QUALITY CONTROL AND IMPROVEMENT BASED ON GROUPED DATA

By

STEFAN HANS STEINER, B.MATH, M.SC.

A Thesis
Submitted to the School of Graduate Studies
in Partial Fulfillment of the Requirements
for the Degree
Doctor of Philosophy (Management Science/Systems)

McMaster University

QUALITY CONTROL AND IMPROVEMENT BASED ON GROUPED DATA

DOCTOR OF PHILOSOPHY (1994) McMASTER UNIVERSITY

(Management Science/Systems) Hamilton, Ontario

TITLE: Quality Control and Improvement Based on Grouped Data

AUTHOR:

Stefan Hans Steiner, B.Math (University of Waterloo)

M.Sc. (University of British Columbia)

SUPERVISOR: Professor G.O. Wesolowsky

NUMBER OF PAGES: ix, 165

Abstract

This thesis develops quality control and improvement techniques based on grouped data. Grouped data occur frequently in industry. However, in the past, most techniques have failed to directly take this grouping into account, and as a result do not perform well in many circumstances.

Two major areas of application are considered. First, acceptance sampling plans, acceptance control charts, and Shewhart control charts based on grouped data are developed. These forms of statistical process control have broad application and are in use widely. The design and implementation methodology is derived assuming either a normal or Weibull process, but is easily adapted to any other underlying distribution. A number of design approaches are presented and their relative advantages and disadvantages are discussed. The second application involves estimating the correlation between destructively measured strength properties. This problem arises in the area of structural design. To obtain an estimate of the correlation censoring of the strength data is required. The censoring or proof-testing results in grouped data. A number of simple estimation procedures are presented and compared.

Acknowledgments

My sincere appreciation goes to my supervisor, Dr. George O. Wesolowsky, for all his encouragement and support.

I would also like to thank the rest of my supervisory committee, Dr. P.D.M Macdonald, and Dr. R. Love, for their careful reading of the text and helpful comments along the way.

Thanks also go to Lee Geyer, a fellow graduate student, for helpful discussions at the beginning of this research, and to anonymous associate editors and referees whose comments and suggestions improved the research articles based on this work. Acknowledgment must also go to Dr. Balakrishnan who pointed out to me at my thesis defense the equivalence of the optimization problem discussed in Section 5.3 with some work in different field. See the beginning of Chapter 5 for a more detailed explanation.

Last, but not least, I thank my wife Anne Marie Mingiardi. Without her this would have seemed a much longer and lonelier journey.

Statement of Contribution

The work presented in this thesis is original. For the most part, the research was done without any outside assistant apart from general helpful comments and suggests from my supervisor. The only exceptions being some contribution from P. Lee Geyer, a fellow graduate student, for the material presented in Sections 3.1 and 3.3.2. The additional contribution for the aforementioned sections consisted of stimulating discussions and the suggestion for the use of the likelihood ratio test with specific alternative hypotheses.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1
Introduction

In the current world of continually increasing global competition it is imperative for all manufacturing and service organizations to improve the quality of their products. Quality has been defined in many ways (Evans and Lindsay, 1992). The American National Standards Institute and the American Society for Quality Control (1978) defined quality as "the totality of features and characteristics of a product or service that bears on its ability to satisfy given needs." The quality of a product or service has always been of interest to both the provider and the customer. In fact, as Duncan (1986) states in the first line of his book, "Quality Control is as old as industry itself." In the ages before the industrial revolution, good craftsmen and artisans learned quickly through intimate contact with their customers that quality products meant satisfied customers, and satisfied customers meant continued business. However, with the industrial revolution came the mass production of products by people who rarely interacted with customers. As a result, although costs decreased, the emphasis on quality also decreased. In addition, as the products made and the services provided became more complex, the need for a formal system to ensure the quality of the final product and all its components became increasingly important.

In modern firms, the quality of their products is dependent on a number of factors such as the organization and control of the firm's employees, and more technical concerns like the quality of design and the quality of production. From a technical perspective, true progress toward improving and monitoring quality on a mass scale did not begin until the advent of statistical quality control usually called statistical process control, or SPC. SPC

was first introduced in the 1920s by the inspection department at Bell Telephone Laboratories led by Walter A. Shewhart, Harold F. Dodge, Donald A. Quarles, and George D. Edwards. SPC refers to the statistical techniques used to control or improve the quality of the output of some production or service process. Interest in quality has recently been growing rapidly in North America in response to the obvious success of the Japanese quality initiative started in the 1950s. Many SPC techniques, especially control charts, are now used in most manufacturing environments. Although service industries have been slower to adopt SPC, mainly due to difficulties in measuring quality, the increased use of SPC in the service sector is now a growing trend.

Most products, even very simple ones, have many characteristics or dimensions, possibly correlated, that affect their quality. For example, a nail is defined by its length, diameter, hardness, etc. However, most SPC techniques restrict attention to one characteristic at a time.

Quality control can be considered from two orientations: we can take a product or a process perspective. Taking a product orientation, the focus is on the parts or units after they are manufactured. Considering a single quality dimension at a time, the quality of a part is defined based on the target value and specification limits for that quality dimension. Specification limits, usually determined by engineering considerations, specify the range of quality dimensions within which it is acceptable for a part's quality dimension to fall. The target value is the most desirable quality dimension value, and is often centred between the specification limits. A non-conforming unit is usually defined as a part whose quality characteristic of interest lies outside the engineering specification limits, whereas if a part's quality dimension falls within specification it is called a conforming unit. In past work a non-conforming item was often called a defective item, but due to legal considerations the term "defective" is no longer recommended. A lot of products is considered acceptable if only a very small proportion of them have quality characteristics outside the specification

limits, and the lot is called rejectable if a fairly large number of units fall outside the specification limits. For example, lots with less than 1 part in a thousand non-conforming may be considered acceptable, whereas lots with 5 parts in a thousand non-conforming are rejectable, with an indifference region in-between. Clearly, it is desirable to produce parts within the specification limits, since from engineering considerations such parts should perform as desired. However, it is even better if all the quality dimensions are at the target. This is especially true for parts that make up large complex final products such as automobiles. When a number of parts must work together the closer each individual part is to the ideal the more likely the final product will work as designed. This idea is formalized by the Taguchi loss function (Taguchi, 1979). Taguchi believes it is best to produce all parts as close to the target as possible, i.e., any deviation from the target is undesirable. With this in mind, Taguchi assigns each part a "loss to society," that reflects how far the part's dimension is from the target. This loss is usually modelled by a quadratic function (Taguchi, 1979).

The process perspective, on the other hand, takes the focus back to the process that is producing the parts. This is an advantage since once a non-conforming part is produced it may be expensive to fix. Better to monitor the system that produces the parts trying to fix problems before many non-conforming products are made. In the process perspective, it is assumed that all the variation in a system arises from one of two sources. First, there is a certain amount of variability inherent in the system that is due to the cumulative effect of many uncontrollable causes. This type of variation is called the "natural variation," and when it is small the process is considered acceptable or "in control." The second source of variation, called an "assignable cause," is usually large compared with the natural variation, and is due to a cause that is feasible to detect, identify and remove. When a process is operating with only natural variation present the process is said to be in a state of statistical control and is called a stable process. When a process is operating with an assignable

cause it is considered "out of control." Notice that an "in control" process does not necessarily produce parts within specifications. An "in control" process produces parts that only differ due to system variability, but the process may still have an inherent variability that is large compared with the spread of the specification limits, and/or the process may not be producing parts with dimensions near the target value.

To address the relationship between the quality of parts produced by an "in control" process and the specification limits we define the process capability measure. There are many different definitions (Ryan, 1989), and all try, in some way, to quantify how many non-conforming (out of specifications) parts the process will produce. The process capability index C_{pk} is defined below (e.g. Ryan, 1989),

$$C_{pk} = \frac{\min(USL-\mu, \mu-LSL)}{3\sigma}$$

where USL and LSL are the upper and lower specification limits respectively, and the variables μ and σ denote the current process mean and standard deviation respectively. Large C_{pk} values indicate better performance, since it is a function of the number of sigma or standard deviation units between the specification limits. A process with a large C_{pk} value is called a capable process, since it is likely to produce the vast majority of parts in specification. In Japan, a decade ago, the minimum acceptable C_{pk} value was 1.33 (Sullivan, 1984). This corresponds, assuming a normal process, to only 6 non-conforming units for every 100,000 produced. However, all manufacturers should be continually working to decrease the variability in their production processes, thus increasing C_{pk} and increasing the quality of their products.

One goal of SPC techniques is to determine whether a process is stable or if a lot of products is acceptable. This involves a test of hypothesis. By convention, the null hypothesis is defined as a stable process or an acceptable lot. Excellent conclusions can

usually be made using 100% inspection, i.e. examining all the items in a lot or coming from a process, but this is very costly and time consuming. As a result, in most applications, it is preferable to try to estimate or infer the state of a process or lot from a small random sample of units. However, using the methods of estimation and inference introduces the possibility of making an incorrect conclusion, since the small sample may not be truly representative of the process or lot. As a result, the hypothesis test may conclude that there is evidence against the null hypothesis when in actuality the process is stable or the lot is acceptable. If this occurs we have made what is called a type I error, also called a false alarm. The probability of making such an error is usually denoted as α. If, on the other hand, we fail to reject the null hypothesis when in fact the process is out of control or the lot is rejectable, then we have made a type II error. The type II error probability is usually denoted by β. Naturally we never wish to make an error, but short of using 100% inspection a certain probability of error must be tolerated. The tolerable error rates depend in each case on a number of factors including the cost of sampling, the cost of investigating false alarms and the cost of failing to detect a deviation from the null hypothesis.

1.1 Grouped Data

Quantification of the quality characteristic is another important consideration in SPC. Traditionally, the measurement strategies considered have resulted almost exclusively of two types of data: variables data and dichotomous pass/fail data. For example, in variables data, the length of a nail could be quantified in centimeters accurate to two significant digits. In dichotomous data, the nails could be classified as greater than or less than 6.2 centimeters. However, grouped data is a third alternative.

Categorical data arise when observations are classified into categories rather than measuring their quality characteristic(s) precisely. In the important special case where the

categories are defined along an underlying continuous scale the data is called grouped. It has long been recognized that even when there is an underlying continuous measurement, it may be more economical to "gauge" observations into groups than to measure their quantities exactly. Exact measurements often require costly skilled personnel and sophisticated instruments (Ladany and Sinuary-Stern 1985), whereas it is usually quicker, easier, and therefore cheaper to classify articles into groups. This is of great practical interest, especially when SPC techniques are used in difficult environments such as factory shop floors. This has clearly been one of the motivational factors behind the development of acceptance sampling plans and control charts based on dichotomous attribute data (see Sections 1.2.1 and 1.2.2).

As Edwards (1972, p. 6) wrote, "data will invariably be either discrete or grouped ... the fineness of the grouping reflecting the resolving power of the experimental technique." In this light, dichotomous attribute data can be thought of as classifying units into one of two groups, and variables data as classifying units into one of many groups, approaching an infinite number of groups as measurement precision increases. With this perspective, assuming some underlying continuous scale, variables data and dichotomous attribute data are not totally distinct, but rather at two ends of a continuum. As a result, it is logical and very appealing to attempt a compromise between the often low data collection costs of dichotomous attribute data and the high information content of variables data. An example is multi-group attribute data where units are classified into one of three or more groups (such as low, medium, and high). This extension to multiple groups has also been suggested very recently in the literature. Pyzdek (1993) suggests there are ways to extract additional information from attribute data:

- Make the attribute less discrete by adding more classification groups;
- Assign weights to the groups to accentuate different levels of quality.

He does not provide a mechanism to formalize this suggestion. This thesis addresses this issue from a statistical perspective.

In industry, when variables measurements are difficult or expensive, step gauges are often used. A k-step gauge is a device that classifies units into one of $k+1$ groups based on some quality characteristic that is theoretically measurable on a continuous scale. An idealization of a four-step gauge is shown in Figure 1.1.

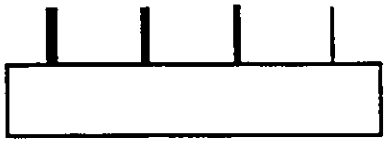

Figure 1.1: A Four-Step Gauge

Step-gauges are used, for example, in the quality control of metal fasteners in a progressive die environment at the Eaton Yale Corporation in Hamilton, Ontario. At Eaton, good control of an opening gap dimension is required; however calipers will distort the measurements since the parts are made of rather pliable metal. As a result, the only way to obtain a measurement of high precision is to use a prohibitively expensive laser. Consequently, the only economical alternative, on the shop floor, is to use a step gauge that has pins of different diameters. The pins classify parts according to which diameter of pin is the smallest that the part's opening gap does not fall through.

Another example of an application of grouped data is the proof-loading used in materials testing. Proof-loads are testing strengths up to which units are stressed; units are thereby classified into groups based on which proof-loads the unit survived. Proof-loading is often used for strength testing, since determining exact breaking strength can be difficult and very expensive due to waste through damaged product and/or the possible need for sophisticated measuring devices. More information on this application is presented in Chapter 4.

In general for reasonable choices of group limits the information content of data increases as the number of groups increases. However, the best possible incremental increase in information decreases as more groups are used. Although variables data maximizes the information content of a sample, as will be shown in Chapter 5, the difference in information content between variables data and multiple group data may be small. This loss in efficiency could easily be compensated for by lower data collection costs. The usual trade-off with respect to information content and data collection costs between two, multiple, and many groups (variables) data is summarized below in Table 1.1.

Table 1.1: Comparison of Two, Multiple and Many Group Data

	Two Groups	Multiple Groups	Many Groups
Information Content	Low	Medium	High
Data Collection Costs	Low	Medium	High

There are few SPC techniques for data classified into three or more groups. Almost all the past work in developing control charts and acceptance sampling plans has emphasized only two types of data, namely exact (variable) measurements, and dichotomous attribute data such as pass/fail or conforming/non conforming. It is true that some researchers have considered the three-group case for acceptance sampling plans and Shewhart control charts. These approaches are covered in Sections 1.2.1 and 1.2.2. However, I am aware of no techniques that have been extended to the general multi-group data case or even the four-group case. In addition, the few techniques that have been developed can not be easily extended to the more general case of multiple groups.

1.2 Areas of Application

This section introduces the areas of application considered in this thesis. First we discuss control charts, explaining their purpose and reviewing past literature in the area with a special emphasis on research pertaining to grouped data. Section 1.2.2 turns to the related application of acceptance sampling, again explaining its purpose and reviewing the literature. Section 1.2.3, introduces a more specific application that arises in material testing: estimation of the correlation between two strength properties whose values can only be determined through destructive testing. Destructive testing leads naturally to grouped data, since to estimate the correlation between two strength modes proof-loading must be used. Proof-loads are specific testing strengths that are applied in order to group units as breaking or not breaking under the proof-load strength.

1.2.1 Shewhart Control Charts

"A statistical (Shewhart) control chart is a graphical device for monitoring a measurable characteristic of a process for the purpose of showing whether the process is operating within its limits of expected variation" (Johnson and Kotz, 1989). The inventor of the first and most common type of control chart was Walter A. Shewhart of Bell Telephone Laboratories. Shewhart made the first sketch of a control chart (now called a Shewhart control chart) in 1924, and published his seminal book *Economic Control of Quality of Manufactured Product* in 1931.

The goal of a Shewhart control chart is to indicate or signal whenever a process is "out of control," i.e. an assignable cause has occurred, but not to signal when a process is "in control," i.e. operating with only natural variation. In other words, we want to know as soon as possible after an assignable cause has occurred, however false alarms are undesirable. In general, there is a tradeoff between a chart's power of detection and its false alarm rate. Shewhart control charts do not use specification limits (as in acceptance

sampling), but rather compare the observed sample to what is expected from the process based on its past performance. Since control charts are used on-line, they provide the user a way of quickly detecting undesirable behaviour in an important quality characteristic, and thus allow for quick corrective action. Control charts are used for two purposes: to prevent bad products from being produced, and for process improvement. Both purposes are achieved due to the timely nature of the provided information. If we are quickly aware of deterioration in quality, the process can be stopped before many bad parts are produced. In addition, many valuable clues are obtained regarding the nature of the problem which may lead to greater understanding of the process and subsequently to improvements.

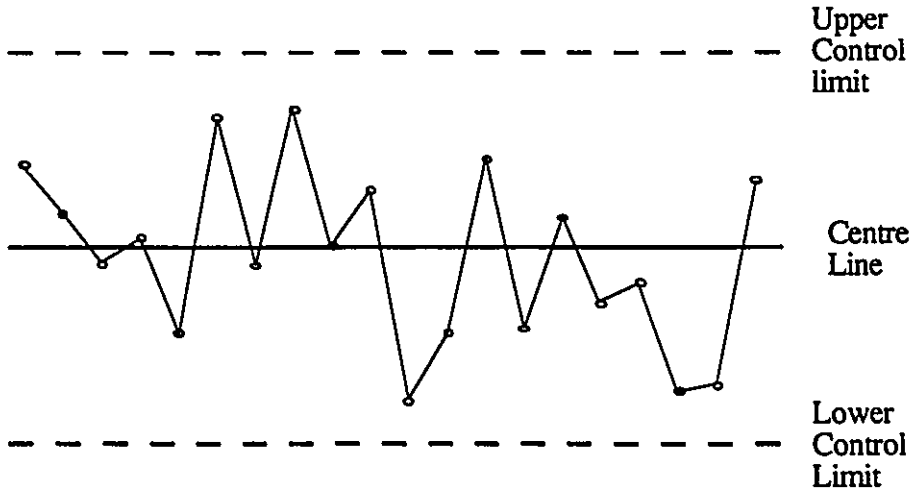

Figure 1.2: Typical Control Chart for a Stable Process

In a control chart, the quality characteristic is monitored by the repeated sampling of the process. Based on each sample a test statistic is calculated and plotted on the control chart. A control chart consists of three lines, the centre line (CL), and the upper and lower control limits (UCL and LCL respectively), see Figure 1.2. A control chart distinguishes between random and assignable causes of variation through its choice of control limits. If a sample test statistic plots outside the control limits we conclude that the process is no longer

stable, and the cause of instability is investigated. The control limits are constructed from confidence intervals so that if the process is "in control" nearly all sample test statistics will plot between them. As such, each point of a control chart can be considered a statistical test of a hypothesis (Box and Kramer, 1992). Considering a control chart a repeated hypothesis test is somewhat controversial the literature. In fact, as eminent a quality control scholar as Deming (1986) feels that this approach may be misleading. Deming believes that it is inappropriate to consider specific alternative hypotheses because the way a process becomes "out of control" is very unpredictable. This thesis takes the view that specific alternative hypotheses are necessary to enable comparisons between the efficiencies of various process control approaches.

Shewhart control charts are used to test the assumption that a process is stable over a period of time. Consequently, by definition, Shewhart charts monitor a process for parameter shifts in both upward and downward directions. As such, a Shewhart chart to detect parameter shifts implicitly tests the hypothesis system defined by:

$$H_0: \quad \theta = \theta_0$$
$$H_1: \quad \theta \neq \theta_0.$$

where, for example, θ_0 is the mean value at which the process is currently stable, and thus the mean value from which we wish to detect any deviation. The parameter θ_0 is typically estimated from the output of the current "in control" process. In this way, the expected type of products are determined, and subsequently a chart can be designed that will detect all significant departures from this "in control" setting.

Shewhart control charts have a "process" orientation. The decision whether or not a sample is acceptable is based solely on what is expected of the process, since the control limits do not depend on the engineering specification limits. Thus, it is determined not

whether the parts made are conforming or non-conforming, but rather whether the process is stable and producing consistent parts.

Shewhart control charts have been developed for a number of different types of data (see Duncan 1986, Juran et al. 1979, and Wadsworth et al. 1986). Attribute control charts apply to binomial data where units are classified as either conforming or non-conforming. Percentage charts (also called p charts) monitor the percentage non-conforming rate of a process, while np charts monitor the number of defectives. Control charts for variables include charts for individual measurements (X charts), charts for sample averages ($\overline{X}$ charts), and charts to monitor the process dispersion, range (R or moving range charts) and standard deviation (s) charts. In all cases, the control limits are derived based on estimates of the mean and standard deviation of the plotted statistic obtained while the process is "in control." For most control charts, it is standard practice for the control limits to be set equal to the average of the statistic ±3 times the standard deviation (σ) of the statistic. If the plotted statistic has a normal distribution, the $\pm 3\sigma$ control limits imply that 99.7% of the charted values will fall within the control limits when only natural variation is present. The remaining 0.3% are type I errors, which are false alarms. Notice that although the graphical aspect of a control chart is unnecessary to simply decide whether to accept or reject the null hypothesis at each sample, it does provide a visual representation of a process' past behaviour, and is easy for production personnel to understand.

A few researchers have considered deriving Shewhart control charts for grouped data. Tippett (1944) and Stevens (1948) were the first to make a strong case for the use of a two-step gauge, which divides observations into three groups. Stevens proposed two simple Shewhart control charts for simultaneously monitoring the mean and standard deviation of a normal distribution using a two-step gauge. He considered testing the hypotheses H_0: $\mu = \mu_0$ vs. H_1: $\mu \neq \mu_0$ and H_0: $\sigma = \sigma_0$ vs. H_1: $\sigma \neq \sigma_0$, and classified observations into one of three groups using a pair of gauge limits placed at x_1 and x_2.

Let p, q and r represent the probability that an observation falls into group 1, 2 or 3, respectively, and let a, b and c equal the actual number of units of a sample of size n that are classified into the three groups. This is summarized below:

Table 1.2: Summary of Stevens' Grouping

Group	Group Probability	Observed Number
1: $x \le x_1$	p	a
2: $x_1 < x < x_2$	q	b
3: $x \ge x_2$	r	c

Stevens proposes monitoring $a+c$ to detect shifts in the process standard deviation, and monitoring $c-a$ to detect shifts in the process mean. He derives control limits by noting that $a+c$ is binomially distributed with probability $p+r$, and that $c-a$ is approximately normal with mean $n(r-p)$ and variance $n\{(p+r)-(p+r)^2\}$. Stevens notes, however, that $a+c$ and $c-a$ are not independent measures of variation in the mean and variance. He also considered the optimal design of the gauge limits by maximizing the expected Fisher's information in a single observation. Stevens concludes that if the gauge limits are properly chosen, grouped data are an excellent alternative to exact measurement. However, it is not straightforward to extend Stevens' methodology to more than three groups, and it is difficult to determine his proposed control charts' operating characteristics.

More recently, advocates of Pre-control, sometimes called Stoplight control, have proposed the use of three classes to monitor the statistical control of a process. See, for example, Traver (1985), Silvia (1988), Shainan and Shainan (1989) and Ermer and Roepke (1991). With Pre-control, classification is commonly based on specification limits. One class consists of the central half of the tolerance range (green), another is based on the remaining tolerance range (yellow), and the third consists of measurements beyond tolerance limits (red). A number of different stopping criteria for on-going process control

have been proposed. For example, take a sample of size two; if either are red, stop and search for an assignable cause, if both are green, continue to run the process, if either are yellow, sample up to an additional 3 units until you get 3 greens (continue process) or either 3 yellows or 1 red (stop process). Pre-control, although appealing due to its simplicity, suffers from a number of shortcomings. The charts are based on specification limits and as such cannot easily detect shifts in the process mean unless the shift is of sufficient magnitude to cause the process to produce a significant number of parts out of specification. As a result, if the process is not very capable (small C_{pk}) many false alarms will register, and if the process is very capable (large C_{pk}) the test will have no power. As a result, the method is not sensitive to changes or improvements in process capability. In addition, Pre-control can not independently monitor for both mean and standard deviation shifts.

In the design of Shewhart control charts it is instructive to consider the chart's operating characteristic curve (OC curve). An OC curve is a plot of the probability that a single sample statistic will fall within the control limits versus true value of a process parameter, and is a useful measure of the chart's effectiveness. The OC curve shows how sensitive a particular chart is in detecting process changes of various degrees. Examining OC curves, one potential pitfall with the classical control chart design methodology (and all tests of pure significance) becomes apparent. The false alarm error rate (type I error) is usually set to be quite small (due to the choice of $\pm 3\,\sigma$ control limits), but the probability of correctly identifying an important shift in the process follows directly from the sample size chosen, and may not be very large. If the sample size is small, only large process shifts will be rapidly apparent. Assuming a normal process, the OC curve for an $\overline{X}$ chart with samples of size 5 is shown in Figure 1.3. Notice that only mean shifts of about 1.3 standard deviation units, also called sigma units, or larger are detected with a probability greater than 50% by the control chart in one sample.

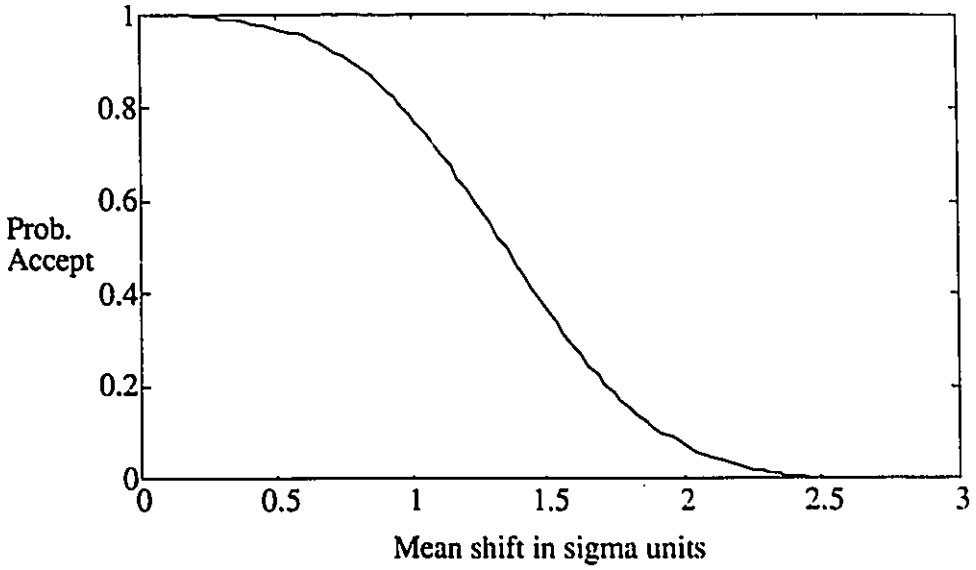

Figure 1.3: Operating Characteristic Curve for $\bar{X}$ chart when n = 5

This problem of lack of power for our statistical test can be avoided if the Shewhart chart is designed to go through two (or more) specific points on its OC curve. In any case, it is unrealistic to try to detect very small deviations from θ_0. In fact, it would be undesirable since there is always some "natural variation" in every process, and we only wish to detect fairly major departures from stability, namely we wish to detect only "assignable causes" of variation. As a result, it makes sense, from a design perspective, to derive the chart's control limits based on specific alternative hypotheses. The alternative hypothesis values for the parameter of interest are significant departures from stability in either an upward or downward direction. This is equivalent to designing a Shewhart chart with a specific alternative hypothesis in mind. For example, a Shewhart chart tests the following hypothesis:

$$H_0: \quad \theta = \theta_0$$

$$H_1: \quad \theta = \theta_1 \quad \text{or} \quad \theta = \theta_{-1}.$$

In other words, the control chart should signal whenever the process mean shifts to θ_1 or θ_{-1}, where without loss of generality, it is assumed that $\theta_1 > \theta_0$, and $\theta_{-1} < \theta_0$. Since, in many cases, Shewhart charts assign equal importance to parameter shifts in both directions, we may also assume $\theta_1 - \theta_0 = \theta_0 - \theta_{-1}$. The analysis that follows could be performed for the case when $\theta_1 - \theta_0 \neq \theta_0 - \theta_{-1}$, but this case is not usually of interest in practice. Considering an alternate hypothesis can ensure that the charts will usually detect important assignable causes quickly, and yet not give many false alarms. This is accomplished by determining what sample size is required so that our chart will be able to detect specified parameter shifts. Another consideration in the design of control charts and acceptance sampling plans is determining desirable levels for the type I and II error rates of the hypothesis tests. In the case of Shewhart charts, given specific type I and II error rates α and β, we wish to find a sample size and control limits such that

$$\begin{aligned} \text{Pr(chart signals | process "in control")} &= \alpha \\ \text{Pr(chart signals | process "out of control")} &= 1 - \beta \end{aligned}$$

A chart's error rates pertain to the performance of the chart based on a single sample. Often interest lies in a chart's performance based on many samples. Since we can assume that each sample is independent, the type I and II error rates are directly related to a chart's average run length (ARL), where the ARL is the average number of sample taken before the chart signals. $\text{ARL} = \dfrac{1}{\text{Pr(signal)}}$, where Pr(signal) is the probability the chart signals based on a single sample. Therefore under the null hypothesis the ARL = $1/\alpha$ which is large, whereas under the alternate hypothesis the ARL = $1/(1-\beta)$ which is small.

