p) = p = x_p \qquad ; p \in (0,1)$$

Reversing this to get the QDF simply gives

$$p = x_p$$

Then the basic quantile distribution function for uniform is

$$Q(p) = x_p = p$$

The range of the distribution from p=0 to p=1 is $(0,1)$

Hence, the uniform quantile distribution is

$$Q(p) = \lambda + \eta\, p$$

The range of the distribution from p=0 and p=1 is $(\lambda, \lambda + \eta)$.

Properties of uniform Quantile Distribution

Some properties of uniform quantile distribution can be seen as follows

- Median

$$M = Q(0.5) = \lambda + 0.5 * \eta \qquad \textbf{4.21}$$

- Inter p - Range

$$R = Q((1-p)\lambda, \eta) - Q(p, \lambda, \eta)$$

$$R = \eta * (1 - 2p) \qquad ; p \prec \frac{1}{2} \qquad \textbf{4.22}$$

- Difference

$$D = Q((1-p),\lambda,\eta) + Q(p,\lambda,\eta) - 2m$$

$$D = \eta\{(1-p) + p - 1\} = 0 \quad 4.23$$

Estimation of Parameters

- Location

$$\hat{\lambda} = m - (0.5)*(\frac{r}{1-2p}) \quad 4.24$$

- Scale

$$\hat{\eta} = \frac{r}{1-2p)} \quad 4.25$$

Measurement of the distance of tails from the median, the Right Percentile Range (RPR_q) and Left Percentile Range (LPR_q) of uniform quantile distribution, can be expressed as follows

$$RPR_q = x_{1-p} - m = \eta\{(1-p) - 0.5\}$$

$$LPR_q = m - x_p = \eta\{0.5 - p\}$$

4.6 Quantile Function for Extreme-value

The density function of extreme-value distribution is

$$f(x) = e^{-e^{-x}} e^{-x} \qquad ; -\infty \prec x \prec \infty$$

and the Cumulative Distribution Function (CDF) is

$$F(x) = e^{-e^{-x}}$$

Then

$$F(x_{(p)}) = p = prob(X \leq x_p) = prob(X \prec Q(p))$$

where $x_p = Q(p)$ is Quantile distribution function (QDF), i.e. the QDF is the inverse of the CDF

$$F(x_p) = p = e^{-e^{-x_p}}$$

Reversing this to get the QDF simply gives

$$x_p = -\ln(-\ln p)$$

Then the basic quantile distribution function for extreme-value quantile is

$$Q(p) = x_p = -\ln(-\ln(p))$$

The range of the distribution from p=0 to p=1 is $(-\infty, \infty)$

Hence, the extreme-value quantile distribution is

$$Q(p) = \lambda + \eta\{-\ln(-\ln p)\}$$

The range of the distribution from p=0 and p=1 is $(-\infty, \infty)$.

Properties of extreme-value Quantile Distribution

Some properties of extreme-value quantile distribution can be seen as follows

- Median

$$M = Q(0.5) = \lambda + \eta\{-\ln(-\ln(0.5))\} \qquad 4.26$$

- Inter p - Range

$$R = Q((1-p)\lambda, \eta) - Q(p, \lambda, \eta)$$

$$R = \eta\{(-\ln(-\ln(1-p))) - (-\ln(-\ln p))\} \qquad 4.27$$

- Difference

$$D = Q((1-p), \lambda, \eta) + Q(p, \lambda, \eta) - 2m$$

$$D = \eta\{(-\ln(-\ln(1-p))) + (-\ln(-\ln p)) - 2(-\ln(-\ln(0.5)))\} \qquad 4.28$$

Estimation of Parameters

- Location

$$\lambda\hat{} = m - r * \left(\frac{(-\ln(-\ln(0.5)))}{\{(-\ln(-\ln(1-p))) - (-\ln(-\ln p))\}}\right) \qquad 4.29$$

- Scale

$$\eta\hat{} = \frac{r}{\{-\ln(-\ln(1-p)) - (-\ln(-\ln p))\}} \qquad 4.30$$

- Skewness (Galton p-Skewness)

δ =Quantile p-Difference/Quantile inter p-Range

$$\delta = \frac{((-\ln(-\ln(1-p))) + (-\ln(-\ln p)) - 2(-\ln(-\ln(0.5)))}{((-\ln(-\ln(1-p))) - (-\ln(-\ln p)))} \qquad 4.31$$

Measurement of the distance of tails from the median, the Right Percentile Range (RPR_q) and Left Percentile Range (LPR_q) of extreme-value quantile distribution, can be expressed as follows

$$RPR_q = x_{1-p} - m = \eta\{(-\ln(-\ln(1-p))) - (-\ln(-\ln(0.5)))\}$$

$$LPR_q = m - x_p = \eta\{(-\ln(-\ln(0.5))) - (-\ln(-\ln p))\}$$

In the next section, we will discuss the Weibull, Power and Pareto distributions in the form $Q(p) = \lambda + \eta R(p;\theta)$, which have got more than two parameters.

4.7 Quantile Function for Weibull Distribution

The density function of Weibull distribution is

$$f(x) = \gamma x^{\gamma-1}\exp(-x^{\gamma}); \; x \geq 0$$

$$F(x) = 1 - \exp(-x^{\gamma})$$

Then the basic quantile distribution function for the Weibull is

$$Q(p) = x_p = (-\ln(1-p))^{\beta} \qquad \text{where } \beta = \frac{1}{\gamma}$$

The range of the distribution from p=0 to p=1 is $(0,\infty)$

Hence, the Generalised Lambda Distribution (GLD) for Weibull quantile distribution is

$$Q(p) = \lambda + \eta(-\ln(1-p))^{\beta} \quad , \beta \succ 0$$

The range of the distribution from p=0 to p=1 is (λ,∞), where λ, η are location and scale parameters and β is the shape of the distribution. It discusses various properties of weibull quantile distribution below.

Properties of Weibull Quantile Distribution

Some properties of the weibull quantile distribution can be seen as follows

- Median

$$M = Q(0.5) = \lambda + \eta(-\ln 0.5)^{\beta} = \lambda + \eta(\ln 2)^{\beta} \qquad 4.32$$

- Inter p - Range

$$R = Q((1-p)\lambda,\eta,\beta) - Q(p,\lambda,\eta,\beta)$$

$$R = \eta[(\ln\frac{1}{p})^{\beta} - (\ln(\frac{1}{1-p}))^{\beta}] \qquad ; p \prec \frac{1}{2} \qquad 4.33$$

- Difference

$$D = Q((1-p),\lambda,\eta,\beta) + Q(p,\lambda,\eta,\beta) - 2m$$

$$D = \eta[(\ln(\frac{1}{p}))^{\beta} + (\ln(\frac{1}{1-p}))^{\beta} - 2(\ln 2)^{\beta}] \qquad 4.34$$

- Inter p- Range (R) > Difference when $\beta \succ 0$, $p \prec \frac{1}{2}$

Parameters estimation of Weibull quantile distribution

Estimated parameters of weibull quantile distribution can be expressed as follows, by using method of percentile (see eq. 4.32, 4.33, 4.34)

- Location

$$\hat{\lambda} = m - \left(\frac{r}{(\ln\frac{1}{p})^{\beta} - (\ln\frac{1}{1-p})^{\beta}}\right)(\ln 2)^{\beta} \qquad 4.35$$

- Scale

$$\hat{\eta} = \frac{r}{(\ln\frac{1}{p})^{\beta} - (\ln\frac{1}{1-p})^{\beta}} \qquad 4.36$$

- Skewness (Galton p-Skewness)

δ =Quantile p-Difference/Quantile inter p-Range

$$\delta = \frac{\{(\ln(\frac{1}{p}))^{\hat{\beta}} + (\ln(\frac{1}{1-p}))^{\hat{\beta}} - 2(\ln 2)^{\hat{\beta}}\}}{(\ln(\frac{1}{p}))^{\hat{\beta}} - (\ln(\frac{1}{1-p}))^{\hat{\beta}}} \qquad 4.37$$

Measurement of the distance of tails from the median, the Right Percentile Range (RPR_q) and Left Percentile Range (LPR_q) of weibull quantile distribution, can be expressed as follows

$$RPR_q = x_{1-p} - m = \eta\{(-\ln p)^{\beta} - (\ln 2)^{\beta}\} \quad ; p \prec \frac{1}{2} \qquad 4.38$$

$$LPR_q = m - x_p = \eta\{(\ln 2)^{\beta} - (-\ln(1-p))^{\beta}\} \; ; p \prec \frac{1}{2} \qquad 4.39$$

Estimation of Shape Parameter (β-Value)

In this section we will consider the quantile distribution as continuous distribution of form

$$Q(p) = \lambda + \eta R(p,\theta)$$

The equation above consists of scale, location and some parameters in $R(p,\theta)$ e.g. skewness and shape.

Exponential and uniform quantile distributions have two parameters only i.e. location and scale which are represented by

$$Q(p) = \lambda + \eta R(p)$$

In this section, we are going to estimate the shape parameter β for the weibull distribution, and by the same method the shape parameter for power and pareto distribution are given.

The Quantile distribution of Weibull can be described as follows :

$$Q(p) = \lambda + \eta\{-\ln(1-p)\}^{\beta} \qquad \textbf{4.40}$$

Here, three parameters needed to be estimated, i.e. λ, η, β. λ, η which have already been estimated in the last section. Hence, estimate of β for the distribution above is needed.

To estimate shape parameter (β) for the distribution with three parameters i.e. Power or Weibull distribution, we can use the following approach, as an iteration processes:

In the past the shape parameter β has been estimated by using many methods, such as maximum likelihood, least square and probability plot and so on. In this section we are going to estimate β shape by using quantile approach. Quantile approach for estimate β shape is done by calculating the β value mathematically, by using the difference (equation 4.34) for each distribution and developing equations of a median, range and difference as exponential terms. This is an iterative process.

By using the percentile method, the form of β-value can be seen as the same for the Weibull and Power distributions, and is given by

$$D = R * (\frac{e^{\theta_1 \beta} + e^{-\theta_2 \beta} - 2}{e^{\theta_1 \beta} - e^{-\theta_2 \beta}}) \qquad \textbf{4.41}$$

Where θ_1 & θ_2 are defined below, R is the inter p-Range of QDF and D is the difference of QDF.

For Weibull distribution

$$\theta_1 = \ln(\frac{\ln(1/p)}{\ln 2}) \qquad \theta_2 = -\ln(\frac{\ln(1/(1-p))}{\ln 2}) \qquad \textbf{4.42}$$

"the proof of the θ_1 and θ_2 values are given in the end of this section"

Now, we try to solve the equation (4.41) for $f(\beta)=0$ in order to obtain the estimate of the β value. Here,

$$f(\beta) = (R-D)Exp(\theta_1 \beta) + (R+D)Exp(-\theta_2 \beta) - 2R \qquad \textbf{4.43}$$

and $\quad f(\beta)=0$ when β=0 or $\beta = \beta^{\wedge}$.

Hence, we seek the solution of $\beta = \beta^{\wedge}$ since $\beta \succ 0$, (see appendix 9).

Proof of θ_1 and θ_2 values for weibull quantile distribution

By using equation 4.17

θ_1-value	θ_2-value
$\frac{1}{p} \geq 2$	$1 \leq \frac{1}{1-p} \leq 2$
$\ln\left(\frac{1}{p}\right) \geq \ln 2$	$0 \leq \ln\left(\frac{1}{1-p}\right) \leq \ln 2$
$\left(\frac{\ln\left(\frac{1}{p}\right)}{\ln(2)}\right) \geq 1$	$0 \leq \left(\frac{\ln\left(\frac{1}{1-p}\right)}{\ln(2)}\right) \leq 1$
$\ln\left(\frac{\ln\left(\frac{1}{p}\right)}{\ln(2)}\right) \geq 0$	$\ln\left(\frac{\ln\left(\frac{1}{1-p}\right)}{\ln 2}\right) \leq 0$
then	then
$\theta_1 = \ln\left(\frac{\ln\left(\frac{1}{p}\right)}{\ln(2)}\right) \text{, } \theta_1 \geq 0$	$\theta_2 = -\ln\left(\frac{\ln\left(\frac{1}{1-p}\right)}{\ln 2}\right) \quad ; \quad \theta_2 \geq 0$

4.8 Quantile Function for Power Distribution

The density function of Power distribution is

$$f(x) = \frac{\alpha}{k}\left(\frac{\mathrm{x}}{\mathrm{k}}\right)^{\alpha-1} \qquad ; \quad 0 \prec x \leq k \;\; , \;\; \mathrm{k} \succ 0 \;\; , \;\; \alpha \succ 0$$

where k is a fixed number.

and the Cumulative Distribution Function (CDF) is

$$F(x) = \left(\frac{x}{k}\right)^{\alpha}$$

Then

$$F(x_{(p)}) = p = prob(X \le x_p) = prob(X \prec Q(p))$$

where $x_p = Q(p)$ is Quantile distribution function (QDF), i.e. the QDF is the inverse of the CDF

$$F(x_p) = p = \left(\frac{x_p}{k}\right)^{\alpha} \quad ; p \in (0,1)$$

Reversing this to get the QDF simply gives

$$x_p = k * p^{1/\alpha}$$

Then the basic quantile distribution function for power is

$$Q(p) = x_p = k * p^{\beta} \qquad \text{where } \beta = \frac{1}{\alpha}$$

The range of the distribution from p=0 to p=1 is $(0, k)$

Hence, the Power quantile distribution is

$$Q(p) = \lambda + \eta\, k * (\mathrm{p}^{\beta}) \quad ; \quad \beta \succ 0$$

The range of the distribution from p=0 and p=1 is $(\lambda, \lambda + \eta k)$.

Properties of Power Quantile Distribution

Some properties of Power quantile distribution can be seen as follows

- Median

$$M = Q(0.5) = \lambda + \eta k(0.5)^{\beta} \qquad 4.44$$

- Inter p - Range

$$R = Q((1-p)\lambda, \eta, \beta) - Q(p, \lambda, \eta, \beta)$$

$$R = \eta k[(1-p)^{\beta} - p^{\beta}] \qquad ; p \prec \frac{1}{2} \qquad 4.45$$

- Difference

$$D = Q((1-p), \lambda, \eta, \beta) + Q(p, \lambda, \eta, \beta) - 2m$$

$$D = \eta k[(1-p)^{\beta} + p^{\beta} - 2(0.5)^{\beta}] \qquad 4.46$$

Estimation of Parameters

- Location

$$\hat{\lambda} = m - (0.5)^{\beta}\left(\frac{r}{((1-p)^{\beta} - p^{\beta})}\right) \qquad 4.47$$

- Scale

$$\hat{\eta} = \frac{r}{k((1-p)^{\beta} - p^{\beta})} \qquad 4.48$$

- Skewness (Galton p-Skewness)

δ =Quantile p-Difference/Quantile inter p-Range

$$\delta = \frac{((1-p)^{\hat{\beta}} + p^{\hat{\beta}} - 2(0.5)^{\hat{\beta}})}{((1-p)^{\hat{\beta}} - p^{\hat{\beta}})} \quad 4.49$$

Measurement of the distance of tails from the median, the Right Percentile Range (RPR_q) and Left Percentile Range (LPR_q) of Power quantile distribution, can be expressed as follows

$$RPR_q = x_{1-p} - m = \eta\{(1-p)^{\beta} - 0.5^{\beta}\}$$

$$LPR_q = m - x_p = \eta\{0.5^{\beta} - p^{\beta}\}$$

4.9 Quantile Function for Pareto Distribution

The density function of Pareto distribution is

$$f(x) = \gamma x^{-(\gamma+1)}, \qquad 1 \le x \le \infty$$

And the Cumulative Distribution Function (CDF) is

$$F(x) = 1 - x^{-(\gamma+1)+1}$$

Then

$$F(x_{(p)}) = p = prob(X \le x_p) = prob(X \prec Q(p))$$

Where $x_p = Q(p)$ is Quantile distribution function (QDF), i.e. the QDF is the inverse of the CDF

$$F(x_p) = p = 1 - x_p^{-(\gamma+1)+1} \qquad ; p \in (0,1)$$

Reversing this to get the QDF simply gives

$$x_p = \frac{1}{(1-p)^{\frac{1}{\gamma}}}$$

Then the basic quantile distribution function for Pareto is

$$Q(p) = x_p = \frac{1}{(1-p)^{\beta}} \qquad \text{where } \beta = \frac{1}{\gamma}$$

The range of the distribution from p=0 and p=1 is $(1, \infty)$

Hence, the Pareto quantile distribution is

$$Q(p) = \lambda + \eta(\frac{1}{(1-p)^{\beta}}) \qquad , \beta \succ 0$$

The range of the distribution from p=0 and p=1 is $(\lambda + \eta, \infty)$.

Where λ and η are location, scale parameters, and β is the distribution's shape.

Properties of Pareto Quantile Distribution

Some properties of Pareto quantile distribution can be seen as follows

- Median

$$M = Q(0.5) = \lambda + \eta \frac{1}{(0.5)^{\beta}} = \lambda + \eta * 2^{\beta} \qquad \mathbf{4.50}$$

- Inter p - Range

$$R = Q((1-p)\lambda, \eta, \beta) - Q(p, \lambda, \eta, \beta)$$

$$R = \eta(\frac{1}{p^{\beta}} - \frac{1}{(1-p)^{\beta}}) \qquad ; p \prec \frac{1}{2} \qquad \mathbf{4.51}$$

- Difference

$$D = Q((1-p), \lambda, \eta, \beta) + Q(p, \lambda, \eta, \beta) - 2m$$

$$D = \eta(\frac{1}{p^{\beta}} + \frac{1}{(1-p)^{\beta}} - 2 * 2^{\beta}) \qquad \mathbf{4.52}$$

- Inter p- Range (R) > Difference when $\beta \succ 0$, $p \prec \frac{1}{2}$

Estimation of Parameters

- Location

$$\lambda^{\wedge} = m - 2^{\beta}(r/(\frac{1}{p^{\beta}} - \frac{1}{(1-p)^{\beta}})) \qquad 4.53$$

- Scale

$$\eta^{\wedge} = (r/(\frac{1}{p^{\beta}} - \frac{1}{(1-p)^{\beta}})) \qquad 4.54$$

- Skewness

δ =Quantile p-Difference/Quantile inter p-Range

$$\delta = \left(\frac{(\frac{1}{p^{\beta^{\wedge}}} + \frac{1}{(1-p)^{\beta^{\wedge}}} - 2*2^{\beta^{\wedge}})}{(\frac{1}{p^{\beta^{\wedge}}} - \frac{1}{(1-p)^{\beta^{\wedge}}})} \right) \qquad 4.55$$

Measurement of the distance of tails from the median, the Right p Percentile Range (RPR_q) and Left p Percentile Range (LPR_q) for Pareto quantile distribution, can be expressed as follows

$$RPR_q = x_{1-p} - m = \eta\{\frac{1}{p^{\beta}} - 2^{\beta}\} \qquad ; p \prec \frac{1}{2}$$

$$LPR_q = m - x_p = \eta\{2^{\beta} - \frac{1}{(1-p)^{\beta}}\} \qquad ; p \prec \frac{1}{2}$$

4.10 Quantile Function for Geometric Distribution

The density function of geometric distribution is

$$f(x) = \theta(1-\theta)^{x-1} \qquad ; \; x = 1,2,3,\ldots$$

and the Cumulative Distribution Function (CDF) is

$$F(x) = \theta\sum_{x=1}^{x}(1-\theta)^{x-1}$$

Then

$$F(x_{(p)}) = p = prob(X \leq x_p) = prob(X \prec Q(p))$$

where $x_p = Q(p)$ is Quantile distribution function (QDF), i.e. the QDF is the inverse of the CDF

$$F(x_p) = p = \theta\sum_{x=1}^{x}(1-\theta)^{x_p-1} \quad ; p \in (0,1)$$

Reversing this to get the QDF simply gives

$$p = \theta\left(\frac{1-(1-\theta)^{x_p-1+1}}{1-(1-\theta)}\right)$$

$$p = 1-(1-\theta)^{x_p}$$

$$x_p = \frac{\ln(1-p)}{\ln(1-\theta)}$$

Here, x_p is a discrete distribution, so the basic quantile function for geometric distribution is

$$Q(p) = x_p = INT\left(\frac{\ln(1-p)}{\ln(1-\theta)}\right)+1$$

The range of the distribution from p=0 to p=1 is $(1,\infty)$

Hence, the GLD for geometric quantile distribution is

$$Q(p) = \lambda + \eta * \left(INT\left(\frac{\ln(1-p)}{\ln(1-\theta)}\right)+1\right)$$

The range of the distribution from p=0 and p=1 is $(\lambda + \eta, \infty)$.

Properties of Geometric Quantile Distribution

Some properties of geometric quantile distribution can be seen as follows

- Median

$$M = Q(0.5) = \lambda + \eta * \left(INT\left(\frac{\ln 0.5}{\ln(1-\theta)}\right)+1\right) \quad 4.56$$

- Inter p - Range

$$R = Q((1-p),\lambda,\eta) - Q(p,\lambda,\eta)$$

$$R = \eta * \left[\left(INT\left(\frac{\ln p}{\ln(1-\theta)}\right)+1\right) - \left(INT\left(\frac{\ln(1-p)}{\ln(1-\theta)}\right)+1\right)\right] \quad 4.57$$

- Difference

$$D = Q((1-p),\lambda,\eta) + Q(p,\lambda,\eta) - 2m$$

$$D = \eta * \left[\left(INT\left(\frac{\ln p}{\ln(1-\theta)}\right)+1\right) + \left(INT\left(\frac{\ln(1-p)}{\ln(1-\theta)}\right)+1\right) - 2 * \left(INT\left(\frac{\ln 0.5}{\ln(1-\theta)}\right)+1\right)\right] \quad 4.58$$

Estimation of Parameters

- Location

$$\hat{\lambda} = m - \left(INT\left(\frac{\ln 0.5}{\ln(1-\theta)} \right) + 1 \right) * \left(r \Big/ \left[\left(INT\left(\frac{\ln p}{\ln(1-\theta)} \right) + 1 \right) - \left(INT\left(\frac{\ln(1-p)}{\ln(1-\theta)} \right) + 1 \right) \right] \right) \qquad 4.59$$

- Scale

$$\hat{\eta} = r \Big/ \left[\left(INT\left(\frac{\ln p}{\ln(1-\theta)} \right) + 1 \right) - \left(INT\left(\frac{\ln(1-p)}{\ln(1-\theta)} \right) + 1 \right) \right] \qquad 4.60$$

- Skewness (Galton p-Skewness)

δ =Quantile p-Difference/Quantile inter p-Range

$$\delta = \frac{\left[\left(INT\left(\frac{\ln p}{\ln(1-\theta)} \right) + 1 \right) + \left(INT\left(\frac{\ln(1-p)}{\ln(1-\theta)} \right) + 1 \right) - 2 * \left(INT\left(\frac{\ln 0.5}{\ln(1-\theta)} \right) + 1 \right) \right]}{\left[\left(INT\left(\frac{\ln p}{\ln(1-\theta)} \right) + 1 \right) - \left(INT\left(\frac{\ln(1-p)}{\ln(1-\theta)} \right) + 1 \right) \right]} \qquad 4.61$$

Measurement of the distance of tails from the median, the Right Percentile Range (RPR_q) and Left Percentile Range (LPR_q) of geometric quantile distribution, can be expressed as follows

$$RPR_q = x_{1-p} - m = \eta * \left(\left(INT\left(\frac{\ln p}{\ln(1-\theta)} \right) + 1 \right) - \left(INT\left(\frac{\ln 0.5}{\ln(1-\theta)} \right) + 1 \right) \right) \qquad 4.62$$

$$LPR_q = m - x_p = \eta * \left(\left(INT\left(\frac{\ln 0.5}{\ln(1-\theta)} \right) + 1 \right) - \left(INT\left(\frac{\ln(1-p)}{\ln(1-\theta)} \right) + 1 \right) \right) \qquad 4.63$$

4.11 Summary

It has provided the theoretical development of quantile approach for non-normal distribution, such as logistic, exponential, Uniform, extreme-value, Weibull, power and pareto distribution for variable measurement and geometric distribution for attribute data. Moreover, it also estimate the parameters of the distributions mentioned above. In the next chapter, we will be discussing the evaluation of quantile control chart for non-normal situation.

Chapter 5: An Evaluation of Quantile Control Chart for Non-Normal Situation

5.1 Introduction

A successful quality improvement process must be based on proper interpretation of statistical data and quality improvement methods. In this chapter we discuss the application of quality improvement process through quantile approach. In doing so we will first discuss the quantile process of monitoring and control and then develop a quality control chart for this purpose using the median rankit, which will be called a median rankit control chart.

Padgett and Spurrier (1990), discussed Shewhart type charts for 100^{th} percentiles of the Weibull and lognormal distributions assuming unknown parameters. Kittlitz (1999) suggested that the long tailed positively skew exponential distribution, could be made into an almost symmetric distribution by taking the fourth root of the data. The transformation data can then be plotted conveniently on an individuals chart, EWMA, or CUSUM chart for statistical process control. For EWMA, Montgomery, Gardiner, and Pizzano (1987) recommend values of λ in the range of $0.05 \le \lambda \le 0.5$, with smaller values of λ being more effective in detecting smaller shift in the mean.

The Shewhart control chart for individual measurement is often used in situations that involve rational subgroups of size n=1 in process monitoring and control. When the assumption of normality is violated, the average run length (ARL) of the individual control chart is adversely affected. For example, ARL for 3σ is 370.4, and ARL for 3.02σ is 395.6 under normality assumptions, the difference between them with 2%

shift in process is obviously not small. Therefore, if we compute the in-control ARL for various non-normal distributions with control limits constructed under the assumption of normality, we will obtained inaccurate results, (see Ryan 2000 and Wheeler 2000). For more details in respect to ARL, see chapter seven. Further, Borror, Montgomery and Runger (1999) showed that EWMA control charts can be designed to be robust to the normality assumption. This implies that, ARL is reasonably close to the normal-theory value for both skewed and heavy-tailed symmetric non-normal distributions.

Box-Cox transformation can be used to transform the data from non-normal to normal situation. However, the Box-Cox transformation is only suitable for non-negative and non-discrete distribution. But it was found that the quantile distribution method can be used for any continuous sets of data.

However, most continuous distributions can be defined very simply in terms of the quantile distribution function. This approach to defining distributions enables the two tails of a distribution to be almost independently modelled. This is a very useful property for handling non-normal distribution

From the above, it is clear that by using the quantile approach, we can easily avoid the use of transformations (e.g. Box-Cox transformation, Quesenberry) for non-normal data in order to obtain a control chart.

Nelson, P. (1979), presented limits for weibull median and range charts, and proposes two additional (location, scale) control charts. According to Nelson, charts constructed with these limits have a risk of 0.003. The centrelines are positioned in such a way, that points have equal probability of falling above or below them. However, a family of weibull distributions approximates many empirical distributions, and provides a model for life and failure situation data. This chapter will present some accurate control limits using median rankit control charts for logistic, exponential, extreme-value, weibull and power distributions. In this chapter we will provide an example of such distributions to indicate how quantile approach could be used to construct control charts for non-normal distribution using median rankit.

5.2 Quantile Control Chart for Non-Normal Distribution

The control chart for quantile distribution can be constructed by using the $Q(p) = \lambda + \eta R(p;\theta)$ for the distribution which has got more than two parameters and using $Q(p) = \lambda + \eta R(p)$ for the distribution which has got two parameters. The action and warning limits can be derived from the formulas above, where the warning limits are Q(0.05) and Q(0.95), and the action limits are Q(0.01) and Q(0.99), and the central point is at Q(0.5).

The steps to construct control limits of quantile distributions are as follows:

- Development of the quantile distribution function

$$Q(p) = \lambda + \eta R(p;\theta)$$

- Estimate the parameters λ, η, δ by using least absolute method then

$$Q(p) = \hat{\lambda} + \hat{\eta} R(p, \hat{\theta})$$

Where $\lambda^{\wedge}, \eta^{\wedge}, \theta^{\wedge}$ are location, scale and skewness respectively.

- The control limits of the quantile distribution function can be obtained by substituting p=0.5 in Q (p) above. It will provide the central point which will be described as median rankit point, and similarly by substituting p=0.05, p=0.95 and p=0.01, p=0.99 will provide both the warning limits and action limits respectively of median rankit point.

5.3 Quantile Control Chart for Logistic Distribution

The median rankit control chart for the Logistic Quantile Distribution (L.Q.D) is described below. The action and warning limits, which are used in the control chart procedures, can be derived as

$$Q(p) = \lambda + \frac{\eta}{2}\{(1-\delta)\ln(p) - (1+\delta)\ln(1-p)\}$$

where the warning limits are Q (0.05) and Q (0.95), the action limits are Q(0.99) and Q (0.01), and the central point (median rankit) is at Q (0.5). Therefore, a typical quantile control chart that can be constructed for logistical quantile distribution is given in the following steps.

The steps to construct control limits of logistic distribution is as follows:

- Development of logistic quantile distribution function

$$Q(p) = \lambda + \frac{\eta}{2}\{(1-\delta)\ln(p) - (1+\delta)\ln(1-p)\}$$

- Estimate the parameters λ, η, δ by using least absolute method (median rankit), then

$$Q(p) = \lambda^{\wedge} + \frac{\eta^{\wedge}}{2}((1-\delta^{\wedge})\ln p - (1+\delta^{\wedge})\ln(1-p))$$

where $\lambda^{\wedge}, \eta^{\wedge}, \delta^{\wedge}$ are location, scale and skewness respectively.

- The control limits of the logistic quantile distribution function can be obtained by substituting p=0.5 in Q (p) above. It will provide the central point which will be described as median rankit point, and similarly by substituting p=0.05, p=0.95 and p=0.01, p=0.99, will provide both the warning limits and action limits respectively of median rankit point.

Example: Real Data. (Chang & Lu. 1994).

Below are 65 observations of the thickness of an oil seal, the sampling distribution is not known:

Table 1

2.4	2.2	2.0	1.9	1.8	1.9	2.0	1.8	2.0	1.6	2.2
2.3	2.4	1.8	1.8	1.9	1.6	2.1	1.8	2.1	1.6	2.0
2.0	2.1	2.3	2.1	2.1	1.9	2.1	1.8	1.8	2.1	2.2
2.2	2.0	2.0	1.8	1.7	2.4	2.0	2.0	2.1	1.9	2.1
2.2	2.2	2.4	2.0	1.6	1.9	1.9	2.0	1.7	1.8	2.3
2.2	2.0	2.4	2.3	2.2	2.1	2.5	1.9	2.0	1.9	

Chang and Lu (1994), mentioned that the data appear to come from a skew distribution. We are interested in finding out whether the thickness of an oil seal is outside the control limits of the production process.

The process of estimation and validation on a real set of data, which is believed to follow the logistic distribution, has been investigated. The data was compared with the model, by comparing the observed value and the fitted value. The observed values are the original set of data under investigation and the fitted values are the values, which are obtained when QDF is fitted to the model together with the scale, location and skewness. For a good fit of the data, the series of points is expected to lie on 45^0 line, which passes through the origin. Where a best model is found, it will then be used to construct a median rankit control chart.

Distribution of Least Absolute

The steps below are required to estimate distribution parameters

1. Find the initial value of $p_{(r)}$, where $p_{(r)} = (BETAINV(0.5, r, n-r+1))$.
2. Sort the skew logistic data which are treated in ascending order.
3. Put initial parameter values using quantile method for location, scale and skewness.

4. Fitted $Q(p) = \lambda + \frac{\eta}{2}\{(1-\delta)\ln(p) - (1+\delta)\ln(1-p)\}$ model,

5. After fitting Q (p), the Solver tool in Excel Package is used to estimate the minimum values of the parameters by using the least absolute method

6. Calculate the residual sum of least absolute

$$\hat{e}_r = x_r - \hat{Q}(\frac{r}{n+1}, \hat{\lambda}, \hat{\eta}, \hat{\delta}).$$

Then the estimation of the parameters λ, η, δ for real data given in table 2 are 2.011055, 0.253986, and 0.04226 respectively. The residual sum of least absolute is **2.041334**.