1.2.2 Acceptance Sampling Plans

The purpose of an acceptance sampling plan is to determine, based on a random sample, whether a particular lot is likely to contain an "acceptable" or "rejectable" quality of products. Based on the results of an inspection of a small sample, we try to surmise the quality of the complete lot, and make a decision to either accept or reject the complete lot based solely on the quality of the sample. For example, if a dimension is such that either smaller is better or larger is better, we may wish to test the one-sided hypothesis test:

$$H_0: \theta = \theta_a$$
$$H_1: \theta = \theta_r$$

where θ_a is a parameter value that leads to only a very small number of non-conforming units (i.e. acceptable level), and θ_r is an unacceptable, also called a rejectable, parameter level. When the quality dimension is acceptable only when it falls into a range a two sided acceptance sampling plan is needed. If the range of acceptable values is θ_a^- to θ_a^+, the implicit hypothesis test for a two-sided acceptance sampling plan is:

$$H_0: \quad \theta_a^- \le \theta \le \theta_a^+$$
$$H_1: \quad \theta \le \theta_r^- \quad \text{OR} \quad \theta \ge \theta_r^+.$$

where θ_r^+ and θ_r^- are the alternate parameter values on the upper and lower sides respectively. Define a lot whose parameter value falls into the range specified by the null hypothesis as an acceptable lot, whereas a lot whose parameter value satisfies the condition specified by the alternative hypothesis is called a rejectable lot.

In the design of acceptance sampling plans, we wish to determine the sample size and limits so that the sampling plan has appropriate error rates. Given type I and II error rates α and β, we must determine the sample size and decision criterion so that:

$$\Pr(\text{reject lot} \mid \text{lot is "acceptable"}) = \alpha,$$
$$\Pr(\text{accept lot} \mid \text{lot is "rejectable"}) = \beta.$$

The first acceptance sampling plans were developed by Harold F. Dodge and Harry G. Romig in their historic paper, "A method of sampling inspection," that appeared in *The Bell System Technical Journal*, in October 1929. The Dodge and Romig plans are based on attribute data. Some time later, Jennett and Welch (1939) developed the first sampling plans based on variables data. As a result, traditional sampling plans are based either on conforming/nonconforming attribute data or on variables data. It is well known that variables based sampling plans may require considerably smaller sample sizes than conformance/non-conformance data based plans to achieve the same error rates (Duncan, 1986).

The simplest type of sampling plan involves a single sample, although more advanced procedures such as double sampling, multiple sampling, item by item sequential sampling, chain sampling and many others have been devised. In single sampling with variables data, a sample of size n is taken, and the decision to accept or reject the lot is based on a comparison of a statistic computed from exact measurements (usually the mean) and a critical value (called A) derived from process specifications and distributional assumptions. With conformance attribute data, the number of nonconforming units in a sample of size n, denoted d, is compared with an acceptance criterion c, and the lot is accepted if $d \leq c$. In both cases the sampling plan design problem is to find the sample size, n, and critical value, A or c, so that the sample plan has the desired error rates, or equivalently, a specified operating characteristic curve (see Schilling, 1981, and Hamaker, 1979).

Acceptance sampling plans have a "product" orientation since the samples are taken from a lot of products. Thus, we make inferences about the quality of the lot, not the

process that made the products. This is in contrast with the "process" orientation of control charts. For this reason, acceptance sampling plans have recently been subjected to valid criticism. Sayings such as "You can't inspect quality into a product" summarize the major complaint. However, sampling plans should not be discarded out of hand; in some circumstances they can still provide a valid and valuable form of SPC. For example, firms engaging new suppliers may wish to use acceptance sampling until the new supplier has established a good track record. Vardeman (1986) and Schilling (1981) provide excellent discussions of how and when acceptance sampling plans should be used in a modern quality environment.

Two-sided acceptance sampling plans are also of interest because they are statistically equivalent to acceptance control charts. Acceptance control charts are a cross between Shewhart control charts and acceptance sampling plans. They are applicable when the process is very capable but the process average is not stable, and may drift due to some explainable yet uncontrollable factor. Tool wear is a good example. As the tool wears the process average begins to shift, but if the process is very capable, the process average has a considerable amount of room to move before an unacceptable number of defective items are produced. Replacing tools can be expensive, so we may wish to tolerate a certain amount of drift in the process mean before taking corrective action. In this situation, a Shewhart type control chart is not applicable. It is no longer desired to detect whenever the process average changes. In acceptance control charts, a process is defined as "out of control" only when the process has drifted too far, and is producing an unacceptably large number of non-conforming units. Unlike acceptance sampling, the samples are taken directly from the production process and do not randomly choose samples from a lot. Therefore, acceptance control charts tell us not about the acceptability or rejectability of lots, but rather when action should be taken to improve the process. As such, an acceptance control chart can be

used for process improvement like a Shewhart control chart, and yet bases its control limits on specification limits like acceptance sampling plans.

The first to try to combine the ideas of Shewhart control charts and acceptance sampling was Winterhalter (1945). He recommended that in addition to the standard control limits one should add "reject limits." These additional limits serve as a guarantee against producing out-of-specification products. As long as the standard control limits lie inside the reject limits, virtually all product will meet specifications. Hill (1956) expanded this idea by suggesting the use of reject limits in place of standard control limits when the process is very capable. Freund (1957) extended the ideas of Winterhalter and Hill to design a chart that also allowed specifying a desired protection from not detecting a significant parameter shift. He coined the term "acceptance control chart," and called his control limits "acceptance control limits." For a full discussion of acceptance control charts and additional references see Duncan (1986) and Wadsworth, Stephens, and Godfrey (1986).

The question of how to design acceptance sampling plans for grouped data has been considered by few researchers. When the standard deviation of the variable of interest is known and an underlying process distribution can be assumed, savings in inspection costs can be realized by using a dichotomous attribute plan with compressed specification limit gauging (also called narrow limit gauging or increased severity testing). Compressed limit sampling plans are a type of grouping and are discussed by Dudding and Jennett (1944), Ott and Mundel (1954), Mace (1952), Ladany (1976) and Duncan (1986). These plans do not classify units in the sample as conforming or non-conforming according to the actual specification limits, but instead classify units as greater than or less than an artificial specification limit. Compressed limit gauging uses assumptions about the process distribution (e.g. normal) and standard deviation to translate the proportion "defective" under the stricter compressed limits to an equivalent true proportion defective based on the

actual specification limits. When the actual proportion defective is small, very few units will be classified as non-conforming using a standard acceptance sampling plan. Thus the classification process will not provide much information. Compressed limit plans, on the other hand, can be designed to have a large number of units fall into each class and thus provide much more information about parameters of interest. Thus compressed limit plans require smaller sample sizes than standard dichotomous attribute plans especially when the actual proportion defective is very small.

Extensions to more than two groups are quite rare in the literature. Beja and Ladany (1974) proposed using three attributes to test for one-sided shifts in the mean of a normal distribution when the process dispersion is known. They consider hypothesis tests such as H_0: $\mu = \mu_0$ versus H_1: $\mu = \mu_1$. They show that optimal partition of the acceptance and rejection regions must be based on the Neyman-Pearson lemma, in other words, on the likelihood ratio. They find the best gauge limit design by assuming that the best limits must be symmetric about the midpoint of the null and alternate means. Ladany and Sinuary-Stern (1985) discuss the curtailment of artificial attribute sampling plans with two or three groups, whereby inspection of a sample is concluded as soon as the number of nonconforming units either exceeds the acceptance number or can not possibly exceed the acceptance number with the remaining unexamined units. They show that their three group curtailed sampling plan requires, on average, a smaller sample size than a variable plan! Unfortunately, the approach of Beja and Ladany (1974) and Ladany and Sinuary-Stern (1985) is not easily extended to more than three groups, where gains in efficiency can be realized.

Bray, Lyon and Burr (1973) consider three-class distribution-free attribute plans. They classify units as good, marginal or bad, and define the rejection region by specifying the critical numbers of marginal and bad units: c_1 and c_2 respectively. To simplify the analysis they focus on the subset of all sampling plans that has $c_2 = 0$, i.e., plans that

tolerate no "bad" units. Their results are not easily applied in practice, since it is difficult to devise a sampling plan using their tables that will attain specific error rates. In addition, the approach is not easily extended to the more general three group case or to more groups. Nevertheless, their methods have found some application in the food sciences area (Ingram et al. 1978), where in testing for food quality a certain level of contamination is totally unacceptable, but moderate levels of contamination can be tolerated to a certain degree. Another similar type of control chart for use with categorized data was first suggested by Duncan (1950), and further developed by Marcucci (1985) and Nelson (1987). The so called chi-square control chart generalizes a p-chart to multiple groups. However, the method does not consider an underlying distribution and it is thus difficult compare the performance of this type chart with more traditional charts. Depending on the way in which the data is categorized the method may be very inefficient, especially if the categorization is done in a manner similar to the traditional p chart. In addition, the chart would require a great deal of prior experience with the process to determine the number of units expected to fall into each category.

No two sided acceptance sampling plans or acceptance control charts have been developed for any type of grouped data, even dichotomous data. However, since the formulation of acceptance sampling plans and acceptance control charts is very similar, it may be possible to combine two of Beja and Ladany (1974) one-sided tests to create a three-group two-sided acceptance sampling plan.

1.2.3 Correlation Estimation from Destructive Testing

Many materials used in construction and other applications can be characterized by two or more important physical strength properties. In assessing the acceptability of the materials, the correlation between the various strength properties can be very important. For physical structures subject to a variety of stresses, large correlations between strength

modes have the effect of increasing the variability of a structure's load-carrying capacity, thus making it less reliable. Suddarth, Woeste and Galligan (1978) and Galligan, Johnson and Taylor (1979), studied the effect of the degree of correlation between bending and tensile strength in metal-plate wood trusses used in the roof structure of most homes. They concluded, based on theoretical and simulated results, that a large correlation may significantly affect the structure's reliability.

In many applications, however, the strength of an item can only be determined through destructive testing. Lumber, for example, has a number of physical properties such as bending strength, tensile strength, shear strength, and compression strength, that can only be determined destructively. As a result, one is able to ascertain the precise breaking strength in only a single mode for each unit. In such situations, the correlations among the various strength properties cannot be measured directly and must be approximated. A number of past studies such as those given by Evans, Johnson and Green (1984), Amorim (1982), Amorim and Johnson (1986), Green, Evans and Johnson (1984), Johnson and Galligan (1983) and Galligan, Johnson and Taylor (1979) have addressed the problem of estimating the correlation between destructively determined variables by using proof-loading. Proof-loading means stressing units only up to a prescribed (proof) load, thereby breaking only the weaker members of a population (Johnson, 1980). This way, although some units break before the proof-load is reached, others survive and can be subjected to further testing in other strength modes. As such, proof-loading leads naturally to grouping data based on whether the unit breaks or survives the testing stress.

The strategy employed in past studies to estimate the correlation (Evans et al., 1984 and Amorim, 1982) involves proof-loading units on the first mode followed by stressing the survivors until failure on a second mode and recording the exact load at failure for each unit. By assuming the strength properties have a bivariate normal distribution with known

means and standard deviations both Evans et al. (1984) and Amorim (1982) were able to solve numerically for the maximum likelihood estimate (MLE) of the correlation for various sample sizes n, and actual correlation value. A simulation study evaluated the mean and standard deviation of the MLE at different proof-load levels. Determining that the MLE was approximately unbiased, they compared the standard deviation of the MLE with the theoretical lower bound given by evaluating the reciprocal of the Fisher information.

Bartlett and Lwin (1984) considered a variation of the correlation estimation problem where a third property, C, can be measured non-destructively. They split the test sample into three groups. For the first group they measured C and the breaking strength on A for all units, thus allowing them to estimate the mean strength in A and C and the correlation between A and C. Similarly, for the second group, they estimate the means and the correlation of C and B. A third test group was first measured on C, then subjected to a proof-load on A and finally failed on test B. This third group then gave information on the property of interest, the correlation between A and B. They then used the Fisher information matrix to determine a lower bound on the variance of their estimate for the correlation between A and B.

Johnson and Galligan (1983) and Galligan, Johnson and Taylor (1979) also present a similar extension. They consider estimating the correlation between two destructively measured properties where each is a function of several properties that can be measured non-destructively. They present results comparing the correlation estimate calculated ignoring the additional dependence on the non-destructively measured properties and estimates obtained utilizing the additional information. The procedure was performed on real data, but the results were inconclusive due to a poor choice of proof-load levels.

All the methods previously developed utilize proof-loading (grouping) in one of the strength modes, but no methods have been developed that utilize grouping in both modes.

This is a logical extension since, as mentioned, grouped data are often easier and cheaper to collect.

One should note that all procedures based on proof-loading implicitly assume that survivors of the proof-load are not damaged. Experimental studies by Madsen (1976), and Strickler et al. (1970) suggest that this may be a reasonable assumption regarding the static strength of lumber, although a few pieces whose strength is only slightly greater than the proof-load stress will likely be weakened. In addition, according to cumulative damage theory, Gerhards (1979), "the theoretical results suggest that some percentage of the population will fail during the proof-load, a very small additional percentage will be weakened, but the remainder will have residual strength virtually equal to original strength." These theoretical results are based on the reasonable assumption that the proof-loading is done at a rapid rate.

1.3 Thesis Outline

The goal of this research is to apply estimation procedures and hypothesis testing based on grouped data in a quality control and improvement context. More specifically, acceptance sampling plans, acceptance control charts, Shewhart control charts, and correlation estimates under destructive testing are developed that are applicable when observations from an underlying distribution are classified into groups. This research fulfills a need from both a theoretical and practical standpoint. Multi-group data can be thought of as a natural compromise between binomial and variables data, and are often collected in industry to monitor the output of a production process. However, statistical process control techniques have been designed only for variables data, and the special case of two and to some extent three group data. Consequently, it is desirable to extend statistical process control methodology to also encompass the general multiple group case.

To handle multi-group data, notice that when observations from a single underlying distribution are classified into groups, the appropriate model is multinomial with group probabilities being known functions of the unknown parameters. Due to the grouping, the multinomial remains the appropriate distribution for any underlying distribution including multivariate distributions. As a result, all the design methodologies presented in this thesis can be very easily adapted for any underlying distribution. To illustrate this point many of the results in the thesis are given for the Weibull distribution as well as the normal.

The thesis is organized in the following manner. Chapter 2 discusses some preliminary items that are necessary for a full understanding of the subsequent work. Much of the notation is defined, and some background literature on maximum likelihood parameter estimation from grouped data is presented. In addition, the problems inherent in utilizing SPC techniques designed for variables data when the data are grouped are illustrated.

Chapter 3 turns to the question of designing acceptance sampling plans, acceptance control charts and Shewhart charts for grouped data. For each application a number of different solutions are considered. The approaches can be classified into two distinct design philosophies; namely, MLE based approaches, and "weights" based approaches. MLE based approaches utilize either the MLE itself or the generalized likelihood ratio as test statistic. Using the MLE of the parameter of interest for grouped data directly may be considered an extension of the methodology used in $\overline{X}$ charts for variables data. The "weights" approach derives from considering the likelihood ratio with specific alternative hypotheses. It is well known that for multinomial data the uniformly most powerful test for comparing simple parameter values is based upon the likelihood ratio of the multinomial probabilities. This is because all the information that a sample provides regarding the relative merits of hypotheses is contained in the likelihood ratio of these hypotheses on the sample (Edwards, 1972). Thus the specific alternative approach, or "weights" approach, is

optimal for one-sided tests, and, as will be shown, is near optimal for some two-sided tests. The various approaches all have their strengths and weaknesses that are discussed in detail. The design of small sample size plans or charts is also discussed in some detail.

Chapter 4 considers estimating the correlation coefficient of a bivariate normal distribution based on destructive testing. Again, a number of different testing procedures are considered. Two simple procedures provide good estimates of the correlation given the individual means and standard deviations are known. The results of these procedures are compared with the results of past studies. Two slightly more advanced procedures extend to the case where none of the five bivariate normal parameter are known.

The question of optimal gauge limit design for acceptance sampling plans and Shewhart control charts and correlation estimates from destructive testing is addressed in Chapter 5. Initially it was assumed that the grouping criterion design is predetermined. However, in some circumstances it is possible to design the step-gauge. If step-gauge design is feasible, there are two decisions to be made in specifying the grouping criteria: how many groups should be used, and how are these groups to be distinguished. As more groups are used, more information becomes available about the parameters of the underlying distribution, but data collection costs increase. The limiting case occurs when the variable is measured to arbitrary precision. Even if the number of groups is fixed, not all gauge limits will provide the same amount of information about the parameters of the underlying distribution. Gauge limits placed very close to one another provide little more information then a single limit, and gauge limits in the extreme tail of the underlying distribution provide almost no information. Also, what may be a beneficial gauge limit placement for estimation of the mean of a distribution is not necessarily very good for estimating a distribution's standard deviation or correlation. It is not intuitively clear how to set the k-gauge limits to optimize the testing procedure.

Finally, in Chapter 6 the major results are summarized and possible extensions are discussed.

Some of this original research has appeared in research papers. Specifically, one-sided acceptance sampling plans for grouped data based on the weights method, Sections 3.1 and 5.1, in Steiner et al. (1994A), Shewhart control charts for grouped data based on the one set and two sets of weights approaches, Sections 3.3.1, 3.3.2 and 5.3.1, in Steiner et al. (1994B), correlation estimation based on grouped data, Sections 4.1 and 5.4, in Steiner and Wesolowsky (1994A) and a review paper on the drawbacks of the *ad hoc* SPC techniques currently in common use, Section 2.3, in Steiner and Wesolowsky (1994B).

CHAPTER 2
Preliminaries

The purpose of this chapter is to set the stage for subsequent work. The three sections are somewhat unrelated, but contain background work that is necessary to fully understand the subsequent chapters.

Section 2.1 introduces much of the notation, and the likelihood function. The normal distribution is the standard choice for work in quality control. However, often the normal distribution provides a poor fit to the data and is not applicable. As the methodology that will be presented in later chapters is easily adaptable to other distributions, Shewhart control charts using the Weibull as underlying distribution are also presented. The Weibull distribution is often a good choice when the normal provides a poor fit, since it allows skewness in either direction. In addition, the exponential distribution is a special case of the Weibull. As a result, the Weibull distribution works especially well for applications where the observations represents a service or waiting time.

Parameter estimation is often a very important part of quality control. To design control charts and to calculate process capability indices we must be able to estimate the parameters of the underlying distribution accurately. The problem of parameter estimation from grouped data has been well studied (Rao, 1973). Section 2.2 gives a short literature survey of the area, discusses the existence conditions for the estimates, and presents algorithms that derive the maximum likelihood estimates of the normal and Weibull parameters from grouped data in our notation. Also presented is an algorithm which when given the mean and variance values, finds the Weibull parameters that

correspond. This procedure is useful since the transformation is not trivial, and is necessary when deriving Shewhart control charts based on the Weibull distribution.

This Chapter concludes in Section 2.3 with an analysis of some *ad hoc* quality control techniques currently used with grouped data. These techniques are used in industry since multiple grouped data is quite common and yet no acceptance sampling plans or control charts have been designed to deal with this type of data. The *ad hoc* procedures assume away the data grouping and use charting techniques designed for variables data. As will be shown these approaches are unreliable.

2.1 Useful Definitions

This section introduces and defines many of the variables that will be used throughout the thesis. See also the glossary in Appendix A.

2.1.1 Normal and Weibull Distributions

I will consider observations that have either a normal or a Weibull distribution. The normal distribution is the standard choice for most quality control applications. The well known normal probability density function (p.d.f.), $f_N(y)$, and cumulative density function (c.d.f.), $F_N(y)$, are given below as equations (2.1). The variable y represents a measurement of the quality characteristic of interest, e.g. the length of a nail.

$$\begin{aligned} f_N(y) &= \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right) \\ F_N(y) &= \int_{-\infty}^{y} f_N(s)ds \end{aligned} \tag{2.1}$$

The two parameter Weibull is also considered. The Weibull has probability density function $f_W(y)$ and cumulative density function $F_W(y)$ given by:

$$
\begin{aligned}
f_W(y) &= \frac{a}{b^a} y^{a-1} \exp\left(-\left(\frac{y}{b}\right)^a\right) \qquad y>0 \\
F_W(y) &= \Pr(0 \le Y \le y) = 1-\exp\left(-\left(\frac{y}{b}\right)^a\right).
\end{aligned} \tag{2.2}
$$

where the parameters a and b are called the shape and scale parameters respectively. The mean μ_W and variance σ_W^2 of the Weibull are given by:

$$
\begin{aligned}
\mu_W &= b\Gamma\left(1+\frac{1}{a}\right) \\
\sigma_W^2 &= b^2\left(\Gamma\left(1+\frac{2}{a}\right)-\Gamma^2\left(1+\frac{1}{a}\right)\right),
\end{aligned} \tag{2.3}
$$

where $\Gamma(x)$ is the Gamma function as defined by Abramowitz and Stegun (1970), 6.1.1. The two-parameter Weibull is a very flexible distribution although it is only defined for $y>0$. In most quality control applications the measurements are positive valued dimensions. For time to failure applications the parameter a has a special interpretation. Namely, if $a > 1$, the failure rate increases with time, whereas if $a = 1$, the Weibull distribution is an exponential distribution, and the failure rate is memoryless, i.e. constant over time, and if $0 < a < 1$ the failure rate decreases with time.

2.1.2 Grouped Data Definition

This thesis considers grouped data. Define the grouping criterion as follows. Let the k interval endpoints of the step-gauge be denoted by x_j, $j = 1, 2, ..., k$; then the probability that an observation is classified as belonging to group j is given by:

$$
\begin{aligned}
\pi_1 &= \int_{-\infty}^{x_1} \phi(y)dy \\
\pi_j &= \int_{x_{j-1}}^{x_j} \phi(y)dy \qquad\qquad j = 2, ..., k
\end{aligned} \tag{2.4}
$$

$$\pi_{k+1} = \int_{x_k}^{\infty} \phi(y)dy$$

where $\phi(y)$ represents the probability density function of the observations.

We can specify explicitly the group probabilities using the normal and Weibull p.d.f.s defined in equation (2.1) and (2.2). For the normal distribution $\phi(y) = f_N(y;\mu,\sigma)$. Then, defining $t_i = (x_i - \mu)/\sigma$, the t_i's are the standardized gauge limits. Also defining $t_0 = -\infty$ and $t_{k+1} = \infty$ for notational convenience, the normal group probabilities can be compactly written as:

$$\pi_j(\mu,\sigma) = \int_{t_{j-1}}^{t_j} f_N(\mu,\sigma)dy. \qquad j = 1,\ldots,k+1 \qquad (2.5)$$

If the observations are Weibull then $\phi(y) = f_W(y;a,b)$, $y > 0$. The Weibull cumulative distribution function can be written explicitly and the $k+1$ group probabilities are

$$\begin{aligned} \pi_1(a,b) &= 1 - \exp\left(-\left(\frac{x_1}{b}\right)^a\right) \\ \pi_j(a,b) &= \exp\left(-\left(\frac{x_{j-1}}{b}\right)^a\right) - \exp\left(-\left(\frac{x_j}{b}\right)^a\right) \qquad j = 2,\ldots,k \qquad (2.6) \\ \pi_{k+1}(a,b) &= \exp\left(-\left(\frac{x_k}{b}\right)^a\right) \end{aligned}$$

where all the x_j's are greater than or equal to zero.

2.1.3 Likelihood and Log-likelihood Ratios

Let **Q** be a (k+1) column vector whose j^{th} element Q_j denotes the total number of observations in a sample of size n that are classified into the j^{th} group. Then, defining θ as the parameter(s) of interest, the likelihood of any hypothesis about θ, given the sample **Q**, is defined as (Edwards, 1972):

$$L(\theta|\mathbf{Q}) = c\prod_{j=1}^{k+1}\pi_j(\theta)^{Q_j}, \qquad \text{where} \qquad \sum_{j=1}^{k+1}Q_j = n, \qquad (2.7)$$

c is the constant of proportionality, and $\pi_j(\theta)$ is the group probability that depends on the underlying distribution, given for example by equation (2.5) or (2.6). As we will be working with the log-likelihood ratio, the constant of proportionality c, is arbitrary, and can with loss of generality be set equal to unity. The log-likelihood using the standardized gauge limits is then

$$\ln L(\theta|\,\mathbf{Q}) = \sum_{j=1}^{k+1}Q_j \ln \pi_j(\theta), \qquad \text{where} \qquad \sum_{j=1}^{k+1}Q_j = n. \qquad (2.8)$$

2.2 Parameter Estimation Based on Grouped Data

Parameter estimation from grouped data has been addressed by a number of researchers (for a detailed review see Johnson and Kotz, 1989, pp. 527-536). Sheppard (1898) gave approximate corrections relating moments from data ungrouped and grouped into equi-spaced intervals. However, Sheppard's corrections are only valid under certain conditions, in particular the first and last frequencies must be small (Kendall, 1938). The corrections apply only to moments of higher order than the mean. Although a number of other moment corrections have been developed, the method of the maximum likelihood estimate (MLE) provide a more widely acceptable principle of estimation.

MLEs for grouped data were first found for equi-spaced grouping by Lindley (1950) utilizing Taylor's expansions. Since a fair number of terms is required to obtain reasonable estimates, MLE methods are numerical in nature. The expansions have been determined and programmed for the location and scale parameters of a number of different distributions. Sections 2.2.1 and 2.2.2 present algorithms to find MLEs for the

normal and Weibull parameters respectively. One potential shortcoming of maximum likelihood estimates is that in some circumstances they may not exist. This is usually not a concern with continuous data. However in our case, due to the discrete nature of grouped data, it is an issue that must be considered. In general, a set of sufficient conditions for the existence and uniqueness of the MLE $\hat{\theta}$ of the likelihood function $L(\theta)$ is (Kulldorff, 1961):

i) $\frac{\partial \ln L}{\partial \theta}$ is continuous for all θ

ii) The sign of $\frac{\partial \ln L}{\partial \theta}$ is different at the two extremes of the parameter space.

iii) $\frac{\partial^2 \ln L}{\partial \theta^2}$ is negative when $\theta = \hat{\theta}$

For each of the different parameters of interest, these general conditions are translated into specific conditions on the observed sample in the following subsections.

2.2.1 Normal Parameter MLEs

A very efficient and general procedure for finding the MLEs, called the expected maximization (EM) technique, was provided by Dempster et al. (1977). Since the EM technique provides excellent estimates in an efficient manner this is the strategy utilized to find MLEs of normal parameters from grouped data. Wolynetz (1979) presents the EM technique to solve for the MLEs of the location ($\hat{\mu}$) and scale parameter ($\hat{\sigma}$) of the normal distribution. The equations to solve are given below in our notation. The MLEs can be found by setting either the derivative of likelihood function itself or the log-likelihood (equations 2.7 or 2.8) equal to zero. Doing so gives the following equations:

$$\left.\frac{\partial \ln L(\mu,\sigma|Q)}{\partial \mu}\right|_{\mu=\hat{\mu},\sigma=\hat{\sigma}} = \frac{1}{\hat{\sigma}}\sum_{j=1}^{t+1} Q_j \frac{f_N(t_{j-1};\hat{\mu},\hat{\sigma}) - f_N(t_j;\hat{\mu},\hat{\sigma})}{\pi_j(\hat{\mu},\hat{\sigma})} = 0$$

$$\left.\frac{\partial \ln L(\mu,\sigma \mid Q)}{\partial \sigma}\right|_{\mu=\hat{\mu},\sigma=\hat{\sigma}} = \frac{1}{\hat{\sigma}}\sum_{j=1}^{k+1} Q_j \frac{t_{j-1} f_N(t_{j-1};\hat{\mu},\hat{\sigma}) - t_j f_N(t_j;\hat{\mu},\hat{\sigma})}{\pi_j(\hat{\mu},\hat{\sigma})} = 0$$

Defining

$$w_j = E\left(y_j \mid t_{j-1} < y_j < t_j;\hat{\mu},\hat{\sigma}\right) \qquad j = 1,\ldots,k+1$$

$$= \hat{\mu} + \hat{\sigma}\frac{f_N(t_{j-1};\hat{\mu},\hat{\sigma}) - f_N(t_j;\hat{\mu},\hat{\sigma})}{\pi_j(\hat{\mu},\hat{\sigma})} \tag{2.9}$$

We get

$$\hat{\mu} = \sum_{j=1}^{k+1}\frac{Q_j w_j}{n}, \tag{2.10}$$

$$\hat{\sigma} = \frac{\sum_{j=1}^{k+1}\left(Q_j w_j - \hat{\mu}\right)^2}{\sum_{j=1}^{k+1}\left[\left(\frac{f_N(t_{j-1};\hat{\mu},\hat{\sigma}) - f_N(t_j;\hat{\mu},\hat{\sigma})}{\pi_j(\hat{\mu},\hat{\sigma})}\right)^2 - \frac{t_{j-1} f_N(t_{j-1};\hat{\mu},\hat{\sigma}) - t_j f_N(t_j;\hat{\mu},\hat{\sigma})}{\pi_j(\hat{\mu},\hat{\sigma})}\right]} \tag{2.11}$$

These equations are solved in an iterative fashion. Initial estimates for μ and σ are substituted into (2.9) to determine the current estimates for each w_i. These values of w_i substituted into (2.10) and (2.11) will yield improved estimates for μ and σ. The performance of the EM algorithm is very good, and iterations converge in a short time even for "wild" initial conditions. In our application, fairly good initial estimates can be ensured by using each group interval's midpoint as a weight in (2.10) and (2.11) to give initial estimates of the underlying distribution's mean and standard deviation. However, as mentioned by Kulldorff (1961), the MLEs for μ and σ only exist if some mild conditions are met. The MLE $\hat{\mu}$ exists if and only if $Q_1 < n$ and $Q_{k+1} < n$ (Kulldorff, 1961, theorem 8.1). In other words, the maximum likelihood estimate of the mean exists only so long as not all the observations fall into one of the outside groups. The MLE $\hat{\sigma}$ exists if and only if either

1) $Q_1 + Q_{k+1} < n$ and $Q_j > 0$ for some j satisfying $\mu < x_{j-1}$ or $\mu > x_j$

2) $Q_1 + Q_{k+1} = n$, $0 < Q_1 < n$ and $Q_1(x_1 - \mu) > Q_{k+1}(x_k - \mu)$.

For fairly large sample sizes, and reasonable grouping criteria, the chances of observing a sample that does not have a MLE is very small. For small sample sizes, on the other hand, nonexistence of the MLEs can be of significant concern.

2.2.2 Weibull Parameter MLEs

The maximum likelihood estimates for the Weibull distribution can be found by differentiating the log likelihood function (2.8) with respect to a and b, and solving the resulting system of two equations for $\hat{a}$ and $\hat{b}$. Substituting into expression (2.6), the two equations to solve are (assuming $x_0 = 0$ and $x_{k+1} = \infty$):

$$\frac{\partial}{\partial a}\sum_{j=1}^{k+1} Q_j \ln\left(\exp\left(-\left(\frac{x_{j-1}}{b}\right)^a\right) - \exp\left(-\left(\frac{x_j}{b}\right)^a\right)\right)\Bigg|_{a=\hat{a},\, b=\hat{b}} = 0$$

$$\frac{\partial}{\partial b}\sum_{j=1}^{k+1} Q_j \ln\left(\exp\left(-\left(\frac{x_{j-1}}{b}\right)^a\right) - \exp\left(-\left(\frac{x_j}{b}\right)^a\right)\right)\Bigg|_{a=\hat{a},\, b=\hat{b}} = 0$$

These equations can not be solved explicitly, but the Newton-Raphson method (Press, et al. 1988) works very well. Use as initial estimates, $\hat{a}_0 = cv(n)^{-1/0.94}$, where cv(n) is the sample coefficient of variation (i.e. cv = 100 times the sample standard deviation over the sample mean) and $\hat{b}_0$ = the sample mean. This approach was suggested by Harlow (1989), and is based on the approximation $cv \approx a^{-0.94}$. The conditions on the existence of these MLE have been determined by Marymont (1975). The MLE $\hat{b}$ exists if and only if $Q_1 < n$ and $Q_{k+1} < n$, whereas the MLE $\hat{a}$ exists if and only if either:

a) $Q_1 + Q_{k+1} < n$ and there exists at least one i such that $Q_i > 0$, and $b < x_{i-1}$ or $b > x_i$.

b) $Q_1 + Q_{k+1} = n$, $0 < Q_{k+1} < n$ and if $x_1 < b$ then we must have $x_1 < b$.

2.2.3 Translating Mean and Variance to Weibull Parameters

When using the Weibull distribution to represent an underlying distribution it is often of interest to work in terms of the mean and standard deviation rather than the Weibull parameters a and b. Translating from the Weibull parameters a and b to the Weibull mean and variance can be done through equations (2.3) above. An algorithm, developed by Macdonald (1993), that does the reverse is presented below. The algorithm works by noticing that the variable b can be eliminated by considering a simple function of the mean and variance. Then using well known properties of the Gamma function gives:

$$\mu = b\Gamma(1+1/a) = \frac{b}{a}\Gamma\left(\frac{1}{a}\right)$$

$$\sigma^2 = b^2\left(\Gamma(1+2/a)-\Gamma^2(1+1/a)\right) = \frac{2b^2}{a}\Gamma\left(\frac{2}{a}\right)-\frac{b^2}{a^2}\Gamma^2\left(\frac{1}{a}\right)$$

$$\frac{\sigma^2+\mu^2}{2\mu^2} = \frac{b^2\left(a\Gamma(2/a)-\Gamma^2(1/a)+\Gamma^2(1/a)\right)}{b^2\Gamma^2(1/a)} = \frac{a\Gamma(2/a)}{\Gamma^2(1/a)}$$

Then it is possible to write:

$$0 = \left(\frac{\mu^2+\sigma^2}{2\mu^2}\right)\frac{\Gamma^2(1/a)}{\Gamma(2/a)}-a$$

Newton's method can be used to interactively solve for the root of the above equation. The derivative of $\Gamma(x)$ can be easily determined in terms of the Digamma or ψ function

(Abramowitz and Stegun, 1970, p. 258). Then once the equation has been solved for the parameter a,

$$b = \frac{a\mu}{\Gamma(1/a)}.$$

This algorithm quickly converges to the correct Weibull parameters. For example, given a Weibull distribution with mean $\mu = 68$, and standard deviation $\sigma = 2.2$, the above algorithm gives corresponding Weibull parameters $a = 38.93$ and $b = 68.98$.

2.3 *Ad hoc* Quality Control Techniques for Grouped Data

As previously mentioned, grouped data is fairly common, but few SPC techniques have been developed that consider such data. As a result, in current industrial practice, grouped data is usually treated as variables data in control charts. All observations in a given sample are assumed to be either at the interval midpoint or at an endpoint. The midpoint approach assigns any observation that falls into a particular group a value equal to the midpoint of the grouping interval. For example, if a unit is gauged to fall somewhere between 6 and 7, the midpoint approach assigns that unit a value of 6.5. The endpoint approach works in a similar manner except that the value assigned to each part will be equal to either the interval's lower endpoint or its upper endpoint. In the above example, the unit is assigned a value of 6 if we are using the lower endpoint strategy and a value of 7 if we are using the upper endpoint method. These representative values for each unit are then used as though they were variable measurements on a continuous scale, and combined into overall measures such as averages to create control charts such as $\overline{X}$ charts. While these *ad hoc* methods have some intuitive appeal, they suffer from a number of shortcomings since they ignore grouping in the data. As stated by Kulldorff (1961), "for reasons of practicability it is customary to treat grouped samples also as if they were non-grouped, giving all the observations in a group an individual value equal to

the central value of that group interval. This procedure represents an approximation that often leads to considerable systematic errors ...."

One concern pertains to the number of groups necessary. For end-groups the interval midpoint does not exist, since the interval extends to plus or minus infinity. Similarly, for the lower (upper) endpoint strategy no lower (upper) interval endpoint exists for the leftmost (rightmost) end-group. These problems can be avoided by utilizing a sufficient number of groups, so that no observations are likely to fall into the end-groups. However, using very many groups increases the measurement costs, and thus may eliminate the advantage of grouped data.

Another major problem especially acute for the endpoint based methods, but also a concern with the midpoint approach, is that parameter estimates derived by using the interval endpoint or midpoint as a representative value will usually be biased. Consider the case where we wish to estimate the process mean. Clearly, assigning each unit a value equal to the lower endpoint of its group will consistently under-estimate the actual values of the process, and vice-versa for the upper endpoint approach. The midpoint approach also introduces a bias, but the source of bias is not as obvious. Consider Figure 2.1 below:

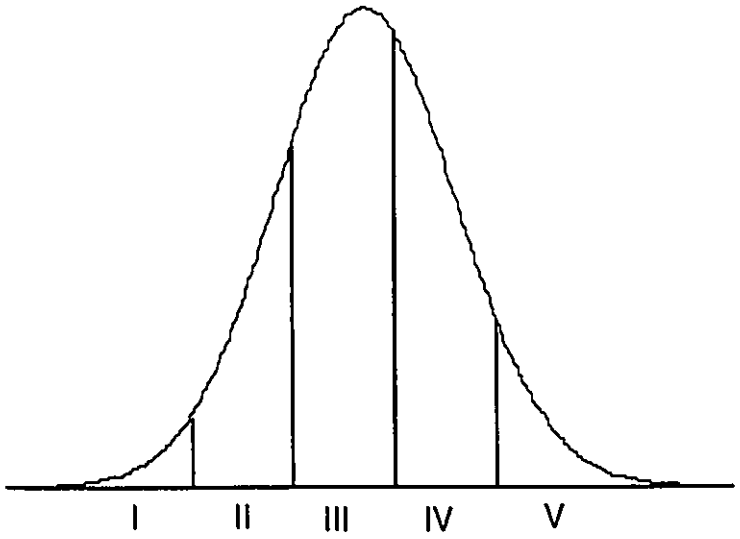

Figure 2.1: Gauge Limits Superimposed on Normal Distribution

Assuming the underlying normal distribution shown on the figure is an accurate approximation of the process' true distribution, any observation that falls into group II is more likely to actually come from the right side of that interval. Using the interval midpoint as a representative value will also introduce a bias in the estimate of the mean and standard deviation (unless the underlying distribution is uniform), though the bias is usually smaller than the bias introduced by the endpoint approaches.

The amount of bias introduced is easily quantified. Suppose we have a normal process with mean μ and variance σ^2, and the gauge limits are placed at $t_1, t_2, ..., t_k$. Then the probability of an observation falling into each group is given by $\pi_j(\mu)$, $j = 1,...,$ (k+1), equation (2.5), where j represents the group, and μ is the mean of the underlying process. Then using the midpoint approach, and assigning units that fall into the end-groups the end gauge limits, the group value representation is given by:

$$\text{V(midpoint)} = \begin{cases} t_1 & \text{if unit falls in 1st group} \\ (t_1 + t_2)/2 & \text{if unit falls in 2nd group} \\ \vdots & \\ (t_{k-1} + t_k)/2 & \text{if unit falls in } k\text{th group} \\ t_k & \text{if unit falls in } (k+1)\text{st group} \end{cases}$$

and the expected value of V(midpoint) is thus

$$\text{E(V(midpoint))} = t_1\pi_1 + \left(\frac{t_1 + t_2}{2}\right)\pi_2 + \ldots + \left(\frac{t_{k-1} + t_k}{2}\right)\pi_k + t_k\pi_{k+1}$$

With the upper endpoint approach, assigning units that fall into the upper end-group the end group limit, the group value representation is given by:

$$\text{V(upper endpoint)} = \begin{cases} t_1 & \text{if unit falls in 1st group} \\ t_2 & \text{if unit falls in 2nd group} \\ \vdots & \\ t_k & \text{if unit falls in kth group} \\ t_k & \text{if unit falls in (k+1)st group} \end{cases}$$

and the expected value of V(upper endpoint) is

$$\text{E(V(upper endpoint))} = t_1\pi_1 + t_2\pi_2 + \ldots + t_k\pi_k + t_k\pi_{k+1}$$

A similar derivation can be performed for the lower endpoint approach. The amount of bias generated by the various approaches clearly depends both on the group limit and the underlying process mean. Figure 2.2 shows the amount of mean bias created by using the various methods for different actual mean values. Note that using many groups also has the advantage of reducing the bias problems discussed previously. Many groups means shorter group intervals which in turn translates into less potential for biasing each observation. However, as mentioned, using many groups often comes at a price, since finer classification usually results in higher data collection costs.

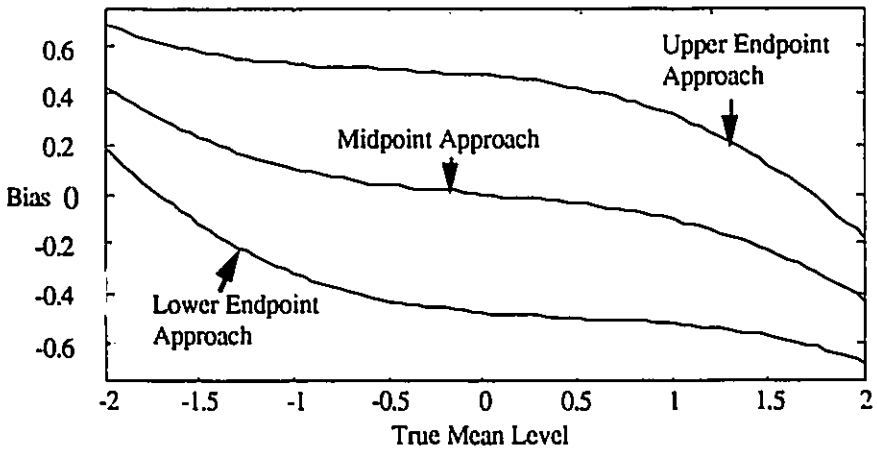

Figure 2.2: Mean Estimate Bias Using Current Approaches
Gauge Limits [-2, -1, 0, 1, 2]

The final concern is that variable based acceptance sampling plans and control charts, such as $\overline{X}$ charts, have been designed to yield particular error rates assuming that the quality characteristic of interest is measured exactly. With discrete data, such as grouped data, these error rates are not likely to still hold precisely. Typically, due to the discreteness of the data, the true error rates are underestimated. As a result, the properties of the control charts and acceptance sampling plans based on these *ad hoc* approaches are not clearly known. As an example, consider creating an $\overline{X}$ chart utilizing either the right endpoint or the midpoint approach. Assume the process is N(0,1), and a sample of size 5 is taken with control limits set at the standard ±3 standard deviations (i.e. LCL = –0.447, UCL = 0.447). Table 2.1 compares the probability of the chart not signaling an out-of-control for both the midpoint point and upper endpoint approaches for a variety of group limits. For comparison purposes, the last row of the table gives the error rates assuming variables data. Note that using the lower endpoint approach would give results that are the mirror image of the results for the upper endpoint approach.

The end-group problem is avoided in this case by deriving logical midpoint and endpoint values used for the end-groups by considering the interval width. For example, when the group limits are [-1, 0, 1] the midpoints used are (-1.5, -0.5, 0.5, 1.5) and the upper-endpoints used are (-1, 0, 1, 2). This method works fairly well in this case, but is more difficult to apply when the group intervals are not all equally wide.

Table 2.1: Comparison of *Ad hoc* Control Chart Approaches

		Probability of "in control" signal				
group limits	method	$\mu = -2$	$\mu = -1$	$\mu = 0$	$\mu = 1$	$\mu = 2$
[-1,0,1]	midpoint	0.5784	0.9687	0.9998	0.9687	0.5784
[-1,0,1]	upper-endpoint	1.0	1.0	0.9723	0.4514	0.0185
[-1.5,-.75,.75,1.5]	midpoint	0.1789	0.8581	0.9983	0.8581	0.1789
[-1.5,-.75,.75,1.5]	upper-endpoint	0.6100	0.9840	0.9610	0.3740	0.0088
[-2,-1,0,1,2]	midpoint	0.1243	0.8301	0.9980	0.8301	0.1243
[-2,-1,0,1,2]	upper-endpoint	0.4519	0.9733	0.9595	0.3405	0.0052
-	variables	0.0705	0.7775	0.9973	0.7775	0.0705

An $\overline{X}$ chart is meant to signal for any major deviation from the target (or current) mean value, but not to signal while the process is still stable. As a result, in Table 2.1, we would like large probability values when $\mu = 0$, and small values otherwise. Clearly, increasing the number of groups has a beneficial effect, but the probability of making an error is still considerably higher with the *ad hoc* approaches than with the variables method. The results are on occasion very striking. For example, the upper-endpoint approach with the 4 group limits given in the table has a 61% chance of not detecting a minus two standard deviation mean shift, whereas the variables based chart has only a 7% chance of failing to detect a mean shift of that magnitude.

When devising charts to detect standard deviation shifts, the poor result exhibited by the *ad hoc* approaches is even more evident. For example, consider using a range (R) chart with either the midpoint or endpoint approaches. An R chart is the standard type of control chart used in practice to detect process standard deviation shifts. Utilizing standard notation, control limits for R charts are set at UCL = $D_4\overline{R}$ and LCL = $D_3\overline{R}$. For a sample size of 5, from Table M in Appendix II of Duncan (1986), $D_3 = 0$ and $D_4 = 2.115$. Assume the group limits are given by (-2, -1, 0, 1, 2) and the process is standard normal. An initial analysis would give, on average, $\overline{R} \cong 2.4$. Thus, we would set LCL = 0 and UCL > 5. With this setup the control chart will never signal an upward standard deviation shifts since based on the group limits the maximum range value is 5.

In summary, the midpoint and endpoint approaches typically used in industry for grouped data introduce a bias into all calculations, require a large number of groups to be effective, and provide no way of determining the true error rates of the resulting control charts. These problems are, of course, especially acute when the number of groups used is fairly small, or if the underlying process distribution only covers a few of the groups. To get more reliable results, the fact that the data is grouped must be taken into account in a more direct way.

CHAPTER 3
Statistical Process Control Based on Grouped Data

This Chapter develops acceptance sampling plans, acceptance control charts, and Shewhart control charts based on grouped data. Section 3.1 discusses one-sided acceptance sampling plans. Sampling plans for one-sided parameter shifts based on the likelihood ratio are derived. Using the likelihood ratio with specific alternatives leads naturally to assigning each unit a weight determined by the group into which the unit is classified. This approach will be called the "weights" method. To detect one-sided parameter shifts this method is optimal in the sense that it will require the smallest sample size to achieve given error rates.

Section 3.2 extends the method to a two-sided test where there is a range of parameter values that are considered acceptable. This type of test is applicable to two-sided acceptance sampling plans and acceptance control charts. Two different approaches are considered. First, the optimal approach based on combining two one-sided tests is described. However, since this method requires the simultaneous use of two sets of weights, another method based on the MLE is also discussed. The MLE based approaches are often nearly optimal and require only one test statistic, but have a number of drawbacks, the most important of these are the possible non-existence of the MLE, the iterative nature of the MLE calculations, and the difficulty of determining the distribution of the MLE for small sample sizes.

In Section 3.3, Shewhart control charts are discussed. Now the purpose of the chart is to detect any parameter shifts from stability. Four different solutions are discussed and contrasted. The two approaches from Section 3.2 are again applicable. In

addition, an approach that utilizes one set of weights, and an approach based on the generalized likelihood ratio (GLR) are presented. Thus, two different weights-based methods and two approaches that require the calculation of the MLE are considered. The various solution methodologies each have certain advantages and disadvantages that are explored in detail.

The design methodology proposed in the first three sections depends on the applicability of the central limit theorem or of some asymptotic property. Often however, the sample sizes used are quite small, especially for control charts. Section 3.4 discusses designing acceptance sampling plans and control charts when the sample size is small.

3.1 One-Sided Acceptance Sampling Plans

Consider a sampling plan to monitor one-sided shifts in the parameter of a distribution using grouped data. Suppose that the quality characteristic of interest is a random variable Y that has probability density function $\phi(y;\theta)$ where θ is the parameter of interest. Assume also, without loss of generality, that large parameter values are undesirable, and that the process parameter is acceptable if less than or equal to θ_a, and that a process parameter greater than or equal to θ_r is considered unacceptable (rejectable). Thus the one-sided sampling plan tests the simple hypothesis system:

$$H_0: \quad \theta = \theta_a$$
$$H_1: \quad \theta = \theta_r.$$

where the desired acceptance and rejection regions are depicted in Figure 3.1.

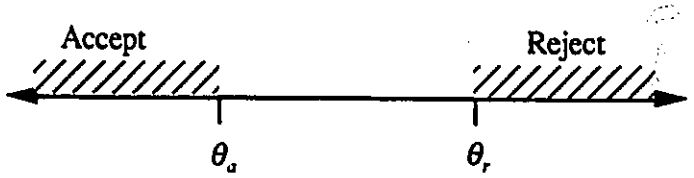

Figure 3.1: Acceptance/Rejection Regions for One-Sided Shifts

Note that one-sided acceptance sampling plans and one-sided control charts are statistically identical. Thus, in the following discussion, although the results are in terms of acceptance sampling plans, they are also applicable to one-sided control charts.

Using the notation introduced in Section 2.1, the likelihood ratio of interest for a grouped sample of size n is given by

$$LR(\theta \mid \mathbf{Q}) \quad = \quad \prod_{j=1}^{k+1}\left(\frac{\pi_j(\theta_r)}{\pi_j(\theta_a)}\right)^{Q_j}, \quad \text{where} \qquad \sum_{j=1}^{k+1} Q_j = n,$$

and $\pi_j(\theta)$ is the probability of a item falling into the j^{th} group when the parameter value is θ, Q_j is the observed number of units from the sample falling into group j, and k equals the number of group limits (i.e. $k+1$ equals the number of groups).

Defining λ as the critical likelihood ratio, our sampling plan will reject the lot if the sample observations are such that $LR(\theta \mid \mathbf{Q}) > e^{n\lambda}$ or equivalently, whenever

$$\sum_{j=1}^{k+1} Q_j \ln\left(\frac{\pi_j(\theta_r)}{\pi_j(\theta_a)}\right) \quad > \quad n\lambda .$$

Let us define $\qquad z_i \quad = \quad \ln\left(\frac{\pi_j(\theta_r)}{\pi_j(\theta_a)}\right), \qquad (3.1)$

where z_i is a random variable that equals $\ln\left(\frac{\pi_j(\theta_r)}{\pi_j(\theta_a)}\right)$ when the i^{th} observation belongs to the j^{th} group. This is equivalent to assigning each observation a "likelihood ratio weight" based on the amount of evidence that observation provides in favour of either hypothesis. From the definition (3.1), a unit's weight will be positive if it would be more likely to arise when $\theta = \theta_r$ than when $\theta = \theta_a$, and negative when the opposite is true. In practice the likelihood ratio weights are rounded off for ease of use.