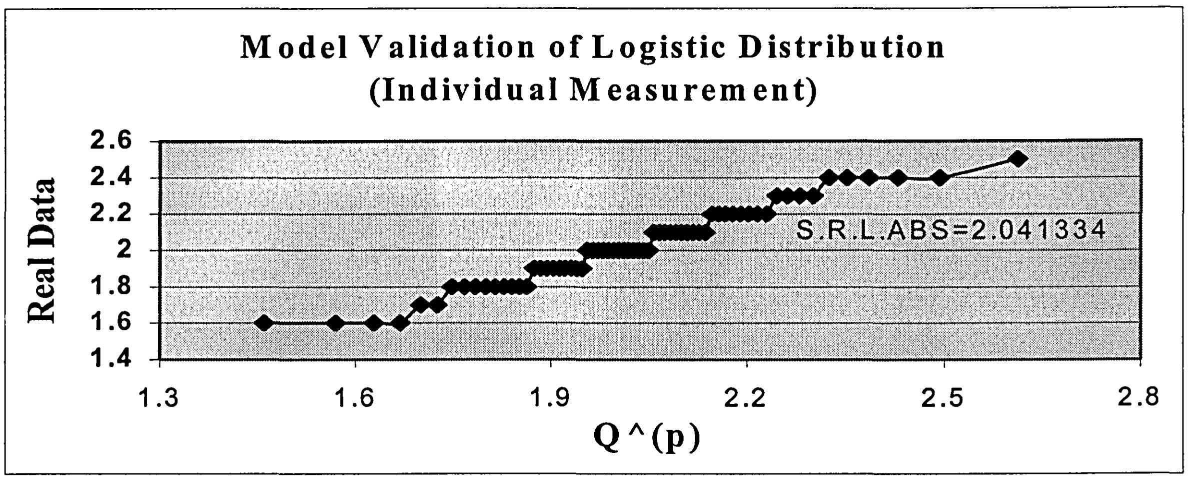


Figure 1: Model Validation Median Rankit for Logistic Distribution

By analysing figure 1, we can see that the model appears to give a reasonable fit. Following the verification of the above data as logistic distribution, the quality control limits of quantile logistic distribution can be calculated for median rankit at various p values and are shown in table 2 . Control limits for median rankit are calculated at p=0.05 and p=0.01 for warning and action limits respectively using the formula

$$Q(p) = \lambda + \frac{\eta}{2}\{(1-\delta)\ln(p) - (1+\delta)\ln(1-p)\} \cdot$$

Here λ, η, δ are given by 2.011055, 0.253986, and 0.04226 respectively and the residual sum of least absolute is 2.0413. This will provide the required control limits as follows. Central point =2.018495, Warning limits = (1.653484, 2.40133) and Action limits = (1.452276, 2.619371). Figure 2 shows the logistic control limits for median rankit at 2.0184 for levels p=0.01 and p=0.05 & p=0.005 and p=0.001 (see figure 3). Here no action is necessary as all the values are within the action limits.

Table 2 Quantile control limits for logistic distribution.

Percentile values (p)	Median Rankit (Least Absolute)		
	Q(p)	**Q(0.5)**	**Q(1-p)**
0.01	1.452276	2.018495	2.619371
0.05	1.653484	2.018495	2.40133
0.001	1.171023	2.018495	2.925241
0.005	1.367304	2.018495	2.711729
0.00135	1.20757	2.018495	2.885477

Figure 3 shows the logistic control limits for median rankit at levels p=0.001 and p=0.005, where the warning limits = (1.367304, 2.711729) and the action limits = (1.171023, 2.925241). Here no action is necessary as all the values are within the warning and action limits.

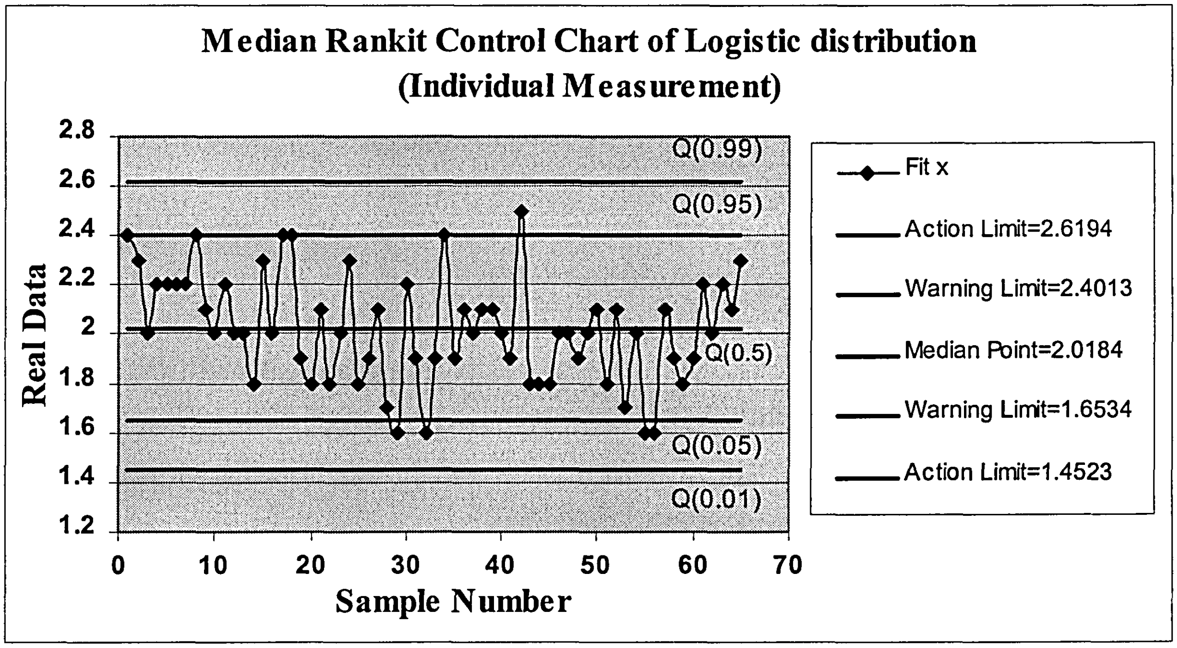


Figure 2 : Median Rankit Control Chart for Logistic Distribution

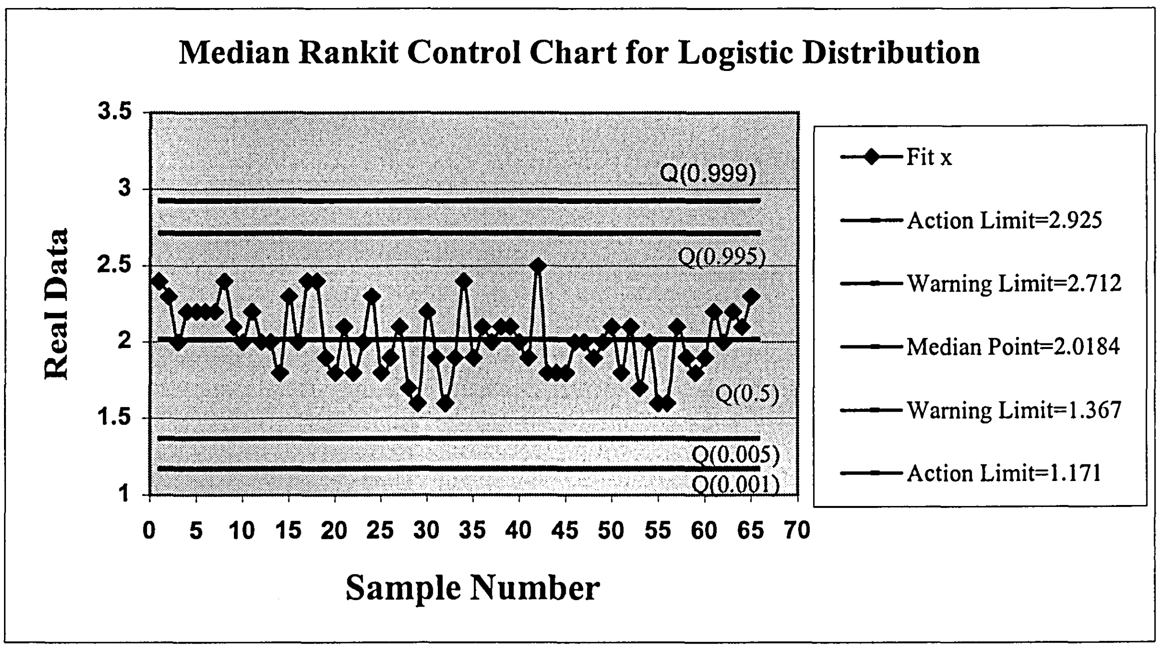


Figure 3: Median Rankit Control Chart for Logistic Distribution at p=0.001& p=0.005.

It is clear from the control chart in figure 2 that the sample numbers 29,32,42,55 and 56 are outside the warning limits respectively. These points must be investigated to see whether an assignable cause can be determined. Furthermore, the chart shows that no single point is outside the action limits i.e. the production process is in control at p=0.01.

In order to compare our methodology with the previous works, we have calculated a control chart for exponentially weighted moving average (EWMA) using the data in table 1. Figure 4 provides the control chart for the data in table 1, using EWMA method.

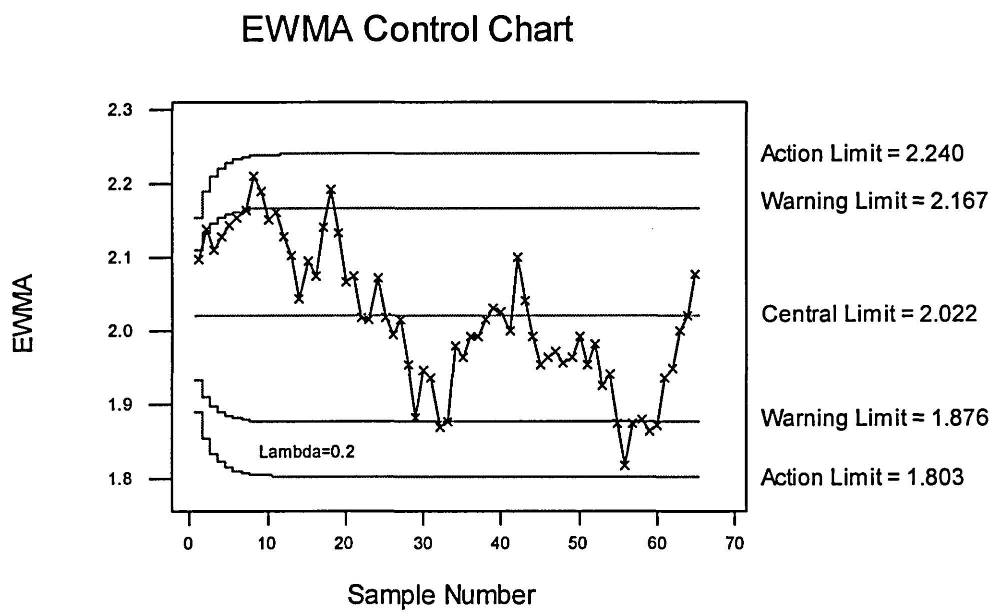


Figure 4: Control Chart for Exponentially Weighted Moving Average

The control limits in both figure 2 and 4 indicate that there are no signal points that lie outside control limits at point p=0.01. Whereas, there are some points that lie outside control limits at p=0.05.

By using EWMA with $\lambda = 1$, we obtained the Shewhart individual control chart of figure 5. The control limits in both figure 2 and figure 5 are nearly the same, due to the fact that the shape of the logistic distribution is close to the shape of normal distribution, especially when the skewness coefficient is small. However, the method of individual control chart is not suitable for non-normal distribution when skewness coefficient is large. Montgomery (1997) has pointed out that moving range and individual measurements charts can provide inappropriate control limits for non-normal data. EWMA is a better alternative to the Shewhart control chart when the aim is to detect a small shift. The individual chart is not robust to the normality assumption, when false alarms are concerned. Both the Shewhart and EWMA charts demonstrate the ability to detect shifts quickly, but the Shewhart chart has a higher false alarm rate (Borror *et al.* 1999). In this research we are addressing these specific issues.

From the results above, we can conclude that the quantile approach is applicable to the dealing with non-normal data.

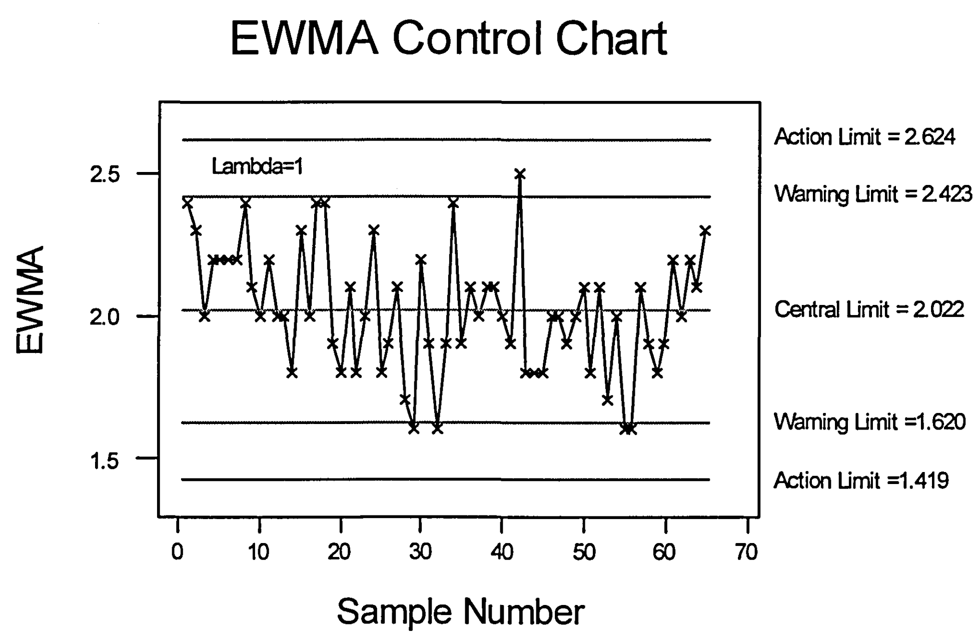


Figure 5: EWMA Control Chart at Lambda =1

5.4 Quantile Control Chart for Exponential Distribution

The median rankit control chart for the exponential quantile distribution can be described as

$$Q(p) = \lambda + \eta(-\ln(1-p))$$

The action and warning limits, which are used in the control chart procedures, can be derived from the $Q(p)$ above. Where the warning limits are Q (0.05) and Q (0.95), the action limits are Q(0.01) and Q (0.99), and the central point (median rankit) is at Q (0.5). A typical quantile control chart for the Exponential Quantile Distribution is given in the following steps.

Required steps for setting up the control limits of exponential distribution are as follows:

- Development of the Exponential quantile distribution function

$$Q(p) = \lambda + \eta(-\ln(1-p))$$

- Estimate the parameters λ, η by using least absolute method then

$$Q(p) = \hat{\lambda} + \hat{\eta}(-\ln(1-p))$$

Where $\lambda^{\wedge}, \eta^{\wedge}$ are location and scale parameters respectively.

- The control limits of the exponential quantile distribution function can be obtained by substituting p=0.5 in Q (p) above. The latter provides the central point, which will be described as median rankit point. Similarly by substituting p=0.05, p=0.95 and p=0.01, p=0.99, both the warning limits and action limits respectively for median rankit point would be obtained.

Example to Apply Quantile Control Chart for Exponential Data.

Data

We have generated a 30 random number from exponential distribution using "Minitab Release 12 and Excel 97" where the mean is equal to one. The process of estimation and validation on a set of data, which is believed to follow the exponential distribution, have been investigated. The data was fitted to the model, to compare the observed value and the fitted value. The observed values are the original set of data under investigation and the fitted values are the values, which are obtained when QDF is fitted to the model. For a good fit of the data, the series of points is expected to lie on 45^0 line, which passes through the origin. Where a best model is found, it will then be used to construct median rankit control chart.

Table 3 : 30 random numbers for exponential distribution.

0.45729	3.35360	1.34826	0.32315	0.08523
0.47807	1.68641	0.73215	1.91830	0.24939
0.57271	0.41638	0.74227	0.37895	0.97464
0.34155	0.84433	1.26890	0.35953	0.95657
0.46069	2.27823	1.12539	0.94095	0.78817
1.40025	0.23362	0.77088	1.03936	3.12027

Estimations of the parameters & residual

The estimate of the parameters of location λ and scale η for the data given in table 3 are 0.173562 and 0.828119 respectively. Here the residual sum of least absolute is equal to 1.862325.

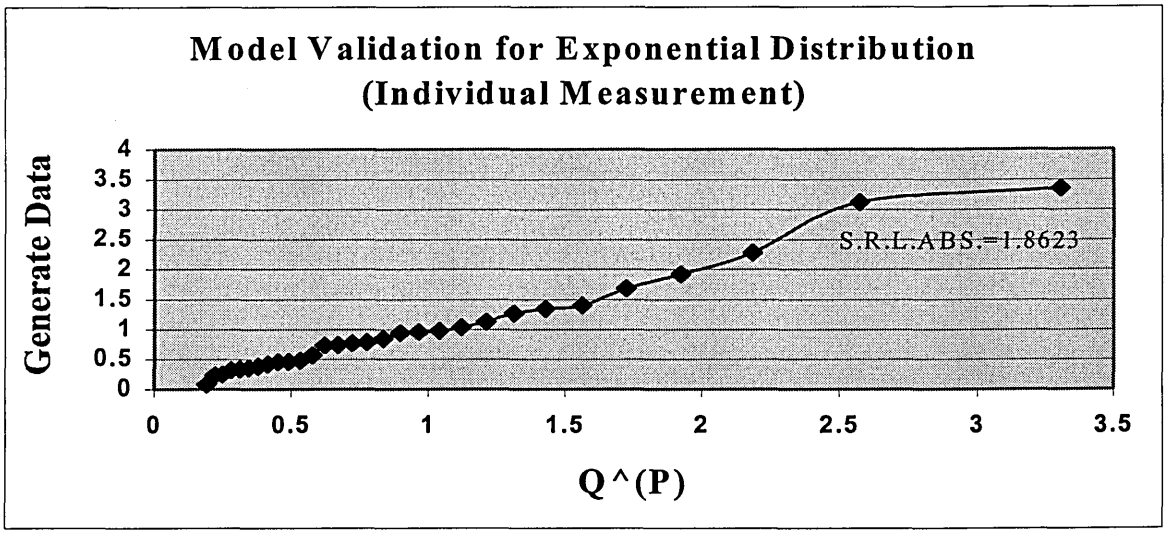


Figure 6: Model Validation using Median Rankit for Exponential

Model Validation

Figure 6, indicates that the data provides a good fit for the exponential distribution, because the points lie approximately on the straight line 45 degrees to the horizontal axis.

On the confirmation that the data follows the exponential distribution, a quality control limits of quantile exponential distribution is provided for median rankit at various p values, in table 4. Various control limits are calculated (p=0.05 and p=0.01) for warning and action limits respectively using the formula

$$Q(p) = \lambda + \eta(-\ln(1-p))$$

Here λ, η are given by 0.173562 and 0.828119 respectively. This will provide the required control limit for the data as follows.

Central point =0.747571, Warning limits = (0.216039, 2.654384) and Action limits = (0.181885, 3.98719).

Table 4: Quantile control limits for exponential distribution.

P	Warning Limit Q(p)	Median Point Q(0.5)	Action Limit Q(1-p)
0.01	0.181885	0.747571	3.98719
0.05	0.216039	0.747571	2.654384
0.001	0.174391	0.747571	5.894004
0.005	0.177713	0.747571	4.561198
0.00135	0.174681	0.747571	5.645482

Median rankit control chart

Figure 7 provides the exponential median rankit control chart at 0.747571. It is clear from control chart in figure 7 that the sample numbers 7,30 are outside the warning limits respectively and the sample number 25 is outside the action limit. These points must be investigated to see whether an assignable cause can be determined.

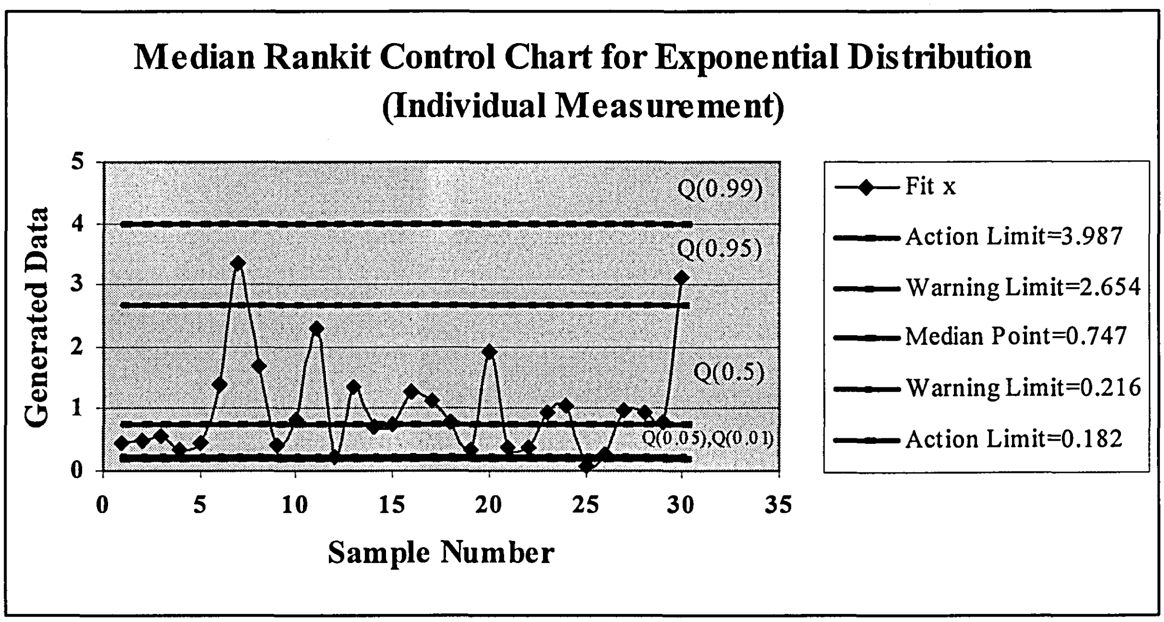


Figure 7: Median rankit control chart for exponential distribution

5.5 Quantile Control Chart for Extreme-value Distribution

The median rankit control chart for the Extreme-value Quantile Distribution is described below. The action and warning limits, which are used in the control chart procedures, can be derived as

$$Q(p) = \lambda + \eta\{-\ln(-\ln p)\}$$

where the warning limits are Q (0.05) and Q (0.95), the action limits are Q(0.01) and Q (0.99), the central point (median rankit) is at Q (0.5). Therefore a typical quantile control chart that can be constructed for Extreme-value quantile distribution is given in the following steps.

The steps to construct control limits of Extreme-value are as follows

- Development the Extreme-value quantile distribution function

$$Q(p) = \lambda + \eta\{-\ln(-\ln p)\}$$

- Estimate the parameters λ, η by using least absolute method (median rankit) then

$$Q(p) = \hat{\lambda} + \hat{\eta}\{-\ln(-\ln p)\}$$

Where $\lambda^{\wedge}, \eta^{\wedge}$, are location and scale respectively.

- The control limits of the Extreme-value quantile distribution function can be obtained by substituting p=0.5 in Q (p) above. The latter provides the central point, which will be describe as median rankit point. Similarly by substituting p=0.05,p=0.95 and p=0.01, p=0.99 will provide both the warning limits and action limits respectively of median rankit point.

5.6 Quantile Control Chart for Weibull Distribution

The median rankit control chart for the weibull quantile distribution (WQD) can be described as

$$Q(p) = \lambda + \eta(-\ln(1-p))^{\beta} \quad ,\beta \succ 0$$

The action and warning limits, which are used in the control chart procedures, can be derived from the (WQD) above. Where the warning limits are Q (0.05) and Q (0.95), the action limits are Q(0.01) and Q (0.99), and the central point (median rankit) is at Q (0.5). A typical quantile control chart for the Weibull Quantile Distribution is given in the following steps.

Required steps for setting up the control limits of weibull distribution are as follows:

- Development of the weibull quantile distribution function

$$Q(p) = \lambda + \eta(-\ln(1-p))^{\beta}$$

- Estimate the parameters λ, η, β by using least absolute method then

$$Q(p) = \lambda^{\wedge} + \eta^{\wedge}(-\ln(1-p))^{\beta^{\wedge}}$$

Where $\lambda^{\wedge}, \eta^{\wedge}, \beta^{\wedge}$ are location, scale and shape parameters respectively.

- The control limits of the weibull quantile distribution function can be obtained by substituting p=0.5 in Q (p) above. It will provide the central point, which will be described as median rankit point and similarly by substituting p=0.05, p=0.95 and p=0.01, p=0.99, which will provide both the warning limits and action limits respectively for median rankit point.

Example to Apply Quantile Control Chart for Weibull Data.

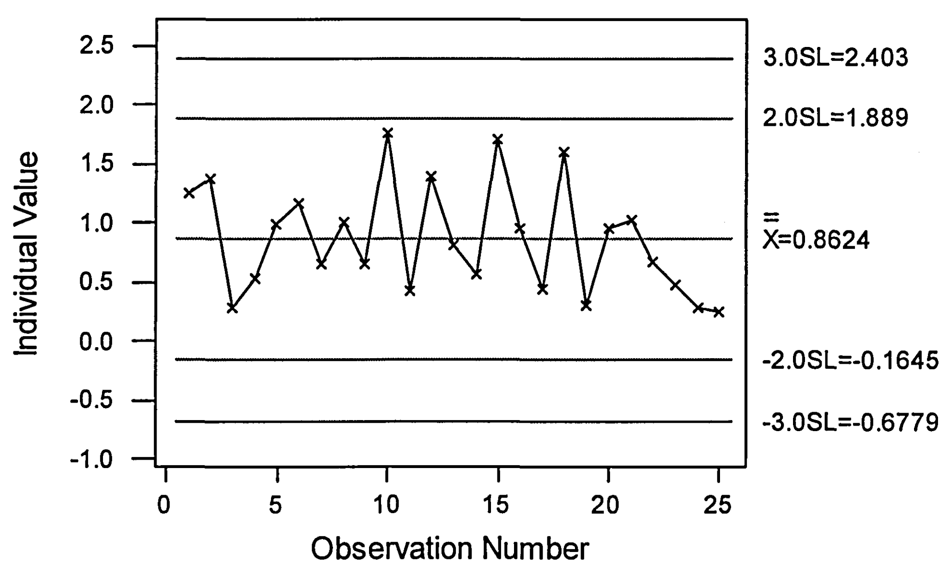
Data

The data in table 5 are the times to failure measurement of 25 light bulbs on accelerated test. The data was taken from Wadsworth (1998, pp: 6.15). Here we are interested in finding out whether the failure times for 25 light bulbs are within the specific acceptance limit of the production process. First of all, the data was fitted to the weibull model, in order to evaluate the suitability of the model and then a control chart was developed to check the conformity of the production process.

Table 5: Times to Failure for Light Bulbs (Months).

1.25	1.17	0.42	0.96	1.03
1.37	0.65	1.39	0.45	0.67
0.28	1.00	0.82	1.61	0.48
0.53	0.66	0.57	0.31	0.29
0.98	1.76	1.71	0.95	0.25

The graph below provide the traditional Shewhart control chart for individual measurement, assuming the quality characteristics are following normal distribution. It can be conclude from this graph that, there is no reason to reject that, the process is in control. In the remain parts of this example, we will construct control chart using quantile approach.



Traditional Shewhart Control Chart for Individual Measurement

Estimations of the parameters & residual

The estimates of the parameters λ, η, β for the data given in table 5 are 0.008078, 0.96979 and 0.563196 respectively. Here, the residual sum of least absolute is equal to 1.142144.

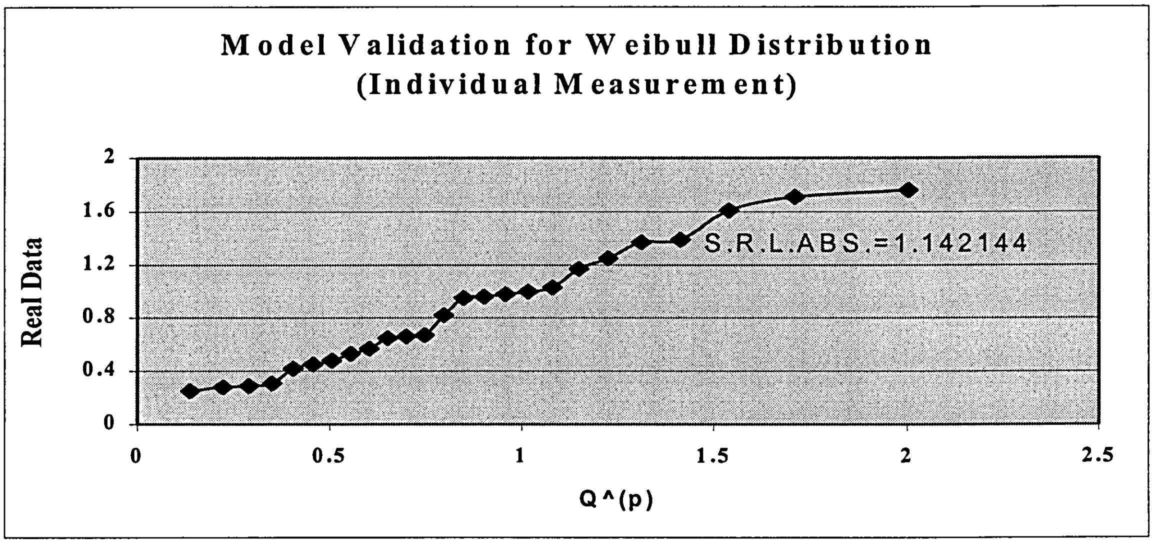


Figure 8: Model Validation Using Median Rankit values for Weibull Distribution

Model Validation

Figure 8 indicates that the data provides a good fit for the weibull distribution, because the points lie approximately on the straight line 45 degrees to the horizontal axis.

On confirmation that the data follows the weibull distribution, a quality control limits of quantile weibull distribution is provided for median rankit at various p values, in table 6. Various control limits are calculated (p=0.05 and p=0.01), for warning and action limits respectively using the formula:

$$Q(p) = \lambda + \eta(-\ln(1-p))^{\beta}$$

Here λ, η, β are given by are 0.008078, 0.96979 and 0.563196 respectively. This will provide the required control limit for the data as follows.

Central point =0.796995, Warning limits = (0.190127, 1.807124) and Action limits =(0.080774, 2.30008).

Table 6: Quantile control limits for weibull distribution.

P	Warning Limit Q(p)	Median Point Q(0.5)	Action Limit Q(1-p)
0.01	0.080774	0.796995	2.30008
0.05	0.190127	0.796995	1.807124
0.001	0.027903	0.796995	2.888053
0.005	0.057209	0.796995	2.488407
0.00135	0.031556	0.796995	2.816904

Median rankit control chart

Figure 9 provides the weibull median rankit control chart at 0.796995. The action and warning control limits in figure 9, indicates no action should be necessary, as all the values are below the action and warning limits at level p=0.01and p=0.05 respectively.

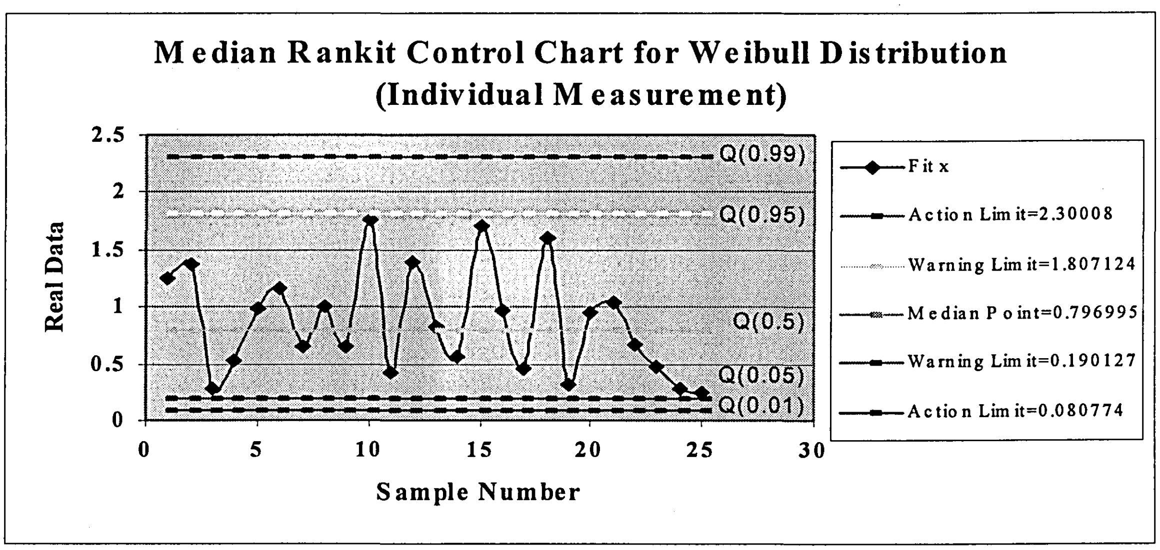


Figure 9: Median rankit control chart for weibull distribution

From the model validation (see figure 8), we are reasonably happy that the model is a good fit for the data with S.R.L.ABS=1.142144. However, for further improvement of the control chart, this data was tested for other distributions. Although, weibull distribution is a good fit for this data, we will now consider whether any other distribution of the weibull family provides an improved control chart. Accordingly, a power distribution is applied as follows in the next section:

5.7 Quantile Control Chart for Power Distribution

The density function of Power distribution is

$$f(x) = \frac{\alpha}{k}\left(\frac{\mathrm{x}}{\mathrm{k}}\right)^{\alpha-1} \quad ; \; 0 \prec x \leq k \;,\; \mathrm{k} \succ 0 \;,\; \alpha \succ 0$$

and the Cumulative Distribution Function (CDF) is

$$F(x_p) = p = \left(\frac{x_p}{k}\right)^{\alpha} \quad ; p \in (0,1)$$

Hence, the Power quantile distribution is

$$Q(p) = \lambda + \eta\, k * (\mathrm{p}^{\beta}) \quad ; \; \beta \succ 0$$

The range of the distribution from p=0 and p=1 is $(\lambda, \lambda + \eta k)$.