Our sampling plan is designed to reject the lot whenever the average likelihood ratio weight $\bar{z} = \sum_{i=1}^{n} z_i/n = \sum_{j=1}^{k+1} \frac{Q_j}{n} \ln\left(\frac{\pi_j(\theta_r)}{\pi_j(\theta_a)}\right)$ of a sample is greater than λ. Then, if α and β are the desired probabilities of type I and II errors respectively, our sampling plan design problem is to find the sample size n, and the critical value or control limit λ so that

$$\alpha = \Pr(\bar{z} > \lambda \mid \theta = \theta_a), \tag{3.2}$$

and

$$1 - \beta = \Pr(\bar{z} > \lambda \mid \theta = \theta_r). \tag{3.3}$$

An approximate solution to (3.2) and (3.3) may be obtained by appealing to the central limit theorem. For large sample sizes, $\bar{z}$ will have an approximate normal distribution with mean $\mu_z(\theta)$, and variance $\sigma_z^2(\theta)/n$, where

$$\mu_z(\theta) = \sum_{j=1}^{k+1} \pi_j(\theta) \ln\left(\frac{\pi_j(\mu_r)}{\pi_j(\mu_a)}\right)$$

$$\sigma_z^2(\theta) = \sum_{j=1}^{k+1} \pi_j(\theta) \ln\left(\frac{\pi_j(\mu_r)}{\pi_j(\mu_a)}\right)^2 - \mu_z^2(\theta).$$

Assuming normality, and solving (3.2) and (3.3) for the required sample size and critical value gives the solution:

$$n(\theta_a, \theta_r) = \left(\frac{\Phi^{-1}(1-\beta)\sigma_z(\theta_r) - \Phi^{-1}(\alpha)\sigma_z(\theta_a)}{\mu_z(\theta_r) - \mu_z(\theta_a)}\right)^2, \tag{3.4}$$

and

$$\lambda(\theta_a, \theta_r) = \frac{\Phi^{-1}(\alpha)\sigma_z(\theta_a)\mu_z(\theta_r) - \Phi^{-1}(1-\beta)\sigma_z(\theta_r)\mu_z(\theta_a)}{\Phi^{-1}(\alpha)\sigma_z(\theta_a) - \Phi^{-1}(1-\beta)\sigma_z(\theta_r)}, \tag{3.5}$$

where $\Phi^{-1}(\ )$ denotes the inverse of the cumulative distribution function of the standard normal distribution. Note however that the value obtained from equation (3.4) for the sample size will not, in general, be an integer. Therefore, in the implementation of a sampling plan, the theoretical sample size suggested by equation (3.4) must be rounded

up to obtain a practical sample size. Rounding up the sample size results in more conservative plans.

It is possible that this rounding up of the sample size is a significant increase. If this occurs it may be desirable to adjust the critical likelihood ratio value obtained from equation (3.5) to take into account this increase in sample size. Using intermediate results from the derivation of equations (3.4) and (3.5), there are two ways to make this adjustment. Denoted the adjustment based on the equations (3.2) and (3.3) as λ_α and λ_β respectively, then

$$\lambda_\alpha = \mu_z(\theta_a) - \Phi^{-1}(\alpha)\frac{\sigma_z(\theta_a)}{\sqrt{\lceil n \rceil}} \tag{3.6}$$

$$\lambda_\beta = \mu_z(\theta_r) - \Phi^{-1}(1-\beta)\frac{\sigma_z(\theta_r)}{\sqrt{\lceil n \rceil}} \tag{3.7}$$

Using λ_α will decrease the critical value, and thus increase the actual α denoted α' and decrease the actual β denoted β'. Using λ_β will have the opposite effect. Which, if either, of these adjustments to use depends on the purpose of the acceptance sampling plan. In most cases, we are more interested in ensuring that the type I error rate α, also called the false alarm rate, is not too large. As a result, using the adjustment λ_β is recommended if the rounded up sample size is much different from the n given by (3.4).

Notice that the design methodology presented can be used for any parameter of interest. For different parameters, only the group probabilities π_j and the group weights change. For grouped data with any underlying distribution the appropriate distribution to use is the multinomial. Also, since in any case the average weight is used, the central limit theorem is applicable.

As an example, consider a one-sided acceptance sampling plan to test for an increase in the standard deviation of a normal process. Assume that the mean is stable at

zero, and that we are interested in comparing the hypothesis H_0: $\sigma = 1$ versus H_1: $\sigma = 2$. If the step-gauge has group limits defined by [-2, -1, 1, 2], then equation (3.1) gives group weights (rounded to one decimal place) of $\mathbf{z} = (1.9, 0.1, -0.6, 0.1, 1.9)$. Then choosing error rates of $\alpha = \beta = 0.05$ equations (3.4) and (3.5) suggest $n = 14.7$ and $\lambda = -0.0354$. Thus the sampling plan will reject the lot if the average weight $\bar{z}$ from a sample of size 15 is greater than –0.0354. As a matter of interest, from Section 3.4, using $n = 15$, with $\lambda = -0.0354$ gives an acceptance sampling plan that has true error rates of $\alpha' = 0.04$ and $\beta' = 0.06$.

3.2 Two-Sided Acceptance Sampling Plans

Often, interest lies not just in parameter shifts in one direction, but also in parameter shifts in both the upward and downward directions. This is the case for two-sided acceptance sampling plans and acceptance control charts. For these applications the hypothesis test of interest can be expressed as:

$$H_0: \quad \theta_a^- \le \theta \le \theta_a^+$$

$$H_1: \quad \theta \ge \theta_r^+ \text{ OR } \theta \le \theta_r^-$$

where θ_a^+, θ_r^+ and θ_a^-, θ_r^- are the acceptable and rejectable parameter levels on the upward and downward shift side respectively. Figure 3.2 shows this situation.

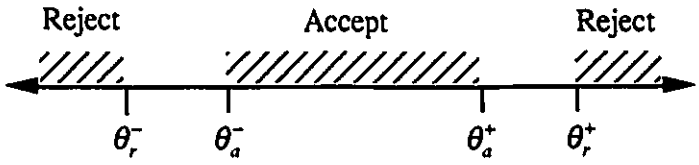

Figure 3.2: Acceptance/Rejection Regions for Two-Sided Shifts

Note that two-sided acceptance sampling plans and acceptance control charts are statistically identical, and thus in terms of the plan or chart design they can be considered interchangeable. For a discussion on their differences see Section 1.3.

Two-sided acceptance sampling plans and acceptance control charts should only be used when an "in control" process can produce the vast majority of its units within specifications. In other words, they are only applicable when a process is capable (see Section 1.2 for a discussion on process capability), and thus the process standard deviation is small compared with the spread of the specification limits, i.e. θ_a^+ is much greater than θ_a^- in terms of the process standard deviation. If the process is not capable, it will be incapable of producing units that are consistently within the specifications, and acceptance sampling will not be effective. Sections 3.2.1 and 3.2.2 present two design approaches.

3.2.1 Two Sets of Weights Approach

If the process is capable, the two sides of the hypothesis tests do not interact, in the sense that if the true parameter value is θ_a^+ or θ_r^+ the chance of concluding that the parameter is θ_r^- is negligible, and if the parameter value is θ_a^- or θ_r^- the chance of concluding that the parameter is θ_r^+ is negligible. As a result, one can design an acceptance sampling plan for two-sided shifts with specified error rates as two one-sided shift acceptance sampling plans with the same error rates. In other words, the two sets of weights for the two-sided test will be based on the following two hypothesis tests:

Upside Hypotheses	Downside Hypotheses
H_0^+: $\theta = \theta_a^+$	H_0^-: $\theta = \theta_a^-$
H_1^+: $\theta = \theta_r^+$	H_1^-: $\theta = \theta_r^-$

The two-sided sampling plan or acceptance control chart is designed so that it rejects a lot or signals "out of control" if either H_0^+ or H_0^- are rejected.

The logistics of combining two one-sided parameter shift hypothesis tests to create one two-sided acceptance sampling plan or acceptance control chart is straightforward. Since the two one-sided shifts can be considered independently, the two sets of weights based on the likelihood ratio (3.1) can be utilized. One set of weights tests for an upward shift, while the other set tests for a downward shift. For each hypothesis test, the results from Section 3.1 can be used to determine the required sample size and critical value. $n^+ = n(\theta_a^+, \theta_r^+)$, $\lambda^+ = \lambda(\theta_a^+, \theta_r^+)$, $n^- = n(\theta_a^-, \theta_r^-)$ and $\lambda^- = \lambda(\theta_a^-, \theta_r^-)$, as expressed by equations (3.4) and (3.5), are the required sample sizes and control limits for the upward and downward shifts respectively.

A two-sided acceptance sampling plan could then be administered as follows:

1. Derive the required sample sizes, critical values and weights for the two one-sided tests, i.e. derive n^+, λ^+, $\mathbf{z}^+$ and n^-, λ^-, $\mathbf{z}^-$. Choose a sample of size n where $n = \max(n^+, n^-)$.
2. Take a random sample of size n from the lot, classify all the units in the sample into the appropriate group, and calculate the average weight of the sample using both sets of weights to get $\bar{z}^+$ and $\bar{z}^-$.
3. Reject the lot if either $\bar{z}^+ > \lambda^+$ or $\bar{z}^- > \lambda^-$, and accept otherwise.

For acceptance control charts, the procedure is very similar. In step 2, rather than take a sample from a lot of products, take the sample directly from the process. Also, the results are presented graphically, either on two separate charts or on one chart with different colours for the two tests. The critical values $\lambda^+$and $\lambda^-$are the control limits and the average samples $\bar{z}^+$ and $\bar{z}^-$ are the plotted statistics. In any case the larger average weight signifies which alternative hypothesis is more likely.

The above procedure is simplified when monitoring the mean of a normal distribution and there is symmetry in the problem; if the group limits are placed symmetrically about the acceptable and rejectable mean values, the required sample sizes and critical values for upward and downward tests will be the same. In this case it is possible to use only one chart, plotting only the larger of the two average weights.

This section concludes with an example. Suppose we are interested in detecting two-sided shifts in the mean of a normal distribution. The process operates with a stable standard deviation of 0.17, and the specification limits are specified as 10 and 12. As the process mean is known to be currently centred between the specification limits, this is a very capable process (C_{pk} = 1.96). As a result, an acceptance control chart or two-sided acceptance sampling plan is applicable. As is typical, the acceptable and rejectable process or lot is defined in terms of the proportion non-conforming, i.e. the proportion of the units falling outside the specification limits. One in 10000 is deemed to be an acceptable proportion of non-conforming units, whereas one in 200 is considered unacceptable. As the process standard deviation is known, these requirements can be translated into acceptable and rejectable mean values on the upper and lower side of the specification range. Performing the calculations gives μ_r^+ = 11.56, μ_a^+ = 11.37, μ_a^- = 10.63, and μ_r^- = 10.44 as acceptable and rejectable mean levels. Assuming the group limits are [10.25, 10.5, 10.75, 11.25, 11.5, 11.75], the equations (3.1), (3.3), and (3.4) will give the group weights, required sample size, and required control limit respectively. Error rates of α = 0.001 and β = 0.005 result in the following for the upper and lower hypothesis tests:

$$\mathbf{z}^+ = (-8.3,\ -6.7,\ -5,\ -2,\ -0.5,\ 0.9,\ 2.4)$$

$$\mathbf{z}^- = (2.4,\ 0.9,\ -0.5,\ -2,\ -5,\ -6.7,\ -8.3)$$

$$n^+ = n^- = 28.7, \quad \lambda^+ = \lambda^- = 0.0478$$

The symmetry is evident. The implementation of this example proceeds as follows: take a sample of size 29, assign weights to each unit based on $\mathbf{z}^+$ and $\mathbf{z}^-$ above, and calculate the average sample weights $\bar{z}^+$ and $\bar{z}^-$. If either average weight is greater than 0.0478, reject the lot or signal "out of control."

3.2.2 Maximum Likelihood Estimate Approach

An alternative design procedure can be derived based on the maximum likelihood estimate. Section 2.2 presents algorithms that calculate the MLEs based on grouped data. The asymptotic distribution of a MLE is normal (Kendall and Stuart, 1979, Chapter 18), and its asymptotic variance equals the minimum variance bound given by the Rao-Cramér inequality. Using these results, it is possible to derive control limits that are appropriate asymptotically.

The variance of the asymptotic distribution of the MLE for the parameter θ can be easily derived. Define $\hat{\theta}$ as the MLE of θ given a sample of grouped data. Using the Rao-Cramér inequality, the variance of the asymptotic distribution of the $\hat{\theta}$, denoted $\mathrm{AsVar}\left(\hat{\theta}\right)$, is:

$$\mathrm{AsVar}\left(\hat{\theta}\right) = \frac{1}{n E\left(\frac{\partial \ln L}{\partial \theta}\right)^2} \tag{3.8}$$

Define $SD\left(\hat{\theta}\right)$ as $\sqrt{\mathrm{n}*\mathrm{AsVar}\left(\hat{\theta}\right)}$. The $SD\left(\hat{\theta}\right)$ value depends on the actual parameter value. As a shorthand, let $SD(\theta_0) = SD\left(\hat{\theta} \mid \theta = \theta_0\right)$. As an example, the derivation of the variance of the asymptotic distribution for the MLE of the mean and standard deviation of the normal distribution from grouped data follows. The variance of the asymptotic distribution of the MLE $\hat{\mu}$, denoted $\mathrm{AsVar}(\hat{\mu})$ can be determined from (3.8):

$$\text{AsVar}(\hat{\mu}) = \frac{1}{n E\left(\frac{\partial \ln L}{\partial \mu}\right)^2} = \frac{\sigma^2/n}{\sum_{i=1}^{k+1} \frac{\left(f_N(t_{i-1}) - f_N(t_i)\right)^2}{\pi_i(\mu,\sigma)}}$$

where μ and σ are the current true values for the mean and standard deviation respectively. Thus, $SD(\hat{\mu})$ is written:

$$SD(\hat{\mu}) = \sqrt{\frac{\sigma^2}{\sum_{i=1}^{k+1} \frac{\left(f_N(t_{i-1}) - f_N(t_i)\right)^2}{\pi_i(\mu,\sigma)}}} \tag{3.9}$$

and similarly,

$$SD(\hat{\sigma}) = \sqrt{\frac{\sigma^2}{\sum_{i=1}^{k+1} \frac{\left(t_{i-1}f_N(t_{i-1}) - t_i f_N(t_i)\right)^2}{\pi_i(\mu,\sigma)}}} \tag{3.10}$$

We will illustrate the design methodology for a two-sided parameter shift through an acceptance control chart. Define the upper and lower acceptance control limits as UACL and LACL respectively. Then, to satisfy the desired error rates α and β, we wish to determine the sample size and control limits such that:

$$\begin{aligned}
\Pr\left(\hat{\theta} > UACL \,\middle|\, \theta = \theta_a^+\right) &= \alpha \\
\Pr\left(\hat{\theta} > UACL \,\middle|\, \theta = \theta_r^+\right) &= 1 - \beta \\
\Pr\left(\hat{\theta} < LACL \,\middle|\, \theta = \theta_a^-\right) &= \alpha \\
\Pr\left(\hat{\theta} < LACL \,\middle|\, \theta = \theta_r^-\right) &= 1 - \beta
\end{aligned} \tag{3.11}$$

Using the asymptotic normality of $\hat{\theta}$ to solve the system of equations (3.11) gives:

$$n^+ = \left(\frac{SD(\theta_r^+)\Phi^{-1}(1-\beta) - SD(\theta_a^+)\Phi^{-1}(\alpha)}{\theta_a^+ - \theta_r^+}\right)^2 \tag{3.12}$$

$$\text{UACL} = \frac{\theta_r^+ SD(\theta_a^+)\Phi^{-1}(\alpha) - \theta_a^+ SD(\theta_r^+)\Phi^{-1}(1-\beta)}{SD(\theta_a^+)\Phi^{-1}(\alpha) - SD(\theta_r^+)\Phi^{-1}(1-\beta)} \tag{3.13}$$

$$n^- = \left(\frac{SD(\theta_r^-)\Phi^{-1}(1-\beta) - SD(\theta_a^-)\Phi^{-1}(\alpha)}{\theta_a^- - \theta_r^-} \right)^2 \tag{3.14}$$

$$\text{LACL} = \frac{\theta_r^- SD(\theta_a^-)\Phi^{-1}(\alpha) - \theta_a^- SD(\theta_r^-)\Phi^{-1}(1-\beta)}{SD(\theta_a^-)\Phi^{-1}(\alpha) - SD(\theta_r^-)\Phi^{-1}(1-\beta)} \tag{3.15}$$

where n^+ and n^- are the sample sizes required to attain the desired error rates for the upper and lower control limits respectively. In an attempt to ensure the desired error rates are attained in both directions simultaneously, the sample size chosen would be $n = \max(n^+, n^-)$. Thus, the procedure would be run as follows: take a sample of size n, classify each of the n items into one of the $k+1$ groups, calculate the MLE of the parameter of interest based on this grouped data, and signal "out of control" if the derived MLE is greater than UACL or less than LACL.

Consider the same example as presented at the end of the section 3.2.1, i.e. t = [10.25, 10.5, 10.75, 11.25, 11.5, 11.75], $\mu_r^+ = 11.56$, $\mu_a^+ = 11.37$, $\mu_a^- = 10.63$, and $\mu_r^- = 10.44$, $\alpha = 0.001$, $\beta = 0.005$ and $\sigma = 0.17$. Using equations (3.9) and (3.12)-(3.15) gives $n^+ = n^- = 31$, UACL = 11.47 and LACL = 10.53. This compares with a required sample size of 29 for the two sets of weights approach.

3.2.3 Comparison of Approaches

To compare the two sets of weights approach and MLE approach consider the sample size required by each method to attain given error rates. Since two-sided acceptance sampling plans are only applicable for very capable processes, the upper side and lower side hypothesis test may be considered in isolation. Table 3.1 compares the sample sizes required to detect mean shifts of a Normal process. The table presents

results for different grouping criteria when the problem has been rescaled so that the acceptable mean value is zero and the standard deviation is unity. Note that since only one side of the hypothesis test was considered when creating Table 3.1, to attain the sample sizes reported, there must also be another set of group limits on the opposite side. For example, if $\mu_r^+ = 11.56$, $\mu_a^+ = 11.37$, $\mu_a^- = 10.63$, $\mu_r^- = 10.44$ and $\sigma = 0.17$, as in the previous example, then the results given under t = [-1, 0, 1] correspond to actual group limits [10.46, 10.63, 10.8, 11.2, 11.37, 11.54].

Table 3.1: Sample Size Comparison - Two Sided Tests

$\mu_a = 0,\ \alpha = 0.001,\ \beta = 0.005$

t	μ_r	n (MLE)	n (weights)
[-1, 1]	0.5	174.7	174.5
"	1.0	44.7	43.5
"	1.5	21.3	19.0
[-1, 0, 1]	0.5	147.5	145.5
"	1.0	38.6	36.3
"	1.5	18.8	16.0
[-1.5, -0.5, 0.5, 1.5]	0.5	141.6	140.8
"	1.0	36.1	35.0
"	1.5	16.8	15.3
[-2, -1, 0, 1, 2]	0.5	139.8	139.5
"	1.0	35.2	34.8
"	1.5	16.0	15.2

Table 3.1 shows that the MLE approach is near optimal in terms of the sample size required to give the desired error rates. However, there are also other important considerations when deciding which method is the best one for a particular application. The two sets of weights approach is optimal in terms of required sample size, and calculations required are very simple and can be done on a factory floor without a computer. But, the weights approach requires the charting or calculation of two test statistics, and for small sample sizes may need a special design procedure (Section 3.4) to result in an a sampling plan with the desired error rates. The MLE approach, on the other hand, requires a computer to find the MLEs, but only one test statistic is required, and the

statistic represents a more meaningful value. In addition, although the required sample size for the MLE approach is near optimal (Table 3.1), for small sample sizes the MLE is fairly likely not to exist.

In summary, the MLE approach is a good choice if the sample size is not small, and computer calculation of the MLE does not present a problem. Otherwise the two sets of weights approach is likely the better one.

3.3 Shewhart Control Charts

For Shewhart control charts detecting two-sided parameter shifts is necessary; however, there is no longer a range of acceptable parameter values. Instead, any deviations from stability or equivalently from the current parameter value are undesirable. Thus, if the current parameter value is $\theta = \theta_0$ where this is a desirable level, the Shewhart control chart is designed based on the hypothesis test:

$$H_0: \theta = \theta_0$$
$$H_1: \theta = \theta_1 \text{ OR } \theta = \theta_{-1},$$

where θ_1 and θ_{-1} are parameter values of interest in the upward and downward direction (from θ_0) respectively. This situation is shown in Figure 3.3. The control chart will be designed with a false alarm rate of α, and will signal with probability $1 - \beta$ if the parameter value has shifted to one of the specific alternatives.

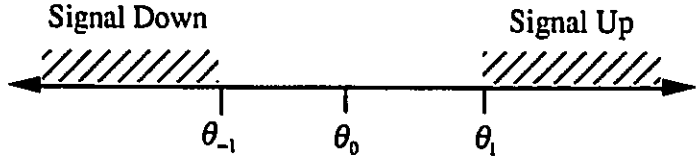

Figure 3.3: Signal Regions for Stability Test

We have developed four possible solution methodologies. The first utilizes two sets of weights and is optimal. The second is not optimal, but adapts the "weights" method so that only one set of weights is required. The MLE approach previously discussed in Section 3.2.2 and another MLE based approach that utilizes the generalized likelihood ratio test are also considered. Each of these different approaches has its own advantages and disadvantages that are compared and contrasted. Due to the relatively small advantage, in terms of required sample size, of the two sets of weights approach in most circumstances where Shewhart charts are applicable, either the one set of weights approach or the MLE approach is recommended for general use.

3.3.1 Two Sets of Weights Approach

One solution strategy is to consider two one-sided tests, i.e. testing

H_0: $\theta = \theta_0$ versus H_1: $\theta = \theta_1$, and

H_0: $\theta = \theta_0$ versus H_{-1}: $\theta = \theta_{-1}$.

Note that this is a different hypothesis test than the one used in Section 3.2, since now both hypothesis tests have the same null hypothesis. Based on a sample, one could calculate an average weight (or likelihood ratio) associated with each hypothesis test from Section 3.1. For each sample, the larger average weight signifies which alternative hypothesis (H_1 or H_{-1}) is more likely. When $\theta = \theta_0$, the chance of either average weight being larger is 50% due to symmetry. When $\theta = \theta_1$, and the difference between θ_{-1} and θ_1 is fairly large with respect to the process standard deviation, the average weight associated with first test will almost certainly be larger, and vice versa when $\theta = \theta_{-1}$. This means that it is possible to design the two sided testing procedure with error rates α and β based one two one-sided procedures with error rates $\alpha/2$ and β. The implementation issues are the same as ones discussed in Section 3.2.1. One option is to maintain two separate charts, one for each hypothesis test. However, if the group limits

are placed symmetrically about θ_0 (as is optimal, see Section 5.3) the critical value for each test is the same, and one could plot both average weights on one chart using different colours or different symbols, signaling if either point lies above the critical value. See Figure 3.4 for a simulated example.

Notice also from Figure 3.4, that the up (down) average weight plotting above the down (up) average weight is equivalent to a point plotting above (below) the centre line in the standard Shewhart chart. Therefore, the runs criterion, which is 8 successive points on one side of the centre line, sometimes recommended for use in addition to the standard control limits for an $\overline{X}$ chart (Juran et al. 1979), can also be applied to this chart.

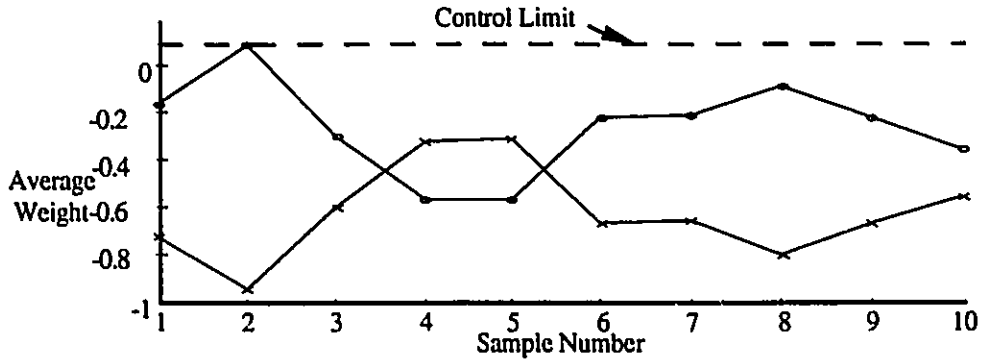

Figure 3.4: Two Sets of Weights Control Chart
$\alpha = 0.01$, $\beta = 0.05$, $n = 18$, $\lambda = 0.081$, $\mu_0 = 0$, $\mu_1 = 1$, $\mu = 0.1$
'o', 'x' = average weight for H_0 vs. H_1 and H_0 vs. H_{-1} respectively

3.3.2 One Set of Weights Approach

An alternative approach is to use one set of weights based on the likelihood of ratio of θ_{-1} versus θ_1. This approach has been used in the sequential analysis of multi-hypothesis problems (Ghosh and Sen, 1991). In other words, consider the following hypothesis test:

$$H_{-1}: \theta = \theta_{-1}$$

$$H_1: \theta = \theta_1$$

Using the notation from section 3.1, this likelihood ratio for a grouped sample of size n is given by

$$LR(\theta \mid \mathbf{Q}) = \prod_{j=1}^{k+1}\left(\frac{\pi_j(\theta_1)}{\pi_j(\theta_{-1})}\right)^{Q_j} \qquad \text{where} \quad \sum_{j=1}^{k+1} Q_j = n.$$

Based on this likelihood ratio we can create a two-sided hypothesis test. Letting λ_U and λ_L equal the critical values or the upper and lower control limits of the plotted statistic respectively, the chart signals the process parameter has shifted upwards whenever $LR(\theta \mid \mathbf{Q}) > e^{n\lambda_U}$, and signals a downward parameter shift whenever $LR(\theta \mid \mathbf{Q}) < e^{n\lambda_L}$. Equivalently, the process is deemed "in control," i.e. the parameter has not shifted, as long as

$$n\lambda_L \le \sum_{j=1}^{k+1} Q_j \ln\left(\frac{\pi_j(\theta_1)}{\pi_j(\theta_{-1})}\right) \le n\lambda_U.$$

Let α and β be the desired probabilities of type one and two errors respectively, and let

$$w_i = \ln\left(\frac{\pi_j(\theta_1)}{\pi_j(\theta_{-1})}\right) \tag{3.16}$$

when the i^{th} observation falls in the j^{th} group. Then each unit that falls into the j^{th} group is assigned the weight $\ln\left(\frac{\pi_j(\theta_1)}{\pi_j(\theta_{-1})}\right)$. Based on these definitions, the chart signals if $\overline{w} < \lambda_L$ or $\overline{w} > \lambda_U$, where $\overline{w} = \sum_{i=1}^{n} w_i/n$, and we wish to find n, λ_U and λ_L so that

$$\Pr\left(\overline{w} > \lambda_U \mid \theta = \theta_0\right) + \Pr\left(\overline{w} < \lambda_L \mid \theta = \theta_0\right) = \alpha, \tag{3.17}$$

$$\Pr\left(\overline{w} > \lambda_U \mid \theta = \theta_1\right) + \Pr\left(\overline{w} < \lambda_L \mid \theta = \theta_1\right) = 1 - \beta, \tag{3.18}$$

and

$$\Pr\left(\overline{w} > \lambda_U \mid \theta = \theta_{-1}\right) + \Pr\left(\overline{w} < \lambda_L \mid \theta = \theta_{-1}\right) = 1 - \beta. \tag{3.19}$$

Assuming that false alarms in the upward and downward direction are equally undesirable, equation (3.17) can divided into two separate equations, namely:

$$\alpha/2 = \Pr\left(\overline{w} > \lambda_U \mid \theta = \theta_0\right), \tag{3.20}$$

$$\alpha/2 = \Pr\left(\overline{w} < \lambda_L \mid \theta = \theta_0\right), \tag{3.21}$$

Also, it is easily shown that the second term of the right hand side of (3.18) and the first term of the right hand side of (3.19) are negligible. Clearly, since $\theta_1 > \theta_0$, $\Pr\left(\overline{w} < \lambda_L \mid \theta = \theta_1\right) << \Pr\left(\overline{w} < \lambda_L \mid \theta = \theta_0\right) = \alpha$, and similarly, since $\theta_{-1} < \theta_0$, $\Pr\left(\overline{w} > \lambda_U \mid \theta = \theta_{-1}\right) << \Pr\left(\overline{w} > \lambda_U \mid \theta = \theta_0\right) = \alpha$. As a result, since $\alpha < 1-\beta$, the two sets of equations (3.18), (3.20) and (3.19), (3.21) can be considered identical to equations (3.2) and (3.3) given in Section 3.1. Then, continuing along similar lines as in Section 3.1, for large sample sizes, $\overline{w}$ will have an approximate normal distribution with mean $\mu_w(\theta)$, and variance $\sigma_w^2(\theta)/n$, where

$$\mu_w(\theta) = \sum_{j=1}^{k+1} \pi_j(\mu) \ln\left(\frac{\pi_j(\theta_1)}{\pi_j(\theta_{-1})}\right)$$

$$\sigma_w^2(\theta) = \sum_{j=1}^{k+1} \pi_j(\theta) \ln\left(\frac{\pi_j(\theta_1)}{\pi_j(\theta_{-1})}\right)^2 - \mu_w^2(\theta).$$

Solving the two sets of equations for n and λ_U and λ_L gives the central limit theorem (CLT) solution:

$$n_U = \left(\frac{\Phi^{-1}(\alpha/2)\sigma_w(\theta_0) - \Phi^{-1}(1-\beta)\sigma_w(\theta_1)}{\mu_w(\theta_0) - \mu_w(\theta_1)}\right)^2 \tag{3.22}$$

$$\lambda_U = \frac{\Phi^{-1}(\alpha/2)\sigma_w(\theta_0)\mu_w(\theta_1) - \Phi^{-1}(1-\beta)\sigma_w(\theta_1)\mu_w(\theta_0)}{\Phi^{-1}(\alpha/2)\sigma_w(\theta_0) - \Phi^{-1}(1-\beta)\sigma_w(\theta_1)}, \tag{3.23}$$

$$n_L = \left(\frac{\Phi^{-1}(\alpha/2)\sigma_w(\theta_0) - \Phi^{-1}(1-\beta)\sigma_w(\theta_{-1})}{\mu_w(\theta_0) - \mu_w(\theta_{-1})} \right)^2 \tag{3.24}$$

$$\lambda_L = \frac{\Phi^{-1}(\alpha/2)\sigma_w(\theta_0)\mu_w(\theta_{-1}) - \Phi^{-1}(1-\beta)\sigma_w(\theta_{-1})\mu_w(\theta_0)}{\Phi^{-1}(\alpha/2)\sigma_w(\theta_0) - \Phi^{-1}(1-\beta)\sigma_w(\theta_{-1})}, \tag{3.25}$$

where n_U and n_L are the solutions for the sample size on the upward side and downward side respectively, and $\Phi^{-1}(\ )$ denotes the inverse of the cumulative distribution function of the standard normal distribution. This set of solutions are identical to the solutions obtained for the one-sided shift chart, except that α is replaced with $\alpha/2$, and the weights used are different. Implement of this control chart requires a sample size equal to $\max(n_U, n_L)$ rounded up to the nearest integer to ensure the desired error rates are attained in both directions. If the group limits are symmetric about μ_0 the above formulation is simplified since $\lambda_U = -\lambda_L$, and $n_U = n_L$.

As in Section 3.1, the rounded up sample size may be a significant increase, especially if $n_U \neq n_L$. One simple method to improve the CLT solution is to adjust the control limits derived by equations (3.23) & (3.25) to reflect the rounded up sample size that will actually be used. This can be accomplished by considering an intermediate result derived directly from equations (3.18)-(3.21) that expresses λ_U and λ_L in terms of the sample size n. These adjustment equations are given below:

$$\lambda_U = \frac{-\sigma_w(\mu_0)\Phi^{-1}(\alpha/2)}{\sqrt{\lceil n \rceil}} + \mu_w(\mu_0) \tag{3.26}$$

$$\lambda_L = \frac{\sigma_w(\mu_0)\Phi^{-1}(\alpha/2)}{\sqrt{\lceil n \rceil}} + \mu_w(\mu_0) \tag{3.27}$$

It is interesting to note that when monitoring the mean of the normal distribution the weights used in this approach can be rescaled to have a physical interpretation in the

original units. Dividing $\ln\left(\frac{\pi_j(\mu_1)}{\pi_j(\mu_{-1})}\right)$, the weight for group j, by twice the shift in σ-units we wish to detect, gives in a weight that is approximately equal to $E\left(y \mid y \in j^{th} \text{ group}, \mu = \mu_0\right)$. This is the expected value of any observation that falls into group j, given that $\mu = \mu_0$. See Appendix B for more details. Using this result, the control chart can be maintained in physically significant units that may be easier for production personnel to interpret.

For an example of the design strategy, consider a control chart whose purpose is to monitor the Weibull mean. The current process produces parts that match a Weibull p.d.f. with mean 101.3 and standard deviation 0.73. Figure 3.5 illustrates the output of the process. Notice that the distribution of the output is skewed right. Clearly, it is easier to detect upward shifts in the mean than downward shifts. If the process mean shifts up, the process is likely to produce parts that are extremely unlikely under the current parameter values. For example, the process may produce many parts with quality dimension greater than or equal to 103 units.

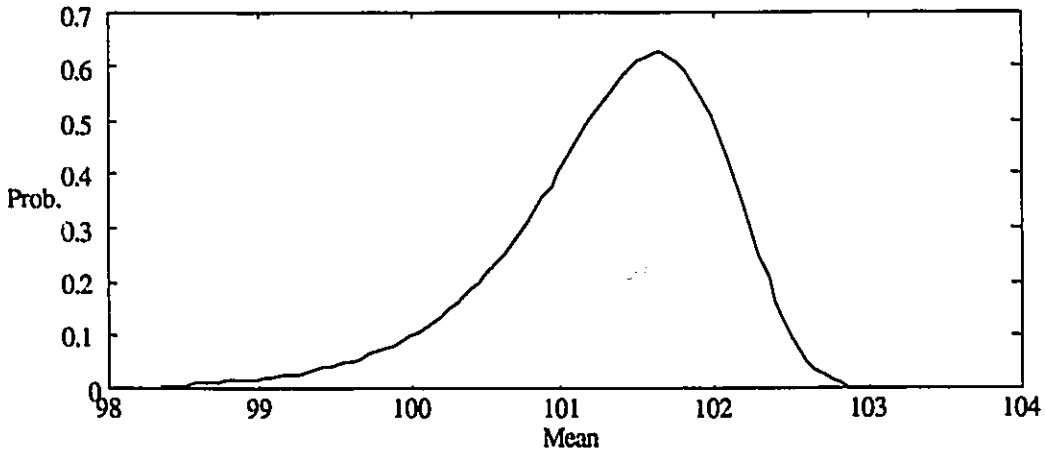

Figure 3.5: Weibull Process Probability Density Function
$\mu = 101.3,\ \sigma = 0.75,\ a = 172.5,\ b = 101.6$

Let us assume we have a step-gauge that classifies parts based on the group limits [99, 100, 101, 102], and that we wish to detect mean shifts of one standard deviation upwards and one and a half standard deviation units downwards with a probability 0.75, and will tolerate only one false alarm in 1000 opportunities. This implies $\mu_0 = 101.3$, $\sigma = 0.75$, $\mu_1 = 102.05$, $\mu_{-1} = 100.175$, $\alpha = 0.001$, and $\beta = 0.25$. Based on comparing the alternative hypothesis μ_1 and μ_{-1}, gives the five group weights: w = (–3.2, –3, –2, 1.1, 11.8). Then, utilizing equation (3.22)-(3.25) gives $n_U = 12.4$, $\lambda_U = 6.08$, $n_L = 16.3$, and $\lambda_L = -1.86$. This shows that although we were less stringent in terms of power on the downward side, we still require a larger sample size to achieve the desired error rates on the downward side due to the long lower tail of the Weibull. In fact a sample of size 17 is required. Since this is a significant increase over n_U, we will adjust the upper control limit to $\lambda_U = 5.53$ using adjustment equation (3.26). Using the results from Section 3.4, shows that this control chart has true error rates of $\alpha' = 0.0014$, $\beta'_U = 0.122$ and $\beta'_L = 0.238$.

3.3.3 Maximum Likelihood Estimate Approach

This section uses methodology similar to that presented in Section 3.2.2; however, the problem is formulated in a slightly different way to conform with the standard notation used for Shewhart charts. Shewhart control charts based on this MLE approach are the logical extension of standard $\overline{X}$ charts. An $\overline{X}$ chart plots the MLE of the mean based on variables data and uses control limits set at ±3 times the standard deviation of the sample mean. These MLE charts are based on plotting the MLE where the control limits are $\pm m$ times the asymptotic standard deviation of the MLE. Using the asymptotic normality of the MLEs, and the hypothesis system suggested for a Shewhart chart, when H_0 holds, $\hat{\theta} \sim N(\theta_0, \text{AsVar}(\theta_0))$, whereas when H_1 holds, $\hat{\theta} \sim N(\theta_1, \text{AsVar}(\theta_1))$ or $\hat{\theta} \sim N(\theta_{-1}, \text{AsVar}(\theta_{-1}))$. These approximate results allow the derivation of the theoretically

required sample size and upper and lower control limits, UCL and LCL. If the type I and II error rates are to be α and β respectively, we wish to find the sample size and control limits such that:

$$\Pr\left(\hat{\theta} > UCL \mid \theta_0\right) + \Pr\left(\hat{\theta} < LCL \mid \theta_0\right) = \alpha \tag{3.28}$$

$$\Pr\left(\hat{\theta} > UCL \mid \theta_1\right) + \Pr\left(\hat{\theta} < LCL \mid \theta_1\right) = 1 - \beta \tag{3.29}$$

$$\Pr\left(\hat{\theta} > UCL \mid \theta_{-1}\right) + \Pr\left(\hat{\theta} < LCL \mid \theta_{-1}\right) = 1 - \beta \tag{3.30}$$

where as defined, $UCL = \theta_0 + m\frac{SD(\theta_0)}{\sqrt{n}}$ and $LCL = \theta_0 - m\frac{SD(\theta_0)}{\sqrt{n}}$. The only unknowns in equations (3.28)-(3.30) are the sample size n, and the multiple of sigma units that the control limits are away from the centre line, namely m. Then, using the same arguments as in Section 3.3.2, this system of equations can be split into two independent systems of equations, one for the upward side and one for the lower side. For the upper side equations (3.28)-(3.30) can be rewritten as:

$$\Phi^{-1}(\alpha/2) = \frac{m_U SD(\theta_0)/\sqrt{n_U} + \theta_0 - \theta_0}{SD(\theta_0)/\sqrt{n_U}} \tag{3.31}$$

$$\Phi^{-1}(1-\beta) = \frac{m_U SD(\theta_0)/\sqrt{n_U} + \theta_0 - \theta_1}{SD(\theta_1)/\sqrt{n_U}} \tag{3.32}$$

where n_U and m_U represent the sample size and sigma multiple required to give the desired error rates on the upper side. A similar system of equations would be given in terms of n_L and m_L with θ_1 replaced by θ_{-1}. The solutions for n and m can be written:

$$m_U = m_L = \Phi^{-1}(\alpha/2) \tag{3.33}$$

$$n_U = \left[\frac{\Phi^{-1}(1-\beta)SD(\theta_1) - \Phi^{-1}(\alpha/2)SD(\theta_0)}{\theta_0 - \theta_1}\right]^2 \tag{3.34}$$

$$n_L = \left[\frac{\Phi^{-1}(1-\beta)SD(\theta_{-1}) - \Phi^{-1}(\alpha/2)SD(\theta_0)}{\theta_0 - \theta_{-1}}\right]^2 \qquad (3.35)$$

In general, the solution given by equations (3.34) and (3.35) will not be the same, since $SD(\hat{\theta})$ depends on the location of the gauge limits.

As an example, assume that the process of interest is currently stable and producing parts based on a normal distribution with a mean of 73.3 and standard deviation of 1.3. We wish to derive a Shewhart control chart to detect mean shifts of one sigma unit (1.3) in either direction with probability 0.75. The step-gauge used has group limits at [73, 74, 75, 76]. With a false alarm rate of 0.005, we have $\alpha = 0.005$, $\beta = 0.25$, $\mu_1 = 75.6$, and $\mu_{-1} = 73.0$. From equation (3.9), under the null hypothesis, $SD(\mu_0) = 1.36$. Equations (3.33)-(3.35) give $m = 2.81$, $n_U = 13.4$, and $n_L = 13.6$. Thus, the charting procedure is as follows: take a sample of size 14, classify each unit into one of 5 groups based on the step-gauge define above, calculate the MLE of the mean $\hat{\mu}$ based on the data, plot $\hat{\mu}$, and signal if $\hat{\mu}$ is greater than 75.3 ($74.3+2.8*1.36/\sqrt{14}$) or less than 73.3 ($74.3-2.8*1.36/\sqrt{14}$).

3.3.4 Generalized Likelihood Ratio Test Approach

An alternative design strategy that also utilizes MLEs is the generalized likelihood ratio (GLR) approach. The GLR method proposed by Neyman and Pearson (1928) is a method of test construction closely allied with maximum likelihood estimation. The methodology is quite general, allowing the testing for shifts in any number of parameters simultaneously. However, the approach is only applicable for omnibus tests, which are tests of the null hypothesis against all other possible parameter values. As such, the GLR test approach can be used to develop Shewhart type control charts, but is not applicable for acceptance sampling plans or acceptance control charts.

Consider a situation with parameters of interest $\theta = (\theta_r, \theta_s)$, $r \geq 1, s \geq 0$, and we wish to test H_0: $\theta_r = \theta_{r0}$ vs. H_1: $\theta_r \neq \theta_{r0}$. Note that both θ_r and θ_s represent vectors. For example, to test for shifts in the mean of a normal distribution we would have $\theta_r = \mu$ and $\theta_s = \sigma$. It is convenient to use minus two times the log of the GLR as the test statistic. Consider the statistic γ defined as:

$$\gamma = -2\ln\left[\frac{L\left(Q \mid \theta_{r0}, \hat{\hat{\theta}}_s\right)}{L\left(Q \mid \hat{\theta}_r, \hat{\theta}_s\right)}\right], \tag{3.36}$$

where $\hat{\theta}_r$ and $\hat{\theta}_s$ are the MLEs of θ_r and θ_s respectively given the data Q, and $\hat{\hat{\theta}}_s$ is the MLE for θ_s given θ_{r0}. In general $\hat{\hat{\theta}}_s$ will not equal $\hat{\theta}_s$; however, the MLE iteration formulas given in Section 2.2 can also be used to find $\hat{\hat{\theta}}_s$

If the null hypothesis holds, the value of γ should be close to zero, as the likelihood ratio will be close to unity. If, on the other hand, the true parameter value(s) is not θ_{r0}, the likelihood ratio will be less than unity and the value of γ will be greater than zero. As a result, a control chart based on the statistic γ will have only an upper control limit, since any parameter shifts in either the upward or downward directions tend to increase γ. This means that a control chart based on the GLR approach is somewhat easier to implement, but requires more investigation to determine the nature of a parameter shift that leads to an "out of control" signal.

The appropriate control limit (or critical likelihood ratio) γ^*, and sample size n, for this chart can be determined in the asymptotic case. When H_0 holds, Wilks (1938) showed that γ is asymptotically distributed as χ^2_r, a chi-square with r degrees of freedom. On the other hand, if the true parameter value(s) is θ_{r1}, γ is asymptotically distributed as $\chi^2_{r,\delta}$, a non-central chi-square with r degrees of freedom, and non-centrality parameter δ (Wald, 1943). For a sample of size n, the non-centrality parameter is given as:

$$\delta = (\theta_{r1} - \theta_{r0})^{T} V_{r1}^{-1} (\theta_{r1} - \theta_{r0}), \tag{3.37}$$

where V_{r1}^{-1} represents minus the matrix of expected information values, from a sample of size n, evaluated at the alternative hypothesis. Thus,

$$V_{r1}^{-1} = -E\left(\frac{\partial^2 \ln L}{\partial\theta_i \partial\theta_j}\right)\Bigg|_{\theta=\theta_{r1}} \qquad i,j \in r \tag{3.38}$$

As information is additive over samples, V_{r1}^{-1} is n times the expected information available in a sample of size one (Edwards, 1972). Therefore, for all parameter values except θ_{r0}, the non-centrality parameter increases linearly in n.

Note that when $i = j$, V_{r1}^{-1} can be written (Kendall and Stuart, 1979, 17.24) as

$$V_{r1}^{-1} = E\left(\left(\frac{\partial \ln L}{\partial\theta}\right)^2\right)\Bigg|_{\theta=\theta_{r1}}. \tag{3.39}$$

This expression is the expected information based on a single parameter. Expressions for (3.39) based on a sample of size one for the parameters of the normal and Weibull distributions are given in Chapter 5. Using the Chapter 5 results, the expression for V_{r1}^{-1} can be calculated when we wish to detect mean and standard deviation shifts from a normal distribution and scale and shape parameter shifts from a Weibull.

To design a Shewhart chart using this approach, the appropriate sample size and critical likelihood ratio value that give type I and II error rates less than or equal to α and β, respectively must be determined. Since the distribution of γ under the null hypothesis does not depend on the sample size, we can derive the critical likelihood ratio value γ^*, based solely only on the desired type I error rate. Given α, we must solve (3.40) for an appropriate γ^* value.

$$\Pr\left(\chi_r^2 > \gamma^*\right) = \alpha \tag{3.40}$$

This can done by inverting the central chi-square (Johnson and Kotz, 1970). Computer algorithms that perform this task are quite common; most simply use a root-finding technique along with a routine that finds chi-square probabilities. However, the power of the chart must also be sufficient. Under the alternative hypothesis, the non-centrality parameter δ increases linearly with the sample size. Thus given γ^* from (3.40), we need only determine the required sample size so that

$$\beta = \Pr\left(\chi_{r,\delta}^2 < \gamma^*\right). \tag{3.41}$$

Inverting this equation for the general non-central chi-square is difficult. However, the non-central chi-square probability integral can be written in terms of the central chi-square and an infinite sum of Laguerre polynomials $P_{(s)}$ (Tiku, 1965):

$$\Pr\left(\chi_{r,\delta}^2 \geq \gamma^*\right) = \Pr\left(\chi_{r/2}^2 \geq \gamma^*/2\right) + \sum_{s=1}^{\infty} P_{(s)}\left(\gamma^*/2; r\right)$$

This series is useful because it converges rapidly to the true value, especially for small values of δ. The Laguerre polynomials themselves can be calculated in an iterative fashion using the equations given below:

$$P_{(0)}(x;r) = 0$$

$$P_{(1)}(x;r) = \frac{\delta \exp(-x) x^{r/2}}{2\Gamma\left(\frac{r}{2}+1\right)}$$

$$P_{(s)}(x;r) = -\frac{\delta(r/2+2*s-3-3x)P_{(s-1)}(x)}{2*s(r/2+s-1)} - \frac{\delta^2(s-2)P_{(s-2)}(x)}{4*s(s-1)(r/2+s-1)}$$

where $\Gamma(x)$ is the gamma function as defined by formula 6.1.1 in Abramowitz and Stegun (1970).

Thus, for any given sample size and γ^* value the corresponding actual error rate $\beta' = \Pr\left(\chi^2_{r,\delta(n)} < \gamma^*\right)$ can be easily found. Since the appropriate γ^* value can be determined from (3.40), we can simply increment n until $\beta' \leq \beta$. One may start initially with $n = 1$, or if it is known, set the initial sample size equal to the sample size required when using variables data.

The design of GLR based control charts thus proceeds as follows

1. Find the critical likelihood ratio value λ^*, by inverting the c.d.f. of the χ^2_r distribution such that equation (3.40) is satisfied.
2. Find the non-centrality parameter at each specified alternative hypothesis using equation (3.37). Choose the smallest one, denoting it as δ.
3. Initially set the sample size equal to unity or to the sample size required with variables data. Increment the sample size, one unit at a time, until $\beta' = \Pr\left(\chi^2_{r,\delta(n)} < \gamma^*\right) < \beta$.

As an example, consider a grouped data control chart to detect simultaneously mean and standard deviation shifts for a normal distribution. Then $\theta_r = \{\mu, \sigma\}$ and $\theta_s = \{\varnothing\}$ and $r = 2$. Assume that the process is standard normal, the group limits are [–2, –1, 0, 1, 2], and we desire a false alarm rate less than or equal to 0.005, and a 50% probability of detecting either a mean shift of one sigma unit in either direction or a standard deviation shift to $\sigma = 2$. Equation (3.40) suggests a control limit of 10.8573. Using (3.37), we can derive expressions for the non-centrality parameters for each of the three specific alternative hypotheses. Since the group limits are symmetric about the null hypothesis ($\mu = 0$), the non-centrality parameters for $\mu = 1$ and $\mu = -1$ are the same. Denoting δ_μ and δ_σ as the non-centrality parameters of the mean and standard deviation

shifts respectively, we get $\delta_\mu = 0.9027n$ and $\delta_\sigma = 0.5599n$. Since δ_σ is smaller, detecting the specified shift in the standard deviation is more difficult than detecting the mean shift. Therefore, to ensure the desired error rates for both the mean and standard deviation shift, the sample size that satisfies equation (3.41) will be found using δ_σ. Incrementing the sample size until the type II error rates are theoretically satisfied for both the specified mean and standard deviation shifts suggests a sample size of 18 units.

3.3.5 Comparison of the Approaches

The four approaches suggested in Sections 3.3.1 to 3.3.4 each have certain advantages and disadvantages. A first comparison can be made on the basis of the sample size each method requires to achieve given error rates. Tables 3.2 and 3.3 present this comparison based on the theoretically required sample size using the asymptotic results from the previous sections. "GLR n," "MLE n," "w n" and "z n" refer to the sample size required by the generalized likelihood approach (§ 3.3.4), the maximum likelihood approach (§ 3.3.3), the one set of weights approach (§ 3.3.2), and the two sets of weights approach (§ 3.3.1) respectively.

Table 3.2: Sample Size Comparison - Shewhart Mean Charts

$\mu_0 = 0,\ \mu_1 = \mu_{-1},\ \alpha = 0.001$

t	β	μ_1	z n	w n	MLE n	GLR n
[-1, 0, 1]	0.005	1.5	17.3	17.7	20.1	24
[-1, 0, 1]	0.005	1	39.0	39.3	41.3	45
[-1, 0, 1]	0.005	0.5	156.1	156.3	158.1	161
[-1, 0, 1]	0.05	1.5	13.1	13.5	13.7	17
[-1, 0, 1]	0.05	1	28.4	28.7	28.8	32
[-1, 0, 1]	0.05	0.5	111.2	111.5	111.5	114
[-2, -1, 0, 1, 2]	0.005	1.5	16.4	16.4	17.1	18
[-2, -1, 0, 1, 2]	0.005	1	37.3	37.3	37.7	39
[-2, -1, 0, 1, 2]	0.005	0.5	149.6	149.6	149.9	151
[-2, -1, 0, 1, 2]	0.05	1.5	11.8	11.8	12.0	13
[-2, -1, 0, 1, 2]	0.05	1	26.5	26.6	26.6	27
[-2, -1, 0, 1, 2]	0.05	0.5	106.0	106.0	106.0	107

Table 3.3: Sample Size Comparison - Shewhart Sigma Charts

$\sigma_0 = 1,\ \sigma_1 = 1.625,\ \sigma_{-1} = 0.5,\ \alpha = 0.001,\ \beta = 0.05$

t	z n	w n	MLE n	GLR n
[–2, –1, 1, 2]	57.8	58.4	61.2	76
[–2.25, –1.5, –0.75, 0.75, 1.5, 2.25]	52.9	53.5	55.8	69

Tables 3.2 and 3.3 suggest that the approaches were presented in the order of increasing sample size requirements. The one set of weights approach requires the smallest sample size, followed closely by the two sets of weights method. The MLE approach requires virtually identical sample sizes as the weights based approaches for moderate β values, but requires larger sample sizes when β is small. In all cases, the GLR approach requires larger sample sizes. Note however that the sample sizes reported in the tables are only accurate when the asymptotic properties used to derive them hold.

In fact, utilizing the small sample size solution methodology from Section 3.4, we can determine (see Table 3.6) that, in most cases, the proposed methods require sample sizes that are quite similar to achieve the desired error rates. In other words, many of the differences in the above tables are due to inaccuracy in the asymptotic solution strategy. This problem is, of course, especially acute when using few group limits. See Section 3.4 for more details.

Another important basis for a comparison of the various approaches is the ease of design and implementation. Sections 3.3.1-3.3.4 show that all the approaches lead to sampling plans and/or control charts where the required sample size and control limits are easily determined with a computer. On the other hand, the approaches differ considerably in their ease of implementation. In general, the two-weights approaches are very easy to use. The weight assigned to each group is pre-determined in the design phase, and on-line the only requirement is the calculation of the sample's average weight(s). Note that the two sets of weights approach has the disadvantage of requiring the calculation of two average weights. This means that the weights methods can be easily applied on the shop floor. In contrast, the MLE and GLR approaches require the on-line use of a computer for the calculation of the MLE. An additional concern when using the MLE or GLR approach with a small sample size and few group limits is the possible non-existence of the MLE. However, the MLE approach is useful since the plotted statistic, the MLE, is a more physically meaningful value to production personnel than an average likelihood weight. Also, the GLR method is the only approach that can test for shifts in more than one parameter simultaneously. Table 3.4 summarizes the major advantages and disadvantages of the four proposed approaches.