Where λ=0.272271, η=0.859574 are location and scale parameters respectively. β=1.574089 is the distribution's shape and k=1.76. Here, the residual sum of least absolute is equal to 1.017205. In order to obtain the properties and estimation parameters of power quantile distribution (see chapter 4).

Model Validation for Power Distribution

We can easily recognise that the data set used in this section follows a Weibull distribution (see probability plot figure 10). However, figure 11 indicates that the data provides a better fit for the Power distribution where most of the points lie an approximately 45 degrees to the horizontal axis.

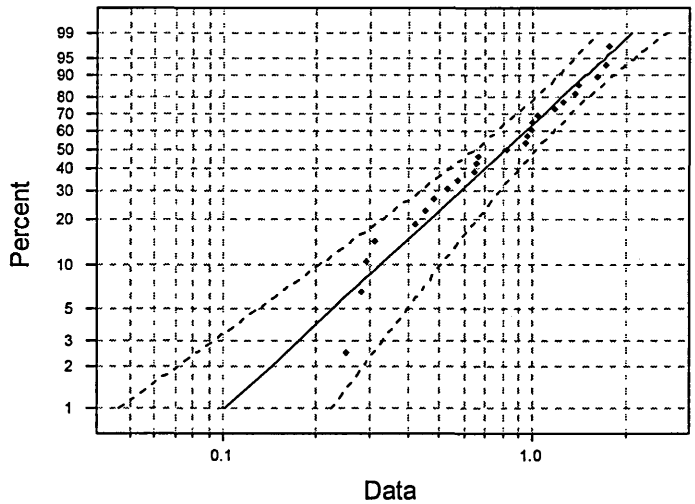


Figure 10: Weibull Probability Plot

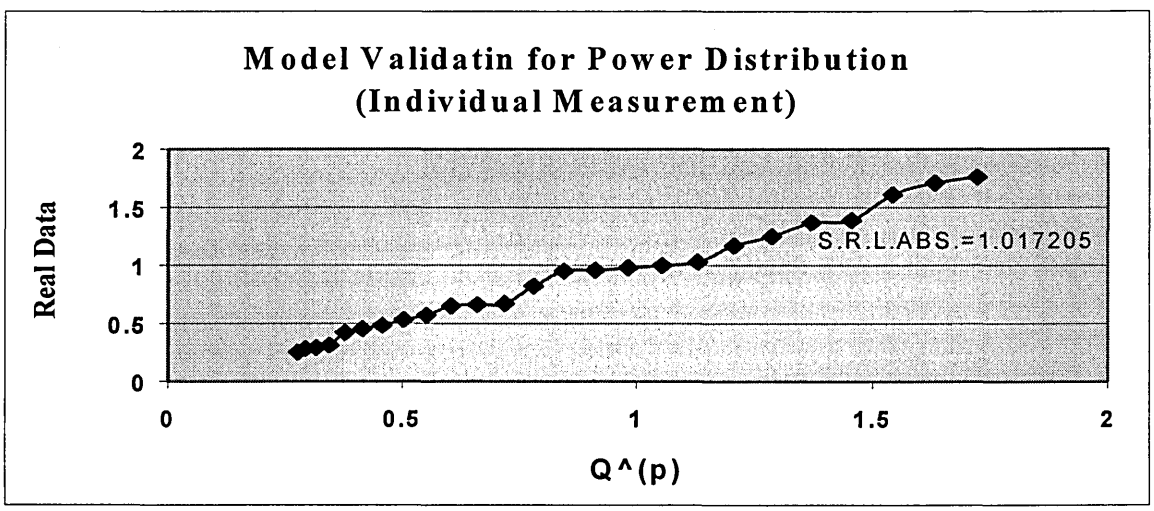


Figure 11: Model Validation using Median Rankit for Power Distribution

It is clear from the control chart in figure 12 that the sample numbers 3,10,15 are outside the warning limits respectively. So the light bulbs failure times is not fully under control at p=0.05. These points must be investigated to see whether an assignable cause can be determined. Moreover, the chart shows that 'sample point 25' is outside the action limits, which indicate that the production process is out of control at p=0.01.

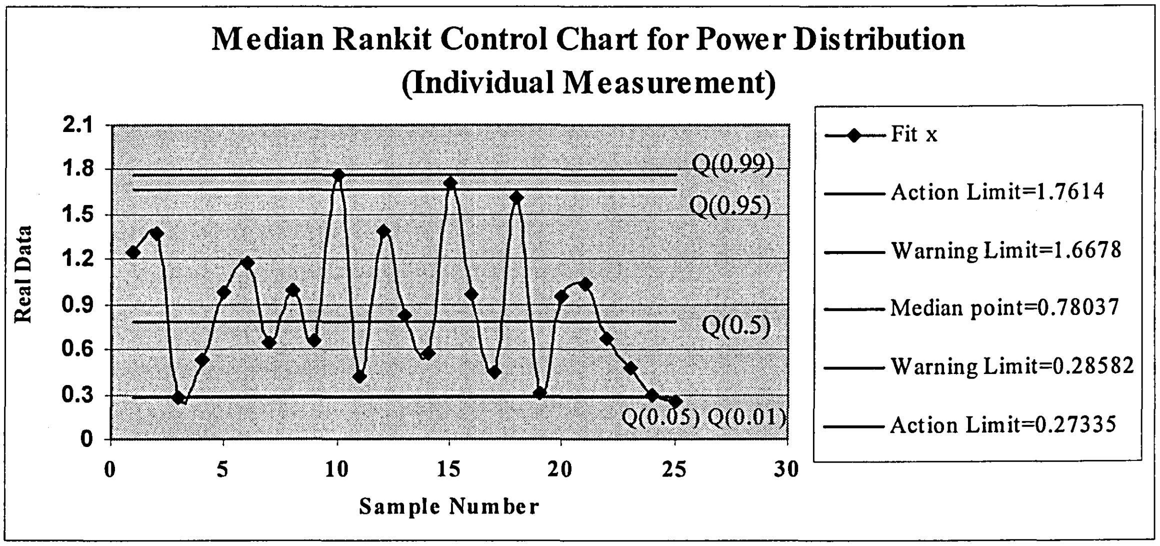


Figure 12: Median rankit control chart for power distribution

Comparison with EWMA Control Chart

In order to compare our methodology with similar work in this area, we will consider the control chart for exponentially weighted moving average (EWMA), using the data given in table 5. Here we will consider the Shewart control chart for individual measurement as a special case of EWMA when $\lambda = 1$ and when $\lambda = 0.2$, which is much popularly used to detect small shifts. Moreover, this value gives the lowest value of residual of least square, (see John 1990). A comparison between figure 13 (EWMA at $\lambda = 1$) and figure 14 (EWMA at $\lambda = 0.2$), shows that the failure times were not significant. On the other hand, Median rankit control chart for power distribution (figure 12) indicates that some points are out of control i.e. significantly different from the median rankit point (0.78037).

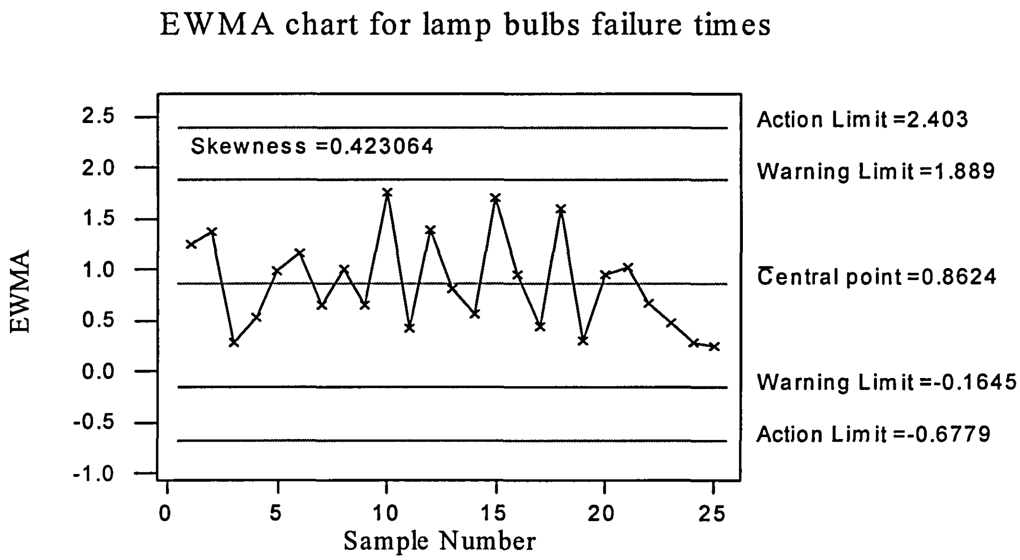


Figure 13: EWMA Control Chart at Lambda =1

Recently, Borror, Montogomery and Runger (1999) suggested that the EWMA control chart is more suitable for dealing with normal and non-normal data, and EWMA is more robust to the normality. However, our analysis in this chapter shows that the quantile control chart for power distribution is more sensitive than the EWMA

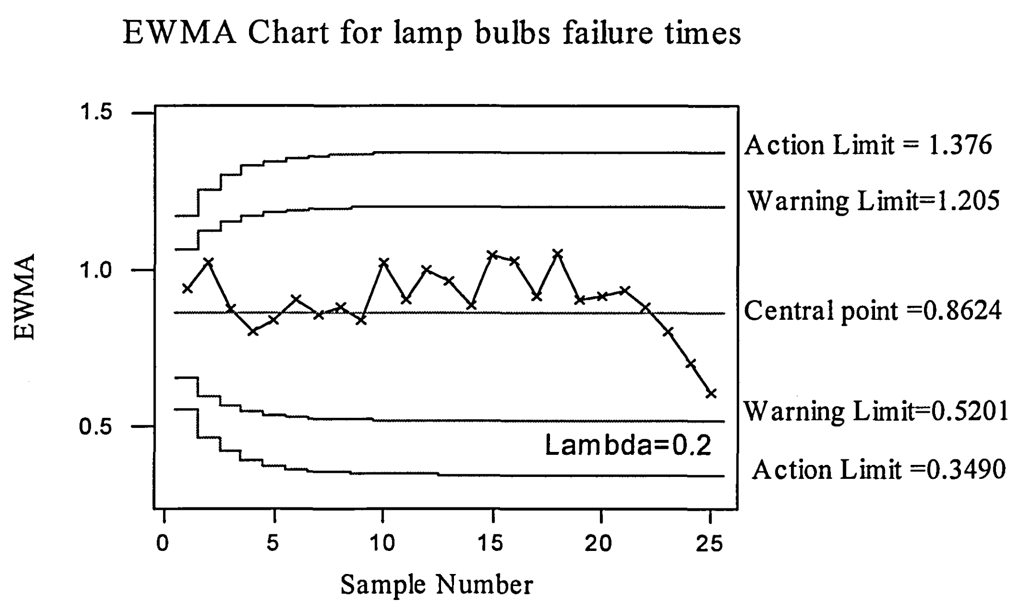


Figure 14: EWMA Control Chart at Lambda =0.2

It can therefore be concluded from figure 12, that the control chart using a power distribution is more appropriate than the weibull control chart (figure 9) and EWMA control chart at $\lambda = 1$ and $\lambda = 0.2$.

5.8 Control chart for non-normal distribution using subgroups of size five

In the previous sections, we present the control chart for non-normal distribution for median rankit using individual measurement. In this section we will apply the control chart for non-normal distribution for median rankit using subgroup five, such as logistic and weibull distributions.

Logistic Distribution

From table 1, we wish to establish statistical control of thickness of an oil seal, using median rankit. Thirty samples each of size five observations, have been generated (from table 1) when we assume the process is in control. The thickness of oil seals shown in table 7. Using the data in table 7, it found that the median point is 2.0143, which is a robust estimate and does not depends on 'n' values.

Table 7: The thickness of oil seals.

Sample Number	Observation 1	Observation 2	Observation 3	Observation 4	Observation 5	Median
1	1.9	1.9	1.8	1.9	1.9	1.9
2	2.0	1.8	1.9	1.9	2.0	1.9
3	2.2	2.1	2.0	2.3	1.8	2.1
4	2.0	2.4	2.1	2.5	2.0	2.1
5	2.0	2.0	2.4	2.0	2.1	2.0
6	1.9	2.3	1.6	1.7	1.9	1.9
7	1.7	2.3	1.6	1.7	1.8	1.7
8	2.0	2.3	1.8	1.8	1.6	1.8
9	2.0	2.3	1.9	2.2	1.6	2.0
10	2.0	2.2	2.0	2.2	2.2	2.2
11	2.1	1.8	2.1	2.3	2.1	2.1
12	2.1	1.6	1.8	2.0	2.1	2.0

Sample Number	Observation 1	Observation 2	Observation 3	Observation 4	Observation 5	Median
13	2.4	1.8	1.8	2.0	2.0	2.0
14	2.2	1.8	1.9	1.9	2.0	1.9
15	2.2	2.1	2.2	1.8	1.9	2.1
16	2.4	2.2	2.0	2.3	2.0	2.2
17	2.2	2.2	2	2.1	1.9	2.1
18	1.8	2.4	1.8	2.2	2.0	2.0
19	2.0	1.8	2.0	1.9	2.0	2.0
20	2.1	1.6	1.6	1.8	2.4	1.8
21	1.9	1.8	2.1	1.8	2.4	1.9
22	2.1	2.0	1.7	2.1	1.9	2.0
23	1.9	2.2	2.3	2.1	1.9	2.1
24	2.4	2.3	1.7	2.0	2.4	2.3
25	1.8	2.3	2.4	2.4	1.9	2.3
26	2.4	2.2	1.9	1.8	1.8	1.9
27	1.9	2.0	1.9	1.6	1.8	1.9
28	1.8	1.8	2	2.2	2.2	2.0
29	2.2	1.8	2.1	2.3	1.8	2.1
30	1.7	2.0	2.0	2.0	1.8	2.0

To find out the control limits on the median point chart, for sample size five, it must find out the values of estimation parameters for location, scale and skewness. Then substitute the estimation parameters in the form

$$Q(p) = \hat{\lambda} + \frac{\hat{\eta}}{2}((1-\hat{\delta})\ln p - (1+\hat{\delta})\ln(1-p))$$

Where $\hat{\lambda} = 2.021332$, $\hat{\eta} = 0.160773$ and $\delta = -0.06306$

Then the control limits of logistic quantile distribution function, can be obtained by substituting p=0.05, p=0.95 and p=0.01, p=0.99, then will provide the warning and action limits respectively, (see figure 15).

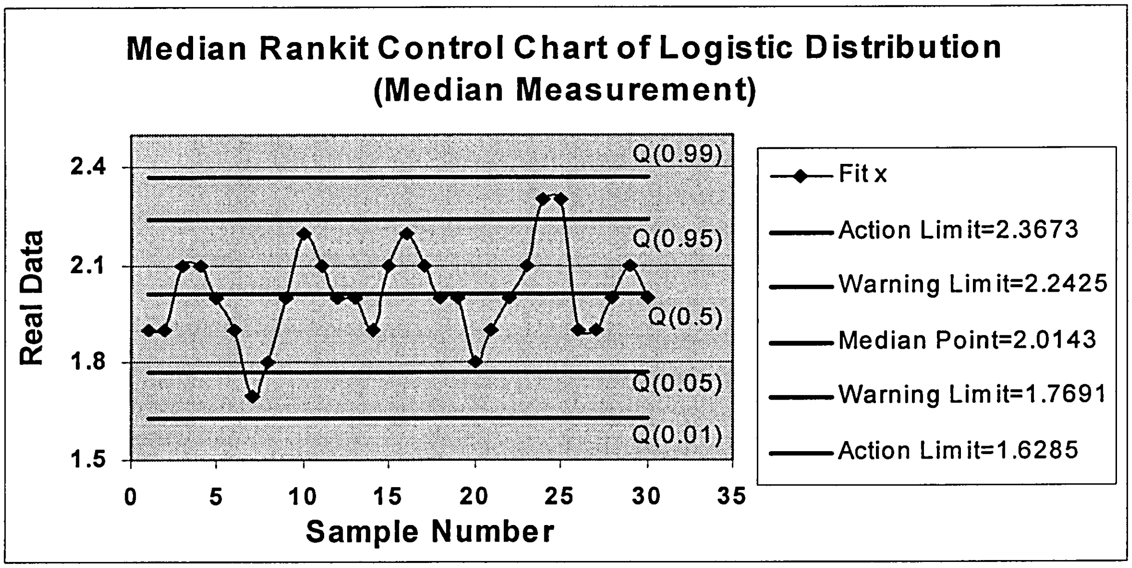


Figure 15: Median Rankit Control Chart for Logistic Distribution (Median Measurement)

When the preliminary sample median are plotted on this chart, there are indications of sample 7, 24, 25 of an out-of-control is observed from warning limits. Whereas, no indication of an out-of control condition is observed from action limits. Also there is no evidences against - the hypothesis that- the process is in control at the level p=0.01.

Weibull Distribution

From the data in table 5 which represents the time to failure measurement of 25 light bulbs on an accelerated test, twenty five samples, each sample of size five have been generated from the original data, table 5. It is assumed that the process is in control. The time to failure measurement are shown in table 8.

Table 8:Time to Failure for Light Bulbs

Sample Number	Observation 1	Observation 2	Observation 3	Observation 4	Observation 5	Median
1	0.66	0.29	1.76	1.17	0.57	0.66
2	0.29	0.57	1.71	0.57	0.96	0.57
3	0.65	1.37	0.45	1.76	0.66	0.66
4	0.95	1.37	1.61	0.65	0.67	0.95
5	1.17	1.39	0.29	0.57	1.76	1.17
6	1.71	0.65	0.28	0.65	1.25	0.65
7	1.61	0.57	0.29	0.28	0.48	0.48
8	0.45	1.03	0.96	0.42	0.29	0.45
9	0.65	0.29	1.71	1.39	0.53	0.65

Sample Number	Observation 1	Observation 2	Observation 3	Observation 4	Observation 5	Median
10	0.57	1	0.82	1.25	0.66	0.82
11	0.95	0.98	0.96	0.25	0.66	0.95
12	0.53	0.96	0.25	1.17	1.71	0.96
13	0.98	0.42	0.48	1.25	0.25	0.48
14	1.17	0.48	0.95	0.28	0.95	0.95
15	1.17	1.61	0.57	1.61	0.57	1.17
16	0.57	1.71	0.95	0.66	0.29	0.66
17	1.61	0.96	0.66	0.25	0.48	0.66
18	0.25	0.29	0.31	1.39	1.39	0.31
19	1.25	1.39	1.76	0.95	0.29	1.25
20	0.57	1.03	1.76	0.25	1.00	1.00
21	0.65	0.31	0.42	1.39	1.39	0.65
22	0.96	1.37	1.03	0.48	0.31	0.96
23	1.00	1.61	0.57	1.03	0.45	1.00
24	1.03	0.45	1.76	1.61	0.48	1.03
25	0.25	1.17	1.76	1.37	0.25	1.17

It is require to compute the weibull control limits on the median point chart for this process. In order to investigate whether 25-subgroups of size 5 process is in control, it needs to find out the values of estimation parameters for location, scale and shape. Then substitute the estimation parameters in the form

$$Q(p) = \lambda^{\wedge} + \eta^{\wedge} * (-\ln(1-p))^{\beta^{\wedge}}$$

Where $\lambda^{\wedge} = 0.002064$, $\eta^{\wedge} = 0.909169$, $\beta^{\wedge} = 0.293165$.

Then the control limits of weibull quantile distribution function can be obtained by substituting p=0.05, p=0.95 and p=0.01, p=0.99, then will provide the warning and action limits respectively, (see figure 16).

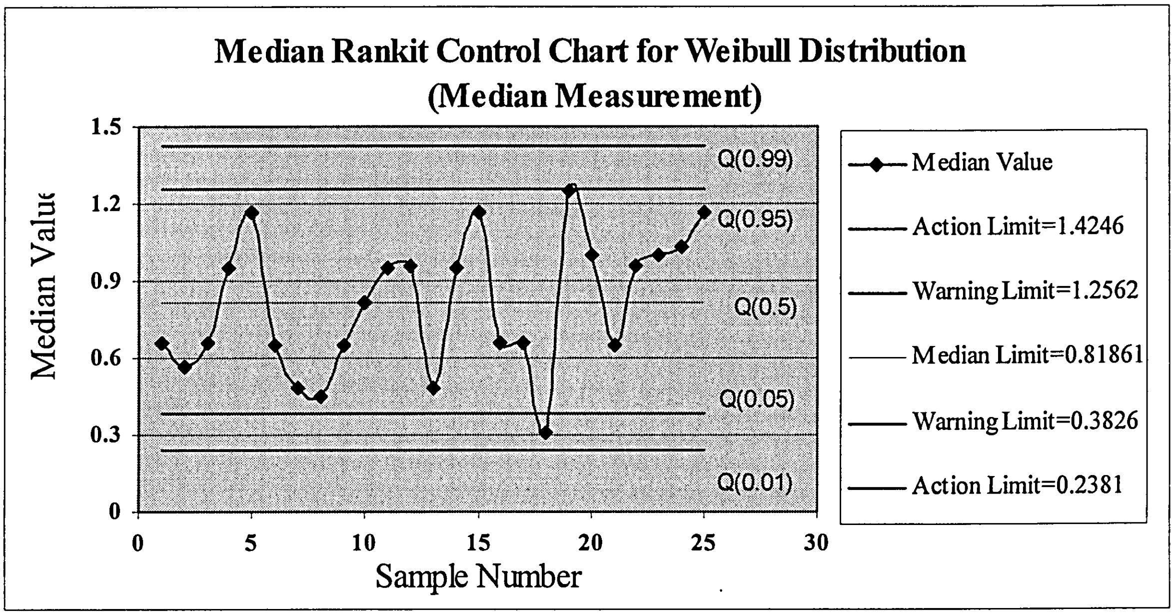


Figure 16: Median Rankit Control Chart for Weibull Distribution (Median Measurement)

A median rankit chart for these data is shown in figure 16. Note that the subgroups median of sample number 18 out-of-control is observed from warning limit. Whereas, the subgroups median process operating in control is observed from action limits.

5.9 *Summary*

In this chapter, we provided an applications of quantile control chart for non-normal situation, such as, logistic, exponential, extreme-value, Weibull and power distribution. Therefore, in the following chapter, we will be discussing the process capability indices using quantile approach for non-normal situation.

Chapter 6: Process Capability Indices using Quantile Approach

6.1 Introduction

Most of the literatures on process capability assume that data follow normal distribution. However, in application, most of the process data is non-normally distributed. Clement (1989) and Gunter (1989) were discussed the process capability for non-normal data and the limitation of C_{pk} with non-normal data respectively. Vannman (1995) suggested a general formula where the four basic indices, C_p, C_{pk}, C_{pm} and C_{pmk} as special case of (6.1). This general formula has been referred to as $C_p(u,v)$, which can be defined as follows:

$$C_p(u,v) = \frac{d - u\,|\,\mu - m\,|}{3\sqrt{\sigma^2 + v(\mu - T)^2}} \qquad 6.1$$

Where μ is the process mean, σ is the process standard deviation, $d = (USL - LSL)/2$ which is half of the length of the specification interval, USL is upper specification limit, and LSL is the lower specification limit, and $m = (USL + LSL)/2$ is the mid point between the two limits, T is the target value, and $u, v \geq 0$. It is easy to verify that the $C_p(0,0) = C_p$, $C_p(1,0) = C_{pk}$, $C_p(0,1) = C_{pm}$ and $C_p(1,1) = C_{pmk}$ as follows:

$$C_p = \frac{USL - LSL}{6\sigma} \qquad 6.2$$

$$C_{pk} = \min\left(C_{pu}, C_{pl}\right) = (1-k)C_p \qquad 6.3$$

where

$$C_{pu} = \frac{USL-\mu}{3\sigma} = \frac{USL-T}{3\sigma}\left(1-\frac{|T-\mu|}{USL-T}\right) \quad ,$$

$$C_{pl} = \frac{\mu-LSL}{3\sigma} = \frac{T-LSL}{3\sigma}\left(1-\frac{|T-\mu|}{T-LSL}\right)$$

$$k = \frac{2\,|\,T-\mu\,|}{USL-LSL} \quad , \qquad 0 \le k \le 1$$

has been suggested for symmetric tolerance i.e. $T = \mu$. If the process is on-target then k=0 ($T = \mu$).

$$C_{pm} = \frac{USL-LSL}{6\sigma'} = \frac{USL-LSL}{6\sqrt{E[(X-T)]^2}} = \frac{USL-LSL}{6\sqrt{\sigma^2+(\mu-T)^2}} \qquad 6.4$$

$$C_{pm} = \frac{C_p}{\sqrt{1+\left(\frac{(\mu-T)}{\sigma}\right)^2}} = \frac{C_{pk}}{\left(1-\frac{|\,\mu-M\,|}{d}\right)\sqrt{1+\frac{(\mu-T)^2}{\sigma^2}}}$$

$$C_{pmk} = \frac{\min(USL-\mu, \mu-LSL)}{3\sqrt{\sigma^2+(\mu-T)^2}} = \frac{C_{pk}}{\sqrt{1+\left(\frac{\mu-T}{\sigma}\right)^2}} \qquad 6.5$$

$$= \left(1-\frac{|\mu-M|}{d}\right)C_{pm} = \frac{d-|\mu-M|}{3\sqrt{\sigma^2+(\mu-T)^2}}$$

Vannman's method (1997) is applied to handling cases with asymmetric tolerances. Vannman's method modified the basic indices by adding a new term $|\mu - T|$ in the numerator of the definitions.

Here the estimates the $C_p(u,v)$ obtained by replacing μ by the sample mean $\bar{x}$ and σ^2 by the sample variance S^2, for normal distribution, both estimators are stable. However, for non-normal situations, these estimators are highly unstable.

Pearn *et al*. (1998) investigated Vannman's method and pointed out that this method is not appropriate for processes with asymmetric tolerances. Pearn and Chen (1997), found that the $C_p(u,v)$ are appropriate indices for processes with normal distributions and inappropriate for non-normal distributions. Pearn and Chen (1998) applied a new method and obtained a generalisation of C_{pk} for asymmetric tolerances. The method takes into account the asymmetry of the corresponding loss function, which is shown to be superior to the other existing methods. Pearn *et al.* (1999) suggested a generalisation of Clement's method for non-normal Pearsonian process with asymmetric tolerances.

6.2 Process capability indices for non-normal distribution

The use of the most common process capability indices assumes normal distribution, despite the fact that, process capability indices often are non-normally distributed (i.e. non-Gaussian) in practice. Here, there are some situations where non-normal process distributions are expected: Skew distributions, Heavy-tailed distributions and Short-tailed distributions, Gunter (1989). It is common to see the data of the process capability is non-normal i.e. more or less skew distributed. Most of the contributions made are assumed that the process is normally distributed, Kane (1986), Bissell (1990), Chou *et al.* (1990), Rodriguez (1992), Chan *et al* (1988), Spiring (1991).

Franklin and Wasserman (1992b) deal with bootstrap confidence limits for C_p, C_{pk}, and C_{pm}, which avoid the assumption of normality. There have been various attempts to extend the definition of standard capability indices to non-normal distribution, e.g. Gilchrist (1995,1993), Clement (1989) and Gunter (1989). Clement proposed a method for calculating estimators of C_p, C_{pk} indices. He assumed that this technique based on the percentage points of the Pearson curves to convert directly normal capability indices to compensate for the non-normality, (for more details see Kotz and Lovelace (1998, pp.145-156)). Clement has given tables for constructing process capability indices, based on the Pearson system of curves. He does not investigate the problem theoretically, or the distributions of estimators of process capability indices under other more general circumstance. Pearn and Kotz (1994) applied Clement's method to obtain estimators for the C_{pm}, C_{pmk} indices. The four indices can be written in the general form, (Vannman, 1995), when a centre target does fall on the midpoint of the specification interval i.e. $T = \dfrac{USL + LSL}{2}$

$$C_p(u,v) = (1-u)\frac{USL-LSL}{6\sqrt{\{[(U_p-L_p)/6]^2 + v(M-T)^2\}}}$$

$$+u^*\min\left\{\frac{USL-M}{3\sqrt{\{[(U_p-M)/3]^2 + v(M-T)^2\}}}, \frac{M-LSL}{3\sqrt{\{[(M-L_p)/3]^2 + v(M-T)^2\}}}\right\} \qquad 6.6$$

The above case is quite common in application.

It is easy to verify that $C_p(0,0) = C_p$, $C_p(1,0) = C_{pk}$, $C_p(0,1) = C_{pm}$, $C_p(1,1) = C_{pmk}$, which can be defined as

$$\hat{C}_p = \left[\frac{USL - LSL}{Up - Lp}\right] \qquad 6.7$$

$$\hat{C}_{pk} = \min\left[\frac{USL - M}{Up - M}, \frac{M - LSL}{M - Lp}\right] \quad 6.8$$

$$\hat{C}_{pm} = \left[\frac{USL - LSL}{6\sqrt{\left(\frac{Up - Lp}{6}\right)^2 + (M - T)^2}}\right] \quad 6.9$$

$$\hat{C}_{pmk} = \min\left[\frac{USL - M}{3\sqrt{\left(\frac{Up - M}{3}\right)^2 + (M - T)^2}}, \frac{M - LSL}{3\sqrt{\left(\frac{M - Lp)}{3}\right)^2 + (M - T)^2}}\right] \quad 6.10$$

Where Up, Lp and M are 0.99865, 0.00135 and 0.5 respectively. Clement's estimators (C_p and C_{pk}) are obtained by replacing the 6σ by $Up - Lp$, Pearn and Kotz's estimators (C_{pm} and C_{pmk}) are obtained by replacing the two 3σ by $Up - M$ and $M - Lp$, for the right and left tails using equations (6.2-6.3) and (6.4-6.5) respectively. $d = \frac{USL - LSL}{2}$ represents half of the length of the specification limits, $m = \frac{USL + LSL}{2}$ is the mid point between the upper and the lower specification limits and T is the target value. U_p and L_p are tabulated values from Clement (1989). The process median is a more robust measure of central tendency than the process mean for skewed distributions with long tails. So that, the process mean here is replaced by the process median M.

Pearn and Chen (1995), applied the new modification to improve the accuracy of Clement's method, by replacing the σ by $(Up - Lp)/6$ for all cases regardless of right or left tail side, when the centre target does not fall on the midpoint of specification interval, i.e. $T \neq \frac{USL + LSL}{2}$. Vannman's superstructure provides the four indices in the general form, (see appendix 3). Vannman (1997) considered an alternative method to handle cases with asymmetric tolerance. The method modified the basic indices by adding a new term $|M - T|$ in the numerator of the definitions. Pearn *et al.* (1998) investigated Vannman's method and pointed out that it can severely understate or overstate process capability. Therefore, Vannman's method is not appropriate for processes with asymmetric tolerances.

Pearn, Chen and Lin (1999) discussed a generalisation of Clement's method for non-normal Pearsonian process (by using the Pearn and Chen method in 1998) where the manufacturing tolerances are asymmetric. They applied a generalisation of Clement's method for non-normal data, when the tolerances are asymmetric i.e. $T \neq m = \frac{USL + LSL}{2}$.

The tolerances symmetric case is a special case from this method, when $T = m$. The generalisation Clement's method is defined as:

$$\hat{C}_p = \left(\frac{2 * d}{U_p - L_p} \right) \qquad \textbf{6.11}$$

$$\hat{C}_{pk} = \min \left(\frac{USL - M}{(U_p - L_p)/2} * \frac{d^*}{d_u}, \frac{M - LSL}{(U_p - L_p)/2} * \frac{d^*}{d_l} \right) \qquad \textbf{6.12}$$

$$\hat{C}_{pm} = \left(\frac{2 * d^*}{6\sqrt{\left(\frac{U_p - L_p}{6} \right)^2 + a^2}} \right) \qquad \textbf{6.13}$$

$$\hat{C}_{pmk} = \min\left(\frac{USL - M}{3\sqrt{\left(\frac{U_p - L_p}{6}\right)^2 + a^2}} * \frac{d^*}{d_u} , \frac{M - LSL}{3\sqrt{\left(\frac{U_p - L_p}{6}\right)^2 + a^2}} * \frac{d^*}{d_l} \right) \qquad 6.14$$

where $d^* = \min(d_u, d_l)$, $d_u = USL - T$, $d_l = T - LSL$, $d = (USL - LSL)/2$ and $a = \max\left(d(M - T)/d_u \ , \ d(T\text{-}M)/d_l\right)$. If $T = m$ then the process capability indices above is reduced to the modified original Clement method, Pearn and Chen (1995), at $a = |M - T|$, (see equations 6.30-6.33 in appendix 4).

The simplest way for dealing with non-normal data is to transform the data to normal or at least closer to normality than the original data, (Box and Cox, 1964). By using a square root transformation, skewed distribution may become normal. If the data passes the test of normality, then the transformed data can be used to estimate capability indices (see, Somerville, Montgomery, 1996). Some practitioners do not like to deal with transform data, because it may create difficulties in translating the results to the original data. Therefore, we suggest a new approach to dealing with non-normal data, which is called Process Capability Indices, using Quantile Approach $C_{p(m.r)}$.

6.3 Clement's methods and its weakness

There have been various attempts to extend formula of capability indices to non-normal distribution. Because, in many applications, processes are not normally distributed. Process departure from normality may be difficult to detect. Gunter (1989) demonstrated the strong impact this has on the sampling distribution of the natural estimator of C_{pk}. Therefore, the natural estimators of four basic indices are inappropriate for non-normal processes.

Clement proposed a method for calculating the estimator C_p and C_{pk} of process capability indices for non-normal distribution, using the Pearson family of curves. Pearn and Kotz (1994) applies Clement's method to obtain estimators C_{pm} and C_{pmk} of the two advance basic process capability indices.

Clement's method is based on a set of available sample data for a well in control process, using estimates of the mean, standard deviation, skewness and kurtosis. Under the assumption that these four parameters determine the type of the Pearson distribution curve, Clement used the tables provided by Gruska *et al.* (1989) for a percentage of the family of Pearson curves, as a function of skewness and kurtosis. The estimators are defined as in equations (6.7)-(6.10), where U_p=0.99865 and L_p=0.00135 percentile determine from the Gruska tables for the four parameters above. For the indices C_p and C_{pm}, Clement's estimators are obtained by replacing the 6σ in (6.2) and (6.4) by $U_p - L_p$. Whereas, the indices for C_{pk} and C_{pmk}, Clement's estimators are obtained by replacing the 3σ in (6.3) and (6.5) by $U_p - M$ and $M - L_p$ respectively for the two sides. The process mean μ is replaced by the process median M, where the median is more robust than the process mean, especially for skewed distributions with long tails.

The application of Clement's method is restricted to the process with symmetric tolerances. Pearn *et al.* (1999) introduces a generalisation of the Clement method to handling cases with asymmetric tolerances.

Clement's technique which is based on the Pearson family of distributions, offers formulas which is in turn based on the percentage points of the Pearson curves to convert directly normal capability. The disadvantage of this technique is the possibility of choosing a distribution that does not fit; the generic family chosen may not offer the best fit possible.

Clement's method depends on the Gruska tables. To determine the U_p, L_p and M, it needs to know the parameters value i.e. mean standard deviation, skewness and kurtosis, and use these values to determine the value of percentile from tables in Clement's (1989). These tables consist of a row which represent the skewness value and a column which represents the kurtosis. The row takes values from 0 to 2 increment 0.1 and the column takes value from -1.4 to 12.2 increment 0.2. If the skewness value is 0.3 and kurtosis is 2, then the value of $U_{0.99865}, L_{0.00135}\ and\ M_{0..5}$ are the exact result we obtained. The disadvantage which I faced here is, if the skewness is 0.33 and the kurtosis is 2.08, then the tables do not give an accurate result. In addition, Clement's approach requires estimates of skewness and kurtosis, that are based on the third and fourth moments, which may be somewhat unreliable for small sample size, (Chang and Lu, 1994).

6.4 Quantile Approach for Non Normal Capability Indices

There are so many methods used with non-normal data, and therefore one needs to distinguish the best method and the advantages and disadvantages of each. Some of these methods, can be used to fit the data sample with a generic family of distributions, then use the percentage points of the fitted distribution to compute equivalent value of C_p and C_{pk}. It has been extracted technique based on the Pearson family of distribution provide formulas based on the percentage points of the indices to compensate for the non-normality, (Clement, 1989). Pearn *et al* (1998) established new techniques for dealing with non-normal data; other techniques followed.

Here, we are going to suggest a new technique for working with non-normal data through quantile approach. This technique is called Quantile Capability Index (QCI), and Capability indices denoted by $C_{p(m.r)}$, $C_{pk(m.r)}$, $C_{pm(m.r)}$ and $C_{pmk(m.r)}$. QCI helps to measure the capability of the process when the process is in control. This is a different method for calculating Q(0.99865), Q(0.00135) and the median Q(0.5), where the percentile 0.99865 and 0.00135 represents the quantity of 3σ. The idea is to generalise the formula for standard indices by replacing 3σ with the percentile above. This method is derived by using the generalisation lambda distribution, is based on percentiles and does not need the tables.

Tukey (1960) was used to derive the Quantile Distribution Function (QDF). If we consider a distribution with parameters λ, η, θ where θ represent one or more parameters then Q(p), is

$$Q(p) = \lambda + \eta R(p;\theta) \qquad 6.15$$

can be defined as a quantile distribution. A standard quantile distribution can be expressed as

$$z_p = \frac{x_p - \lambda}{\eta} = R(p,\theta)$$

where λ and η as location and scale parameters and $R(p, \theta)$ depends on the parameters (e.g. skewness, shape).

Further, a quantile distribution, which requires only two parameters (i.e. location and scale parameters), can be expressed as

$$Q(p) = \lambda + \eta R(p) \qquad 6.16$$

where $R(p)$ does not depend on the parameter (θ). Distributions such as Exponential, Extreme value and Uniform, fall under this category. For more details about quantile approach, see Kanji & Arif (2000).

For the indices C_p and C_{pm}, QCI estimators are obtained by replacing the 6σ by Q(1-p)-Q(p). For the indices C_{pk} and C_{pmk}, QCI estimators are obtained by replacing 3σ by Q(1-p)-Q(0.5) and Q(0.5)-Q(p) respectively for the left and right side tail. The process mean is replaced by the process median Q(0.5).

Case 1: The central target does fall on the midpoint of the specification interval i.e. $T = \frac{USL + LSl}{2}$

$$C_{p(m.r)} = \left[\frac{USL - LSL}{Q(0.99865) - Q(0.00135)} \right] \qquad 6.17$$

$$C_{pk(m.r)} = \left[\frac{\min(USL - Q(0.5), Q(0.5) - LSL)}{(Q(0.99865) - Q(0.00135))/2} \right] \qquad 6.18$$

$$C_{pm(m.r)} = \left[\frac{USL - LSL}{6\sqrt{\left(\frac{Q(0.99865) - Q(0.00135)}{6}\right)^2 + (Q(0.5) - T)^2}} \right] \qquad 6.19$$

$$C_{pmk(m.r)} = \left[\frac{\min(USL - Q(0.5), Q(0.5) - LSL)}{3\sqrt{\left(\frac{Q(0.99865) - Q(0.00135)}{6}\right)^2 + (Q(0.5) - T)^2}} \right] \qquad 6.20$$

Case 2: The central target does not fall on the midpoint of the specification interval

i.e. $T \neq \frac{USL + LSL}{2}$

$$\hat{C}_{p(m.r)} = \left(\frac{2 * d}{Q(0.99865) - Q(0.00135)} \right) \qquad 6.21$$

$$\hat{C}_{pk(m.r)} = min\left(\frac{USL - Q(0.5)}{(Q(0.99865) - Q(0.00135))/2} * \frac{d^*}{d_u}, \frac{Q(0.5) - LSL}{(Q(0.99865) - Q(0.00135))/2} * \frac{d^*}{d_l} \right)$$

6.22

$$\hat{C}_{pm(m.r)} = \left(\frac{2 * d^*}{6\sqrt{\left(\frac{Q(0.99865) - Q(0.00135)}{6}\right)^2 + a^2}} \right) \quad 6.23$$

$$\hat{C}_{pmk(m.r)} = \min \left(\frac{USL - Q(0.5)}{3\sqrt{\left(\frac{Q(0.99865) - Q(0.00135)}{6}\right)^2 + a^2}} * \frac{d^*}{d_u} , \frac{Q(0.5) - LSL}{3\sqrt{\left(\frac{Q(0.99865) - Q(0.00135)}{6}\right)^2 + a^2}} * \frac{d^*}{d_l} \right) \quad 6.24$$

where $d^* = \min(d_u, d_l)$, $d_u = USL - T$, $d_l = T - LSL$, $d = (USL - LSL)/2$ and $a = \max\left(d(M-T)/d_u \ , \ d(T\text{-}M)/d_l\right)$. Q(0.99865), Q(0.00135 and Q(0.5) are represented by using quantile approach explained above. If $T = m$ then the process capability indices above are reduced to the original Clement method using, quantile approach at $a = |M - T|$, (see equations 6.17-6.20 above).

6.5 Methodology

The aim here is to measure the process capability indices. First of all, we investigate whether the data is normal or not. If the data follows normal data i.e. the mid-point equals the process mean, then apply the 4 basic capability indices and take the decision whether the data is capable or not. If the data does not follow normal distribution i.e. the mid-point is not equal to the process mean, then there are two methods which can be used to calculate the capability indices.

The first method is Generalisation Clement's method through Pearsonian process, and the other method is Generalisation Clement's method through Quantile Approach. "The second method was proposed by the author". The next step is to calculate the capability indices, and take the decision whether the data is capable or not, (see figure 6.1).

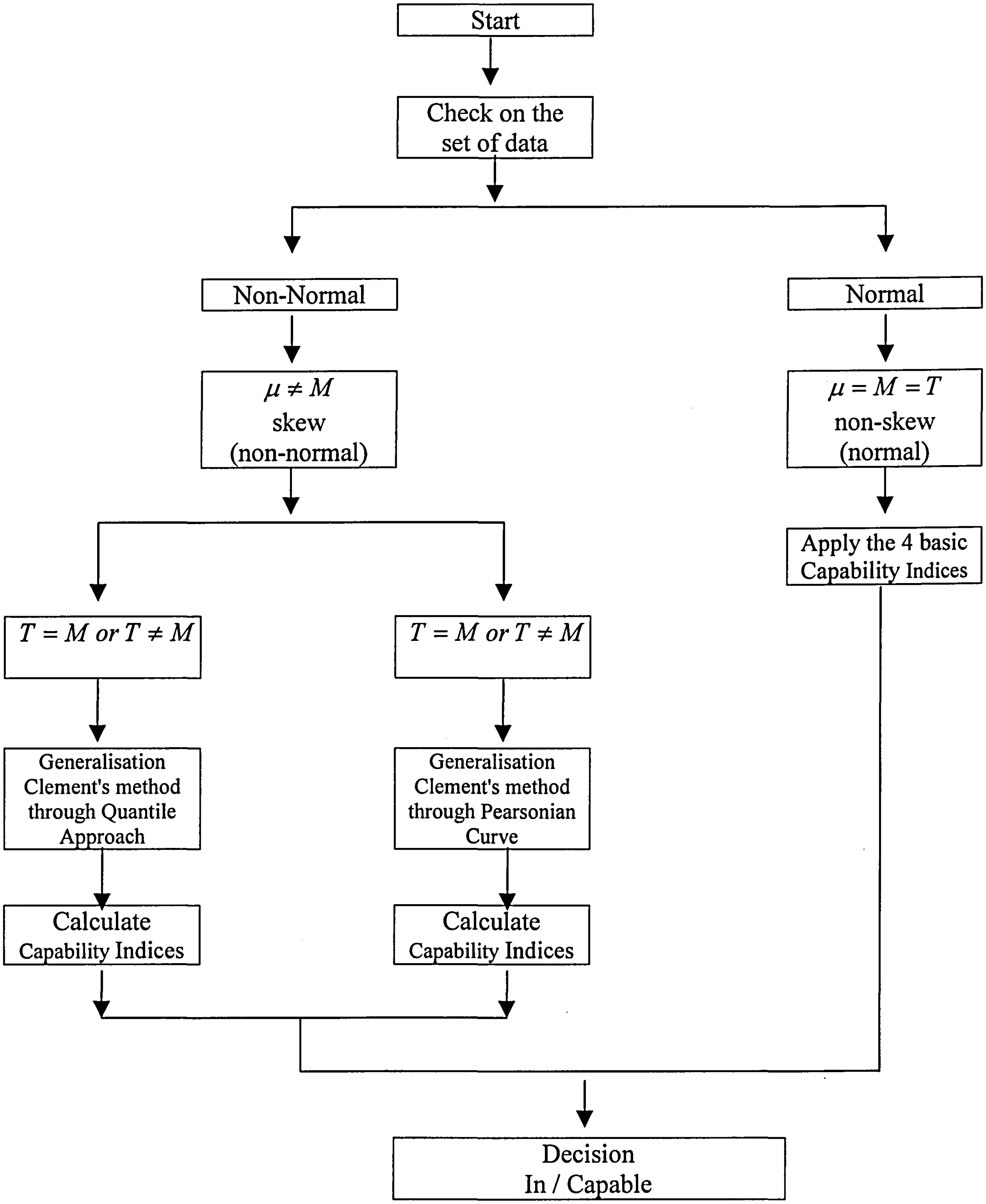


Figure 6. 1 Flow-Chart of Process Capability Indices.

Here, we are going to estimate the capability indices, by using the Generalisation Clement's method through Pearsonian process (GCMPP) e.g. C_p and calculate the estimators of capability indices by using the Generalisation Clement's method through Quantile approach (GCMQA) e.g. $C_{p(m.r)}$.

- To calculate the values of the estimators $\hat{C}_p, \hat{C}_{pk}, \hat{C}_{pm}, \hat{C}_{pmk}$ using (GCMPP) as follows

We first proceed with calculating the following and check tables in Gruska *et al.* (1989) to find out U_P, U_L and the sample median, for more details, (see Clement, 1989)

Table 6. 1 Describe data for GCMPP

Mean	St. Deviation	Skewness	Kurtosis	Median	U_p	L_p
2.0215	0.2190	0.0568	0.047	2.00	2.6025	1.44049

The upper specification USL=3.2, the lower specification=1, the target value T is 2.1, then substitute the results above in equations (6.11-6.14) for $T \neq \frac{USL+LSL}{2}$, and for $T = \frac{USL+LSL}{2}$, substitute the process capability indices in equations (6.11-6.14) will reduce to the modified Clement's method, (see equations 6.30-6.33, appendix 4).

- The values of the estimators $\hat{C}_{p(m.r)}$, $\hat{C}_{pk(m.r)}, \hat{C}_{pm(m.r)}, \hat{C}_{pmk(m.r)}$ are calculated using (GCMQA) as follows

We first proceed with calculating the following and check Quantile Approach to find out Q(0.99865), Q(0.00135), Q(0.5), for more details see chapter 5

Table 6. 2 Describe data for GCMQA

Mean	St. Deviation	Skewness	Kurtosis	Q(0.5)	Q(0.99865)	Q(0.00135)
2.0215	0.2190	0.0568	0.047	2.0184	2.885477	1.20757

Then substitute the results above in equations (6.21-6.24) for $T \neq \frac{USL+LSL}{2}$, and for $T = \frac{USL+LSL}{2}$, substitute the process capability indices in equations (6.21-6.24) will be reduced to the modified Clement's method through quantile approach, (see equations, 6.17-6.20).

Using MINITAB released 12.1, provided two programs to calculate the process capability indices for two method GCMPP and GCMQA. By using these programs it makes the process indices capability values much easier . The program required six observations, which are USL, LSL, T, Mean or Q(0.5), U_p or Q(0.99865) and U_l or Q(0.00135), then gives all capability indices C_p, C_{pk}, C_{pm} and C_{pmk} respectively, (see appendix 5 and appendix 6).

Example:

Logistic Data

Process Capability Indices	Generalisation Clement's Method through Pearsonian Curve		Generalisation Clement's Method through Quantile approach	
	$T = \frac{USL+LSL}{2}$	$T \neq \frac{USL+LSL}{2}$	$T = \frac{USL+LSL}{2}$	$T \neq \frac{USL+LSL}{2}$
C_p	1.89327	1.89327	1.31116	1.31116
C_{pk}	1.75816	1.60528	1.21389	1.10834
C_{pm}	1.75461	1.52585	1.25867	1.14127
C_{pmk}	1.62940	1.35536	1.16530	1.01067

Where USL=3.2 , LSL=1,

$T = \frac{USL+LSL}{2}$=2.1 and $T \neq \frac{USL+LSL}{2}$=2.15

Up=2.6025 , Lp=0.144049 , M=2.0215

Q(0.99865)=2.885477, Q(0.00135)=1.20757, Q(0.5)=2.0184.

We note that the all four index values are greater than one in both cases and methods which is quite good. Hence, conclude that the process is capable. From the table above, we can also conclude that the indices values using quantile approach are less than the indices values using Pearsonian curve, because the quantile approach gives more precise results than the Pearsonian curve. Capability process indices depends on factors such as, how far the target value from the medium value, the range of specification limits and the spread or variation of the process. A change in one of the these factors using the same data above, would yield other indices values.

6.6 Summary

Clement (1989) proposed the use of Pearsonian processes method for non-normal data. He proposed the first two estimators of the process capability. Pearn and Kotz (1994) extended the application of Clement's method to other two estimators of the process capability. The disadvantage of this method is covering symmetric tolerances i.e. $\mu = M$. Pearn *et al.* (1999) proposed the Generalisation Clement's Method using Pearsonian Process (GCMPP). This method covers asymmetric tolerances, and the symmetric tolerances is a special case when $T = \mu$. We considered the Generalisation Clement's Method using Quantile Approach (GCMQA). GCMQA which is more accurate than GCMPP, because it gives accurate percentile results and does not depend on statistical tables. In order to measuring the performance of control chart for non-normal distribution, average run length (ARL) is required. Therefore, ARL for non-normal situation will be discuss in the next chapter.

Chapter 7: Determination of Average Run Length (ARL) for Non-Normal Data

7.1 Introduction

It is assumed that using the Shewhart individual control chart is carried out under the assumption that the measures of quality characteristics follow normal distribution. When these measures actually come from non-normal distribution, the result will be a considerable deterioration of the in-control ARL performance of the chart.

The performance of a control chart for monitoring a process can be measured by the run length distribution and its mean, i.e. the average run length (ARL). When there is a significant change in the process, it is desirable to have a small ARL, so that the change could be detected quickly. However, if the process is in control, then it is preferable to have a large ARL corresponding to a low false alarm rate. A large ARL is desired when the mean has not shifted (or the shift is within an acceptable limit), whereas a small ARL is preferable when the size of a shift is unacceptable.

Rational subgroups of size one are frequently encountered in process monitoring and control. The Shewhart control chart for individuals is often used in this situation. It is known that the in-control average run length of this chart at 3σ is 370.4, under the assumption that the observations are selected at random from a normal population. When the assumption normality is violated, the ARL of the individuals control chart is adversely affected (Borror *et al.* 1999) Therefore, the ARL is constructed using quantile distribution for non-normal populations.

Crowder (1987) evaluated the joint performance of the x chart and the MR chart by using the computation of their joint ARL. He noted that the usefulness of this conventional procedure. Lucas and Saccucci (1990) and Roberts (1966) conclude that the EWMA and CUSUM charts are known to have very similar performances in monitoring a normal mean. Del Castillo and Montgomery (1994) investigated the ARL performance of Q-chart for variables, and found out that in some cases it is inadequate. They suggest some modifications to the Q-Chart procedure and alternate methods based on the EWMA.

Ng and Case (1992) provide a formula for the ARL of $\bar{x}$ charts with unknown parameters. Burroughs *et al.* (1993) uses the formula given by Ghosh *et al.* (1981) to find ARL of Shewhart control charts with run rules and unknown process parameters. Quesenberry (1991) proposed the Q-chart technique for process with unknown parameters. Albin *et al. (1997)* estimated the ARL to false alarms and to detection of shifts in the process mean and standard deviation. Amin *et al.* (2000) proposed a EWMA control chart based on the smallest and largest observations in each sample, which called is MaxMin EWMA control chart. The latter has good ARL properties for simultaneous changes in the mean and standard deviation.

The aim of this chapter is to derive the Average Run Length (ARL) for non-normal distributions, which will be concerned in the light of quantile development in Ch.4, with the distributions of Exponential, Extreme-value, Logistic, Weibull, Pareto and Power. In order to calculate the ARL "i.e. the average number of points that must be plotted before a point indicates an out of control", when k is a fixed number, K>1, and the probabilities are varied, i.e. p has been generated from U(0.001,0.003). In addition, ARL is calculated, when p is fixed, i.e. p=0.00135 and k is variable. In the next section, the theory underpinning ARL for such distributions will be discussed.

7.2 ARL for non-normal distribution

Quantile distribution is defined as

$$Q(p) = \lambda + \eta * R(p)$$

where λ *and* η represents a location and scale parameters

let

$$x = Q(u) = \lambda + \eta * R(u)$$

So

$$u = F\left(\frac{x-\lambda}{\eta}\right)$$

Here, we consider a shift of scale from η to $k\eta$ in process, where k>1.

The probability limit for quantile distribution of a signal at a given time period, when the process is in control is

$$p = prob(x \succ x_u \text{ or } x \prec x_l)$$

$$= 1 - F(x_u) + F(x_l) = 1 - F\left(\frac{\lambda + \eta * R(p_{(U)}) - \lambda}{k\eta}\right) + F\left(\frac{\lambda + \eta * R(p_{(L)}) - \lambda}{k\eta}\right)$$

$$p = 1 - F\left(\frac{R(p_{(U)})}{k}\right) + F\left(\frac{R(p_{(L)})}{k}\right)$$

As we know in normal situation, Montgomery (1997)

$$ARL = \frac{1}{p}$$

then, ARL for non-normal distribution using quantile approach can be found as follows

$$ARL = \frac{1}{\left(1 - F\left(\frac{R(p_{(U)})}{k}\right) + F\left(\frac{R(p_{(L)})}{k}\right)\right)}$$

In the next section, we prove such probabilities, which are $F\left(\frac{R(p_{(U)})}{k}\right)$ and $F\left(\frac{R(p_{(L)})}{k}\right)$ in ARL for some non-normal distribution, such as exponential, extreme, power, Weibull ,pareto and logistic distribution.

7.3 ARL for Exponential Distribution

Quantile function for exponential distribution is

$$R(p) = -\ln(1-p)$$

By using this formula, the probability of $p_{(L)}$, i.e. lower limit, is calculated.

$$R(p_{(L)}) = -\ln(1 - p_{(L)})$$

$$\frac{1}{k}R(p_{(L)}) = -\frac{1}{k}\ln(1 - p_{(L)})$$

$$= -\ln(1 - p_{(L)})^{\frac{1}{k}}$$

Let $\quad 1 - p^*_{(L)} = (1 - p_{(L)})^{\frac{1}{k}}$

then

$$= -\ln(1 - p^*_{(L)})$$

$$F\left(\frac{1}{k}R(p_{(L)})\right) = F\left(-\ln(1 - p^*_{(L)})\right)$$

$$= F(R(p^*_{(L)}))$$

$$= p^*_{(L)}$$

Where $p^*_{(L)} = 1 - (1 - p_{(L)})^{\frac{1}{k}}$, i.e. $p^*_{(L)}$ represents the lower limit, when k>1.

On the other hand, the probability of $p_{(U)}$, i.e. upper limit is calculated

$$R(p_{(U)}) = -\ln(1 - p_{(U)})$$

$$\frac{1}{k}R(p_{(U)}) = -\frac{1}{k}\ln(1 - p_{(U)})$$

$$= -\ln(1 - p_{(U)})^{\frac{1}{k}}$$

Let $\quad 1 - p^*_{(U)} = (1 - p_{(U)})^{\frac{1}{k}}$

then

$$= -\ln(1 - p^*_{(U)})$$

$$F\left(\frac{1}{k}R(p_{(U)})\right) = F\left(-\ln(1 - p^*_{(U)})\right)$$

$$= F(R(p^*_{(U)}))$$

$$= p^*_{(U)}$$

Where $p^*_{(U)} = 1 - (1 - p_{(U)})^{\frac{1}{k}}$, i.e. $p^*_{(U)}$ represents the upper limit, when k>1.

at $p_u = 1 - p_L$

$$p^*_{(U)} = 1 - (1 - (1 - p_{(L)}))^{\frac{1}{k}}$$

then

$$ARL = \frac{1}{\left(1 - p^{*}_{(U)} + p^{*}_{(L)}\right)}$$

Table 7.1 provides the ARL's results for various distributions. For more details, proof etc. of the ARL, see appendix 7.

Table 7.1 ARL for non-normal distribution using quantile approach

Distribution	$p^*_{(L)}$	$p^*_{(U)}$	ARL
Exponential	$p^*_{(L)} = 1-(1-p_{(L)})^{\frac{1}{k}}$	$p^*_{(U)} = 1-(1-(1-p_{(L)}))^{\frac{1}{k}}$	$ARL = \frac{1}{\left(1-p^*_{(U)}+p^*_{(L)}\right)}$
Extreme-value	$p^*_{(L)} = \exp(-(-\ln p_{(L)})^{\frac{1}{k}})$	$p^*_{(U)} = \exp(-(-\ln(1-p_{(L)}))^{\frac{1}{k}})$	$ARL = \frac{1}{\left(1-p^*_{(U)}+p^*_{(L)}\right)}$
Weibull	$p^*_{(L)} = 1-\exp(-(\left(\frac{1}{k}\right)^{\frac{1}{\beta}}(-\ln(1-p_{(L)}))))$	$p^*_{(U)} = 1-\exp(-(\left(\frac{1}{k}\right)^{\frac{1}{\beta}}(-\ln p_{(L)})))$	$ARL = \frac{1}{\left(1-p^*_{(U)}+p^*_{(L)}\right)}$
Power	$p^*_{(L)} = \left(\frac{1}{k}\right)^{\frac{1}{\beta}} p_{(L)}$	$P^*_{(U)} = \left(\frac{1}{k}\right)^{\frac{1}{\beta}}(1-P_{(L)})$	$ARL = \frac{1}{\left(1-p^*_{(U)}+p^*_{(L)}\right)}$
Pareto	$p^*_{(L)} = 1-\left(\frac{1}{k}\right)^{\beta}(1-p_{(L)})$	$P^*_{(U)} = 1-\left(\frac{1}{k}\right)^{\beta} P_{(L)}$	$ARL = \frac{1}{\left(1-p^*_{(U)}+p^*_{(L)}\right)}$
Logistic	$p^* = \frac{\left(\frac{p}{1-p}\right)^{\frac{1}{k}}}{\left(1+\left(\frac{p}{1-p}\right)^{\frac{1}{k}}\right)}$		$ARL = \frac{1}{2p^*}$

Where $p^*_{(L)}$ *and* $p^*_{(U)}$ represents lower and upper limits respectively, and k represents the shift in the process

7.4 Application

The ARL for non-normal distribution in the two cases was calculated mainly using ARL theory. In **case 1**, ARL is calculated, when k is variable and p-value is fixed, where k is generated from U(1,2), k value was generated from U(1,2), because when k>2 , ARL value will be very low, (see tables 7.2a for exponential, extreme-value and logistic and see table 7.2b for Weibull, power and pareto distribution). In **case 2**, ARL was calculated when p is variable and k is fixed, and where p-value is generated from U(0.001,0.003), (see table 7.3a for exponential, extreme-value, logistic and table 7.3b for Weibull, power and pareto distribution).

Graphs of the ARL for n=1 for these distributions, are shown in figures (7.1-7.6) for **case 1** and figure (7.7-7.10) for **case 2**.

Note from the tables and graphs that the ARL, when there is no shift for scale (η) we get an average of one false signal in 370.37 plotted points, i.e. when k=1. The ARL for quantile control chart for such distributions at k=1, is the same for a 3-sigma $\bar{x}$-chart, which is 370.37, when the probability of a false signal is 0.00135, (see tables 7.2a-7.2b).

Once k is shifted from k=1 to k>1, the values of the ARL for n=1 decreases very quickly. For example, at p=0.00135, ARL value for extreme-value when k=1.1 is 159.032, while ARL value for extreme-value when k=1.2 is 82.696, see table 7.2a. The shift in the process for just 0.1 when k changes from 1.1 to 1.2 , cost about 76.336 samples. Another example, at p=0.00135, ARL value for weibull when k=1.4 is 151.507, while ARL value for weibull when k=1.5 is 123.131, see table 7.2b. The shift in the process for just 0.1, costs about 28.376 samples. From tables 7.2a and 7.2b, ARL value for such distributions are different from each other, at fixed k or at fixed p, see table 7.3a and 7.3b.

On the other hand, it can be observed from the tables 7.3a and 7.3b, that the ARL gives the different value when k is fixed and p is change. For example, for exponential distribution at k=1.6, the ARL value for probabilities 0.001, 0.00135, 0.0027, 0.005, 0.01 and 0.05 are 71.6315, 59.0655, 37.7402, 25.2581, 16.001 and 5.3962 respectively, see table 7.3a.

By using the ARL Formula in table 7.1, the information of ARL values in tables 7.2a, 7.2b, 7.3a and 7.3b are given. It can be observed that, when there is no shift, the scale will average of one false signal in 370 plotted points, i.e. we will get an average of one false signal in 370 subgroups.

Figure 7.1-7.6 for such Distribution display the relationship between ARL value and k, when p is fixed and k is generated from U(1,2). It was found that the relationship between ARL and k is that when k increases, ARL decreases, and vice versa.
Figure 7.7-7.10 for such distribution display the relationship between ARL value and p-value, when k is fixed and p is generated from (0.001,0.003). It was found that the relationship between ARL and p-value is that when ARL increases, p-value decreases and vice versa.

The values in tables 7.2a, 7.2b, 7.3a and 7.3b were obtained by creating a program to calculate ARL for each distribution, using Minitab 12.1. For example, we will discuss the ARL program for exponential distribution here, and for other distributions, see appendix 8.

ARL for Exponential

R(p)= -ln(1-p)

P*(u)=(1-((1-(1-pl))**(1/k)))

P*(l)=(1-((1-pl)**(1/k)))

ARL=1/(1-P*(u)+P*(l))

Case 1

#exponential

#p is fixed , k is variable

#k is generated from (1,2)

```
# c1=k, c2=p
let c3=(1-((1-c2)**(1/c1)))
let c4=(1-((1-(1-c2))**(1/c1)))

let c5=(1)/(1-c4+c3)
# c5=ARL
```

Case 2

```
# exponential
# k is fixed, p is variable
#Generate p from uniform (0.001,0.003)

#c1=p,c2=k
let c3=(1-((1-c1)**(1/c2)))
let c4=(1-((1-(1-c1))**(1/c2)))

let c5=(1)/(1-c4+c3)
#c5=ARL
```

Case 1 help us to calculate the ARL for exponential distribution, when k is variable and p is fixed. The procedures to run this program using Minitab 12.1 are as follows:

First

- Open Worksheet 1
- Generate n observation (n=60, say) from Uniform(1,2), put it in C1
- Put the p-value which you want to use (0.00135, say) in C2

Second

You need now to follow these procedures to write down the program

- Chose Edit
- Choose Command Editor

Now open window to write down the program,

- let c3=(1-((1-c2)**(1/c1)))

- let c4=(1-((1-(1-c2))**(1/c1)))
- let c5=(1)/(1-c4+c3)

then submit command.

You will find the results in Worksheet 1, where C1=k, C2=p, C3= p_L^*, $C4 = p_U^*$ and C5=ARL.

If you want to calculate ARL using another probability, you need to change the p-value in C2, and so on.

Case 2 helps us to calculate the ARL for exponential distribution when p is variable and k is fixed. The procedures to run this program using Minitab 12.1 are as follows:

First

- Open Worksheet 2
- Generate n observation (n=60, say) from Uniform(0.001,0.003), put it in C1
- Put the k value, which you want to use (0.00135, say) in C2

Second

You need now to follow these procedures to write down the program

- Choose Edit
- Choose Command Editor

Now it will open a window to write down the program,

- let c3=(1-((1-c1)**(1/c2)))
- let c4=(1-((1-(1-c1))**(1/c2)))

- let c5=(1)/(1-c4+c3)

then submit command.