Table 3.4: Advantages and Disadvantages - Shewhart Charts

Method	Advantages	Disadvantages
two sets of weights	optimal sample size no computer required on-line small sample size analysis	two test statistics weights have no physical meaning
one set of weights	no computer required on-line small sample size analysis one test statistic	not optimal sample size weights have no physical meaning
MLE	direct extension of variables approach plot meaningful values	not optimal sample size MLE may not exist computer required on-line
GLR	direct extension of variables approach extendible to multiple parameters	not optimal sample size likelihood ratio may be undefined statistic has no physical interpretation computer required on-line one control limit

Based on Table 3.4, the one set of weights approach seems the best compromise. It is very easy to use, does not require computer calculations on-line, and requires a near optimal sample size.

3.4 Small Sample Size Plans and Charts

The solutions presented in the previous sections assumed that the resulting sample size would be large enough so that either the central limit theorem is applicable or some asymptotic results hold. This is not true in general, especially since, traditionally, small sample sizes have been used for Shewhart control charts. Also, since sample sizes must be discrete, only a discrete number of different error rates is possible, none of which are likely to match precisely the desired error rates. The question of how large a sample is required for the asymptotic solutions to be sufficiently accurate is an important one.

Clearly the answer depends on many factors, including the number of group limits used, the location of group limits, the magnitude of parameter shift we wish to readily detect, and the accuracy required for a particular application.

For example, when the sample size is small, the CLT solutions for the weights based approaches will often have error rates that are larger than desired. This can be due to the inherent discreteness, but can also result from a skewed distribution of the average weight. The distribution of the average weight will be significantly skewed if the sample size is small and the group limits are not placed symmetrically about the true parameter value. This skewness problem is, of course, most pronounced when using few group limits, or if the group limits are poorly chosen. As an illustration, consider Figures 3.6-3.7. The figures show the distribution of $\bar{z}$, the average weight for the one-sided test, when using a sample size of 9, with group limits placed at 0.5 and 1.5. This example was created by considering a normal process with standard deviation equal to unity, where the one-sided hypothesis test is H_0: $\mu = 0$ vs. H_1: $\mu = 2$. With this setup, we get from (3.1), group weights equal to (-2.2, 0, 2.2). If the desired error rates are α = 0.001 and β = 0.001, the CLT solution from equations (3.4) and (3.5) suggest a sample size of 9 with critical likelihood ratio $\lambda = 0$.

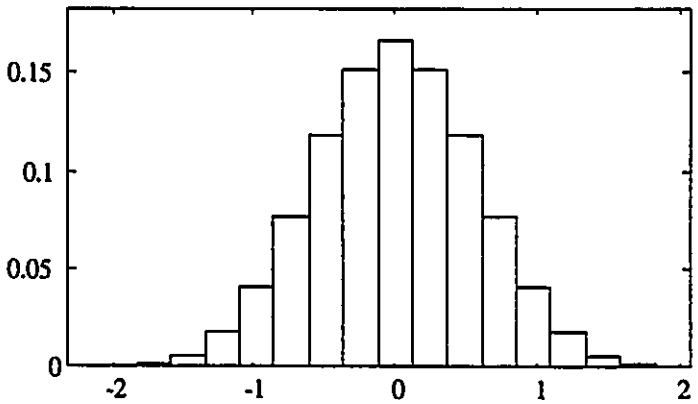

Figure 3.6: Distribution of $\bar{z}$ when $\mu = 1$

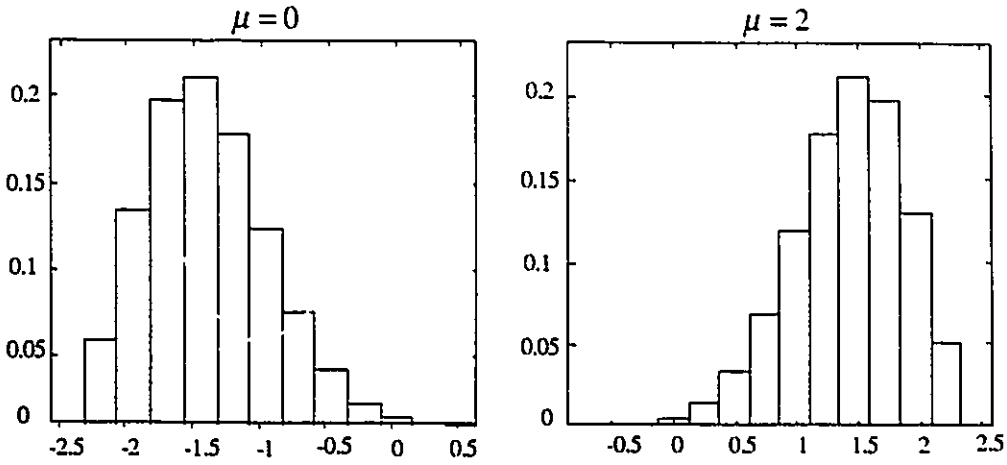

Figure 3.7: Distribution of $\bar{z}$ when $\mu = 0$ and 2

At the mean value $\mu = 1$ (see Figure 3.6) the group limits are symmetrically placed, and as a result the distribution of $\bar{z}$ is approximately normal, although still quite discrete. However, when the mean value is not centred between the group limits, as in Figure 3.7, we see a definite skewing of the distribution of $\bar{z}$. Typically, as in this example, the long tail of the distribution overlaps the critical likelihood ratio value, and as a result the true error rates will be larger than desired.

Therefore, to design sampling plans or control charts for small sample sizes it is necessary to take into account the discreteness and possible skewness of the distribution. For moderate sample sizes and number of group limits it is possible, using a computer, to determine the distribution of $\bar{z}$ (one-sided test) or $\overline{w}$ (two-sided test) by using enumeration. For example, the plots in Figures 3.6 and 3.7 were created by considering each possible sample in turn, calculating the corresponding average weight value, and ordering the samples based on their $\bar{z}$ values. For the MLE based approaches a similar analysis can be performed, however it involves much more work. For each possible sample it would be necessary to iterate to find the corresponding MLE. In addition, we must determine and eliminate all possible samples where the MLE does not exist.

Nevertheless, enumeration is possible for the MLE and GLR approaches if the sample size and number of group limits are sufficiently small (see Table 3.6).

3.4.1 Determining Actual Error Rates

Using the distribution of the plotted statistic ($\bar{z}$, $\overline{w}$, $\hat{\theta}$ or γ), it is possible to compute the actual α and β values, denoted α' and β', for a sampling plan or control chart. Given the sample of size n and k group limits all partitions, without regard to order, of the n observations into $(k+1)$ groups must be determined. Let $\{Q\}$ represent the set of all such partitions. The number of elements in the set $\{Q\}$ grows exponentially as the number of groups increases and polynomially as the sample size increases.

It is of interest to calculate the total number of different possible samples in the set $\{Q\}$. For a sample size of n units, Table 3.5 gives the total number of ways to partition n units into $k+1$ groups. The expressions in sum notation are derived by induction, and then translated into polynomials in n by well-known series formulae. Thus, the number of partitions is given by the expression $(n+k)!/(n!k!)$. From Table 3.5, the histograms in Figures 3.6 and 3.7 have a total number of possible samples equal to 55, since $n = 9$ and there are 3 groups.

Table 3.5: Number of Ways to Partition a Sample of Size n

number of groups	sum notation	polynomial in n
2	$\sum_{i=1}^{n+1} 1$	$n+1$
3	$\sum_{i=1}^{n+1} i$	$\frac{(n+1)(n+2)}{2}$
4	$\sum_{i=1}^{n+1}\sum_{j=1}^{i} j$	$\frac{(n+1)(n+2)(n+3)}{6}$
5	$\sum_{i=1}^{n+1}\sum_{j=1}^{i}\sum_{k=1}^{j} k$	$\frac{(n+1)(n+2)(n+3)(n+4)}{24}$

Define $\{Q_r\}$ as the subset of $\{Q\}$ that contains all the elements that lead to rejection of the lot or an "out of control" signal. Given $\{Q\}$ and the control limit(s) it is easy to find the subset $\{Q_r\}$. Then determining the probability of occurrence for each element in $\{Q_r\}$ under the null and alternative hypothesis the actual error rates α' and β' can be determined. See Steiner et al. (1993A) for a Branch and Bound algorithm that finds $\{Q_r\}$ and the actual error rates for the one-sided test without having to explicitly consider all possible partitions. Unfortunately, the Branch and Bound algorithm is not applicable when dealing with two rejection regions.

In general, there are two competing influences on the actual error rates. Rounding up the sample size to the nearest integer will tend to improve the error rates, especially the α' level, whereas the discreteness of the test statistic tends to increase the error rates, especially the β' levels.

For one-sided acceptance sampling plans with critical value λ the probability that a lot is rejected when the true parameter value is θ, is given by the following multinomial sum:

$$\Pr\left(LR(\theta|\mathbf{Q}) > e^{n\lambda}\right) = \sum_{\{Q\}} \frac{n!}{Q_1!Q_2!\ldots Q_{k+1}!}\prod_{j=1}^{k+1}\pi_j(\theta)^{Q_j}.$$

Therefore, the true error rates are $\alpha' = 1-\Pr\left(LR(\theta_u|\mathbf{Q}) > e^{n\lambda}\right)$ and $\beta' = \Pr\left(LR(\theta_r|\mathbf{Q}) > e^{n\lambda}\right)$. Similarly, for the one set of weights approach to two-sided tests with control limits λ_U and λ_L, the probability that the chart signals is given by

$$\Pr\left(LR(\theta\,|\,\mathbf{Q}) > e^{n\lambda_U} \text{ or } LR(\theta\,|\,\mathbf{Q}) < e^{n\lambda_L}\right) = \sum_{\{Q\}} \frac{n!}{Q_1!Q_2!\ldots Q_{k+1}!}\prod_{j=1}^{k+1}\pi_j(\theta)^{Q_j}.$$

Then the false alarm rate is $\alpha' = 1-\Pr\left(LR(\theta_0\,|\mathbf{Q}) > e^{n\lambda_U} \text{ or } LR(\theta_0\,|\mathbf{Q}) < e^{n\lambda_L}\right)$, and the probability of not immediately detecting upward parameter shifts to θ_1 and downward

parameter shifts to θ_{-1} are $\beta_1' = \Pr\left(LR(\theta_1|Q) > e^{n\lambda_U}\right)$ and $\beta_{-1}' = \Pr\left(LR(\theta_{-1}|Q) < e^{n\lambda_L}\right)$ respectively. Expressions for the true error rates for the MLE and GLR methods are written similarly.

As an example, consider Figures 3.8, 3.9 and 3.10. The figures show α' and β' for sampling plans designed with the CLT solution (equations (3.4) and (3.5)) to detect one-sided mean shifts with 2, 3 and 5 gauge limits respectively and desired error rates $\alpha = \beta = 0.005$.

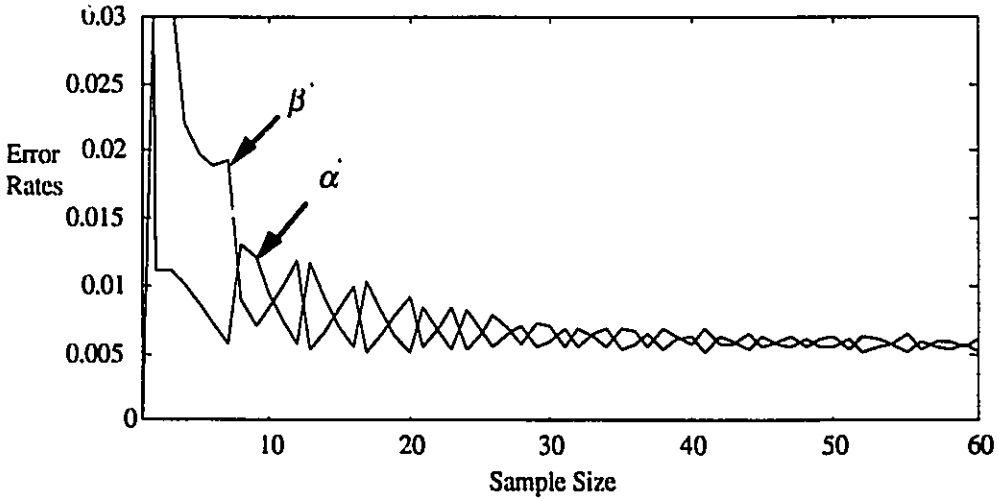

Figure 3.8: True Error Rates with 2 Gauge Limits
t = [0.25 1.25]

Figures 3.8-3.10 all show that, due to skewness in the distribution of $\bar{z}$, the true error rates are always somewhat larger than expected. This underestimation of the error rates by the CLT solution is expected, and the CLT solution may be considered is appropriate if the deviations of the true error rates from the desired rates is small. With this evaluation criterion, the effect of the number of gauge limits, as expressed in the figures, is significant. In the two gauge limit case, the fluctuations in the true error rates are still apparent at sample size 50. However, for more gauge limits the CLT solution performs much better. When using three gauge limits, the error rates become quite

stable, and close to the desired levels by sample size 20. In the five gauge limit case, the same is true at sample size 15. These results are, of course, dependent on the location of the gauge limits.

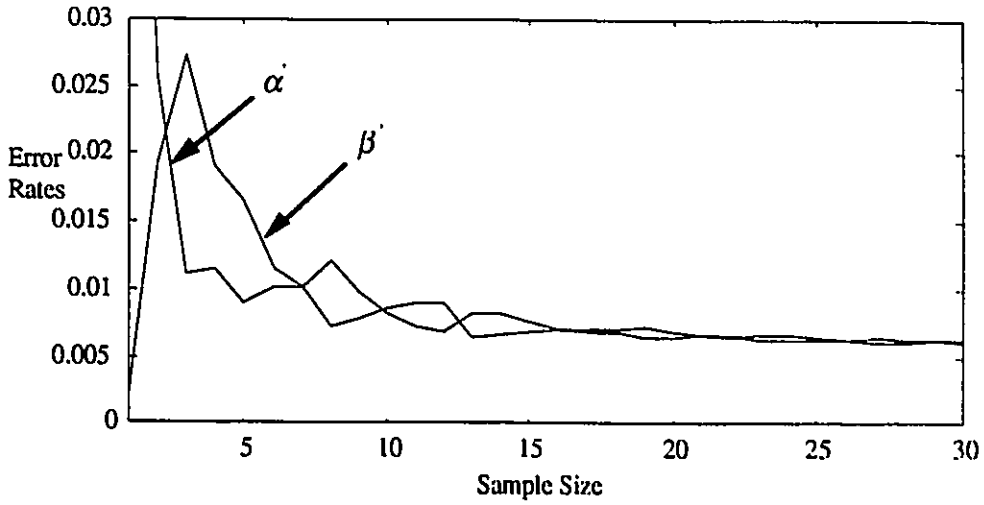

Figure 3.9: True Error Rates with 3 Gauge Limits
t = [0, 0.75 1.5]

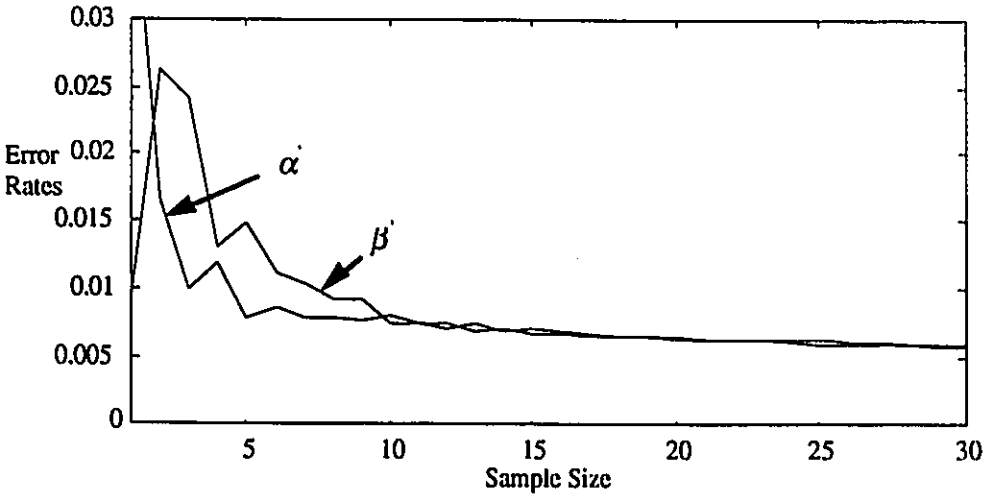

Figure 3.10: True Error Rates with 5 Gauge Limits
t = [–0.25, 0.25, 0.75 1.25, 1.75]

A similar analysis may be performed on the asymptotic results that compare the various approaches for grouped Shewhart charts presented in Table 3.2. Consider the

case where we are interested in detecting a 1.5 sigma shift in the mean. This leads to sufficiently small sample sizes so that our small sample size analysis can be performed. Table 3.6 shows the actual error rates for the various approaches based on the sample size and control limits recommended by the asymptotic results. Note that due to computational limitations, for the MLE and GLR approaches we are restricted to considering the case with only three group limits.

Table 3.6: Actual Error Rate Comparison
$\alpha = 0.001,\ \mu_1 = 1.5$

t	Method	β	n	α'	β'_1, β'_{-1}
[-1, 0, 1]	z weights	0.005	18	0.00075	0.0091
[-1, 0, 1]	w weights	0.005	18	0.00062	0.0106
[-1, 0, 1]	MLE	0.005	20	0.0015	0.0025
[-1, 0, 1]	GLR	0.005	24	0.0011	0.0005
[-1, 0, 1]	z weights	0.05	14	0.00076	0.049
[-1, 0, 1]	w weights	0.05	14	0.00077	0.049
[-1, 0, 1]	MLE	0.05	14	0.0015	0.038
[-1, 0, 1]	GLR	0.05	17	0.0011	0.0128
[-2, -1, 0, 1, 2]	z weights	0.005	17	0.00094	0.0046
[-2, -1, 0, 1, 2]	w weights	0.005	17	0.00098	0.0046
[-2, -1, 0, 1, 2]	z weights	0.05	12	0.0011	0.044
[-2, -1, 0, 1, 2]	w weights	0.05	12	0.0012	0.043

Table 3.6 shows that the weights based methods do very well in attaining the desired false alarm rate; however, when the desired β value is small, the weights based methods underestimate the required sample size. The table also shows that the asymptotically designed MLE approach gives quite conservative β values, but does not attain the desired false alarm rate. For the GLR method, the false alarm rate is only marginally larger than desired, but the β value is much lower. In other words, the required sample size was overestimated. As a conclusion, Table 3.6 suggests that much

of the differences in sample sizes reported in Table 3.2 are due to problems with the asymptotic solutions, and do not reflect on the true power of the various tests.

3.4.2 Designing Small Sample Size Plans and Charts

It is also possible to use the set $\{Q\}$ and subset $\{Q_r\}$ in a more proactive manner to design the sampling plans and control charts to have desired actual error rates. The following algorithm is recommended to determine the best sample size and control limits when designed control charts for grouped data with small sample sizes. Assume the group limits, the alternative parameter values, and the desired type I and type II error rates are given. The algorithm for the determination of the best n and λ is as follows:

1. Estimate n using the asymptotic results from Sections 3.3.1-3.3.4. Use this n as an initial estimate for the best sample size.

2. Using the current sample size, derive the set $\{Q\}$ by determining the test statistic under all possible sample outcomes. Also find the probability of occurrence of each element assuming the null hypothesis (i.e. derive the distribution of test statistic, e.g. $\bar{z}$ or $\bar{w}$, under the null hypothesis). Using the distribution of test statistic determine the control limit(s) (given as a range) that yield a type I error rate less than or equal to α, and the largest possible power. Note that for charts with two control limits determine the limits such that the false alarm rate is less than or equal to $\alpha/2$ in either direction.

3. Using the determined control limit(s) determine the subset $\{Q_r\}$, and the probability of occurrence of each element under the alternative hypothesis. From this information, calculate the actual type II error rate(s). For the one-sided test denote the type II error rate as β', for the two-sided test denote β'_1 and β'_{-1} as the type II error rates for upward and downward shifts respectively.

4. If β' or $\beta' = \max(\beta'_1, \beta'_{-1})$ is less than or equal to β the required sample size is n, otherwise increment n by one and repeat the algorithm starting at step 2.

Note that finding the set $\{Q\}$ by determining the test statistic under all possible sample outcomes is much easier for the weights based approaches than the MLE and GLR approaches. For the weights based approaches each sample is assigned a predetermined weight depending on the group it falls into. As a result, the set $\{Q\}$ is found simply by determining all the possible average weight values. For the MLE and GLR approaches we must determine the MLE under each possible sample. Finding the MLE requires an iterative process that is much more computationally intensive than simply finding an average. Consequently, although for each approach the same number of possible sample must be considered, the weights based approaches are more readily analyzed using the above algorithm.

However, an important question is: when is this more computationally intensive method warranted? For the CLT solution of the weights based approaches the sample size required to give good answers is highly dependent on the magnitude of β. For large β levels ($0.05 < \beta < 0.5$), the CLT solution, or the CLT solution with adjusted control limits (equations (3.6), (3.7), (3.26) and (3.27)), provided conservative error rates in most cases. This is due to the influence of rounding up the required sample size. For small β levels ($\beta < 0.05$), the skewness of the distribution of $\bar{z}$ or $\bar{w}$ is more important, since we are now trying to match the normal distribution to the extreme tail of a skewed distribution. Small deviations from normality of $\bar{z}$ or $\bar{w}$ can have a significant effect, and thus usually the required sample size must increment to create a control chart with conservative error rates. For MLE and GLR methods similar problems occur, although they are more related to the sample size and number of group limits than the β level.

To illustrate the use of this solution methodology for a two sided test consider the following example with three cases. Assume we desired a Shewhart control chart to detect a shift of 2 σ units in a standard normal mean with false alarm rate less than α =

0.001, utilizing a three step gauge defined by t = {-1, 0, 1} (i.e. group weights w = -6.4, -1.8, 1.8 and 6.4). Since, in this example, the gauge limits are symmetric about $\mu_0 = 0$, $n_U = n_L$ and $\lambda_U = -\lambda_L$ and as a result only the values for n_U and λ_U are reported. Table 3.7 presents the results for β = 0.5, 0.1 and 0.001. For each proposed design we determine the actual error rates α' and β' using enumeration. First use equations (3.22) and (3.23) to determine the CLT solution for the required sample size and control limit. If the resulting actual error rates are not sufficiently small (i.e. if $\alpha' > \alpha$ or $\beta' > \beta$) use (3.26) to adjust the control limit. If the error rates are still too large, use the enumeration algorithm presented to find the optimal location for the control limit, and increment n until the desired error rates are obtained.

Table 3.7: Error rates for different designs

β	solution type	n	λ_U	α'	β'
0.5	CLT	6	5.6	0.0004	0.3
0.1	CLT	8	4.7	0.0004	0.13
0.1	(3.26) adjust	8	4.58	0.0008	0.07
0.001	CLT	12	3.8	0.0005	0.0058
0.001	(3.26) adjust	12	3.74	0.0008	0.0033
0.001	enumeration	12	3.6	0.0009	0.003
0.001	enumeration	13	3.5	0.0009	0.0013
0.001	enumeration	14	3.35	0.0009	0.0005

Note that in all cases the α' level was sufficiently small. However, due to skewness, when β = 0.001, although the CLT solution suggested n = 12, we actually required n = 14 for both α' and β' to be less than 0.001.

CHAPTER 4
Parameter Estimation under Destructive Testing

This chapter turns to a different but related application of grouped data to SPC. We wish to estimate the correlation between two strength properties that can only be measured destructively, as described in Section 1.4. Previous researchers (De Amorim, 1982, Evans et al. 1984, De Amorim and Johnson, 1986) considered a strategy where breaking strength recorded only as pass/fail relative to a proof-load in one strength mode and measured precisely in the second strength mode. This strategy will be refered to as the De Amorim method in this thesis. This extends the idea of grouping data to multiple dimensions through the use of proof-loads, possibly more than one, in each strength modes. As in the De Amorim method, it is assumed that the variables have a bivariate normal distribution and that survivors of proof-loads are not significantly damaged.

Section 4.1 presents two correlation estimation procedures that utilize a single proof-load in each strength mode. These procedures are comparable to De Amorim method, but are easier and cheaper to use since they require only grouped data and no precise measurements and also avoid some wastage of material. Recording only either pass or fail at a set load is very natural for destructive strength tests as it is often difficult to measure breaking strength precisely. Moreover, Proof-loading generally does not require sophisticated measuring equipment, and can be done more quickly and easily. Section 4.2 extends the concept of two dimensional grouping further. The two procedures presented in Section 4.2 utilize two proof-loads in each of the strength modes. In this way, the number of groups used is increased. The resulting additional information allows the estimation of not only the correlation, but also the estimation of the means and

standard deviations of the individual strength variables. These procedures are thus more flexible than De Amorim method since they do not require as much prior information, and may also provide better correlation estimates.

4.1 Estimation of the Correlation Only

First consider the case where we wish only to estimate the correlation. This is the situation of interest in most past work in the area. Section 4.1.1 presents Procedure I, the simplest adaptation of the work of Evans et al. (1984). As such, in Section 4.1.1 I also adopt the assumption from Evans et al. (1984) that good prior estimates of the means and standard deviations of the bivariate distribution are available.

Section 4.1.2 introduces a slightly more complicated procedure, called Procedure II, where part of the sample has the proof-tests applied in the reverse order. In this way more information is available, and MLEs for the probabilities of failing the proof-loads are obtainable. As a result, the MLE of the correlation is obtainable without estimates for the means and standard deviations. Thus for Procedure II prior estimates for the means and standard deviations are not needed.

The two procedures are analyzed in detail, and compared with past results from the literature in Section 4.1.3. In general, Procedure I provides slightly better correlation estimates than Procedure II. However, this is only true if the prior estimates of the individual means and standard deviations are accurate. In practice, Procedure II is recommended.

4.1.1 Procedure I: One-way Estimation

In developing Procedure I to estimate the correlation between two variables that can individually only be measured destructively it is assumed that the individual strength properties have known means and standard deviations, μ_a, μ_b, σ_a, σ_b that have been

determined from prior data. However, to simplify the testing required, rather than recording the precise load at failure for each unit in the sample as in previous studies, a grouped data approach is adopted and the number of units that fail each of the two proof-loads is recorded. Procedure I is performed as follows:

1) Start with a sample of size n. Load each unit up to a proof-load of PL_a in variable A. Let p_a equal the probability of failure on this load, i.e. $p_a = \Phi\left((PL_a - \mu_a)/\sigma_a\right)$, where Φ is the cumulative standard normal probability. Denote the number of units that break under this first proof-load as n_a.