You will find the results in Worksheet 2, where C1=p, C2=k, C3= p_L^*, $C4 = p_U^*$ and C5=ARL.

If you want to calculate ARL using another k value, you need to change the k value in C2, and so on.

Table 7.2a: ARL (p-fixed and k-variable)

k	Exponential		Extreme-value		Logistic	
	p		p		p	
	0.00135	0.0027	0.00135	0.0027	0.00135	0.0027
1	370.370	185.185	370.370	185.185	370.370	185.185
1.1	271.082	141.296	159.032	89.743	203.373	108.401
1.2	192.830	105.418	82.696	51.234	123.490	69.448
1.3	138.101	79.059	49.711	33.027	81.029	47.700
1.4	101.195	60.387	33.214	23.258	56.516	34.609
1.5	76.248	47.190	23.976	17.480	41.397	26.240
1.6	59.066	37.740	18.331	13.793	31.556	20.620
1.7	46.939	30.842	14.639	11.296	24.859	16.690
1.8	38.163	25.700	12.091	9.523	20.129	13.846
1.9	31.657	21.789	10.254	8.215	16.681	11.727
2	26.725	18.757	8.882	7.219	14.099	10.110
2.1	22.912	16.365	7.828	6.440	12.120	8.848
2.2	19.909	14.447	6.997	5.818	10.572	7.845
2.3	17.506	12.888	6.328	5.311	9.339	7.035
2.4	15.555	11.603	5.781	4.893	8.342	6.371
2.5	13.950	10.531	5.327	4.541	7.524	5.821
2.6	12.614	9.629	4.944	4.243	6.845	5.358
2.7	11.490	8.861	4.617	3.987	6.276	4.966
2.8	10.536	8.202	4.337	3.766	5.792	4.630
2.9	9.718	7.632	4.093	3.572	5.379	4.340
3	9.011	7.135	3.879	3.402	5.022	4.087
3.1	8.397	6.700	3.691	3.251	4.712	3.867
3.2	7.858	6.315	3.524	3.117	4.441	3.672
3.3	7.384	5.974	3.375	2.997	4.202	3.499
3.4	6.963	5.669	3.241	2.888	3.990	3.345
3.6	6.253	5.150	3.011	2.701	3.633	3.083
3.8	5.679	4.726	2.821	2.546	3.344	2.869
4	5.208	4.374	2.662	2.415	3.108	2.692

Table 7.2b: ARL (p-fixed and k-variable) for weibull, power and pareto $\beta = 1.4$

$\beta = 1.4$	Weibull		Power		Pareto	
k	p		p		p	
	0.00135	0.0027	0.00135	0.0027	0.00135	0.0027
1	370.370	185.185	370.370	185.185	370.370	185.185
1.1	298.807	153.671	14.634	14.113	7.857	7.714
1.2	237.510	126.105	8.034	7.884	4.398	4.358
1.3	188.852	103.545	5.776	5.702	3.233	3.214
1.4	151.507	85.643	4.635	4.590	2.650	2.638
1.5	123.131	71.590	3.945	3.914	2.301	2.292
1.6	101.514	60.552	3.483	3.460	2.068	2.062
1.7	84.896	51.828	3.151	3.133	1.903	1.898
1.8	71.967	44.867	2.902	2.887	1.779	1.776
1.9	61.776	39.254	2.707	2.694	1.684	1.681
2	53.638	34.679	2.550	2.539	1.607	1.605
2.1	47.056	30.909	2.422	2.412	1.545	1.543
2.2	41.670	27.772	2.314	2.306	1.494	1.492
2.3	37.213	25.135	2.223	2.215	1.451	1.449
2.4	33.486	22.899	2.144	2.138	1.414	1.412
2.5	30.341	20.987	2.076	2.070	1.382	1.381
2.6	27.663	19.339	2.016	2.011	1.355	1.353
2.7	25.365	17.909	1.963	1.958	1.330	1.329
2.8	23.377	16.659	1.916	1.911	1.309	1.308
2.9	21.647	15.561	1.873	1.869	1.290	1.289
3	20.130	14.590	1.835	1.831	1.273	1.272
3.1	18.794	13.727	1.800	1.796	1.257	1.256
3.2	17.610	12.956	1.768	1.765	1.243	1.243
3.3	16.556	12.264	1.739	1.736	1.231	1.230
3.4	15.612	11.640	1.713	1.709	1.219	1.218
3.6	13.999	10.564	1.665	1.662	1.199	1.198
3.8	12.676	9.671	1.624	1.621	1.182	1.181
4	11.576	8.919	1.589	1.586	1.167	1.167

Table 7.3a: ARL for p-variable and k-fixed

Dist.	p	k												
		1	1.1	1.2	1.4	1.6	1.8	2	2.2	2.4	2.6	2.8	3	4
Exponential	0.001	500.000	359.331	250.270	126.402	71.6315	45.2488	31.1304	22.8612	17.6520	14.1733	11.7383	9.96677	5.61552
	0.00135	370.370	271.082	192.830	101.195	59.0655	38.1631	26.7254	19.9092	15.5552	12.6143	10.5358	9.01135	5.20777
	0.0027	185.185	141.296	105.418	60.387	37.7402	25.7002	18.7573	14.4473	11.6026	9.6287	8.2020	7.13529	4.37395
	0.005	100.000	79.112	61.502	38.032	25.2581	18.0306	13.6586	10.8411	8.9248	7.5621	6.5565	5.79149	3.74298
	0.01	50.000	41.162	33.462	22.508	16.0010	12.0490	9.5227	7.8221	6.6243	5.7477	5.0851	4.57064	3.13738
	0.05	10.000	8.992	8.050	6.508	5.3962	4.5994	4.0172	3.5807	3.2451	2.9809	2.7686	2.59501	2.05926
Extreme-value	0.001	500.000	203.399	101.424	38.5797	20.6357	13.3472	9.67683	7.55068	6.19426	5.26655	4.59822	4.09708	2.77319
	0.00135	370.370	159.032	82.696	33.2142	18.3313	12.0910	8.88241	6.99682	5.78134	4.94363	4.33657	3.87928	2.66176
	0.0027	185.185	89.743	51.234	23.2583	13.7929	9.5230	7.21860	5.81786	4.89254	4.24300	3.76560	3.40194	2.41462
	0.005	100.000	53.723	33.194	16.7379	10.5745	7.6088	5.93722	4.88955	4.18171	3.67627	3.29979	3.00993	2.20773
	0.01	50.000	30.013	20.182	11.4057	7.7327	5.8336	4.70932	3.97939	3.47312	3.10425	2.82506	2.60730	1.99007
	0.05	10.000	7.608	6.128	4.4463	3.5462	2.9976	2.63294	2.37504	2.18407	2.03754	1.92190	1.82851	1.54563
Logistic	0.001	500.000	267.092	158.482	69.9251	37.9713	23.6950	16.3035	12.0454	9.38769	7.62277	6.39174	5.49833	3.31100
	0.00135	370.370	203.373	123.490	56.5161	31.5559	20.1292	14.0991	10.5716	8.34187	6.84545	5.79226	5.02199	3.10759
	0.0027	185.185	108.401	69.448	34.6092	20.6203	13.8456	10.1095	7.8451	6.37146	5.35798	4.62972	4.08749	2.69197
	0.005	100.000	61.995	41.679	22.4284	14.1695	9.9648	7.5534	6.0450	5.03759	4.32952	3.81123	3.41914	2.37795
	0.01	50.000	33.097	23.514	13.8175	9.3357	6.9218	5.4749	4.5372	3.89223	3.42767	3.08046	2.81303	2.07717
	0.05	10.000	7.769	6.316	4.5961	3.6491	3.0668	2.6794	2.4064	2.20523	2.05167	1.93109	1.83420	1.54390

Table 7.3b:ARL for p-variable and k-fixed. for weibull, power and pareto $\beta = 1.4$

Dist.	p	k												
		1	1.1	1.2	1.4	1.6	1.8	2	2.2	2.4	2.6	2.8	3	4
Weibull	0.001	500.000	398.439	312.267	193.769	126.823	88.2071	64.7200	49.6282	39.4439	32.2775	27.0527	23.1276	12.9539
	0.00135	370.370	298.807	237.510	151.507	101.514	71.9670	53.6380	41.6702	33.4865	27.6632	23.3771	20.1304	11.5757
	0.0027	185.185	153.671	126.105	85.643	60.552	44.8670	34.6790	27.7718	22.8988	19.3390	16.6594	14.5898	8.9195
	0.005	100.000	85.050	71.739	51.435	38.122	29.3760	23.4558	19.3026	16.2880	14.0321	12.2987	10.9358	7.0645
	0.01	50.000	43.696	37.977	28.886	22.552	18.1558	15.0376	12.7628	11.0565	9.7434	8.7100	7.8804	5.4217
	0.05	10.000	9.286	8.612	7.443	6.515	5.7873	5.2124	4.7524	4.3788	4.0710	3.8139	3.5964	2.8779
Power	0.001	500.000	14.7751	8.07350	4.64662	3.48913	2.90556	2.55289	2.31618	2.14598	2.01750	1.91694	1.83599	1.58921
	0.00135	370.370	14.6337	8.03364	4.63477	3.48305	2.90168	2.55012	2.31404	2.14425	2.01606	1.91571	1.83491	1.58855
	0.0027	185.185	14.1128	7.88352	4.58960	3.45979	2.88682	2.53946	2.30584	2.13763	2.01053	1.91097	1.83077	1.58602
	0.005	100.000	13.3058	7.64029	4.51465	3.42088	2.86185	2.52151	2.29200	2.12644	2.00118	1.90295	1.82376	1.58174
	0.01	50.000	11.8348	7.16004	4.35987	3.33922	2.80902	2.48334	2.26247	2.10252	1.98114	1.88575	1.80871	1.57250
	0.05	10.000	6.2802	4.76427	3.42145	2.80381	2.44757	2.21511	2.05109	1.92892	1.83423	1.75859	1.69670	1.50229
Pareto	0.001	500.000	7.89483	4.40865	2.65316	2.06974	1.78024	1.60816	1.49462	1.41438	1.35487	1.30908	1.27286	1.16727
	0.00135	370.370	7.85683	4.39814	2.65009	2.06819	1.77926	1.60747	1.49410	1.41397	1.35453	1.30880	1.27262	1.16713
	0.0027	185.185	7.71364	4.35804	2.63830	2.06223	1.77552	1.60483	1.49210	1.41239	1.35323	1.30771	1.27168	1.16661
	0.005	100.000	7.48134	4.29139	2.61846	2.05214	1.76917	1.60036	1.48871	1.40970	1.35102	1.30585	1.27008	1.16571
	0.01	50.000	7.02165	4.15331	2.57634	2.03056	1.75553	1.59071	1.48140	1.40389	1.34625	1.30183	1.26663	1.16376
	0.05	10.000	4.70758	3.30306	2.28261	1.87299	1.65355	1.51753	1.42539	1.35908	1.30924	1.27052	1.23965	1.14841

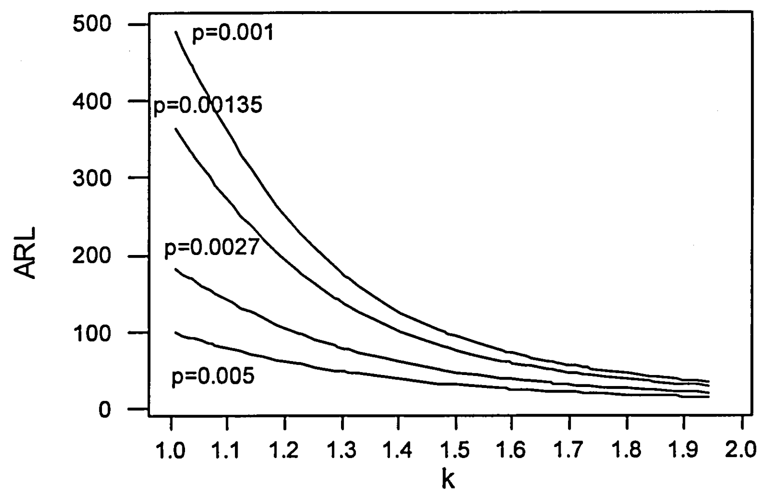


Figure 7.1 ARL for Exponential Dist. when p-fixed and k is generated from U(1,2).

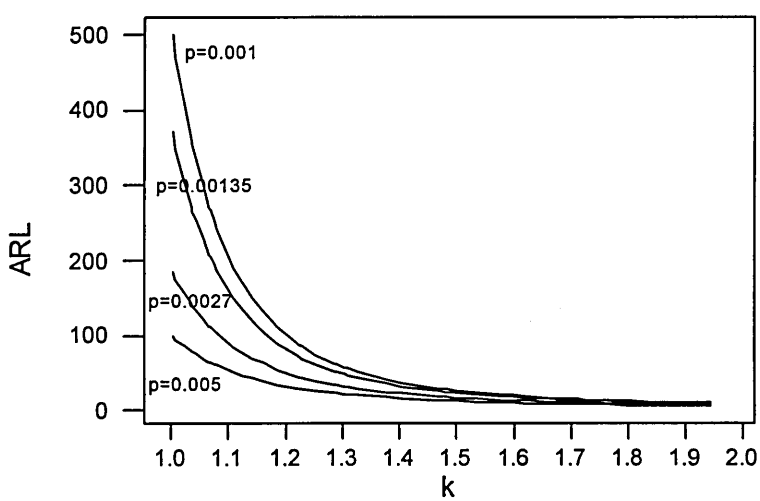


Figure 7.2 ARL for Extreme-value when p-fixed and k is generated from U(1,2)

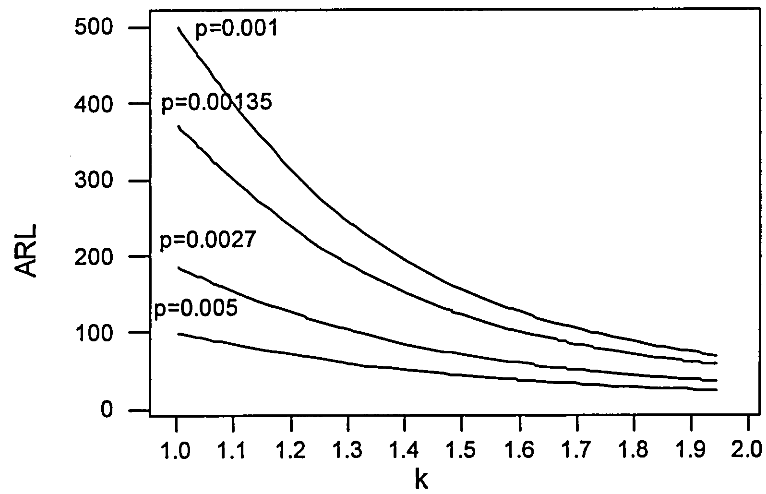


Figure 7.3 ARL for Weibull Dist. when p-fixed and k is generated from U(1,2)

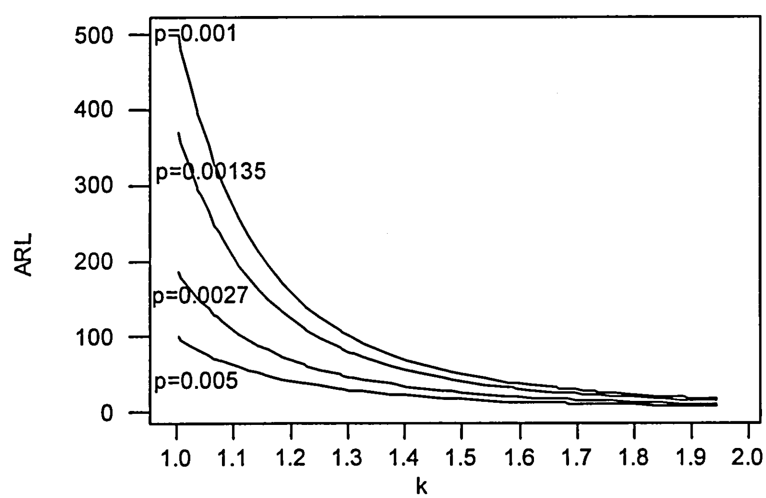


Figure 7.4 ARL for Logistic Dist. when p-fixed and k is generated from U(1,2).

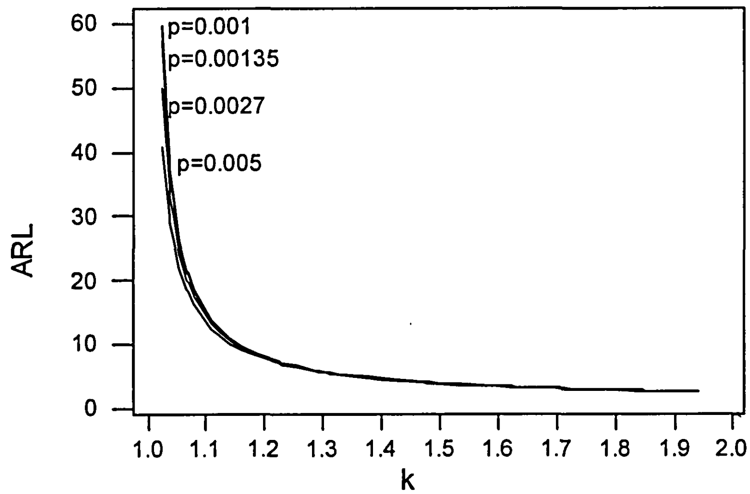


Figure 7.5 ARL for Power Dist. when p-fixed and k is generated from U(1,2).

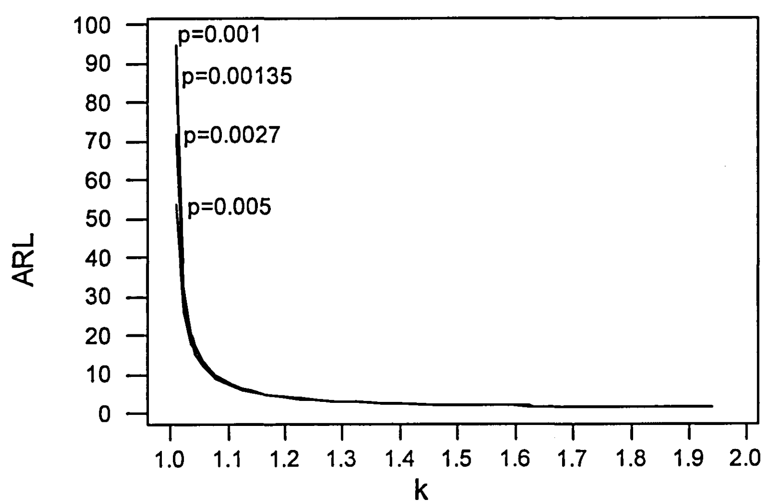


Figure 7.6 ARL for Pareto Dist. when p-fixed and k is generated from U(1,2).

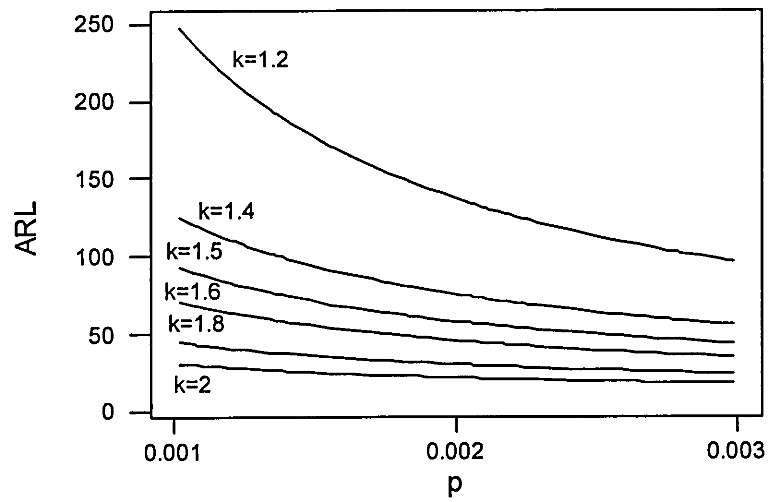


Figure 7.7 ARL for Exponential Dist. when k is fixed and p is generated from U(0.001,0.003)

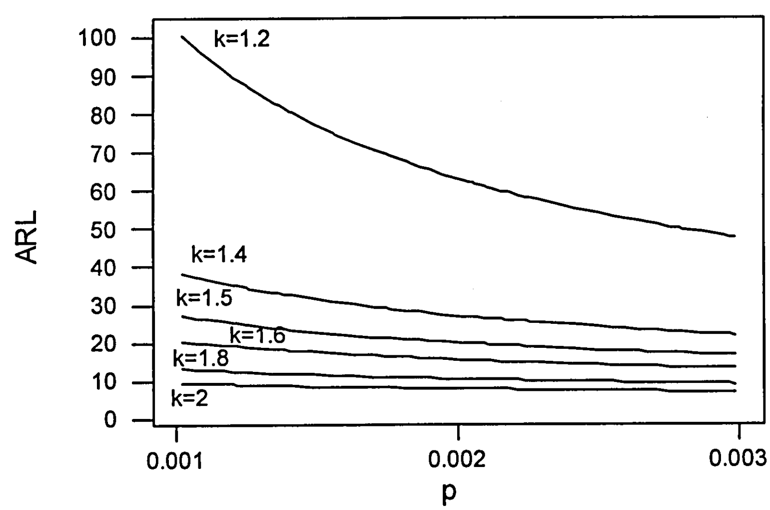


Figure 7.8 ARL for Extreme-value when k is fixed and p is generated from U(0.001,0.003).

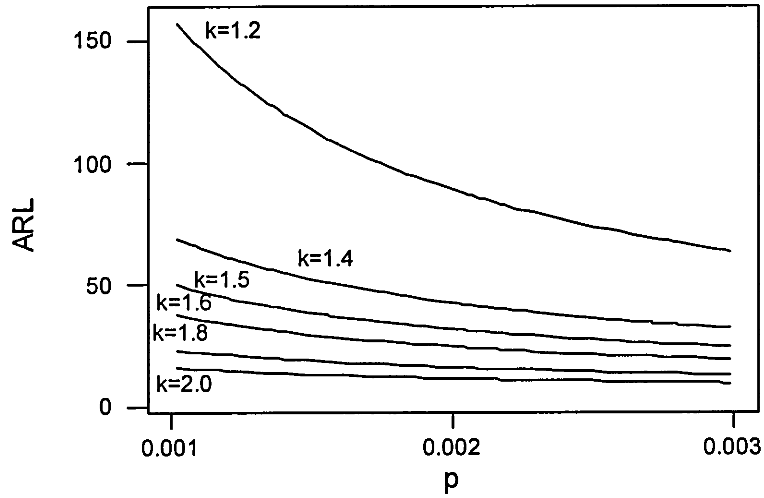


Figure 7.9 ARL for Logistic Dist. when k is fixed and p is generated from U(0.001,0.003)

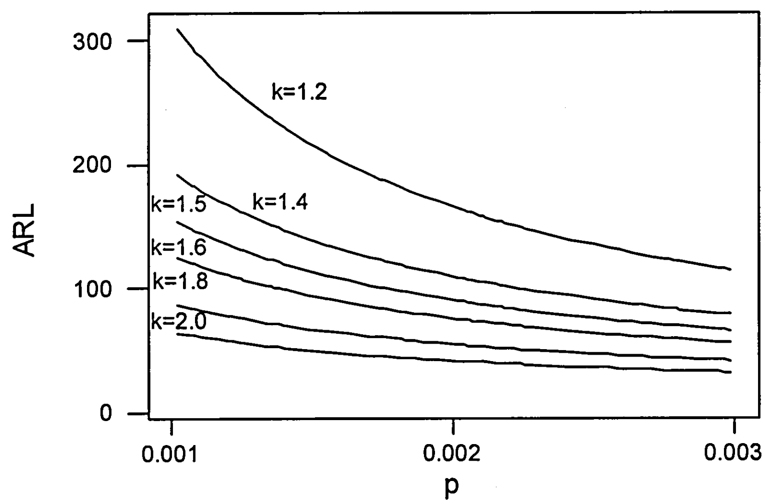


Figure 7.10 ARL for Weibull Dist. when k is fixed and p is generated from U(0.001,0.003)

7.5 Summary

The performance of a control chart for monitoring a process can be measured by the run length distribution and its mean, i.e. the average run length (ARL). Here, it provides the theoretical approach for exponential, extreme-value, pareto, power, Weibull and logistic distributions using quantile approach, (see table 7.1 and appendix 7). Moreover, the application of ARL has provided two cases, where case1, calculated ARL when k is variable and p-value is fixed, (see tables 7.2a-7.2b), and case 2, calculate ARL when p is variable and k is fixed, (see tables 7.3a-7.3b). From these cases, it has been found that, when k increases, ARL decreases. On the other hand, in case 2, when p increases, ARL decreases.

In the earlier chapters, we have introduced the evaluation of quantile control chart, process capability index and average run length for one variable. But, in reality, quality characteristic can depend on more than one variable, when a multivariate control chart is required. Therefore, in the next chapter, we will introduce the evaluation of multivariate control chart using quantile approach.

Chapter 8: Evaluating Multivariate Control Chart using Quantile Approach

8.1 Introduction

Control charts for one quality characteristic are used in industrial applications to observe whether a process is in control. This procedure has been discussed in previous chapters. In order to control and monitor the process mean of two or more quality characteristics simultaneously, multivariate control charts are required, (Sullivan and Woodall, 1996).

Traditional multivariate method depends on data vector being a random sample from multivariate normal distribution. One advantage of this method is the sampling distributions of most multivariate statistics which are approximately normal, regardless of the form of the parent population, because of a central limit theory, (see Johnson and Wichern 1998). This method plays an important role in practice when dealing with more than one factor. Multivariate normal distribution is based on normality, but in reality, most sets of data do not follow normal distribution. Therefore, the distributional shape is different from one distribution to another. For example, if we generate two sets of data, one from normal with mean 1 & standard deviation 1, and another from exponential with mean equal 1, both would have the same mean and standard deviation. However, the distributional shape would not be the same, especially if a small sample was used. So, the multivariate normal method does not gives accurate results when the quality characteristic comes from non-normal distribution.

Featured in the literature reviewed, trimmed mean control charts for univariate case has been presented. The use of these control charts are advantageous if the distribution is non-normal, (see White and Shroeder (1987), Iqlewicz and Hoaglin (1987) and Langenberg and Iglewicz (1986)). Alloway and Raghavachari (1990 , 1991) extended trimmed mean control charts for multivariate case. Abu-Shawiesh and Abdullah (1997) proposed control charts based on the α-trimmed mean. Cox T. F., (2000) suggested the use of Multidimensional Scaling used in Multivariate statistical process control, e.g. classical scaling, nonmetric scaling and biplots, when data are not normally distributed. Beirlant, Mason and Vynckier (1999) proposed a technique for description and analysis of non-normality data based on generalised quantiles of minimum volume ellipsoids.

Therefore, constructing multivariate control charts using multivariate normal method when the set of data are not normal would give inaccurate results. Grimshaw and Alt (1997), proposed a control chart using values of the quantile function. Estimating quantile function values requires a random sample to construct the sample quantile function,

$$\tilde{Q}(u) = n\left(\frac{2i+1}{2n} - u\right)x\,(i;n) + n\left(u - \frac{2i-1}{2n}\right)x\,(i+1;n)$$

for

$$\frac{2i-1}{2n} \leq u \leq \frac{2i+1}{2n} \qquad ; \qquad i = 1,2,\ldots,n-1$$

where

$x\,(1,n) \leq x\,(2,n) \leq \ldots \leq x\,(n;n)$ denote the order statistics.

This approach deals with the data as general data, i.e. it is not concerned with the distribution of quality characteristics. It concludes that the control charts for quantile values are quite effective at detecting changes in the distributional shape which would be difficult to detect in $\bar{x}$ *and* R charts. Moreover, this technique concentrates on the distributional shape in case the sets of data are normal or not, regardless of what the distribution of the sets of data are when it is non-normal.

Discussed here is the multivariate control chart using quantile approach (MCCQA). This technique is concerned with the distribution of the quality characteristics. For example, if the quality characteristics follows Weibull distribution, then MCCQA is dealing with the Weibull distribution properties to construct a multivariate control chart.

Multivariate control charts using quantile approach (MCCQA) is explained in the next section with an example that illustrates how the technique is used.

8.2 Multivariate Control Chart using Quantile Approach (MCCQA)

In order to improve the quality of a product, quality characteristics should be tested for causes of variation. The stability of variation means that it is produced by the common causes which is often present. These common causes do not always represent a major source of variation. The main role of control charts are to find out the occurrence of special causes of variation that affect from outside of the process. These variations can be used to repair the defect and improve the process.

This chapter discusses the use of quantile function defined by

$$Q(u) = F^{-1}(u) = [x : F(x) = u], \qquad 0 < u < 1$$

to construct multivariate control charts for non-normal data.

In order to estimate quantile distribution for non-normal data, take a sample (quality characteristic) for non-normal distribution and then calculate the quantile distribution Q(p) where

$$Q(p) = \lambda + \eta * R(p)$$

where λ *and* η are location and scale parameters, respectively, and $R(p)$ is quantile function.

The steps of construct multivariate control charts are as follows:

1. Calculate $Q_i(p)$ for each component in all factors
2. Calculate the mean of $Q_i(p)$ for each factor
3. Calculate the overall mean of step 2
4. Calculate the variance of each raw in each factor

5. Calculate the mean of variance in step 4

6. Calculate the covariance of the data. This can be calculated by taking the whole raw from each factor

7. For sufficient large n sample, the random factors are approximately multivariate normal, then calculate T-square value.

$$T^2 = n\,(\bar{x} - \bar{\bar{x}})'\,\Sigma^{-1}\,(\bar{x} - \bar{\bar{x}})$$

where $\bar{x}$ is substituted by $\bar{Q}(p)$, $\bar{\bar{x}}$ is substituted by $\bar{\bar{Q}}(p)$, Σ is substituted by $Cov(Q_i(p), Q_j(p))$ and n is the sample number.

8. $UCL = \dfrac{p\,(m-1)\,(n-1)}{mn - m - p + 1} F_{\alpha, p, mn-m-p+1}$ and $LCL = 0$; for **n>1**

9. $UCL = \dfrac{(m-1)^2}{m} * \dfrac{(p/(m-p-1)\,F(\alpha/2; p, m-p-1)}{1 + (p/(m-p-1))\,F(\alpha/2; p, m-p-1)}$ and $LCL = 0$; for **n=1**

10. Plot T^2-value on the control chart using UCL and LCL shown in steps 8 and 9.

11. Then take the design whether there is a special cause or not.

Table 8.1 provides the whole process for constructing multivariate control charts using quantile approach, where p=2. Moreover, for p>2 the process uses the same method and extending the table 8.1 to the number of factors.

Table 8.1 multivariate control chart using quantile approach, where p=2

Steps	p_1				p_2			
1	x_{111}	x_{121}	$\cdots$	x_{1j1}	x_{112}	x_{122}	$\cdots$	x_{1j2}
	x_{211}	x_{221}	$\cdots$	x_{2j1}	x_{212}	x_{222}	$\cdots$	x_{2j2}
	x_{311}	x_{321}	$\cdots$	x_{3j1}	x_{312}	x_{322}	$\cdots$	x_{3j2}
	. .	. .	. .	. .	. .	. .	. .	. .
	x_{i11}	x_{i21}	$\cdots$	x_{ij1}	x_{i12}	x_{i22}	$\cdots$	x_{ij2}
2	Calculate quantile function for each preliminary sample of both factors							
	$Q(x_{i11)})$ $Q(x_{i21)})$ $Q(x_{ij1})$ where $i=1,2..,m$; $j=1,2..,n$ and $k=1,2,..,p$ are represented by preliminary sample, sample number and factors number, respectively				$Q(x_{i12)})$ $Q(x_{i22)})$ $Q(x_{ij2)})$ where $i=1,2..,m$; $j=1,2..,n$ and $k=1,2,..,p$ are represented by preliminary sample, sample number and factors number, respectively			
3	p_1				p_2			
	$Q(x_{111})$	$Q(x_{121})$	$\cdots$	$Q(x_{1j1})$	$Q(x_{112})$	$Q(x_{122})$	$\cdots$	$Q(x_{1j2})$
	$Q(x_{211})$	$Q(x_{221})$	$\cdots$	$Q(x_{2j1})$	$Q(x_{212})$	$Q(x_{222})$	$\cdots$	$Q(x_{2j2})$
	$Q(x_{311})$	$Q(x_{321})$	$\cdots$	$Q(x_{3j1})$	$Q(x_{312})$	$Q(x_{322)}$	$\cdots$	$Q(x_{3j2})$
	. .	. .	. .	. .	. .	. .	. .	. .
	$Q(x_{i11})$	$Q(x_{i21})$	$\cdots$	$Q(x_{ij1})$	$Q(x_{i12})$	$Q(x_{i22})$	$\cdots$	$Q(x_{ij2})$

4	Calculate the mean of the each raw in step 3 which called $\overline{Q}(x_{i\,j1)})$	Calculate the mean of the each raw in step 3 which called $\overline{Q}(x_{i\,j2)})$
5	Calculate the over all mean in step 4 which called $\overline{\overline{Q}}(x_{i\,j1})$	Calculate the over all mean in step 4 which called $\overline{\overline{Q}}(x_{i\,j2})$
6	Calculate the variance of $Q(x_{i\,j1})$ from each raw in step 3	Calculate the variance of $Q(x_{i\,j2})$ from each raw in step 3
7	Calculate the mean of variance in step 6	Calculate the mean of variance in step 6
8	Covariance of the data. can be calculated by taking the whole raw from each factor. e.g. $Cov\,(Q(x_{1j1}), Q(x_{1j2}))$	
9	Calculate Hotelling T^2-test	
10	$UCL = \frac{p(m-1)(n-1)}{mn-m-p+1} F_{\alpha, p, mn-m-p+1}$ $LCL = 0$	
11	Plot T^2 values using control limits in step 10	

Note: A program using Minitab 12.2 is required in order to calculate T^2-value at p=2

```
#In order to calculate Hotelling T-square
#c1=n
#c2=x1-bar
#c3=x2-baar

#c4=x1-bar-bar
#c5=x2-bar-bar
#c6=var(x1)
#c7=var(x2)
#c8=cov(x1,x2)
```

```
#c13=Hotelling T-square

let c10=(c1/((c6*c7)-(c8**2)))

let c11=(c7*((c2-c4)**2))+(c6*((c3-c5)**2))

let c12=(2*c8*((c2-c4)*(c3-c5)))

let c13=c10*(c11-c12)
```

8.3 Application

In this section, the concept of multivariate control charts is applied using a quantile approach. In industrial application, the sample size (subgroup size) is mostly n=1. Therefore, two quality characteristics i.e. p=2, with individual samples from Weibull distribution, with each factor having 30 preliminary samples, were generated. The reason or choosing these data is in order to construct multivariate control charts almost dealing with the data as a normal data dependent on the central limit theory. Whereas, in reality, most sets of data are not following normal distribution. On the basis of a Weibull data shown in table 8.2, the Hotelling T-square value, using the original method and quantile approach were calculated. The calculation quantile approach in table 8.1 is used based when n=1. The original method can be found in the literature on multivariate quality control of chapter 2, see table 8.2. The calculation for the original method is also based on n=1

Hotelling T-square control charts are presented on figure 8.1 for original method and figure 8.2 for quantile approach. The calculation of statistical control limit and α can be discussed next.

For quality characteristics multivariate charts, the probability that the chart indicates the control when the process is in control is $1-\alpha$, which equal 1-0.0027p, where α can be calculated by $\frac{\alpha}{2p} = 0.00135$, Alt (1982a). In this example, p=2, then α=0.0054.

Control limits and T-square values are needed to construct a multivariate control chart. The T-square value are given in table 8.2. The corresponding control limits are as follows

$$UCL = \frac{(m-1)^2}{m} * \frac{(p/(m-p-1)\,F(\alpha/2; p, m-p-1)}{1+(p/(m-p-1))\,F(\alpha/2; p, m-p-1)}$$

where m=30, p=2 and α /2=0.0027

then UCL=9.9447 and LCL=0, because any shift in the mean will lead to an increase in the statistics T-square.

These control limits are shown on the chart in figure 8.1 for original method and figure 8.2 for quantile approach. Note that there are two points that exceed the limits, which are point 9 and point 21 in both graphs. Therefore, it can be concluded that the process is out-of-control, using both methods. In addition, from figure 8.2 and table 8.2, the T-square value using quantile approach is more sensitive than the original method.

Table 8.2: Generated Weibull Data (p=2, m=30, n=1)

First Variable	Second Variable	$Q_1(p)$	$Q_2(p)$	T^2 Using Original Method	T^2 Using Quantile Method
0.49768	1.60048	0.124518	0.187012	0.8208	1.2917
0.27421	1.29882	0.149376	0.252207	0.7126	0.953
0.73444	0.91155	0.171868	0.306842	0.6907	0.4989
0.60762	2.44292	0.193241	0.35638	4.8634	7.0668
0.34679	0.46853	0.214037	0.402955	0.8493	1.085
0.1588	1.30821	0.234564	0.447685	1.4213	1.6815
0.37905	1.98865	0.255026	0.49126	2.3448	3.5387
0.59516	0.76282	0.275574	0.534157	0.2507	0.1965
1.51664	1.11038	0.296336	0.57674	11.7773	10.3288
0.29411	0.4076	0.317423	0.619302	1.1985	1.5019
0.85034	0.56975	0.338942	0.6621	1.8682	1.6486
0.68036	0.93597	0.360998	0.70537	0.4192	0.28
0.40606	1.27413	0.383701	0.749341	0.2475	0.4275
0.48438	0.41187	0.407173	0.794246	0.77	0.9928
0.83928	0.12253	0.431545	0.840332	3.09	3.3483
0.75684	1.24276	0.456969	0.887871	0.9386	0.84
0.24784	1.3752	0.483625	0.937168	0.9624	1.2722
0.11747	0.82459	0.511724	0.98858	1.591	1.6907
0.21567	1.37701	0.541525	1.042536	1.1485	1.4642
0.53503	0.09482	0.573353	1.099559	1.8551	2.4326
0.59071	0.79566	0.607619	1.160313	0.2082	0.1502
0.45503	3.10855	0.644866	1.225658	10.0444	14.7286
0.66969	1.40307	0.685828	1.29675	0.7378	0.8537
0.40722	0.6712	0.731533	1.3752	0.311	0.3992
0.19265	0.7108	0.7835	1.463372	1.1526	1.2791
0.22723	0.34137	0.844107	1.564944	1.7216	2.1025
0.74164	0.57204	0.917414	1.686158	1.1203	1.0189
0.09058	0.7487	1.011265	1.838994	1.8921	2.0048
0.09706	0.45101	1.144311	2.051693	2.3671	2.6463
0.63011	0.57463	1.384605	2.425972	0.6253	0.6238

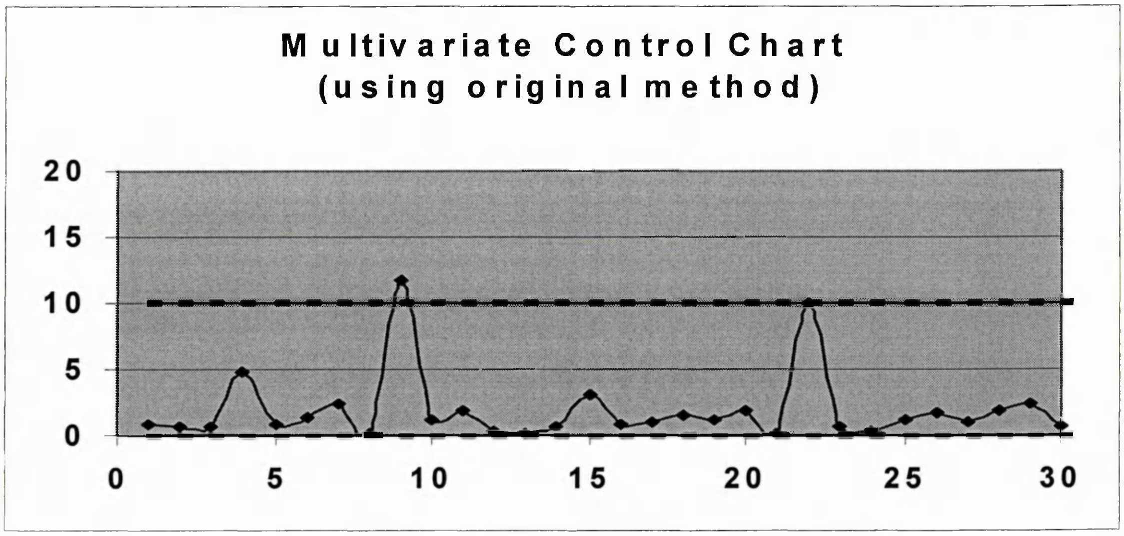


Figure: 8.1

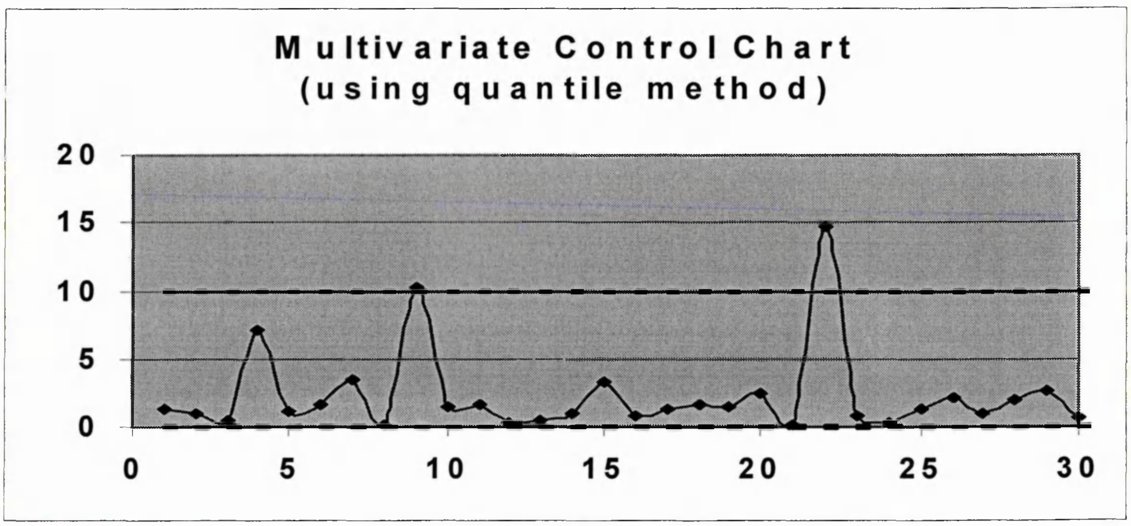


Figure: 8.2

8.4 Summary

In the literature, multivariate control chart was discussed for normal distribution. Whereas, in the reality, most quality characteristics do not follow normal distribution. Therefore, it produces a new technique, in order to dealing with non-normal situation, which called Multivariate Control Chart using Quantile Approach (MCCQA). The latter was compare with the traditional multivariate control chart (MCC), it founds that the MCCQA is more sensitive than the (MCC).

Multivariate control chart used when quality characteristics depends on more than one variable. On the other hand, when quality characteristics depends on one variable, but this variable rely on more than one quality distribution, then a mixture distribution is needed. Therefore, in the next chapter, will be discuss the quantile control chart for mixture distribution.

Chapter 9: Quantile Control Chart for Mixture Distribution: AN INNOVATIVE APPROACH

9.1 Introduction

A mixture distribution can be considered when the data are represented by two or more kinds of distribution. Suppose that a random variable takes values in a sample space and its distribution can be represented by probability density function of the form

$$f(x) = \sum_{i=1}^{k} \alpha_i f(x_i)$$

where $\alpha_i > 0$, i=1,2,...,k , $\sum_{i=1}^{k} \alpha_i = 1$

$f_i(x) \geq 0$ $\quad \int f_i(x)dx = 1$

where α_i denotes the mixing weight.

In general, mixture distributions have a wide application, for example, in biology it is sometimes used to measure certain characteristics in natural populations. Bhattacharya (1967) and Cassie (1954) discussed the length distribution of a certain type of fish, and found it useful to split their observations into age categories, where each category contributes a normal component distribution, to yield an overall mixture. Ashton (1971) used a mixture of a gamma distribution with a displaced exponential distribution to model the frequency distribution of time gaps in road traffic and fitted mixture with Weibull components, (see Kao 1959). A mixture distribution can be used to investigate whether a sample of blood pressure data can be separated into two normal populations (Clark, 1968). George (1969) also applied a mixture of normal distributions to data arising from measuring the content of DNA in the nuclei of liver cells of rats. In industrial context, a mixture distribution can be observed as in time failure to bulb life.

Kanji (1985) discussed a mixture of Laplace and normal density function with variable proportionality constant to describe a Wind Shear. He suggested that there was a systematic difference in the values mixing parameters for different bands widths. Kapoor and Kanji (1990) apply the characterisation theory to develop a model selection to choose the mixing parameter. Jones and McLachlan (1990) propose Kanji's technique for fitting a mixture of Laplace and normal distributions, but without the constraint that the component densities should have equal variance. Scallan (1992) discussed Kanji (1985) using composite link formulation to be fitted to data which may be cross-classified by one or more factors. Titterington, *et al.* (1985), Everitt and Hand (1981) give some applications of finite mixture distribution and its properties.

Certain difficulties are associated with statistical analysis of mixture data due to mainly two reasons. Firstly, explicit formulae generally does not exist for the estimates of the various parameters, and therefore numerical methods are required. Secondly, there are theoretical difficulties which arise in certain aspects of the statistical analysis, that create some common mixture problems which are of non-standard nature. As a result, detailed investigation of the analysis of finite mixture problems provides more than just a catalogue of straightforward applications of standard methods to a particular class of statistical approach.

In quality control research, the output of a production process of quality characteristics are sometimes obtained from one or more distributions, that have different statistical properties. In such situations a mixture distribution approach helps us to find out the properties of the production process.

The purpose of this chapter is to develop median rankit control charts using quantile approach for individual measurements arising from a process which follow a mixture distribution.

It is observed from the reviewed literature that most of the work on mixtures with continuous components which are non-normal uses exponential components, (Everitt and Hand, 1981). Such mixtures arise in industrial applications, especially in the analysis of failure time data, and have important mathematical properties. Practical applications for mixture distribution seem to be rare in the literature.

Mixed failure population are encountered in many fields of applied science, such as in engineering applications. The engineer may divide the failures of a system or a device into two or more different types of distribution. An example, provided here, is to discussed the time of failure which follows two distributions. It provides the mixture model for these distributions, and the estimates of the parameters of the model. A control chart for the quantile mixture distribution will be discussed.

In reality, the construction of control charts for mixture distribution have some difficulties. In this research the author distinguishes between an unstable process and a mixture process, i.e. if the process is unstable then charting is not appropriate. Another reason for difficulties, is to identify the components of the mixture. Finally, the problem lies in the mathematical complexities of the mixture distribution.

9.2 Characterisation Theory

In this section characterisation theory will be discuss briefly, using Bhattacharyya bounds to understand a particular issue of distribution theory, used by Fosam and Kanji (1994) and Kapoor and Kanji (1990).

A series of lower bound for the variance of an unbiased estimator of a function parameter was established by Bhattacharyya (1946), a special case of which is the Cramer-Rao bound. The Bhattacharyya matrix is defined as the covariance matrix $J_{r,s}$ where r,s=1,2,...,k given by

$$J_{r,s}=E\left(\frac{L^{(r)}(\alpha)}{L(\alpha)}\ \frac{L^{(s)}(\alpha)}{L(\alpha)}\right)$$

where $L^{r}(\alpha)$ denotes the rth derivatives with respect to α of the likelihood function. Whittaker (1973) showed that for a mixture of two known densities,

$$p(x)=\alpha f_1(x)+(1-\alpha)f_2(x), \qquad 0\prec\alpha\prec 1$$

where f_1 and f_2 are such that they differ at each point of some set of positive Lebesgue measure (or positive counting measure in the discrete case), the Bhattacharyya matrix for the mixture distribution is given by;

$$J_{r,s}=\left\{\begin{matrix}(r!)^2\binom{k}{r}\left(\frac{1-I(\alpha)}{\alpha(1-\alpha)}\right)^{\mathrm{r}} & ,\mathrm{r}=\mathrm{s}=1,2,...\mathrm{k} \\ 0 & ,\text{otherwith}\end{matrix}\right\}$$

where

$$I(\alpha)=\int\frac{f_1(x)f_2(x)}{p(x)}$$

is the Fisher information. The diagonal nature of the matrix is particularly useful since construction of the lower bound involves inverting it. Shanbhag (19972,1979) has obtained characterisations based on the diagonality property. Kapoor and Kanji (1990) showed that for these two components mixture, the Bhattacharyya matrix is diagonal if and only if the mixing parameter is linear in the parameter of interest. Fosam and Kanji

(1994) generalised this result to the case of mixtures with more than two components and constructed the Bhattacharyya matrices for such cases.

The latter authors applied the theoretical aspect mentioned above to wind shear data. The models considered included mainly mixtures of Normal, Lognormal and Laplace distributions. This resulted in two competing models, Normal-Lognormal and Normal-Laplace. The Normal-Laplace mixture was selected on the basis of goodness of fit to the data depending on the method used to estimate the mixing proportion.

In this section, the author will try to apply this method to Time of Failure for Light Bulbs, (see table 5, in chapter 5). It was found that this data can follow Weibull family, such as Weibull, Power and Pareto distributions. It can also follow a mixture of one of the following; Weibull-Power, Weibull-Pareto and Power-Pareto. The reason for applying mixture distribution through characterisation theory is to compare this methodology with the quantile approach, which is discussed in the next section.

Relying on robustness in estimating the mixing parameter, it is possible to use the special version of Theorem 3 (from Fosam and Kanji 1994), to select the better of the two models. This can be achieved by comparing the Bhattacharyya bound for the two models. The better model will have a bigger bound.

Let $p_1(x)$ and $p_2(x)$ be the two competiting models, where $p_1(x)$ represents a mixture of Weibull-Pareto model and $p_2(x)$ represents a mixture of Weibull-Power model.

$$p_1(x) = \alpha\,\eta\,x^{\eta-1}\,e^{-x\eta} + (1-\alpha)\,\eta\,x^{-(\eta+1)}$$

and

$$p_2(x) = \alpha\,\eta\,x^{\eta-1}\,e^{-x\eta} + (1-\alpha)\,\eta\,x^{\eta-1}$$

Suppose the Fisher information for the two models are $I_1(\alpha)$ and $I_2(\alpha)$ respectively, then the model $p_2(x)$ to be considered the more robust require $I_1(\alpha) < I_2(\alpha)$.

This is so if

$$\frac{\alpha\,\eta\,x^{\eta-1}\,\mathrm{e}^{-x\eta} + (1-\alpha)\,\eta\,x^{\eta-1}}{\alpha\,\eta\,x^{\eta-1}\,\mathrm{e}^{-x\eta} + (1-\alpha)\,\eta\,x^{-(\eta+1)}} < 1$$

Such robustness of $p_2(x)$ is achieved by solving the above equation. Here, we can see that, it is not so easy to solve this equation as it requires a knowledge of complex mathematics. Moreover, most practitioners working in quality areas such as in quality engineering and quality management, will find difficulties to deal with such mathematics. Therefore, we will offer another simple approach which is more practical, and provide the required answer for quantile mixture distribution as follows.

9.3 Quantile Mixture Distribution

Let us assume that quality characteristics are produced from two distributions. Using quantile approach, the first distribution can be represented by $Q_1(p)$ and the other one can be represented by $Q_2(p)$. So, using the quantile rule, a linear model can be developed as follows

$$Q(p) = \alpha Q_1(p) + (1-\alpha) Q_2(p) \qquad \textbf{9.1}$$

which can be called a Quantile Mixture Distribution.

From the analysis conducted in chapter 5 (table 5), this study found that quality characteristics follow Weibull distribution, where the residual of sum of least absolute is 1.142144. Whereas, for further improvement to investigate which better distribution can be represent quality characteristics, Power distribution was found to be more accurate than Weibull distribution, where the residual of sum of least absolute is 1.017205, (see table 1, page 163).

Here, we discuss quantile mixture distribution for Weibull-Power model, where $Q_1(p) = \lambda + \eta(-\ln(1-p))^{\beta}$ represents weibull quantile distribution and $Q_2(p) = \lambda + \eta p^{\beta}$ represents power quantile distribution and α is a mixing weight, so the quantile mixture distribution for Weibull-Power is as follows

$$Q(p) = \lambda + \eta[\alpha(-\ln(1-p))^{\beta} + (1-\alpha)p^{\beta}] \qquad 9.2$$

In the next sections, we discuss the estimate of the parameters of the Weibull-Power mixture model, and then construct a median rankit control chart for mixture of Weibull-power distribution. Finally, we will apply the present theory an average run length (ARL) for Weibull-Power distribution.

9.4 Quantile Control Chart Theory for Mixture Distribution

9.4.1 Estimation of the Parameters for the Weibull-Power mixture Distribution

Some of the properties of the mixture of Weibull-Power distributions can be derived from equation 9.2 as follows

- Median

$$M = Q(0.5) = \lambda + \eta(\alpha * (-\ln 0.5)^{\beta} + (1-\alpha) * 0.5^{\beta})$$

- Inter p- Range

$$R = Q((1-p),\lambda,\eta,\beta,\alpha) - Q(p,\lambda,\eta,\beta,\alpha)$$

$$R = \eta(\alpha((-\ln p)^{\beta} - (-\ln(1-p))^{\beta}) + (1-\alpha)((1-p)^{\beta} - p^{\beta}))$$

- Difference

$$D = Q((1-p),\lambda,\eta,\beta,\alpha) + Q(p,\lambda,\eta,\beta,\alpha) - 2m$$

$$D = \eta(\alpha((-\ln p)^{\beta} + (-\ln(1-p))^{\beta} - 2(-\ln 0.5)^{\beta}) + (1-\alpha)((1-p)^{\beta} + p^{\beta} - 2(0.5)^{\beta}))$$

Where λ, η and β in the equations above represent the median, scale and shape of a sample population, respectively. α is a mixing weight of the model, the initial value of α is 0.5. In order to estimate these parameters λ, η, β and α, it is recommended that the method of least absolute be used. This method is more robust than the method of least square, because it utilises median rankit. The best model that represents quality characteristics has the smallest residual sum of least absolute.

9.4.2 Control Chart for Mixture of Weibull-Power Distribution

Described below is the median rankit control chart for the mixture of quantile Weibull-power Distribution. The action and warning limits, which are used in the control chart procedures, can be derived from:

$$Q(p) = \lambda + \eta\,[\alpha(-\ln(1-p))^{\beta} + (1-\alpha)p^{\beta}]$$

where the warning limits are Q (0.05) and Q (0.95), the action limits are Q(0.99) and Q (0.01), and the central point (median rankit) is at Q (0.5). Therefore a typical quantile control limits that can be constructed for mixture quantile of Weibull-power distribution by following the steps below:

- Develop the mixture of Weibull-power quantile distribution

$$Q(p) = \lambda + \eta[\alpha(-\ln(1-p))^{\beta} + (1-\alpha)p^{\beta}]$$

- Estimate the parameters $\lambda, \eta, \alpha, \beta$ by using the least absolute method (median rankit), then

$$Q(p) = \lambda^{\wedge} + \eta^{\wedge}[\alpha^{\wedge}(-\ln(1-p))^{\beta^{\wedge}} + (1-\alpha^{\wedge})p^{\beta^{\wedge}}]$$

Where $\lambda^{\wedge}, \eta^{\wedge}, \alpha^{\wedge}, \beta^{\wedge}$ are location, scale, mixing weight and skewness respectively.

- The control limits of the mixture quantile of Weibull-power distribution function can be obtained by substituting p=0.5 in Q (p) above. It will provide the central point which will be described as median rankit point, and similarly by substituting p=0.05, p=0.95 and p=0.01, p=0.99, will provide both the warning limits and action limits of median rankit point, respectively.

9.4.3 ARL for Mixture of Weibull-Power Distribution

Quantile distribution is defined as

$$Q(P) = \lambda + \eta R(p)$$

Quantile distribution for mixture distribution is defined as

$$Q(P) = \lambda + \eta[R_1(p) + R_2(p)]$$

Here, we discuss the mixture of Weibull-power distribution, which is defined as

$$Q(p) = \lambda + \eta[\alpha(-\ln(1-p))^{\beta} + (1-\alpha)p^{\beta}]$$

So, quantile function of mixture of Weibull-power distribution is defined as

$$R(p) = [\alpha(-\ln(1-p))^{\beta} + (1-\alpha)p^{\beta}] \qquad 9.3$$

where α is the mixing weight, β is the distribution's shape and p is probability of lower limit $p_{(L)}$ and upper limit $p_{(U)}$.

By using equation 9.3, **Lower Probability Limit** can be found as

$$R(p_{(L)}) = [\alpha(-\ln(1-p_{(L)}))^{\beta} + (1-\alpha)p_{(L)}^{\beta}]$$

$$\frac{1}{k}R(p_{(L)}) = \frac{1}{k}[\alpha(-\ln(1-p_{(L)}))^{\beta} + (1-\alpha)p_{(L)}^{\beta}]$$

Let

$$[\alpha(-\ln(1-p_{(L)}^{*}))^{\beta} + (1-\alpha){p_{(L)}^{*}}^{\beta}] = \frac{1}{k}[\alpha(-\ln(1-p_{(L)}))^{\beta} + (1-\alpha)p_{(L)}^{\beta}] \qquad 9.4$$

$$= [\alpha(-\ln(1-p_{(L)}^{*}))^{\beta} + (1-\alpha){p_{(L)}^{*}}^{\beta}]$$

$$F(\frac{1}{k}R(p_{(L)}) = F[\alpha(-\ln(1-p_{(L)}^{*}))^{\beta} + (1-\alpha){p_{(L)}^{*}}^{\beta}]$$

$$= F(R(p_{(L)}^{*}))$$

$$= p_{(L)}^{*}$$

By using equation 9.3 again, **Upper Probability Limit** can be found as

$$R(p_{(U)}) = [\alpha(-\ln(1-p_{(U)}))^{\beta} + (1-\alpha)p_{(U)}^{\beta}]$$

$$\frac{1}{k}R(p_{(U)}) = \frac{1}{k}[\alpha(-\ln(1-p_{(U)}))^{\beta} + (1-\alpha)p_{(U)}^{\beta}]$$

Let

$$[\alpha(-\ln(1-p_{(U)}^{*}))^{\beta} + (1-\alpha){p_{(U)}^{*}}^{\beta}] = \frac{1}{k}[\alpha(-\ln(1-p_{(U)}))^{\beta} + (1-\alpha)p_{(U)}^{\beta}] \qquad 9.5$$

$$= [\alpha(-\ln(1-p_{(U)}^{*}))^{\beta} + (1-\alpha){p_{(U)}^{*}}^{\beta}]$$

$$F(\frac{1}{k}R(p_{(U)}) = F[\alpha(-\ln(1-p_{(U)}^{*}))^{\beta} + (1-\alpha){p_{(U)}^{*}}^{\beta}]$$

$$= F(R(p_{(U)}^{*}))$$

$$= p_{(U)}^{*}$$

then from 7.2

$$ARL = \frac{1}{(1-p_{(U)}^{*} + P_{(L)}^{*})}$$

In order to find out the value of ARL, we need to know the values of lower probability limit $p_{(L)}^{*}$ and upper probability limit $p_{(U)}^{*}$ in equation 9.4 and equation 9.5. It is not possible to solve these equations mathematically, but they could be solved, using iteration method. Excel 97, helps to make this job much easier.

In order to calculate the $p_{(L)}^{*}$ and $p_{(U)}^{*}$ of ARL, using Excel 97, to generate 10000 numbers from U(0.0001,1), which are called p-value. Then, solve the right hand side of equation 9.4 for lower limit or equation 9.5 for upper limit, and look for which p-value is present the result of right hand side of equation 9.4 or equation 9.5. This p-value represents $p_{(L)}^{*}$ of equation 9.4 and $p_{(U)}^{*}$ of equation 9.5.

9.5 Application

It will discuss quality characteristics of the time to failure measurement of light bulbs in table 5 in chapter 5. We are interested here in finding out whether the failure time for 25 light bulbs are within the specific acceptance limit of production process. The data has been considered either for the Weibull and Power models separately. It has been found that quantile power distribution is better fit than quantile Weibull distribution. However, here we are interested in fitting a quantile mixture model for the light bulbs data.

9.5.1 Model Validation

The process estimation and validation of the data which is believed to follow quantile mixture distribution of Weibull and power, has been investigated. For a good fit of the data, the series points is expected to lie on 45^0 line, see figure 1.

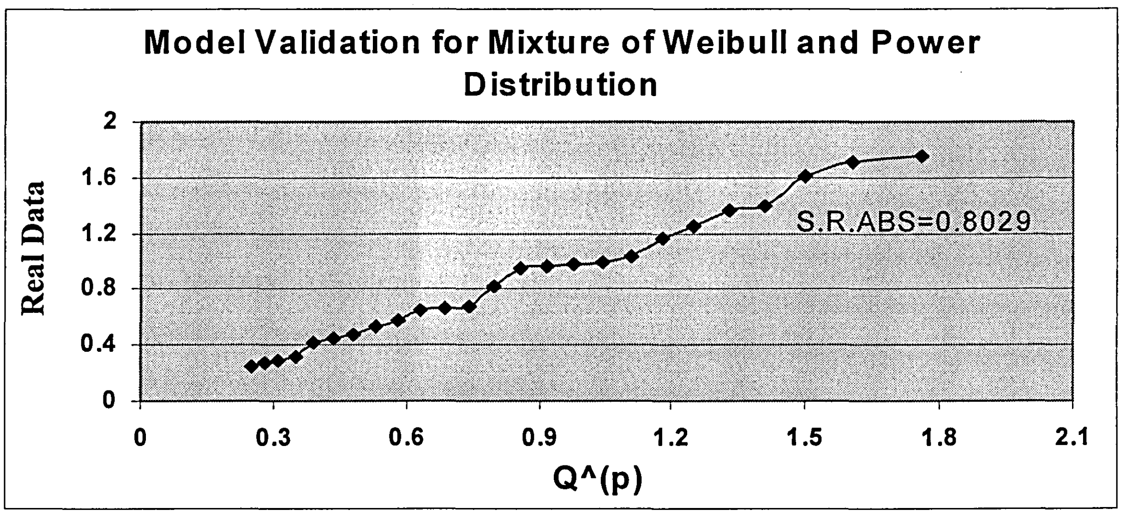


Figure 1: Model Validation using Median Rankit Value for Mixture of Weibull-Power Distribution

It is shown from the result given in table 1 and figure 1, that the Weibull-power mixture distribution fits the data very well, as the sum of residual of least absolute is 0.8029, which is better than the other two distributions. Here, the sum of residual of least absolute of Power distribution is smaller than the sum of residual of least absolute of Weibull distribution, when dealing with these distributions as a single distribution, see figure 8 and figure 11 in chapter five. Therefore, the best model for the data, can be represented by a quantile mixture distribution of Weibull-power.

In order to estimate the parameters, it deals with the same technique used in chapter 5. In addition, use the mixing weight is 0.5 as initial value, then estimate the whole parameters in order the get the best parameters represent the mixture model, see table 1.

Table 1: Estimate Parameters and Residual of Sum Least Absolute Value.

	λ	η	β	S.L.ABS.
Weibull	0.008078	0.96979	0.563196	1.142144
Power	0.272271	0.859574	1.574089	1.017205
Weibull-Power	0.236137	0.775063	1.265184	0.8029

9.5.2 Mixture of Weibull-Power Distribution Control Limits for Individual Measurement

Median rankit control limits for mixture of Weibull-power distribution are calculated at p=0.05 and p=0.01 for warning and action limits respectively using the following formula

$$Q(p) = \lambda^{\wedge} + \eta^{\wedge} [\alpha^{\wedge}(-\ln(1-p))^{\beta^{\wedge}} + (1-\alpha^{\wedge})p^{\beta^{\wedge}}]$$

Where $\lambda^{\wedge}, \eta^{\wedge}, \beta^{\wedge}$ and $\alpha^{\wedge}$ are given by 0.236137, 0.775063, 1.265184 and 0.079598 respectively and the residual sum of least absolute is 0.8029. This will provide the required control limits as, Control point=0.7973, Warning limits = (0.2659, 1.66) and Action limits =(0.24, 1.9017). Table 2, provides the quantile mixture control limits for Weibull-Power Distribution at different p levels.

Table 2: Quantile Mixture Control Limits for Weibull-Power Distribution.