2) Subject the remaining $n - n_a$ units to a proof-load of PL_b on mode B, where $p_b = \Phi\left((PL_b - \mu_b)/\sigma_b\right)$ equals the probability of failure. Let n_b equal the number of units that fail this second proof-load.

With this procedure there are $n - n_a - n_b$ units that fail neither proof-load. As a result, the one-way procedure divides the observations into 3 distinct groups. The one-way procedure subjects n units to a proof-load in mode A, followed by a proof-load in mode B for the survivors of the first proof-load. Normally, using one proof-load in each of two modes would allow us to classify each unit into one of four groups. However, since our measurement is destructive, it is not possible to observe units that fail in both proof-loads. Figure 4.1 represents the groups created where the horizontal axis represents the breaking strength in mode A and the vertical axis represents the breaking strength in mode B. The values n_a, n_b and n - n_a- n_b are the observed number of units falling into each of the groups.

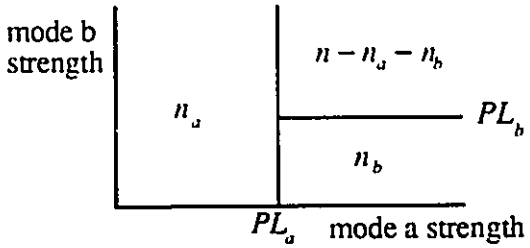

Figure 4.1: Grouping for One-way Procedure

The likelihood of obtaining any given n_a and n_b can be written as:

$$L(n,n_a,n_b) = p_a^{n_a}\left((1-p_a)p_{b|\bar{a}}\right)^{n_b}\left((1-p_a)\left(1-p_{b|\bar{a}}\right)\right)^{n-n_a-n_b}$$

$$L(n,n_a,n_b) = p_a^{n_a}\left(p_b - p_{a\cap b}\right)^{n_b}\left(1-p_a-p_b+p_{a\cap b}\right)^{n-n_a-n_b}, \qquad (4.1)$$

where $p_{b|\bar{a}}$ equals the probability of failure on mode B given that the unit did not fail on mode A, and $p_{a\cap b}$ equals the theoretical probability that a unit would have failed both proof-loads PL_a and PL_b. Solving for $p^*_{a\cap b}$, the maximum likelihood estimate (MLE) of $p_{a\cap b}$, we get

$$p^*_{a\cap b} = p_b - \frac{n_b(1-p_a)}{n-n_a} \qquad (4.2)$$

Using equation (4.2) it is possible, although unlikely unless proof-load levels are very small, that the estimate $p^*_{a\cap b}$ is negative. This makes no physical sense, and it is recommended that if $p^*_{a\cap b} < 0$, we set $p^*_{a\cap b} = 0$ and proceed.

Since $p_{a\cap b}$ is a function of p_a, p_b, and ρ_{ab} it is possible to solve for the MLE of the correlation, ρ^*_{ab}, that corresponds to the given values of p_a, p_b and the MLE $p^*_{a\cap b}$. Defining $z_a = (x-\mu_a)/\sigma_a$ and $z_b = (y-\mu_b)/\sigma_b$ we have $p_{a\cap b}$ equal to the bivariate normal integral:

$$p_{a\cap b} = \int_{y=PL_b}^{\infty}\int_{x=PL_a}^{\infty} \frac{1}{2\pi\sigma_a\sigma_b\sqrt{1-\rho_{ab}^2}}\exp\left(-\frac{1}{2}\frac{z_a^{\,2}-2\rho z_a z_b+z_b^{\,2}}{1-\rho_{ab}^2}\right)dxdy$$

After standardizing the variables, and defining $h = (PL_a - \mu_a)/\sigma_a$ and $k = (PL_b - \mu_b)/\sigma_b$ or equivalently finding h and k such that $p_a = \Pr(z \geq h)$ and $p_b = \Pr(z \geq k)$ where z is a standard normal variate, we get

$$\begin{aligned} p_{a\cap b} &= \int_{y=k}^{\infty}\int_{x=h}^{\infty} \frac{1}{2\pi\sqrt{1-\rho_{ab}^2}} \exp\left(-\frac{1}{2}\frac{x^2-2\rho xy+y^2}{1-\rho_{ab}^2}\right)dxdy \\ &= L(h,k,\rho_{ab}) \\ &= f(p_a,p_b,\rho_{ab}), \end{aligned} \tag{4.3}$$

where $L(h,k,\rho_{ab})$ is the standard bivariate normal integral, as defined by formula 26.3.3 of Abramowitz and Stegun (1972). For any given values of h, k, and ρ_{ab} a simple numerical procedure given in Drezner and Wesolowsky (1990) will determine $p_{a\cap b}$.

The slope of $L(h,k,\rho)$ is given by Drezner and Wesolowsky (1990) as:

$$\frac{\partial L(h,k,\rho)}{\partial \rho} = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{h^2-2\rho hk+k^2}{2(1-\rho^2)}\right) \tag{4.4}$$

This expression is strictly positive, therefore $f(p_a,\ p_b,\ \rho_{ab})$ is a strictly increasing function as ρ_{ab} ranges from -1 to +1. Thus we can search for the MLE ρ_{ab}^* value that corresponds to the determined $p_{a\cap b}^*$ using the method of bisection. The search stops when changes in ρ_{ab} are less than 0.00001.

In summary, for Procedure I, the determination of the MLE ρ_{ab}^* is performed as follows:

1. Choose a sample size n, and failure probabilities in the strength modes A and B, p_a and p_b respectively.

2. Perform experiment as outlined to obtain n_a and n_b.

3. Use expression (4.2) to obtain the MLE $p_{a\cap b}^*$. If $p_{a\cap b}^* < 0$ set $p_{a\cap b}^* = 0$.

4. Determine the values for h and k such that $p_a = \Pr(z \geq h)$ and $p_b = \Pr(z \geq k)$ where z is a standard normal variate.

5. Using the calculated $p^*_{a\cap b}$, h, k and the method of bisection find the value of ρ^*_{ab} that corresponds using equation (4.3).

The MLE $p^*_{a\cap b}$, as derived in equation (4.2), is defined only if $n_a < n$, in other words, if not all units fail under the first proof-load. If $n_a = n$, none of the units are subjected to any proof-loading in the second strength mode, and as a result, no information is obtained about the correlation between strength modes. We shall proceed with the analysis assuming that $n_a \neq n$, concluding that if $n_a = n$ the procedure provided no information. Fortunately, since the proof-load levels can be set at any desired values, the likelihood of observing a situation where $n_a \neq n$ is very rare even for small sample sizes. For example, given a proof-load level p_a equal to 0.6, the probability of observing n failures on this proof-load from a sample of size n is $(0.6)^n$, which is 8×10^{-12} for $n =$ 50.

The expected value and standard deviation of $p^*_{a\cap b}$ can be determined using conditioning arguments. The following Lemma will be useful in the upcoming theorem.

Lemma: $$E\left(\frac{n_b}{n-n_a}\right) = \frac{p_b - p_{a\cap b}}{1-p_a}, \qquad \text{assuming } n_a \neq n$$

We know $$E\left(n_b \mid n_a\right) = (n-n_a)p_{b|\bar{a}}$$

Therefore, $$E\left(\frac{n_b}{n-n_a} \mid n_a\right) = p_{b|\bar{a}} = \frac{p_b - p_{a\cap b}}{1-p_a}$$

which does not depend on n_a so the unconditional expectation is as given.

Theorem: The estimate $p^{*}_{a\cap b}$ is unbiased, assuming $n_a \neq n$.

Proof:
$$\begin{aligned} E\left(p^{*}_{a\cap b}\right) &= E\left(p_b - n_b(1-p_a)/(n-n_a)\right) \\ &= p_b - (1-p_a)E\left(n_b/(n-n_a)\right) \end{aligned}$$

by the above Lemma
$$\begin{aligned} E\left(p^{*}_{a\cap b}\right) &= p_b - (1-p_a)\frac{p_b - p_{a\cap b}}{1-p_a} \\ &= p_{a\cap b} \end{aligned}$$

The variance of the MLE $p^{*}_{a\cap b}$ can also be determined. The following ignores the recommended truncation of the estimate $p^{*}_{a\cap b}$ at zero, and as result is only valid when such truncation is unlikely. In any event, the practical effect of the truncation is to reduce the variance of the estimate.

Using the conditional variance formula gives

$$\begin{aligned} \mathrm{Var}\left(p^{*}_{a\cap b}\right) &= \mathrm{Var}\left(p_b - \frac{n_b(1-p_a)}{n-n_a}\right) \\ &= (1-p_a)^2\,\mathrm{Var}\left(\frac{n_b}{n-n_a}\right) \\ &= (1-p_a)^2\left(\mathrm{Var}\left(E\left(\frac{n_b}{n-n_a}\middle| n_a\right)\right) + E\left(\mathrm{Var}\left(\frac{n_b}{n-n_a}\middle| n_a\right)\right)\right) \\ &= (1-p_a)^2\left(\mathrm{Var}\left(p_{b|\bar{a}}\right) + E\left(\frac{p_{b|\bar{a}}\left(1-p_{b|\bar{a}}\right)}{n-n_a}\right)\right) \\ &= (1-p_a)^2\left(0 + p_{b|\bar{a}}\left(1-p_{b|\bar{a}}\right)E\left(\frac{1}{n-n_a}\right)\right) \\ &= (p_b - p_{a\cap b})(1-p_a-p_b+p_{a\cap b})E\left(\frac{1}{n-n_a}\right) \end{aligned}$$

Expressions of the form $E\left(x^{-1}\right)$, where x is a binomial variate bounded away from zero have been studied (Johnson and Kotz, 1969, Section 3.10). The approximation suggested

by Grab and Savage (1954), namely $E(x^{-1}) \equiv (np-q)^{-1}$ gives two significant figures of accuracy for $np > 10$, and is more than adequate for our application. Bearing in mind the restriction that $n_a \neq n$, the variable $n - n_a$ is such a binomial variate with sample size n, and probability of success $1 - p_a$. Thus using the Grab and Savage (1954) approximation, $p^*_{a\cap b}$ can be closely approximated when $n(1-p_a) > 10$, by:

$$\mathrm{Var}(p^*_{a\cap b}) \equiv \frac{(p_b - p_{a\cap b})(1 - p_a - p_b + p_{a\cap b})}{n(1-p_a) - p_a} \tag{4.5}$$

The estimates $p^*_{a\cap b}$ are translated to ρ^*_{ab} through the relation $p_{a\cap b} = f(\rho_{ab}; p_a, p_b)$. However, the function $f(\rho_{ab})$ is not a simple linear relation. Thus, unfortunately, the MLE ρ^*_{ab} is not unbiased in general. However, simulation studies presented later in this section suggest that the bias of ρ^*_{ab} is approximately zero, and that the bias is an insignificant part of the estimate's mean squared error for most proof-load levels. Fortunately, the standard deviation of ρ^*_{ab} can be estimated. For ρ_{ab} values away from the extremes ± 1, $f(\rho_{ab})$ can be closely approximated by a linear function. As a result, for ρ_{ab} away from ± 1, the standard deviation of ρ^*_{ab} can be closely approximated using the δ-method, also called the method of statistical differentials, (Johnson and Kotz, 1989). Letting $g = f^{-1}$, so that $\rho_{ab} = g(p_{a\cap b}; p_a, p_b)$, and using the Taylor expansion we can derive:

$$\rho^*_{ab} = g(p^*_{a\cap b}) \equiv g(p_{a\cap b}) + \frac{\partial g}{\partial p_{a\cap b}}(p^*_{a\cap b} - p_{a\cap b})$$

$$\text{so} \qquad \mathrm{Var}(\rho^*_{ab}) = \mathrm{Var}(g(p^*_{a\cap b})) \equiv (\partial g/\partial p_{a\cap b})^2 \mathrm{Var}(p^*_{a\cap b})$$

Assuming that $f(p_a, p_b, \rho_{ab})$ can be locally approximated by a straight line gives

$$\mathrm{Var}(\rho^*_{ab}) \equiv (\partial f/\partial \rho_{ab})^{-2} \mathrm{Var}(p^*_{a\cap b}).$$

where h and k are such that for $z \sim N(0,1)$, $\Pr(z \geq h) = p_a$ and $\Pr(z \geq k) = p_b$. Thus, using equations (4.4) and (4.5), the standard deviation of ρ^*_{ab} can be approximated as:

$$sd(\rho^*_{ab}) = \frac{\text{sqrt}(\text{Var}(p^*_{a\cap b}))}{\partial f / \partial \rho_{ab}} \qquad (4.6)$$

$$\equiv \frac{2\pi\sqrt{1-\rho^2_{ab}}}{\exp\left(-\dfrac{h^2 - 2hk\rho_{ab} + k^2}{2(1-\rho^2_{ab})}\right)} \sqrt{\frac{(p_b - p_{a\cap b})(1 - p_a - p_b + p_{a\cap b})}{n(1-p_a) - p_a}}$$

Notice that equation (4.6) cannot be computed unless the true correlation level ρ_{ab} is known. However, we have found that using the computed MLEs ρ^*_{ab} and $p^*_{a\cap b}$ in equation (4.6) provides a good estimate of $sd(\rho^*_{ab})$ based solely on the sample data.

We will now explore the effect of proof-load levels, actual correlation values, and sample sizes on the variability and bias of our correlation estimate in more detail. For the purposes of illustration, assume without loss of generality that variables A and B are distributed as standard normals, i.e. assume that $\mu_a = \mu_b = 0$ and $\sigma_a = \sigma_b = 1$. The estimate ρ^*_{ab} does not depend directly on μ_a, μ_b, σ_a and σ_b, the means and standard deviations of the strength in modes A and B, but depends only on the failure proportions p_a and p_b. As a result, given any mean and standard deviation an appropriate proof-load level can always be chosen such that any desired proportion of units would be expected to fail.

The sensitivity of the estimate ρ^*_{ab} to changes in proof-load levels p_a and p_b can be shown in 3-D surface plots. Figures 4.2 and 4.3 show surface plots, given $n = 300$ and $\rho_{ab} = 0.6$, of the bias and standard deviation of ρ^*_{ab} respectively, where for each combination of p_a and p_b 10000 trials are used. Figure 4.2 suggests that the correlation estimate obtained with Procedure I is unbiased for a large range of p_a and p_b values; only for extreme combinations where one proof-load is high and the other low does the bias

deviate significantly from zero. In addition, the bias of ρ^*_{ab} is very small compared with its standard deviation for most proof-load levels. This implies, that in most cases, the bias is an insignificant part of the estimate's mean squared error. Figure 4.3 illustrates an important property of the proposed procedure: the standard deviation of our estimate is relatively insensitive to changes in proof-load levels near the optimal values for p_a and p_b. This is shown by the large flat section near the minimum.

The question of how large a sample size is required can be addressed from two standpoints. First of all, we would like our estimate ρ^*_{ab} to be approximately unbiased. Second, we would like the standard deviation of our estimate to be sufficiently small. The effect of the sample size n on the standard deviation of ρ^*_{ab} is clearly demonstrated through equation (4.6). The standard deviation of ρ^*_{ab} decreases as a function of $1/\sqrt{n}$ as n increases. The effect of n on the bias of ρ^*_{ab} is more difficult to quantify. However, through a simulation study, the absolute value of the bias of ρ^*_{ab} decreases approximately as a function of $1/n$ as n increases. In any case, for near optimal proof-load levels the bias of ρ^*_{ab} is insignificant compared with its standard deviation for any sample size. Figure 4.4 shows the simulated bias of the estimate for two extreme cases. One curve represents the case $p_a = 0.2$, $p_b = 0.2$, $\rho_{ab} = 0.6$, in other words, when the proof-load levels are far from optimal. The other curve shows the bias when $p_a = p_b = \rho_{ab} = 0.6$, i.e. near optimal proof-load levels. The figure suggests that even for poor proof-load levels the bias is virtually eliminated by sample sizes greater than 100.

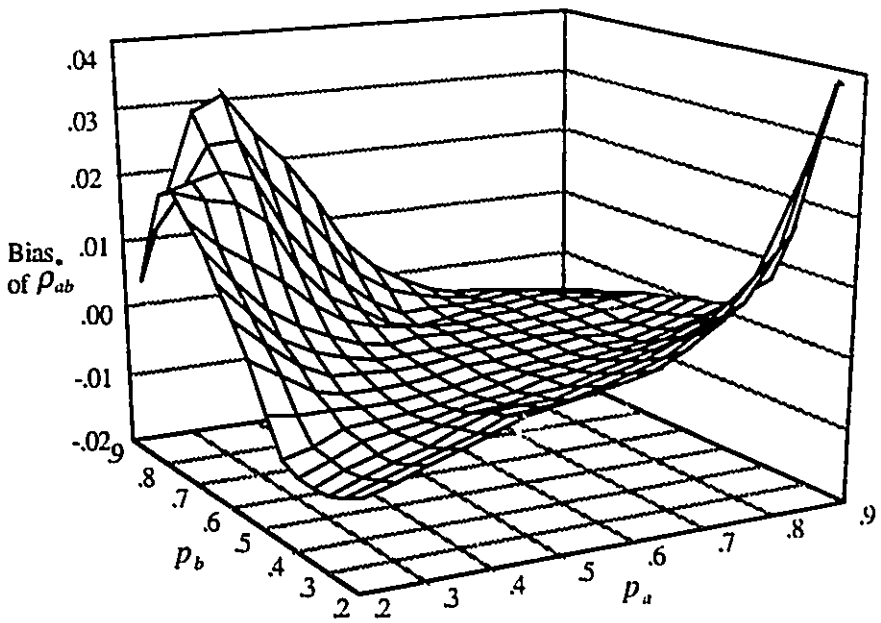

Figure 4.2: Bias of Simulated ρ^*_{ab} - Procedure I
$n = 300$, $\rho_{ab} = 0.6$, 10000 samples

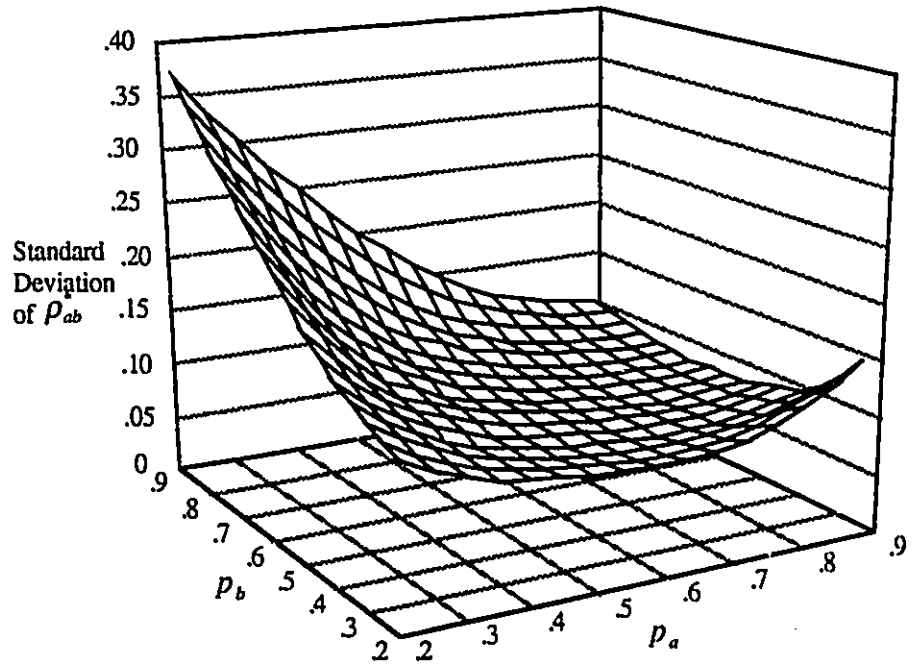

Figure 4.3: Standard Deviation of Simulated ρ^*_{ab} - Procedure I
$n = 300$, $\rho_{ab} = 0.6$, 10000 samples

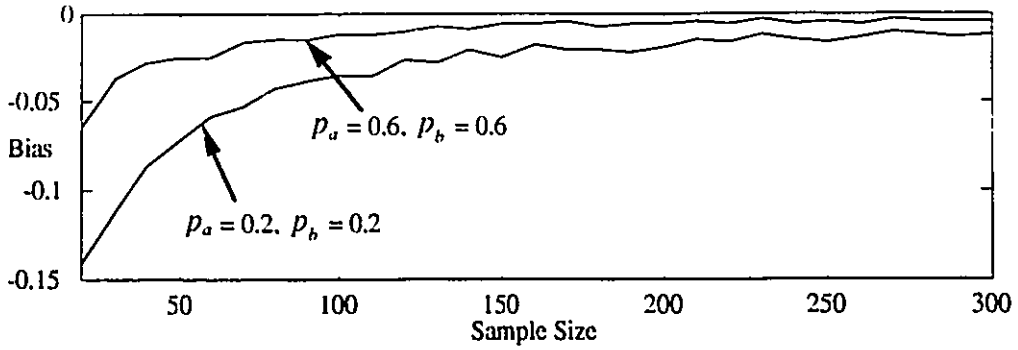

Figure 4.4: Bias of Simulated ρ^*_{ab} versus n - Procedure I, $\rho_{ab} = 0.6$

Another factor that has a significant influence on the standard deviation of the derived correlation estimate is the true correlation value. Figure 4.5 explores, using our theoretical result (4.6), how changes in sample size and the actual ρ_{ab} affect the standard deviation of ρ^*_{ab}, and can guide the practitioner in choosing an appropriate sample size. The estimation procedure works best when the real correlation is strongly positive or negative. The curves for negative correlations correspond almost exactly to the curves for positive correlations with the same absolute value. Simulation studies suggest that the actual correlation value has little effect on the bias of ρ^*_{ab} for near optimal proof-load levels.

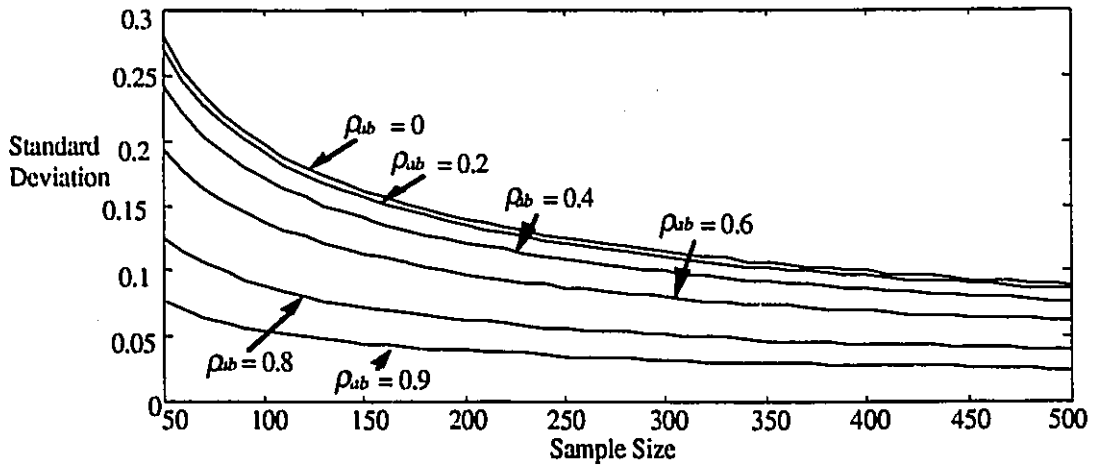

Figure 4.5: Standard Deviation of ρ^*_{ab} versus n - Procedure I
optimal proof-load levels are used, see Section 5.4

Another interesting concern relates to the effect on the estimation procedure of incorrect initial values for p_a and p_b. This question is explored in Figure 4.6, which shows how Procedure I will bias its estimate for the correlation if prior estimates for p_a and p_b are incorrect. For example, say we believe our current proof-load level will result in $p_b = 0.4$, but in reality $p_b = 0.5$, i.e. $\Delta p_b = -0.1$. Figure 4.6 shows that this incorrect assumption leads to correlation estimates that are on average around 0.19 lower than the true value. As shown, the effect of using incorrect p_a or p_b values can be very significant, and this is a major problem with any estimation procedure that relies on prior estimates.

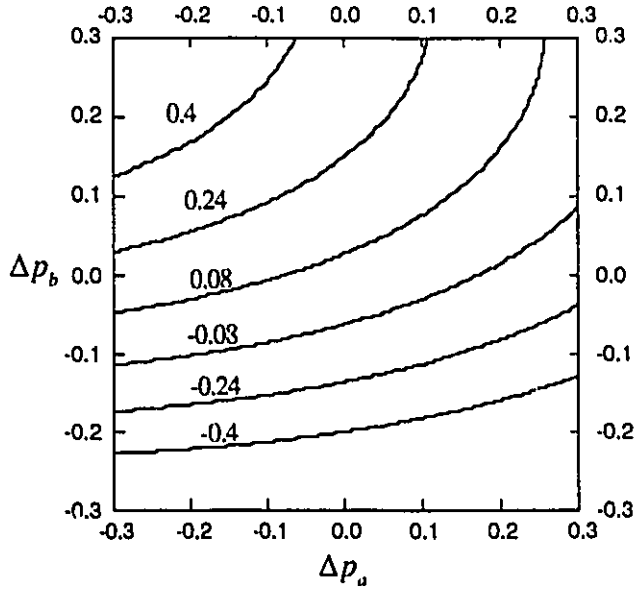

Figure 4.6: Contour Plot of the ρ^*_{ab} Bias - Procedure I
When Prior Estimates of p_a and p_b are Incorrect
Δp_a and Δp_b are deviations from the actual values for p_a and p_b
$n = 300$, $p_a = 0.7$, $p_b = 0.5$, $\rho_{ab} = 0.6$

This Section concludes with an exampie. Suppose we are interested in estimating the correlation between the bending and tension strength of lumber. From prior testing, or experience, we can set proof-load level in the two strength modes at any probability of

failure level desired. Following the summary of Procedure I resulted in the following steps:

- A sample of 300 lumber specimens were taken (i.e. n = 300), and the proof-load levels in bending and tension were chosen such that $p_a = 0.65$, $p_b = 0.45$.
- The experiment resulted in 190 units failing under the proof-load in bending, and 23 units of the remaining 110 units failing under the proof-load in tension. Therefore, n_a = 190 and $n_b = 23$.
- Using these variables and equation (4.2) gave $p^*_{a\cap b} = 0.377$
- Solving 0.65 = Pr($z \geq h$) and 0.45 = Pr($z \geq k$) where z is a standard normal variate for h and k gave: $h = -0.385$ and $k = 0.126$.
- Solving 0.377 = L(-.385, 0.126, ρ^*_{ab}) for the correlation estimate using the method of bisection gaves $\rho^*_{ab} = 0.557$. From equation (4.6), and our estimate ρ^*_{ab} = 0.557, we estimated $sd(\rho^*_{ab}) = 0.084$.