Percentile value	Q(p)	Q(0.5)	Q(1-p)
0.01	0.240023	0.79729	1.901763
0.05	0.265942	0.79729	1.660008
0.001	0.23634	0.79729	2.201549
0.005	0.237753	0.79729	1.99236
0.00135	0.236446	0.79729	2.162115

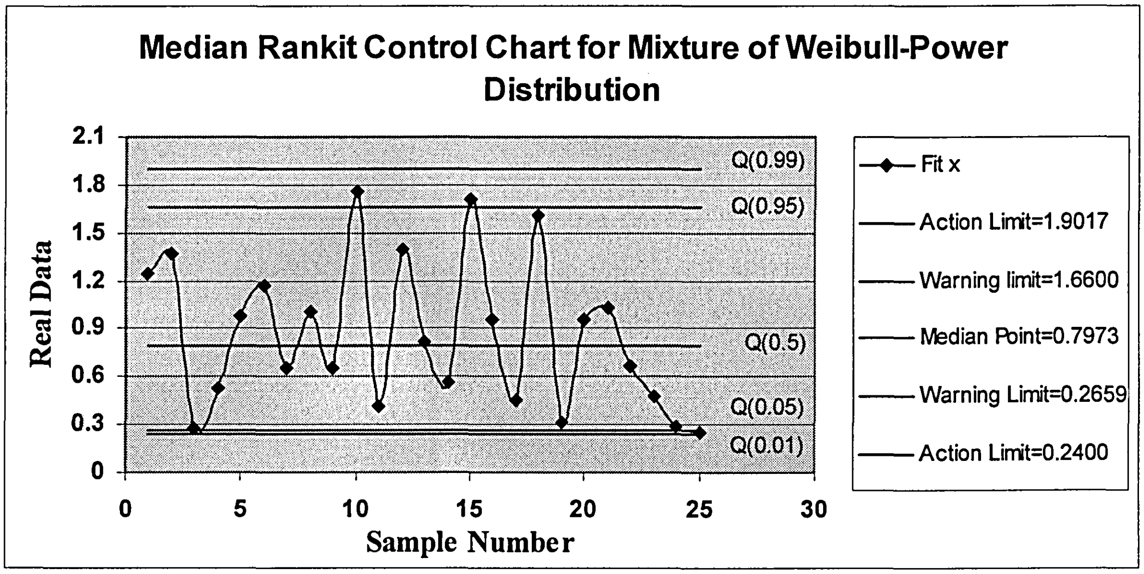


Figure 2: Median Rankit Control Chart for Mixture of Weibull-Power Distribution

It is clear from figure 2 that the sample number 10, 15, and 25 respectively are outside the warning limit. Therefore, the light bulbs failure time is out of control at p=0.05.

In order to compare quantile control chart for Weibull-power distribution, where the mixing weight is 0.079598, with the quantile control chart for Weibull and power as a single case, it was found in general that there are four samples needed to be investigated, which are samples 3, 10, 15 and 25.

It has been found that in Weibull control chart, (see figure 9 in chapter 5) the four samples in question are in control. On the other hand, in power control chart, (see figure 12 in chapter 5) it was found that the samples number 3, 10 and 15 are outside warning limit and sample 25 are outside action limit. While, the present mixture of Weibull-power control chart, indicates that the sample number 3 is with in control, and the samples number 10, 15 and 25 are outside warning limit, respectively.

Therefore, we can say that the power control chart is more sensitive than Weibull control chart, and a mixture control chart of Weibull-power distribution is much better than the a single distribution, i.e. Weibull & Power.

9.5.3 ARL for Mixture of Weibull-Power Distribution

By using average run length (ARL), we can found out the average number of points that must be plotted before a point which indicates an out-of-control. It has used the ARL formula which has been shown in section 9.4.3, in order to calculate ARL when $p_L = 0.00135, \alpha = 0.079598$ and $\beta = 1.265184$. It has been found that the ARL is equal to 370.37, when k=1, i.e. there is no shift in the process. Whereas, when the process shifted, say at k=1.1, the ARL is decreased to 205.9474, (see table 3). It can be noted from table 3 and figure 3, a relationship between the ARL and the shift in the process is, when k increases ARL decreases and vice versa.

Table 3 ARL for Mixture of Weibull-Power Distribution

K	p_L^*	p_U^*	ARL
1	0.00135	0.99865	370.37
1.1	0.00125204	0.9963964335	205.9474
1.2	0.00116883	0.991712156	105.7454
1.3	0.001097175	0.983299618	56.1874
1.4	0.001034759	0.970214078	32.4457
1.5	0.000979843	0.952431949	20.5982
1.6	0.000931115	0.930870222	14.2733
1.7	0.000887552	0.906868796	10.63617
1.8	0.00084835	0.881676203	8.3912
1.9	0.00081288	0.856227701	6.9163
2.0	0.000780576	0.831140777	5.8948

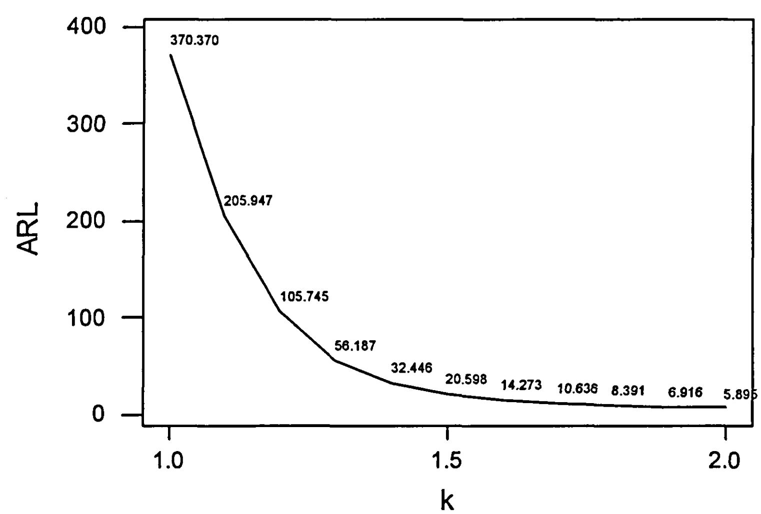


Figure 3. ARL for Mixture of Weibull-Power Distribution

9.6 Summary

It has constructed the mixture control chart using quantile approach for Weibull-power model and calculate average run length for this model. From the analysis, it found that the power control chart is more sensitive than Weibull control chart, and a mixture control chart of Weibull-power distribution is much better than a single distribution, i.e. Weibull & Power. Therefore, we can conclude that the mixture distribution plays very important role in quality control measurement of control chart.

In next chapter, it will discuss the conclusion of the thesis and the future work could be do, in the area of statistical process control by quantile approach.

Chapter 10 Conclusion and Future Work

10.1 Conclusion

Statistical process control (SPC) can be divided into two types. These are on-line SPC and off-line SPC. This thesis has focused on the former.

One of the underlying assumptions of SPC is the use of the normal distribution. Such an assumption is implicit in the construction of control charts and process capability studies. It has long been realised that the variability associated with many engineering processes does not have a normal distribution.

This assumption investigated the effect of non-normality on control charts and concluded that the non-normality is usually not problematic for subgroup sizes of four or more. For smaller subgroup sizes, and especially for individual measurements, non-normality can be a serious problem.

Control charts and process capability calculations remain fundamental techniques for statistical process control. However, it has long been realised that the accuracy of these calculations can be significantly affected when sampling is drawn a non-normal population. Many quality practitioners are conscious of these problems but are not aware of the effects of such problems on the validity of their results.

The above information indicates that there are real problems in dealing with statistical process control for non-normal distributions and mixture distributions. The main purpose of this thesis was to develop quality control charts and capability index for non-normal distribution and mixture distribution which can be easily adopted by practitioners of statistical process control.

Therefore, quantile approach was developed & used because it offered relatively new and generally powerful techniques for non-normal and mixture situations in the area of SPC. Moreover, it provided quantile distribution to construct median rankit control charts for non-normal distributions, e.g. Logistic, Exponential, Weibull, etc.

It compared the Shewhart control chart for individual measurements with median rankit control chart (MRCC) for individual measurement of logistic distribution, it achieved nearly the same results. This indicates that the median rankit control chart using quantile distribution plays an important role for quality improvement in industrial applications. In addition, it constructed this control chart for the rest of other distributions, such as Weibull and Power distributions.

In addition, it discussed and provided process capability indices using quantile approach. This is a new approach which is called, Generalisation Clement's Method using Quantile Approach (GCMQA). The latter was compared with the Generalisation Clement's Method using Pearsonian Process (GCMPP). It found that GCMQA is more accurate than GCMPP, because it gives accurate percentile results and does not depend on statistical tables.

Moreover, it provided the average run length using quantile distribution for non-normal distribution, and provided tables to make it easy to know the average number of sample which can be plotted before a sample indicating an out of control. Nonetheless, use of quantile approach to construct multivariate control chart was discussed. It was found that the T-square values using quantile approach are more sensitive than the original method.

Finally, in some practical cases, the quality characteristics follows more than one distribution. Therefore, a mixture distribution using quantile distribution was considered. It discussed the mixture model of Weibull-Power distribution and found that this model gives good results compared to when Weibull and Power distributions as a single model are discussed.

10.2 Future Work

In general, quantile approach are not widely available in the literature. Therefore, it can be used properly to improve many areas of statistical applications, especially, in statistical quality control.

One of these areas can be implemented is in six-sigma technique, (see, Buckley and Caine, 2000). Six sigma is a data driven process methodology which resolves problem by five stages which are Define the problem, Measure, analysis, Improve and Control. Most of the research done on six sigma approach dealing with quality characteristics using the assumption of normality. Therefore, quantile approach can be used to develop the idea of six-sigma approach in a robust method.

In general, six sigma through quantile approach can be used for the development of health care quality improvement. Health care can be looked at a wide variety of processes such as emergency room treatment, surgery and clinical testing. There are number of important problems, e.g. improve patient care, reduce costs, reduce patient treatment cycle time, reduce treatment improvement, etc. can be solved. Moreover, health care has a number of well defined metrics that measure the performance of health care processes such as length of stay in hospital, treatment error, speed of recovery, cost, etc. which can be dealt with quantile approach.

Shewhart control charts are constructed under the normality assumptions, which depends on the mean and variance. In this research, we provide control charts for non-normal distribution. However, there one can construct a control chart for parametric and non-parametric distribution using Bootstrap method, (see Seppala 1995). A comparison between control charts using quantile approach and control chart using Bootstrap method will be useful for further research, (see Hutson 2000).

References

Abu-Shawiesh, Moustafa, O. and Abdullah, Mokhtar, B. (1997). On the Hodges-Lehmann Control Chart based on the α-Trimmed Mean. Bulletin of the International Statistical Institute. 51^{th} Session ISI. Contribution paper, Book 1, pp: 285-286.

Albin, Susan L.; Kang, Lan; and Shea, Gerald. (1997). An X and EWMA Chart for Individual Observations. Journal of Quality Technology 29, pp.41-48.

Alloway, J. A. Jr. (1990). Multivariate Control Charts Based on Trimmed Means. ASQC Quality Congress Transactions. San Francisco, pp:449-453.

Alloway, J. A. Jr. (1991). Control chart Based on the Hodges-Lehmann Estimator. Journal of Quality Technology, 23, pp:336-347

Alloway, James A. , Jr. (1994). Dynamic Graphics Improve Multivariate SPC. 48th Annual Quality Congress, May 1994, Las Vegas NV, pp. 25-31.

Alt, F. B. (1982). Multivariate Quality Control: State of the Art. ASQC Quality Congress Transactions, pp. 886-893.

Alt, F. B. (1985). Multivariate Quality Control. Encyclopaedia of Statistical Sciences 6, edited by N. L. Johnson and S. Kotz. John Wiley, New York.

Amin, Raid W. and Ethridge, Ronald A. (1998). A Note on Individual and Moving Range Control Charts. Journal of Quality Technology. Vol. 30, No. 1, pp: 70-74.

Amin, Raid W; Wolf Hans; Besenfelder Werner and Baxley Robert (2000). EWMA Control Charts for the Smallest and Largest Observation. Journal of Quality Technology 31, pp.189-206.

Ashton, W. D. (1971). Distribution for gaps in road traffic. J. Inst. Maths. and its Applics, 7, pp.37-46.

Beirlant, J., Mason, D. M. and Vynckier, C. (1999). Goodness-of fit analysis for multivariate normality based on generalised quantile. Computational Statistics & Data Analysis, 30, pp:119-142.

Bhattacharya, C. G. (1967). A Simple Method of Resolution of a Distribution into Gaussian Components. Biometrics, 23, pp:115-135.

Bhattacharyya, A. (1946). On Some Analogues of the Amount of Information and Their use in Statistical Estimation. Sankhya B, pp:1-14.

Bissell, A. F. (1990). How Reliable is Your Capability Index? *Appl. Statist*, Vol. 39, No. 3, pp. 331-340.

Blache, K.M., Stewart, K.C., Zimmerman, R.L., Shaull, J.E., Benner, R.D., & Humphrey, G.P. (1988). Process-Control and People at General-Motors Delta Engine Plant. *Industrial Engineering, 20, pp.*24-30.

Bloomfield, P. and Steiger, W. L. (1983). Least Absolute Deviations. Theory, Applications, and Algorithms, Birkhauser Verlag, Boston, MA.

Borror, C. M., Montgomery, D. C., & Runger, G. C. (1999). Robustness of the EWMA control chart to non-normality. Journal of Quality Technology, 31, No. 3. pp: 309-316.

Box, G. E. and Cox, D. R. (1964). An analysis of transformation. J. Roy. statist. Soc., B26, pp.211-243.

Box, G. E. P. and Cox, D. R.(1964).An Analysis of Transformations. Journal of the Royal Statistical Society, Series B, 26, pp.211-252.

Box, George and Kramer, Tim. (1992). Statistical Process Monitoring and Feedback Adjustment- A Discussion. Technometrics. Vol. 34, No.3. pp:251-267.

Box, George E., Coleman, David E. and Jr. Baxley, Robert V. (1997). Journal of Quality Technology. Vol. 29, No. 2, pp: 128-130.

Boyles, Russell A. (1991). The Taguchi Capability Index. Journal of Quality Technology, Vol. 23, pp. 17-26.

Buckley, Bob and Caine, Peter (2000). Improved Process Capability Analysis using 6σ tools. Proceedings of the Industrial Statistical in Action 2000, International Conference, paper number 63. University of Newcastle Conference.

Burr, Irving W. (1967). The Effect of Non-Normality on Constants for $\bar{x}$ and R Charts. Industrial Quality Control, Vol. 23, pp:563-569.

Burroughs, T. E.; Rigdon, S. E.; and Champ, C. W. (1993). Analysis of Shewhart Charts with Runs Rules when No Standards are Given. 1993 Proceedings of the Section on Quality and Productivity. American Statistical Association, Alexandria, VA. pp. 16-19.

Carlyle, W. Matthew, Montgomery, Douglas C. and Runger, George C. (2000). Optimisation Problem and Methods in Quality Control and Improvement. Journal of Quality Technology. Vol. 32, No. 1, pp: 1-18.

Cassie, R. M. (1954). Some Uses of Probability Paper in the Analysis of Size Frequency Distributions. Austral. J. of Marine and Freshwater Res., 5, pp.513-522.

Chan, Lai K., Cheng, Simley W. and Spiring, Frederick A. (1988). A New Measure of Process Capability: C_{pm}. Journal of Quality Technology, Vol. 20, pp. 162-175.

Chang, P., & Lu, K. (1994). PCI Calculations for Any Shape of Distribution with Percentile. *QWTS, pp.*110-114.

Chen, Wen-Hsien. (1996). The Effects of SPC on the target of Process Quality Improvement. Journal of Quality Technology. Vol. 28, No. 2, pp: 224-232.

Chou, Youn-Min, Owen, D. B. and Borrego, A. Salvador A. (1990). Lower Confidence Limits on Process Capability Indices. Journal of Quality Technology, Vol. 22, pp. 223-229.

Chou, Youn-Min, Polansky, Alan M. and Mason, Robert L. (1998). Transforming Non-Normal Data to Normality in Statistical Process Control. Journal of Quality Technology, No.30, pp: 133-141.

Clark, V. A., Chapman, J. M., Coulson, A. H., and Hasselblad, V. (1968). Dividing the blood pressures from the Los Angeles heart study into two normal Distributions. John Hopkins. Medical Journal., 122, pp.77-83.

Clements, John A. (1989). Process Capability Calculations for Non-Normal Distributions. Quality Progress, Vol. 22, pp. 95-100.

Cox T. F., (2000). Multidimensional Scaling used in Multivariate Statistical Process Control. Industrial Statistics In Action 2000. International Conference. University of Newcastle Upon Tyne, UK.

Crowder, S. V. (1987). Computation of ARL for Combined Individual Measurement and Moving Range Charts. Journal of Quality Technology, Vol.19.

Crowder, S. V.(1997). A Simple Method for Studying Run-Length Distribution of Exponentially Weighted Moving Average. Technometrics 29, pp.398-399.

Davis, D. J. (1952). An Analysis of Some Failure Data. Journal of American Statistical Association, 47, pp.113-150.

Del Castillo, E. and Montgomery, D. C. (1994). Short Run Statistical Process Control: Q-Chart Enhancements and alternative Methods. Quality and Reliability Engineering International, Vol. 10.

Dodge, Y.(1987). Statistical Data Analysis Based on the L_1-Norm and Related Methods, North-Holland, Amsterdam.

DuBois, D., Stephenson, W.R., Gibberd, R., & Shaw, C. (1991). Control Charts for Hospital Length of Stay Data. *Proceedings of the Section on Statistical Education. American Statistical Association, Alexandria VA, pp.*120-125.

Dudewicz, E. J., Ramberg, J. S., & Tadikamalla., P. R. (1974). A Distribution for Data fitting and Simulation. Annual Technical Conference Transactions of the American Society for Quality Control, 28, pp.402-418.

Everitt, B, S. and Hand, D. J. (1981). Finite Mixture Distribution. Chapman and Hall, London.

Faltin, Frederick W.; Mastrangelo, Christina M.; Runger, George C.; Ryan, Thomas P. (1997). Considerations in the Monitoring of Autocorrelated and Independent Data. Journal of Quality Technology, Vol. 29, pp:131-133.

Fosam, E. B. and Kanji, G, K. (1997). Quality Control Charts for Mixture Distributions. Bulletin of the International Statistical Institute (ISI). Session 51, Book1, pp:289-290.

Franklin, L. A. and Wasserman, G. S. (1992b). Bootstrap Lower Confidence Limits for Capability Indices. Journal of Quality Technology 24, pp. 196-210.

Gan, F. F. (1996). Average Run Length for Cumulative Sum Control Charts Under Linear Trend. Appl. Statist. 45, No.4,pp. 505-526.

George, J. (1969). An Algorithm for the Decomposition of a Distribution into Gaussian Components. Biometrics, 25, pp.79-93.

Ghosh, B. K.; Reynolds, M. R. JR.; and Van Hui, y. (1981). Shewhart $\bar{x}$ charts Estimated Variance. Communications in Statistics- Theory and Methods 18, pp.1797-1822.

Gilchrist, W. (1997). Modelling with quantile distribution functions. Journal of Applied Statistics, Vol. 24, No.1, pp:113-122.

Gilchrist, W. G. (1993). Modelling Capability. Journal of Operation Research Society, No. 44, pp:909-923.

Gilchrist, W. G. (1995). Capability and Non-Normality. Total Quality Management, Proceedings of the First World Congress, pp:395-398. Edited By Gopal K. Kanji. Chapman & Hall.

Gilchrist, W. G. (2000). Capability, entry in Kotz, S. and Johnson, N. L. Encyclopaedia of Statistical Sciences. Updating Volume.

Gnanadesikan, R. and Kettenring, J. R. (1972). Robust Estimates, Residuals, and Outlier Detection with Multiresponse Data. Biometrics 28, pp. 81-124.

Grimshaw, S.D., & Alt, F.B. (1997). Control charts for quantile function values. *Journal of Quality Technology, 29, pp.* 1-7.

Gruska, G. F., Mirkhani, L. and Lamberson, L. R. (1989). Non Normal Analysis (St. Clair Shores, MI: Applied Computer Solutions, Inc.).

Gunter, B. (1989). The Use and Abuse of C_{pk}. Quality Progress, 22(1), pp.72-73.

Gunter, B.(1989). The Use and Abuse of C_{pk}, Part 2. Quality Progress, 22(3), pp.108-109.

Gunter, B.(1989). The Use and Abuse of C_{pk}, Part 3. Quality Progress, 22(5), pp.79-80.

Gunter, B.(1989). The Use and Abuse of C_{pk}, Part 4. Quality Progress, 22(7), pp.86-87.

Herman, John T. (1989). Capability Index Enough for Process Industries?, Proceedings, ASQC 43rd, AQC.

Hotelling, H. (1947). Multivariate Quality Control. Techniques of Statistical Analysis, Eisenhart, Hastay, and Wallis, Eds. McGraw-Hill York.

Hsiang, T. C. and Taguchi, G. (1985). A tutorial on quality control and assurance - The Taguchi methods. ASA Annual Meeting. Las Vegas, Nevada.

Hunter, J. stuart. (1999) Statistical Process Control. A glimpse at its past and future. Quality Progress December.

Hutson, Alan. D. (2000). A Composite Quantile Function Estimator with Applications in Bootstrapping. Journal of Applied Statistics, Vol. 27, No. 5, pp:567-577.

Iglewicz, B. and Hoaglin, D., (1987). Use of Box Plots for Process Evaluation. Journal of Quality Technology. 19; pp.180-190.

Jackson, J. E. (1980). Principal Components and Factors Analysis: Part I- Principal Components. Journal of Quality Technology, No.12, pp: 201-213.

Jackson, J. E. (1981a). Principal Components and Factors Analysis: Part II- Additional Topics Related to Principal Components. Journal of Quality Technology, No.13, pp: 46-58.

Jackson, J. E. (1981b). Principal Components and Factors Analysis: Part III- What is Factor Analysis? Journal of Quality Technology, No.13, pp: 125-130.

Jackson, J. E. (1985). Multivariate Quality Control. Communications in Statistics, No.14, pp:2657-2688.

Janacek, G. J, and Meikle, S. E. (1997). Control Chart Based on Medians. Statistician, 46, pp:19-31.

John, Peter W. M..(1990). Statistical Methods in Engineering and Quality Assurance. John Wiley &Son, Inc.

Johnson, Richard A. and Wichern, Dean W. (1998). Applied Multivariate Statistical analysis. Prentice Hall, Fourth edition.

Joiner, B.L., & Rosenblatt, J.R. (1971). Some Properties of the Range in Samples from Tukey's Symmetric Lambda Distributions. *Journal of the American Statistical Association, 66, pp.*394-399.

Jones, P. N. and McLachlan, G. J. (1990). Laplace-Normal Mixtures Fitted to Wind Shear Data. Journal of Applied Statistics, Vol. 17, No. 2, pp: 271-276.

Juran, J. M. (1974). Quality Control Handbook, third edition. McGraw-Hill, New York, NY.

Juran, J. M. and Gryna, Frank M. (1993). Quality Planning and Analysis. 3rd Edition, McGraw-Hill, Inc.

Juran, J. M., Gryna, F. M. and Bingham, R. S. (1979). Quality Control Handbook. McGraw-Hill, New York, New York.

Kackar, Raghu. (1985). " Off-line Quality Control, Parameter Design, and the Taguchi Method", Journal of Quality Technology, Vol. 17, No.4, pp. 176-188.

Kane, Victor E. (1986). Process Capability Indices. Journal of Quality Technology, 18. No. 1, pp.41-52.

Kanji, G. K. (1985). A Mixture Model for Wind Shear Data. Journal of Applied Statistics, Vol. 12, No. 1, pp:49-58.

Kanji, G. K. (1995). Quality and statistics concepts. Total Quality Management, Proceedings of the First World Congress. Edited by Gopal L. Kanji. Chapman & Hall.

Kanji, G. K. and Arif, Osama Hasan. (2000). Median Rankit Control Chart by Quantile Approach. Journal of Applied Statistics, 27, No. 6, pp: 757-770.

Kao, J. H. K. (1959). A Graphical Estimation of Mixed Weibull Parameters in Life-Testing Electron tubes. Technometrics, 1, pp.389-407.

Kapoor, S. and Kanji, G. K. (1990). Application of the Characterisation theory to the Mixture Model. Journal of Applied Statistics, Vol. 17, No. 2, pp:263-270.

Keats, J. Bert, Castillo, Enrique Del, Collani, Elart Von and Saniga, Erwin M. (1997). Journal of Quality Technology. Vol. 29, No. 2, pp: 144-147.

Kittlitz, R. G. (1999). Transforming the exponential for SPC applications. Journal of Quality Technology, 31, No. 3. pp: 301-308.

Kotz, Samuel and Johnson, Norman L. (1993). Process Capability Indices. Chapman & Hall.

Kotz, Samuel and Lovelace, Cynthia R. (1998). Process Capability Indices in Theory and Practice. Arnold.

Kourti, T. and MacGregor, J. F. (1996). Multivariate SPC Method for Process and Product Monitoring. Journal of Quality Technology 28, pp. 409-428.

Langenberg, P. and Iglewicz, B. (1986). Trimmed Mean $\bar{x}$ and R Charts. Journal of Quality Technology, 18; pp. 152-161.

Lucas, J. M. (1976). The design and Use of Cumulative Sum Quality Control Schemes. Journal of Quality Technology, Vol.8.

Lucas, J. M. and Saccucci, M. S. (1990). Exponential Weighted Moving Average Control Charts: Properties and Enhancements (with discussion). Technometrics 32, pp.1-29.

Mason, Robert L., Champ. Charles W., Tracy. Nola D., Wierda. Siebrand J. and Young. John C. (1997). Assessment of Multivariate Process Control Techniques. Journal of Quality Technology. Vol. 29, No. 2, pp: 140-1143.

Mendenhall, W. and Harder, R. J. (1958). Estimation of Parameters of Mixed Exponentially Distributed Failure Time Distributions From Censored life test Data. Biometrika, 45, pp.504-520.

Montgomery, D. C. (1997). Introduction to Statistical Quality Control. Third edition. John Wiley & Son, New York, NY.

Montgomery, D. C., Gardiner, J. S., & Pizzano, B. A. (1987). Statistical Process Control Methods for Detecting Small Process Shift, in Frontiers in Statistical Process Control, Vol. 3, H.-J. Lenz, G.B. Wetherill, and P.-Th. Wilrich, editors, Physica-Verlag, Heidelberg, West Germany, pp 163-170.

Montgomery, D. C., Keats, J. B., Runger, G. C. and Messina, W.S. (1994). Integrating Statistical Process Control and Engineering Process Control. Journal of Quality Technology, Vol. 26.

Montgomery, D. C., Keats, J. Bert, Runger, George C. and Messina, William S. (1994). Integrating Statistical Process Control and Engineering Process Control. Journal of Quality Technology. Vol. 26, No. 2, pp: 79-87.

Montgomery, Douglas C. and Woodall, William H. (1997). A Discussion on Statistically-Based Process Monitoring and Control. Vol. 29, No. 2, pp. 121.

Moore, P. G. (1958). Some Properties of Runs in Quality Procedures. Biometrika, Vol. XLV, pp: 89-95.

Nelson, L. S. (1983). Transformation For Attribute Data. Journal of Quality Technology 15, pp.55-56.

Nelson, L. S. (1984). The Shewhart Control Chart- Tests for Special Causes. Journal of Quality Technology 16, pp.237-239.

Nelson, L. S. (1989). Standardisation of Shewhart Control Charts. Journal of Quality Technology 21, pp.287-289.

Nelson, P. R. (1979). Control Charts for Weibull Process with Standard Given. IEEE Transactions on Reliability, Vol.R-28, No. 4, pp: 283-288.

Nelson, W. (1985). Weibull Analysis of Reliability Data with Few or No Failures. Journal of Quality Technology, 17, No. 3. pp: 140-146.

NG, C. H. and Case, K. E. (1992). Control Limits and the ARL: Some Surprises. First Industrial Engineering Research Conference Proceedings. Institute of Industrial Engineers, Norcross, GA. pp. 127-129.

Padgett, W. J., Spurrier, John D. (1990). Shewhart-Type Charts for Percentiles of Strength Distributions. Journal of Quality Technology, 22, No.4. pp:283-288.

Palar M and Wesolowsky GO (1999). Specification limits, Capability Indices, and Process Centering in Assembly Manufacture. Journal of Quality Technology. Vol. 31, No. 3, pp. 317-325.

Palm, Andrew C., Rodriguez, Robert N., Spiring, Fred A. and Wheeler, Donald J. (1997). Some Perspective and Challenges for Control Chart Methods. Journal of Quality Technology. Vol. 29, No. 2, pp: 122-127.

Parzen, E. (1979). Non-parametric Statistical Data Modelling. *Journal of the American Statistical Association, 74, pp.*105-122.

Patel, B. M. (1993). Quality and Productivity Improvement Through Effective Management of Control Charts and Analysis of Statistical Data. Quality Forum, Vol. 19, No.2.

Pearn, W. L and Chen, K. S. (1997). Capability Indices for Non-Normal Distributions With an Application in Electrolytic Capacitor Manufacturing. Microelectronic. Reliab, Vol. 37, No. 12, pp. 1853-1858.

Pearn, W. L. and Chen, K. S. (1995). Estimating Process Capability Indices for Non-Normal Pearsonian Populations. Quality and Reliability Engineering International. Vol. 11. pp. 386-388.

Pearn, W. L. and Chen, K.S. (1998). A generalisation of the process capability index C_{pk} for asymmetric tolerances. Journal of Applied Statistics, Vol.25, No. 6. pp. 801-10.

Pearn, W. L. and Kotz, S. (1994). Application of Clement's method for calculating second and third generation process capability indices for non-normal Pearsonian populations. Quality Engineering, Vol. 7, No. 1, pp. 139-145.

Pearn, W. L., Chen, K. S. and Lin, G. H. (1999). A generalisation of Clement's method for non-normal Pearsonian processes with asymmetric tolerances. International Journal of Quality & Reliability Management, Vol. 16, No. 5.

Pearn, W. L., Kotz, S. and Johnson, N. L. (1992). Distributional and Inferential Properties of Process Capability Indices. Journal of Quality Technology, Vol. 24, pp. 216-231.

Pearn. W. L., Lin, B. C. and Chen, K. S. (1998).Generalisations of the process capability index C_{pm} for asymmetric tolerances. Working paper, National Chiao Tung University.

Quesenberry, C. P. (1991a). SPC Q Charts for Start-Up Processes and Short or Long Runs. Journal of Quality Technology 23, pp 213-224.

Quesenberry, C. P. (1991b). SPC Q Charts for a Binomial Parameter p: Short or Long Runs. Journal of Quality Technology 23, pp. 239-246.

Quesenberry, C. P. (1991c). SPC Q Charts for a Poisson Parameter λ: Short or Long Runs. Journal of Quality Technology 23, pp.296-303.

Quesenberry, C. P. (1993). The Effect of Sample Size on Estimated Limits for $\bar{x}$ and X Control Chart. Journal of Quality Technology 25, pp. 237-247.

Quesenberry, C. P. (1995a). On Properties of Binomial Q Charts for Attributes. Journal of Quality Technology 27, pp. 204-213.

Quesenberry, C. P. (1995b). On Properties of Poisson Q Charts for Attributes. Journal of Quality Technology 27, pp.293-303.

Quesenberry, C. P. (1995c). Geometric Q Charts for High Quality Processes. Journal of Quality Technology 27, pp.304-315.

Quesenberry, C.P. (1995). On Properties of Q-Charts for Variables. *Journal of Quality Technology, 27, pp.*184-203.

Ramberg, J.S., Dudewicz, E.J., Tadikamalla, P.R., & Mykytka, E.F. (1979). A Probability Distribution and Its Uses in Fitting Data. *Technometrics, 21, pp.*201-209.

Ramberg, John. S. and Schmeiser, Bruce W. (1972). An Approximate Method for Generating Symmetric Random Variable. *Communications of the ACM, 15, pp.* 78-82.

Ramberg, John. S., & Schmeiser, Bruce W. (1974). An Approximate Method for Generating Asymmetric Random Variable. *Communications of the ACM, 17, pp.*78-82.

Rigdon, Steven E, Cruthis, Emma N. and Champ, Charles W. (1994). Design Strategies for Individuals and Moving Range Control Charts. Journal of Quality Technology. Vol. 26, No. 4, pp: 274-287.

Robert, S. W. (1966). A Comparison of Some Control Chart Procedures. Technometrics 8, pp. 411-430.

Rodriguez, Robert N. (1992). Recent Developments in Process Capability Analysis. Journal of Quality Technology, Vol. 24, pp. 176-187.

Ryan, T. P. (1989). Statistical Methods for Quality Improvement. John Wiley & Son, New York.

Ryan, T. P. (2000). Discussion. Journal of Quality Technology. Vol. 32, No. 4, pp: 367-369.

Ryan, Thomas P.; Schwertman, Neil C. (1997). Attributes control charts; Optimal control limits. Journal of Quality Technology, Vol. 29, No. 1, pp. 86-98.

Ryan, T. P. and Howley, P.P (1999). Correspondence: Control Charts Based on Medians. The Statistician, 48, pp: 441-442.

Sakia, R. M. (1992). The Box-Cox transformation technique: a review. The Statistician,41, pp.169-178.

Scallan, A. J. (1992). Maximum Likelihood Estimation for a Normal / Laplace Mixture Distribution. The Statistician, 41, pp:227-231.

Schilling, Edward G. & Nelson, Peter R. (1976). The Effect of Non-Normal on the Control Limits of $\bar{x}$ Charts. *Journal of Quality Technology, 8, pp.*183-188.

Seppala, Tomi, Moskowitz, Herbert, Plante, Robert and Tang, Jen. (1995) Statistical Process Control via the Subgroup Bootstrap. Journal of Quality Technology. Vol. 27, No. 2, pp: 139-153.

Shanbhag, D. N. (1972). Diagonality of Bhattacharyya Matrix as a Characterisation. Theory prob. Appl. No.24, pp:430-433.

Shanbhag, D. N. (1972). Some Characterisations based on the Bhattacharyya matrix. Journal of Applied Probability. No.9, pp:580-587.

Shapiro, S. S., & Gross, A.J. (1981). *Statistical Modelling Techniques.* Marcel Dekker, New York.

Shewhart, W. A. (1931). Economic Control of Quality of Manufactured Product, D. van Nostrand Co., New York.

Somerville, S. E. and Montgomery, D. C. (1996). Process Capability Indices and Non-Normal Distributions. Quality Engineering, Vol. 9, No. 2, pp. 305-316.

Spedding, T. A. and Rawlings, P. L. (1994) Non-normality in Statistical Process Control Measurements. International Journal of Quality & Reliability Management, Vol. 11, No 6.

Spiring, Fred A. (1991). The C_{pm} Index. Quality Progress. February, pp.57-61.

Spiring, Fred A.(1991). Assessing Process Capability in the Presence of Systematic Assignable Cause. Journal of Quality Technology, Vol. 23, pp. 125-134.

Sullivan, Joe, H., and Woodall, William, H. (1996). A Comparison of Multivariate Control Charts for Individual Observations. Journal of Quality Technology, 28, pp: 398-408.

Sullivan, Joe, H., and Woodall, William, H. (1996). A Control Chart for Preliminary Analysis of Individual Observations. Journal of Quality Technology, 28, pp: 265-278.

Taguchi, G. (1978a). off-line and on-line quality control system. pp.B4-1 to B4-5, ICQC 1978, Tokyo.

Taguchi, G., Elsayed, E. and Hsiang, T. (1989). Quality Engineering in Production System, Mcgraw Hill, New York, N. Y.

Tarcy, Nola D., Young, John, C. and Mason, Robert, L.(1992). Multivariate Control Charts for Individual Observations. Journal of Quality Technology. Vol. 24, No. 2, pp: 88-95.

Tatum, Lawrence. G., (1996). Control Charts for the Detection of a Periodic Component. Technometrics, May, Vol.38, No.2.

Thomas, E. A. C. (1966). Mathematical models for the clustered firing of single cortical neurones. Brit. J. Math. Stat. Psychol., 19, pp.151-162.

Titterington, D. M., Smith, A. F. M. and Makov, U. E. (1985). Statistical Analysis of Finite Mixture Distributions. John Wiley & Son.

Tukey, J. W. (1957). The Comparative anatomy of transformation. Annals of Mathematical Statistics, 28, 602-632.

Tukey, J.W. (1960). The practical Relationship between the Common Transformations of Percentages of Counts and of Amounts. Technical Report 36, Statistical Techniques Research Group. Princeton University.

Vannman, K.(1995). A unified approach to capability indices. Statistica Sinica 5, pp. 805-820.

Vannman, K.(1997). A general class of capability indices in the case of asymmetric tolerance. Communications in statistics-Theory and Methods, Vol.26, No.8. pp: 2049-72.

Wadsworth, H. M. (1998). Handbook of Statistical Methods for Engineers and Scientists, second edition. MCGRAW-HILL.

Wallgren, E. (1996). Properties of the Taguchi capability index for Markov dependent quality characteristics. University of *Örebro*, Sweden.

Wheeler, Donald. J. (2000). Discussion. Journal of Quality Technology. Vol. 32, No. 4, pp: 361-363.

White, E. M. and Shroeder, R., (1987). A Simultaneous Control Chart. Journal of Quality Technology, 19; pp. 1-10.

Wood, Michael. (1995). Three suggestions for improving control charting procedures. International Journal of Quality & Reliability Management. Vol.12, No.5. pp:61-74.

Woodall, William. H. (2000). Controversies and Contradictions in Statistical Process Control. Journal of Quality Technology. Vol. 32, No. 4, pp: 341-350.

Woodall, William. H. and Montgomery, Douglas, C. (1999). Research Issues and Ideas in Statistical Process Control. Journal of Quality Technology. Vol. 31, No. 4, pp: 376-386.

Yourstone, S. A. and Zimmer, W. J. (1992). Nonnormality and the Design of Control Charts for Averages. Decision Sciences, No. 23, pp. 1099-1113.

Zhang, Nien Fan. (1998) A Statistical Control Chart for Stationary Process Data. Technometrics. Vol. 40, No.1. pp:24-38.

Appendix

Appendix 1

Glossary of Terms

Off-line quality control: Quality control applied before production, at the product-development stage, or during installation of a process.

On-line quality control: Quality control applied during full production, for instance process capability index, SPC, control chart, reliability studies, cause and effect diagram are all known as on-line quality control methods.

SQC: The quality of a product is maintained by ensuring that the process is operating properly.

SPC: Work under the SQC assumption that if a process is operating properly, it will produce consistently high quality products.

Appendix 2

Relationship among the four indices C_p, C_{pk}, C_{pm} and C_{pmk}

- If $C_p = C_{pk}$, the process is centred at the midpoint of the specifications.
- If $C_p < C_{pk}$, the process is off-centre.
- C_p measures potential capability in the process.
- C_{pk} Measures actual capability.
- If the process mean is exactly equal to one of the specification limits, leading to C_{pk}=0.

- C_{pk} <0, the implication is that the process mean lies outside the specification.
- Some authors define C_{pk} as non negative, so that values less than zero are defined as zero.
- $C_{pk} = \left(1 - \frac{|\mu - M|}{d}\right) C_p$ where $d = \frac{USL - LSL}{2}$
- $C_{Pm} = C_{pk} = C_p$ when $\mu = T$
- $C_{pmk} \leq C_{pk}$
- $C_{pmk} = C_{pk}$, if the process is on target ($\mu = T$)
- $C_{pmk} < C_{pk}$, if the process is not on target ($\mu \neq T$)
- C_{pmk} and C_{pm} are related in the same thing as C_p and C_{pk},

  $C_{pmk} = (1-k)C_{pm}$

There is a unified relationship between C_{pk} and C_{pm} for fixed values of C_p. The indices C_{pm} and C_p will be identical when the process mean and the target value coincide. This implies that $\hat{C}_{pm}$ and $\hat{C}_p$ estimate the same value when $T = \mu$, (see, Chan *et al.* (1988) for more details).

In general, the relationships among the four indices can be established as the following: $C_{pm} = C_P\{1+[(\mu - T)/\sigma]^2\}^{-1/2}$, and $C_{pmk} = C_{pk}\{1+[(\mu - T)/\sigma]^2\}^{-1/2}$.

The four most sensitive indices to the departure of the process mean from the target value, from the upper sensitive to the lower sensitive, are as follows C_{pmk} , C_{pm} , C_{pk} and C_p.

For the symmetric tolerances, $C_{pk} = (1-K)C_p$ and $C_{pmk} = (1-K)C_{pm}$, where $K = |\mu - T|/d$ is the departure ratio. If the process is on target, then K=0 and $(\mu = T)$ then $C_p = C_{pk} = C_{pm} = C_{pmk} = d/3\sigma = (USL - LSL)/6\sigma$.

Appendix 3

When the centre target does not fall on the midpoint of specification interval, i.e. $T \neq \frac{USL + LSL}{2}$. Vannman superstructure provides the four indices in the general form :

$$C_p(u,v) = (1-u)\frac{\min\{USL-T, T-LSL\}}{3\sqrt{\{[(U_p - L_p)/6]^2 + v(M-T)^2\}}} + u * \min\left\{\frac{(USL-T)-|M-T|}{3\sqrt{\{[(U_p - M)/3]^2 + v(M-T)^2\}}}, \frac{(T-LSL)-|M-T|}{3\sqrt{\{[(M-L_p)/3]^2 + v(M-T)^2\}}}\right\} \qquad 6.25$$

The four indices are obtained by setting $C_p(0,0) = C_p$, $C_p(1,0) = C_{pk}$, $C_p(0,1) = C_{pm}$ and $C_p(1,1) = C_{pmk}$ which are:

$$C_p = \frac{\min\{USL - T, T - LSL\}}{(U_p - L_p)/2} \qquad 6.26$$

$$C_{pk} = \min\left\{\frac{(USL - T) - |M - T|}{U_p - M}, \frac{(T - LSL) - |M - T|}{M - L_p}\right\} \qquad 6.27$$

$$C_{pm} = \frac{\min\{USL - T, T - LSL\}}{3\sqrt{\{[(U_p - L_p)/6]^2 + (M - T)^2\}}} \qquad 6.28$$

$$C_{pmk} = \min\left\{\frac{(USL-T)-|M-T|}{3\sqrt{\{[(U_p - M)/3]^2 + (M-T)^2\}}}, \frac{(T-LSL)-|M-T|}{3\sqrt{\{[(M-L_p)/3]^2 + (M-T)^2}}\right\} \qquad 6.29$$

Appendix 4

Modified original Clement's method by Pearn and Chen (1995) at $T = \frac{USL + LSL}{2}$

$$C_p = \left[\frac{USL - LSL}{Up - Lp} \right] \quad 6.30$$

$$C_{pk} = \left[\frac{\min(USL - M, M - LSL)}{(U_p - L_p)/2} \right] \quad 6.31$$

$$C_{pm} = \left[\frac{USL - LSL}{6\sqrt{\left(\frac{Up - Lp}{6}\right)^2 + (M - T)^2}} \right] \quad 6.32$$

$$C_{pmk} = \left[\frac{\min(USL - M, M - LSL)}{3\sqrt{\left(\frac{Up - L_p}{6}\right)^2 + (M - T)^2}} \right] \quad 6.33$$

Appendix 5

Generalization Clement's Method through Pearsonian curve(symmetric and asymmetric tolerances i.e. when T is equal M and when T is not equal M, respectively).

```
MTB > #Pearn et al. (1999).
MTB > # c1=USL
```

```
MTB > # c2=LSL
MTB > # c3=T
MTB > # c4
MTB > # c5=Up
MTB > # c6=Lp
MTB > # c7=M or X0.5 (medium)


MTB > let c10=c1-c3
MTB > let c11=c3-c2
MTB > let c12=((c1-c2)/2)
MTB > stack c10 c11 c13
MTB > minimam c13 c14


Minimum of stack (c10,c11) = 0.80000


MTB > let c18=((c12)*(c7-c3))/c10
MTB > let c19=((c12)*(c3-c7))/c11
MTB > stack c18 c19 c20
MTB > maximam c20 c21
Maximum of C20 = 0.34064
MTB > # if T=M then a=|M-T|, Go back to modified Clement's
method (Pearn &Chen (1995).


MTB > let c22=c21**2


MTB > Let c25=c1-c7
MTB > let c26=c7-c2
MTB > let c27= c14/c10
MTB > let c28=c14/c11
MTB > let c30=c5-c6
MTB > let c31=c30/2
```

```
MTB > let c32=c30/6
MTB > let c34=c32**2

MTB > let c36=(c34+c22)**(0.5)


MTB > let c38=(2*c12)/c30
MTB > let c40=(c25/c31)*(c27)
MTB > let c41=(c26/c31)*(c28)
MTB > stack c40 c41 c42
MTB > minimam c42 c43


Minimum of C42 = 0.55416


MTB > let c45=(2*c14)/(6*c36)


MTB > let c47=(c25/(3*c36))*c27
MTB > let c48=(c26/(3*c36))*c28
MTB > stack c47 c48 c49
MTB > minimam c49 c50


Minimum of C49 = 0.37210

MTB > stack c38 c43 c45 c50 c52
```

Appendix 6

Generalization Clement's Method through Quantile approach (symmetric and asymmetric tolerances i.e. when T is equal M and when T is not equal M, respectively), Pearn et al. (1999).

```
MTB > # c1=USL
MTB > # c2=LSL
MTB > # c3=T
MTB > # c4
MTB > # c5=Q(0.99865)
MTB > # c6=Q(0.00135)
MTB > # c7=Q(0.5)


MTB > let c10=c1-c3
MTB > let c11=c3-c2
MTB > let c12=((c1-c2)/2)
MTB > stack c10 c11 c13
MTB > minimam c13 c14


Minimum of stack (c10,c11) = 0.95000


MTB > let c18=((c12)*(c7-c3))/c10
MTB > let c19=((c12)*(c3-c7))/c11
MTB > stack c18 c19 c20
MTB > maximam c20 c21


Maximum of C20 = 0.26598
```

```
MTB > # if T=M then a=|M-T|, Go back to modified clement
method (Pearn &Chen (1995)).


MTB > let c22=c21**2

MTB > Let c25=c1-c7
MTB > let c26=c7-c2
MTB > let c27= c14/c10
MTB > let c28=c14/c11

MTB > let c30=c5-c6
MTB > let c31=c30/2
MTB > let c32=c30/6
MTB > let c34=c32**2


MTB > let c36=(c34+c22)**(0.5)

MTB > let c38=(2*c12)/c30

MTB > let c40=(c25/c31)*(c27)
MTB > let c41=(c26/c31)*(c28)
MTB > stack c40 c41 c42
MTB > minimam c42 c43
Minimum of C42 = 0.88013

MTB > let c45=(2*c14)/(6*c36)

MTB > let c47=(c25/(3*c36))*c27
MTB > let c48=(c26/(3*c36))*c28
MTB > stack c47 c48 c49
MTB > minimam c49 c50
Minimum of C49 = 0.61409
MTB > stack c38 c43 c45 c50 c52
```

Appendix 7

ARL for Extreme-value

Quantile function for extreme-value is

$$R(p) = -\ln(-\ln p)$$

By using this formula, the probability of $p_{(L)}$, i.e. lower limit, is calculated.

$$R(p_{(L)}) = -\ln(-\ln p_{(L)})$$

$$\frac{1}{k}R(p_{(L)}) = -\frac{1}{k}\ln(-\ln p_{(L)})$$

$$= -\ln(-\ln p_{(L)})^{\frac{1}{k}}$$

Let

$$-\ln p^{*}_{(L)} = (-\ln p_{(L)})^{\frac{1}{k}}$$

then

$$= -\ln(-\ln p^{*}_{(L)})$$

Where

$$p^{*}_{(L)} = \exp(-(-\ln p_{(L)})^{\frac{1}{k}})$$

$$F\left(\frac{1}{k}R(p_{(L)})\right) = F(-\ln(-\ln p^{*}_{(L)})$$

$$= F(R(p^{*}_{(L)}))$$

$$= p^{*}_{(L)}$$

On the other hand, the probability of $p_{(U)}$, i.e. upper limit, is calculated.

$$R(p_{(U)}) = -\ln(-\ln p_{(U)})$$

$$\frac{1}{k}R(p_{(U)}) = -\frac{1}{k}\ln(-\ln p_{(U)})$$

$$= -\ln(-\ln p_{(U)})^{\frac{1}{k}}$$

Let

$$-\ln p^{*}_{(U)} = (-\ln p_{(U)})^{\frac{1}{k}}$$

then

$$= -\ln(-\ln p^{*}_{(U)})$$

Where

$$p^*_{(U)} = \exp(-(-\ln p_{(U)})^{\frac{1}{k}})$$

$$F\left(\frac{1}{k}R(p_{(U)})\right) = F(-\ln(-\ln p^*_{(U)})$$

$$= F(R(p^*_{(U)}))$$

$$= p^*_{(U)}$$

At $p_U = 1 - p_L$

$$p^*_{(U)} = \exp(-(-\ln(1 - p_{(L)}))^{\frac{1}{k}})$$

then

$$ARL = \frac{1}{\left(1 - p^*_{(U)} + p^*_{(L)}\right)}$$

ARL for Pareto Distribution

Quantile function for pareto distribution is

$$R(p) = \frac{1}{(1-p)^{\beta}}$$

By using this formula, the probability of $p_{(L)}$, i.e. lower limit, is calculated.

$$R(p_{(L)}) = \frac{1}{(1 - p_{(L)})^{\beta}} = (1 - p_{(L)})^{-\beta}$$

$$\frac{1}{k}R(p_{(L)}) = \frac{1}{k}(1 - p_{(L)})^{-\beta}$$

Let

$$(1 - p^*_{(L)})^{-\beta} = \frac{1}{k}(1 - p_{(L)})^{-\beta}$$

$$(1 - p^*_{(L)}) = \left(\frac{1}{k}\right)^{\beta}(1 - p_{(L)})$$

then

$$p^*_{(L)} = 1 - \left(\frac{1}{k}\right)^{\beta}(1 - p_{(L)})$$

$$F\left(\frac{1}{k}R(p_{(L)})\right) = F((1 - p^*_{(L)})^{-\beta})$$

$$= F(R(p^*_{(L)}))$$
$$= p^*_{(L)}$$

On the other hand, the probability of $p_{(U)}$, i.e. upper limit, is calculated.

$$R(p_{(U)}) = \frac{1}{(1-p_{(U)})^{\beta}} = (1-p_{(U)})^{-\beta}$$
$$\frac{1}{k}R(p_{(U)}) = \frac{1}{k}(1-p_{(U)})^{-\beta}$$

Let

$$(1-p^*_{(U)})^{-\beta} = \frac{1}{k}(1-p_{(U)})^{-\beta}$$
$$(1-p^*_{(U)}) = \left(\frac{1}{k}\right)^{\beta}(1-p_{(U)})$$

then

$$p^*_{(U)} = 1-\left(\frac{1}{k}\right)^{\beta}(1-p_{(U)})$$
$$F\left(\frac{1}{k}R(p_{(U)})\right) = F((1-p^*_{(U)})^{-\beta})$$
$$= F(R(p^*_{(U)}))$$
$$= p^*_{(U)}$$

At $p_U = 1 - p_L$
then

$$P^*_{(U)} = 1-\left(\frac{1}{k}\right)^{\beta}P_{(L)}$$

$$ARL = \frac{1}{\left(1-p^*_{(U)}+p^*_{(L)}\right)}$$

ARL for Power Distribution

Quantile function for power distribution is

$$R(p) = p^{\beta}$$

By using this formula, the probability of $p_{(L)}$, i.e. lower limit, is calculated.

$$R(p_{(L)}) = p_{(L)}^{\beta}$$

$$\frac{1}{k}R(p_{(L)}) = \frac{1}{k}p_{(L)}^{\beta}$$

Let

$$p_{(L)}^{*\ \beta} = \frac{1}{k}p_{(L)}^{\beta}$$

then

$$p_{(L)}^{*} = \left(\frac{1}{k}\right)^{\frac{1}{\beta}} p_{(L)}$$

$$F\left(\frac{1}{k}R(p_{(L)})\right) = F(p_{(L)}^{*\ \beta})$$
$$= F(R(p_{(L)}^{*}))$$
$$= p_{(L)}^{*}$$

On the other hand, the probability of $p_{(U)}$, i.e. upper limit, is calculated.

$$R(p_{(U)}) = p_{(U)}^{\beta}$$

$$\frac{1}{k}R(p_{(U)}) = \frac{1}{k}p_{(U)}^{\beta}$$

Let

$$p_{(U)}^{*\ \beta} = \frac{1}{k}p_{(U)}^{\beta}$$

then

$$p_{(U)}^{*} = \left(\frac{1}{k}\right)^{\frac{1}{\beta}} p_{(U)}$$

$$F\left(\frac{1}{k}R(p_{(U)})\right) = F(p_{(U)}^{*\ \beta})$$
$$= F(R(p_{(U)}^{*}))$$
$$= p_{(U)}^{*}$$

At $p_U = 1 - p_L$
then

$$P_{(U)}^{*} = \left(\frac{1}{k}\right)^{\frac{1}{\beta}} (1 - P_{(L)})$$

$$ARL = \frac{1}{\left(1 - p_{(U)}^{*} + p_{(L)}^{*}\right)}$$

ARL for Weibull Distribution

Quantile function for Weibull distribution is

$$R(p)=(-\ln(1-p))^{\beta}$$

By using this formula, the probability of $p_{(L)}$, i.e. lower limit, is calculated.

$$R(p_{(L)})=(-\ln(1-p_{(L)}))^{\beta}$$

$$\frac{1}{k}R(p_{(L)})=\frac{1}{k}(-\ln(1-p_{(L)}))^{\beta}$$

Let

$$(-\ln(1-p_{(L)}^{*}))^{\beta}=\frac{1}{k}(-\ln(1-p_{(L)}))^{\beta}$$

then

$$p_{(L)}^{*}=1-\exp(-(\left(\frac{1}{k}\right)^{\frac{1}{\beta}}(-\ln(1-p_{(L)}))))$$

$$F(\left(\frac{1}{k}\right)R(p_{(L)}))=F((-\ln(1-p_{(L)}^{*}))^{\beta})$$

$$=F(R(p_{(L)}^{*}))$$

$$=p_{(L)}^{*}$$

On the other hand, the probability of $p_{(U)}$, i.e. upper limit, is calculated.

$$R(p_{(U)})=(-\ln(1-p_{(U)}))^{\beta}$$

$$\frac{1}{k}R(p_{(U)})=\frac{1}{k}(-\ln(1-p_{(U)}))^{\beta}$$

let

$$(-\ln(1-p_{(U)}^{*}))^{\beta}=\frac{1}{k}(-\ln(1-p_{(U)}))^{\beta}$$

then

$$p_{(U)}^{*}=1-\exp(-(\left(\frac{1}{k}\right)^{\frac{1}{\beta}}(-\ln(1-p_{(U)}))))$$

$$F(\left(\frac{1}{k}\right)R(p_{(U)}))=F((-\ln(1-p_{(U)}^{*}))^{\beta})$$

$$=F(R(p_{(U)}^{*}))$$

$$=p_{(U)}^{*}$$

At $p_U=1-p_L$

then

$$p^*_{(U)} = 1 - \exp(-(\left(\frac{1}{k}\right)^{\frac{1}{\beta}}(-\ln p_{(L)})))$$

$$ARL = \frac{1}{\left(1 - p^*_{(U)} + p^*_{(L)}\right)}$$

ARL for Logistic Distribution

Quantile function for logistic distribution is

$$R(p) = \ln\left(\frac{p}{1-p}\right)$$

By using this formula, the probability of $p_{(L)}$, i.e. lower limit, is calculated.

$$\frac{1}{k}R(p) = \frac{1}{k}\ln\left(\frac{p}{1-p}\right)$$

Let

$$\ln\left(\frac{p^*}{1-p^*}\right) = \frac{1}{k}\ln\left(\frac{p}{1-p}\right)$$

$$\ln\left(\frac{p^*}{1-p^*}\right) = \ln\left(\frac{p}{1-p}\right)^{\frac{1}{k}}$$

$$\left(\frac{p^*}{1-p^*}\right) = \left(\frac{p}{1-p}\right)^{\frac{1}{k}}$$

then

$$p^* = \frac{\left(\frac{p}{1-p}\right)^{\frac{1}{k}}}{1 + \left(\frac{p}{1-p}\right)^{\frac{1}{k}}}$$

$$= \ln\left(\frac{p^*}{1-p^*}\right)$$

$$
\begin{aligned}
F\left(\frac{1}{k}R(p)\right) &= F\left(\ln\left(\frac{p^*}{1-p^*}\right)\right) \\
&= F(R(p^*)) \\
&= p^*
\end{aligned}
$$

Logistic distribution is a symmetric, then

$$ARL = \frac{1}{2p^*}$$

Appendix 8

ARL for Extreme-value

$$R(p) = -\ln(-\ln p)$$

$$p^{*}_{(L)} = \exp(-(-\ln p_{(L)})^{\frac{1}{k}})$$

$$p^{*}_{(U)} = \exp(-(-\ln(1 - p_{(L)}))^{\frac{1}{k}})$$

$$ARL = \frac{1}{\left(1 - p^{*}_{(U)} + p^{*}_{(L)}\right)}$$

Case 1

```
# extreme
# p is fixed , k is variable
# k is generated from uniform (1,2)

# c1=p , c2=k

loge c1 c3
#c3=ln p

let c4=expo(-((-c3)**(1/c2)))
# c4=pl*

Let c5=1-c1
loge c5 c6
#c5=1-pl
#c6=ln(1-pl)

let c7=expo(-((-c6)**(1/c2)))
#c7=pu*

let c8=1/(1-c7+c4)
#c8=ARL
```

Case 2

```
# extreme
# k is fixed , p is variable
#Generate p from uniform (0.001,0.003)

# c1=p , c2=k

loge c1 c3
```

```
#c3=ln p

let c4=expo(-((-c3)**(1/c2)))
# c4=pl

Let c5=1-c1
loge c5 c6
#c5=1-pl
#c6=ln(1-pl)

let c7=expo(-((-c6)**(1/c2)))
#c7=pu

let c8=1/(1-c7+c4)
#c8=ARL
```

ARL for Logistic

R(p)=p/(1-p)

$$p^* = \frac{\left(\dfrac{p}{1-p}\right)^{\frac{1}{k}}}{1+\left(\dfrac{p}{1-p}\right)^{\frac{1}{k}}}$$

ARL=(1/(2* p^*))

Case 1

```
# Logistic
# p is fixed , k is variable
# Generate k from uniform(1,2)

# c1=k,c2=p

let c3=(((c2)/(1-c2))**(1/c1))/(1+(((c2)/(1-c2))**(1/c1)))

# c3=p*

Let c4=(1/(2*c3))
# ARL=c4
```

Case 2

```
# Logistic
# k is fixed, p is variable
# Generate p from uniform(0.001,0.003)

#c1=p,c2=k

let c3=(((c1)/(1-c1))**(1/c2))/(1+(((c1)/(1-c1))**(1/c2)))

#c3=p*

Let c4=(1/(2*c3))
# ARL=c4
```

ARL for Weibull

$$R(p) = (-\ln(1-p))^{\beta}$$

$$p^{*}_{(L)} = 1 - \exp(-(\left(\frac{1}{k}\right)^{\frac{1}{\beta}}(-\ln(1-p_{(L)}))))$$

$$p^{*}_{(U)} = 1 - \exp(-(\left(\frac{1}{k}\right)^{\frac{1}{\beta}}(-\ln p_{(L)})))$$

$$ARL = \frac{1}{\left(1 - p^{*}_{(U)} + p^{*}_{(L)}\right)}$$

Case 1

```
# p is fixed, k is variable
# Generate k from uniform(1,2)

# c1=k,c2=p, c3=beta

let c4=1-c2
loge c4 c5
loge c2 c6

# c5=ln (1-c2)
# c6=ln c2

let c7=1-expo(-(((1/c1)**(1/c3))*(-c5)))

# c7=pl*

let c8=1-expo(-(((1/c1)**(1/c3))*(-c6))
```

```
# c8=pu*

let c9=1/(1-c8+c7)

# c9=ARL
```

Case 2

```
#Weibull
#k is fixed, p is variable
#Generate p from uniform (0.001,0.003)

#c1=p,c2=k, c3=beta

let c4=1-c1
loge c4 c5
loge c1 c6

#c5=ln (1-c1)
#c6=ln c1

let c7=1-expo(-(((1/c2)**(1/c3))*(-c5)))

#c7=pl*

let c8=1-expo(-((1/c2)**(1/c3))*(-c6))

#c8=pu*

let c9=1/(1-c8+c7)

#c9=ARL
```

ARL for Pareto

$$R(p)=\frac{1}{(1-p)^{\beta}}$$

$$p^{*}_{(L)}=1-\left(\frac{1}{k}\right)^{\beta}(1-p_{(L)})$$

$$P^{*}_{(U)}=1-\left(\frac{1}{K}\right)^{\beta}P_{(L)}$$

$$ARL = \frac{1}{\left(1 - p^{*}_{(U)} + p^{*}_{(L)}\right)}$$

Case 1

#Pareto
#p is fixed, k is variable
#Generate k from uniform (1,2)

#c1=k,c2=p, c3=beta

let c4=(1-(((1/c1)**(c3))*(1-c2)))

#c4=pl*

let c5=(1-(((1/c1)**(c3))*(c2)))
#c5=pu*

let c6=1/(1-c5+c4)
#c6=ARL

Case 2

#Pareto
#k is fixed, p is variable
#Generate p from uniform (0.001,0.003)
#c1=p,c2=k, c3=beta

let c4=(1-(((1/c2)**(c3))*(1-c1)))

#c4=pl*

let c5=(1-(((1/c2)**(c3))*(c1)))

#c5=pu*

let c6=1/(1-c5+c4)
#c6=ARL

ARL for Power

$$R(p) = p^{\beta}$$

$$p^{*}_{(L)} = \left(\frac{1}{k}\right)^{\frac{1}{\beta}} p_{(L)}$$

$$P^{*}_{(U)} = \left(\frac{1}{K}\right)^{\frac{1}{\beta}} (1 - P_{(L)})$$

$$ARL = \frac{1}{\left(1 - p^{*}_{(U)} + p^{*}_{(L)}\right)}$$

Case 1

```
#Power
#p is fixed, k is variable
#Generate k from uniform(1,2)

#c1=k,c2=p, c3=beta

let c4=((1/c1)**(1/c3))*(c2)

#c4=pl*

let c5=(((1/c1)**(1/c3))*(1-c2))

#c5=pu*

let c6=1/(1-c5+c4)

# c6=ARL
```

Case 2

```
#Power
#k is fixed, p is variable
#Generate p from uniform (0.001,0.003)

#c1=p,c2=k, c3=beta

let c4=((1/c2)**(1/c3))*(c1)

#c4=pl*

let c5=(((1/c2)**(1/c3))*(1-c1))

#c5=pu*

let c6=1/(1-c5+c4)
 # c6=ARL
```

Appendix 9

The Newton method is a method for solving equation $f(\beta)$ where $f(\beta)$ is assumed to have a continuous derivative $f'(\beta)$. The underlying idea is that we approximate the graph of $f(\beta)$ by suitable tangents, using an approximate value β_0 obtained from the graph of $f(\beta)$, we let β_1 be the point of intersection of the x-axis and the tangent to the curve of $f(\beta)$ at β_0 (see figure below), then

$$\tan z = f'(\beta_0) = \frac{f(\beta_0)}{\beta_0 - \beta_1}$$

Hence,

$$\beta_1 = \beta_0 - \frac{f(\beta_0)}{f'(\beta_0)} \qquad \textbf{4. 64}$$

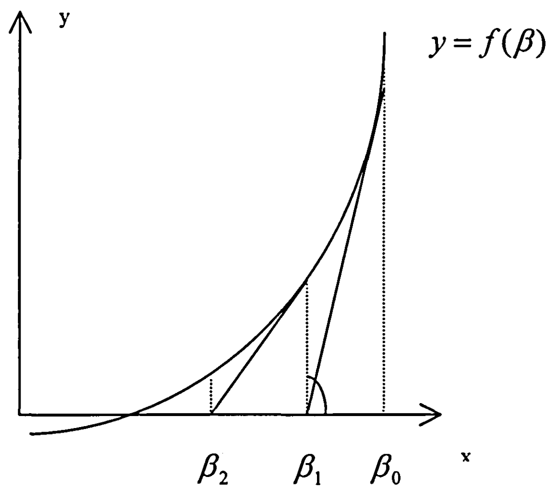
$$f'(\beta_0) = \theta_1 * (R - D) Exp(\theta_1 \beta) - \theta_2 * (R + D) Exp(-\theta_2 \beta)$$

and

$$f''(\beta_0) = \theta_1^2 * (R - D) Exp(\theta_1 \beta) + \theta_2^2 * (R + D) Exp(-\theta_2 \beta)$$

Where R>D and $f''(\beta_0)$ is positive, then the minimum value of $f(\beta)$ for distribution (Weibull and Power) is

$$\beta_{\min} = \left(\frac{1}{\theta_1 + \theta_2}\right) * \ln\left(\frac{\theta_2 (R + D)}{\theta_1 (R - D)}\right) \qquad \textbf{4. 65}$$



Newton Method

Then

- Choose a starting value β_0=(roughly value * $\beta_{\min}$)
- Calculate β_1 by using equation (4.64)
- Repeat steps above until convergence i.e. $\beta_{n+1} - \beta_n = 0$

When $\beta_{n+1} = \beta_n$ then this β is called $\beta^\wedge$ or the root of equation.

Hence, substituting $\beta^\wedge$ in the equations 4.35 and 4.36, then we can obtain the values of $\lambda^\wedge$ and $\eta^\wedge$.