These results are quite good, since this example was generated randomly using a bivariate normal distribution with p_a = 0.65, p_b = 0.45, and ρ_{ab} = 0.5 (thus $p_{a\cap b}$ = 0.368). Simulation results we generated in MATLAB (MathWorks, 1991) suggest that $sd(\rho^*_{ab})$ = 0.091.

4.1.2 Procedure II: Symmetric Procedure

In developing the one-way estimation procedure of the last section, it was assumed that good prior estimates of the means and variances of the two characteristics are available. In practice, these estimates may either be inaccurate due to quality changes or unavailable due to a lack of sufficient prior knowledge. The symmetric procedure outlined below alleviates this difficulty by using the results of the proof-loads to estimate not only $p_{a\cap b}$, but also p_a and p_b. The simplicity of the one-way procedure is retained by still considering only one proof-load level in each mode. However reversing the order

of the proof-loads for some of the units results in more information. A similar type of procedure has been suggested by Green and Evans (1983). They propose extending the Amorim (1982) procedure to be symmetric. In other words, they split the sample in half, subjecting half the sample to a proof-load in mode A followed by stress until failure in mode B, and subjecting the other half to proof-load in mode B followed by stress until failure in mode A. Green and Evans (1983) report good results in estimating all five parameters of a bivariate normal distribution, but do not present their results or analysis.

The proposed symmetric procedure is outlined below:

1) Start with a sample of size $n+m$.
2) Perform the one-way procedure of Section 4.1.1 on n units.
3) Perform the one-way procedure in reverse order on m units.

Following this procedure, the first n units are subject to a proof-load in mode A followed by a proof-load in mode B, whereas the second m units are tested first in mode B, followed by a test in mode A. Denote the number of the first n units that break under proof-loads PL_a and PL_b as n_a and n_b respectively, and denote the number of the m units, in the reverse order test, that fail under the proof-loads PL_b and PL_a as m_b and m_a respectively. Group the first n observations as in the one-way procedure (see Figure 4.1), and classify the second m units into groups as shown in Figure 4.7 below.

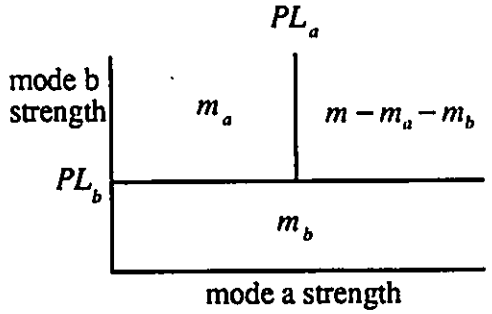

Figure 4.7: Grouping for Symmetric Procedure

Using the same notation as for equation (4.1), the likelihood function for the symmetric procedure is

$$\begin{aligned} L(n,n_a,n_b,m_a,m_b) &= p_a^{n_a}\left((1-p_a)p_{b|\bar{a}}\right)^{n_b}\left((1-p_a)\left(1-p_{b|\bar{a}}\right)\right)^{n-n_a-n_b} \\ &\quad p_b^{m_b}\left((1-p_b)p_{b|\bar{a}}\right)^{m_a}\left((1-p_b)\left(1-p_{b|\bar{a}}\right)\right)^{n-m_a-m_b} \\ &= p_a^{n_a}\left(p_b-p_{a\cap b}\right)^{n_b}\left(1-p_a-p_b+p_{a\cap b}\right)^{n-n_a-n_b} \\ &\quad p_b^{m_b}\left(p_a-p_{a\cap b}\right)^{m_a}\left(1-p_a-p_b+p_{a\cap b}\right)^{m-m_a-m_b} \end{aligned}$$

With this symmetric procedure more information is available, and are now it is possible to derive the MLE for not only for $p_{a\cap b}$ but also for p_a and p_b. We get the MLEs:

$$\begin{aligned} p_a^{\bullet} &= \frac{n_a(n_a+n_b+m_a+m_b)}{(n+m)(n_a+n_b)}, \\ p_b^{\bullet} &= \frac{m_b(n_a+n_b+m_a+m_b)}{(n+m)(m_a+m_b)}, \\ p_{a\cap b}^{\bullet} &= \frac{(m_b n_a-m_a n_b)(n_a+n_b+m_a+m_b)}{(n+m)(n_a+n_b)(m_a+m_b)}. \end{aligned} \tag{4.7}$$

Note, that it is possible that $p_{a\cap b}^{\bullet}$ is less than zero, although this makes no physical sense. However, unless p_a and p_b are quite small, it is very unlikely; normally we will observe $n_a > m_a$ and $m_b > n_b$. This is because n_a and m_b represent the number of units that fail the first proof-load, and any units that fail in the first proof-load are not available for testing in the second proof-load. If the experiment results in $p_{a\cap b}^{\bullet} < 0$, it is recommended that $p_{a\cap b}^{\bullet}$ is set to zero.

To obtain the estimate $\rho_{ab}^{\bullet}$ use the same methodology as presented in Section 4.1.1. However, now rather than using known values for p_a and p_b the MLEs $p_a^{\bullet}$ and $p_b^{\bullet}$

are used: i.e. solve $p^*_{a\cap b} = f\left(p^*_a, p^*_b, \rho^*_{ab}\right)$ (equation 4.3) for ρ^*_{ab} using the method of bisection. The determination of the MLE ρ^*_{ab} using Procedure II is performed as follows:

1. Choose sample sizes n and m, and proof-load levels in mode A and B.
2. Perform an experiment as outlined to obtain n_a, n_b, m_a, m_b.
3. Use (4.7) to obtain the MLEs p^*_a, p^*_b and $p^*_{a\cap b}$. If $p^*_{a\cap b} < 0$ set $p^*_{a\cap b} = 0$.
4. Determine the values for h and k such that $p^*_a = \Pr(z \geq h)$ and $p^*_b = \Pr(z \geq k)$ where z is a standard normal variate.
5. Using the calculated $p^*_{a\cap b}$, h, k and the method of bisection find the value of ρ^*_{ab} that corresponds by equation (4.3).

If the analysis is restricted to the case where $n_a + n_b \neq 0$ and $m_a + m_b \neq 0$, the MLEs given by equations (8) are all unbiased. This restriction is necessary, since if either $n_a + n_b = 0$ or $m_a + m_b = 0$, i.e. no units fail in either proof-load, not enough information is obtained, and two or more of the MLEs are undefined. The likelihood of observing either $n_a + n_b = 0$ or $m_a + m_b = 0$ depends on the probabilities of failure p_a and p_b. The probabilities of failure can not be chosen but even for the case where p_a and p_b are small, say 0.2, and the sample size is only 50, the likelihood of no units failing is 7×10^{-8} if $\rho_{ab} = 0.6$.

In appendix C, it is shown that the MLEs, given by equation (4.7) are all unbiased, in other words:

$$\mathrm{E}\left(p^*_a\right) = p_a$$
$$\mathrm{E}\left(p^*_b\right) = p_b$$
$$\mathrm{E}\left(p^*_{a\cap b}\right) = p_{a\cap b}$$

Expressions that closely approximate the standard deviation of the MLEs can also be derived. However, they are very complicated and have thus been omitted.

Determining properties of the correlation estimate ρ^*_{ab} is more difficult. The estimate ρ^*_{ab} depends on the derived MLEs, p^*_a, p^*_b and $p^*_{a\cap b}$, through the nonlinear relation represented by equation (4.3). In fact, even though the MLEs p^*_a, p^*_b and $p^*_{a\cap b}$ are unbiased (given $n_a + n_b \neq 0$ and $m_a + m_b \neq 0$), the MLE ρ^*_{ab} is not in general unbiased. Developing an estimate for the standard deviation of ρ^*_{ab} for Procedure II similar to the one derived for Procedure I proved to be too difficult. Instead, the Jackknife method (Efron, 1981) is used.

The Jackknife method provides a simple nonparametric way to estimate the standard deviation and bias of our estimate based solely on the observed sample. Bootstrap methods (Efron, 1981) could also be used, and may provide improved estimates, but are more complicated and require more calculations. The Jackknife method works on the premise of extending the formula for sample standard deviation to a more general case (Efron and Gong, 1983). This is done, in our case, by considering the correlation estimate that would have been obtained from each sub-sample of the original sample that has one observation removed. In most applications, the number of sub-samples is a function of the sample size. Fortunately, in our case, due to the discreteness of our data, the number of distinct subgroups depends on the number of possible outcomes. The estimate ρ^*_{ab} can be thought of as a function of the experimental outcome $\rho^*_{ab} = d(n_a, n_b, m_a, m_b, n+m)$, where $d()$ is a function that represents the result of performing steps 3-5 of the procedure to determine ρ^*_{ab} using Procedure II. Due to the discreteness of our problem only five distinct cases ρ^i_{ab} with corresponding weights w_i need be considered:

$$\begin{aligned}
\rho^1_{ab} &= d(n_a - 1, n_b, m_a, m_b, n+m-1), \qquad w_1 = n_a \\
\rho^2_{ab} &= d(n_a, n_b - 1, m_a, m_b, n+m-1), \qquad w_2 = n_b \\
\rho^3_{ab} &= d(n_a, n_b, m_a - 1, m_b, n+m-1), \qquad w_3 = m_a
\end{aligned} \tag{4.8}$$

$$\rho_{ab}^{4} = d(n_a, n_b, m_a, m_b - 1, n+m-1), \qquad w_4 = m_b$$

$$\rho_{ab}^{5} = d(n_a, n_b, m_a, m_b, n+m-1), \qquad w_5 = n+m-n_a-n_b-m_a-m_b$$

Based on these additional correlation estimates, the jackknife estimate of the bias ξ, and standard deviation S, of ρ_{ab}^{*} can be written:

$$\xi = \frac{n+m-1}{n+m}\sum_{k=1}^{5} w_k(\rho_{ab}^{*} - \rho_{ab}^{k})$$

$$S = \sqrt{\frac{n+m-1}{n+m}\sum_{k=1}^{5} w_k\left(\rho_{ab}^{*} - \rho_{ab}^{k}\right)^2} \qquad (4.9)$$

Using these jackknife estimates a bias adjusted estimate for the correlation can be derived, namely $\rho_{ab}^{\xi} = \rho_{ab}^{*} + \xi$. It is also possible to derive crude confidence intervals for ρ_{ab}^{*} based on the above results that ignore any possible deviations from normality.

It is of interest to determine the effect of sample size, proof-load levels and actual correlation values on the bias and standard deviation of the correlation estimate ρ_{ab}^{*} derived by Procedure II. The simulation results show that the bias of the symmetric procedure follows a very similar pattern to that exhibited by Procedure I. In other words, the bias is very small unless one or both the proof-loads are small (p_a or $p_b < 0.2$). Figure 4.8 explores the simulated standard deviation of ρ_{ab}^{*} for various proof-load levels given a sample size of 300 ($n = m = 150$), and a true correlation value of 0.6.

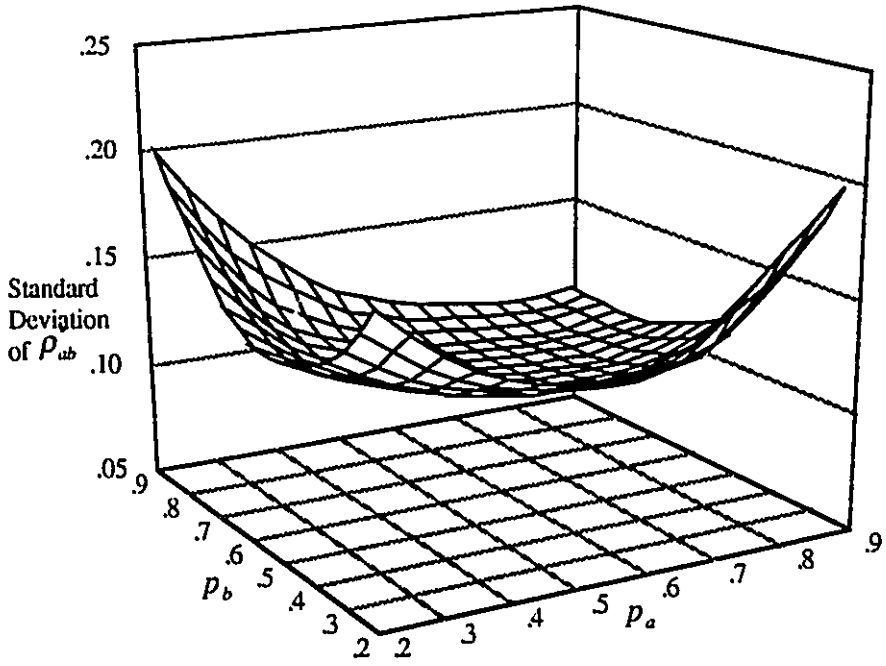

Figure 4.8: Standard Deviation of Simulated ρ^*_{ab} - Procedure II
$n = m = 150$, $\rho_{ab} = 0.6$, 10000 samples

Based on Figure 4.8, it is recommended that one try to choose proof-load levels that result in about 60% failures. Near these proof-load levels the standard deviation of ρ^*_{ab} is relatively insensitive to changes in p_a and p_b. This insensitivity is very important, since when the individual means and standard deviations are not known with certainty it is impossible to set proof-load levels exactly at a desired level.

The symmetric procedure also exhibits performance similar to Procedure I with regards to the effect of sample size (see Figures 4.4 and 4.5). However, since for the symmetric procedure it is not possible to set the proof-load probabilities p_a and p_b we are very interested in the effect they have on the procedure's estimates. Figures 4.9 and 4.10 show simulated results of standard deviation and bias of ρ^*_{ab} as a function of proof-load probabilities and true correlation levels.

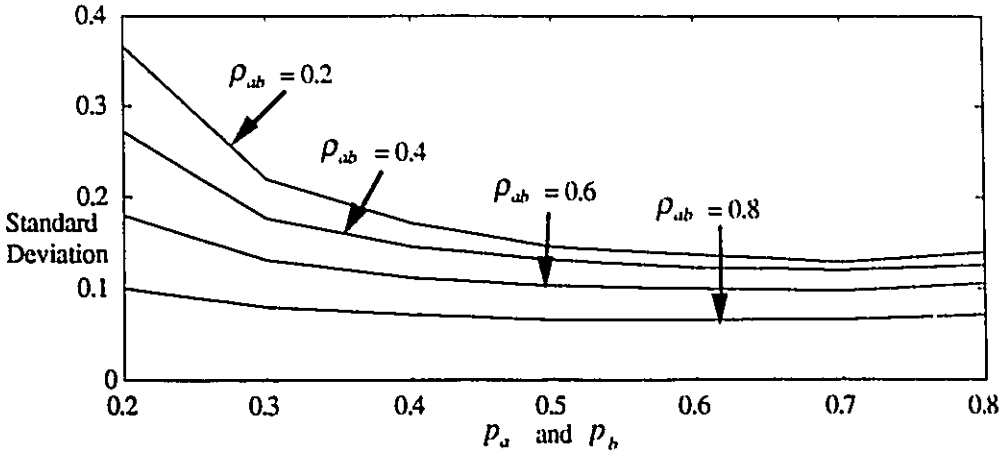

Figure 4.9: Standard Deviation of ρ^*_{ab} versus p_a and p_b - Procedure II
$n+m = 300$

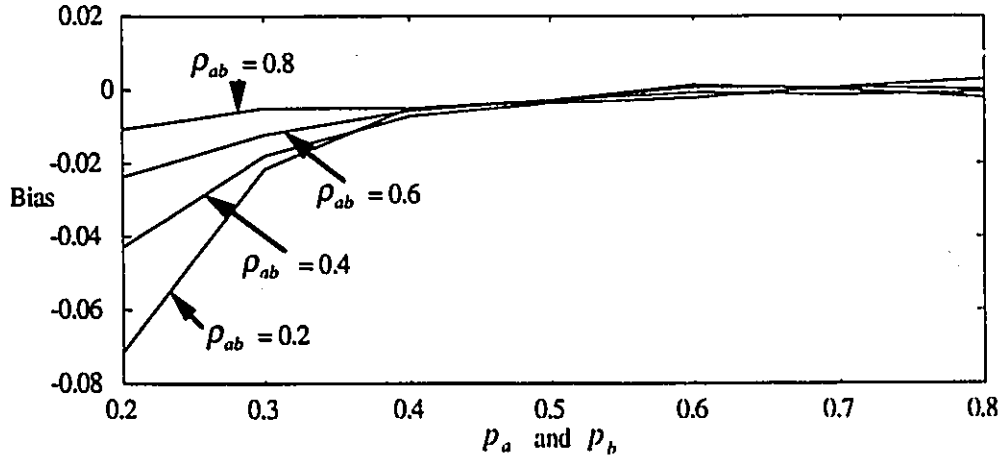

Figure 4.10: Bias of ρ^*_{ab} versus Proof-load Levels - Procedure II
$n+m = 300$

Clearly, the estimation procedure works best when ρ_{ab} is large, and p_a and p_b are not small. Based on the simulation results, it seems the estimation procedure works very well unless $p^*_{a\cap b}$ (or $m_b n_a - n_b m_a$) is likely to be negative or close to zero. When $p^*_{a\cap b}$ is close to zero, the procedure is not very stable, since small changes in the experimental outcome (n_a, n_b, m_a, m_b) results in large changes in the corresponding ρ^*_{ab} value. In this case, the small, yet discrete, changes in experimental outcome have the effect of introducing bias unless very large sample sizes are used. Fortunately, $m_b n_a - n_b m_a$ is only

likely to be close to zero when p_a or p_b or both are small, and ρ_{ab} is not strongly positive.

Unfortunately, the jackknife bias adjusted estimate provides no improvement when $m_b n_a - n_b m_a$ is close to zero. In that case, the jackknife estimation procedure encounters the same discreteness problems and tends to over-adjust. In fact, since in any other case the correlation estimate's bias is quite small, the jackknife bias adjustment is of little value. On the other hand, our simulation results suggest that the jackknife estimates for the standard deviation of the correlation estimate are usually very good. The estimate S, given by equation (4.9), matches closely with estimates obtained by repeated simulation, and exhibited only moderate variability.

This Section concludes with an example that illustrates the estimation procedure. Suppose we are interested in estimating the correlation between the bending and tension strength of lumber. Following the summary of Procedure II we performed the following steps:

- A sample of 300 lumber specimens was taken.
- Choosing $n = m = 150$ and running the experiment gave $n_a = 96$ and $n_b = 11$, $m_a = 42$ and $m_b = 65$.
- Substituting into equations (4.7) gave $p_a^* = 0.64$, $p_b^* = 0.43$, and $p_{a \cap b}^* = 0.36$.
- Solving $0.64 = \Pr(z \geq h)$ and $0.43 = \Pr(z \geq k)$ where z is a standard normal variate for h and k gave: $h = -0.359$ and $k = 0.164$.
- Solving $0.36 = L(-.359, 0.168, \rho_{ab}^*)$ for the correlation estimate using the method of bisection gave $\rho_{ab}^* = 0.547$. Using the jackknife procedure, equations (4.8) and (4.9) gave: $\rho_{ab}^1 = 0.546$, $\rho_{ab}^2 = 0.573$, $\rho_{ab}^3 = 0.556$, $\rho_{ab}^4 = 0.549$, $\rho_{ab}^5 = 0.544$. Which resulted in $\xi = -0.027$, $S = 0.113$, and $\rho_{ab}^{\xi} = 0.520$.

This example was generated randomly using a bivariate normal distribution with p_a = 0.65, p_b = 0.45 and ρ_{ab}= 0.5 (thus $p_{a\cap b}$= 0.368).

4.1.3 Comparison of Results with the De Amorim Method

Table 4.1 compares the best standard deviations of ρ^*_{ab} obtained by simulating the Evans et al. (1984) method and the proposed one-way and symmetric procedures. As expected, the standard deviations of ρ^*_{ab} for our procedures I and II are higher (about 20-25% and 40-60% respectively). However, this decrease in efficiency will be compensated by the lower cost and greater ease of implementation. Both Procedure I and II have the advantage of only requiring proof-loads in both strength modes. The advantage of Procedure II, the symmetric procedure, lies in its lack of need for accurate prior mean and standard deviation estimates. When such prior estimates are unavailable or incorrect the symmetric procedure may well outperform the one-way procedure and previously developed procedures like Evans et al. (1984). The one-way procedure is dependent on knowing the values p_a and p_b. The symmetric procedure, on the other hand, derives its own estimates for p_a and p_b.

Table 4.1: Comparison of Procedures I, II and the De Amorim Method

actual ρ_{ab}	De Amorim method $n = 300$	Proposed Procedure I $n = 300$	Proposed Procedure I $n = 365$	Proposed Procedure II $n+m = 300$
0.2	0.0895	0.1101	0.0972	0.1328
0.6	0.0649	0.0804	0.0709	0.0982
0.8	0.0400	0.0520	0.0456	0.0701
0.9	0.0257	0.0310	0.0284	0.0444

Moreover, since the proposed procedures use proof-loading in both modes some of the units tested are not destroyed and can be returned to the population of unused samples. The average number of units not destroyed is easily approximated assuming

that the p_a and p_b values used are close to optimal. For example, assuming $p_a = 0.65$ and $p_b = 0.5$ on average 17.5% of the units will not be destroyed. These units not only survive the testing, but are in fact the strongest units in the sample since they withstood two proof-loads. This argument relies on the assumption that the proof-loads do not significantly damage the units. For comparison, Table 4.1 also shows results for Procedure I with $n = 365$ (if 17.8% of the units do not fail, the procedure will destroy 300 units). The standard deviations of the correlation estimates for Procedure I are now only about 10% higher than those obtained with the De Amorim method.

4.2 Estimating all Five Bivariate Normal Parameters

The procedures suggested in Section 4.1 utilize one proof-load in each strength mode. A logical extension of this work, considering the idea of grouping, is to use more groups. This section presents two extensions that utilize two different proof-loads in each mode. The additional groups created provide more information about the underlying parameters of the distribution. Assuming a bivariate normal distribution, this additional information allows estimation of the means and standard deviations as well as the correlation. Also, for Procedure IV, the additional information available due to the extra groups provides for a better correlation estimate. This is an important improvement since often little reliable prior information is available and yet estimates for the correlation as well as the means and standard deviations of the two strength modes are desired. Procedure II gives estimates for the probability of failure under each proof-load, but does not allow estimation of the means and standard deviations of the individual strength modes.

Section 4.2.1 describes Procedure III. This procedure extends the symmetric procedure to allow the estimation of the means and standard deviations; however, it yields the same correlation estimates. Section 4.2.2 considers a further extension,

denoted Procedure IV, which provides the same mean and standard deviation estimates and also yields an improved correlation estimate. Both procedures require a sample of $n+m$ units. Define PL_{a1} and PL_{a2} as the two proof-load levels in mode A, where $PL_{a1} < PL_{a2}$. Similarly define PL_{b1} and PL_{b2} as the two proof-load levels in mode B, with $PL_{b1} < PL_{b2}$. Also denote p_{a1}, p_{a2}, p_{b1} and p_{b2} as the unknown probabilities of failure on the proof-loads PL_{a1}, PL_{a2}, PL_{b1} and PL_{b2} respectively.

4.2.1 Procedure III

Procedure III is performed as follows:

1) For the first n units perform the following proof-loads in order until either the unit breaks or all the tests are completed. Stress each unit up to a load of PL_{a1}, then apply a load of PL_{a2} followed by a load of PL_{b2}.

2) For the second m units perform the tests in the order PL_{b1}, PL_{b2} followed by PL_{a2} until the unit either breaks or all the tests are completed.

Let n_{a1}, $n_{a2}-n_{a1}$ and n_{b2} equal the number of units that fail or break under the each of the three proof-load tests applied on the first n units, and let m_{b1}, $m_{b2}-m_{b1}$ and m_{a2} equal the number of units that fail under the three proof-tests on the second m units. This classification scheme results in four distinct groups, and is represented graphically in Figure 4.11.

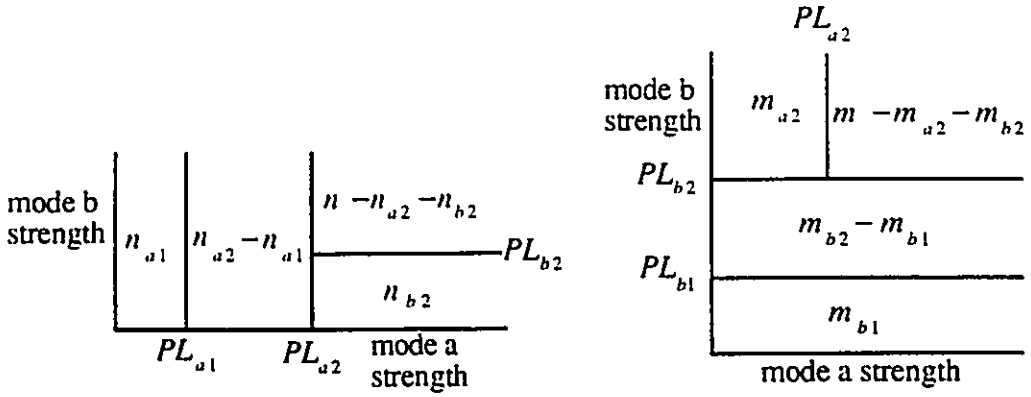

Figure 4.11: Grouping for Procedure III
classification of the n and m units is shown in the left and right figure respectively

The probability of observing any given n_{a1}, n_{a2}, n_{b2}, m_{b1}, m_{b2} and m_{a2} values depends on the probabilities of failure under each proof-load conditional on not having broken on any previous tests. This likelihood can be written:

$$\begin{aligned} L_{III} \;=\; & p_{a1}^{n_{a1}}\left(p_{a2}-p_{a1}\right)^{n_{a2}-n_{a1}}\left(p_{b2}-p_{a2\cap b2}\right)^{n_{b2}} \\ & p_{b1}^{m_{b1}}\left(p_{b2}-p_{b1}\right)^{m_{b2}-m_{b1}}\left(p_{a2}-p_{a2\cap b2}\right)^{m_{a2}} \\ & \left(1-p_{a2}-p_{b2}+p_{a2\cap b2}\right)^{n+m-n_{a2}-n_{b2}-m_{a2}-m_{b2}} \end{aligned} \qquad (4.10)$$

For example, $p_{b2} - p_{a2\cap b2}$ equals the probability that one of the first n units fails under the proof-load PL_{b2} given that it did not fail under either proof-load PL_{a1} or PL_{a2}. The likelihood function (4.10) depends on five unknowns: p_{a1}, p_{a2}, p_{b1}, p_{b2} and $p_{a2\cap b2}$. Taking derivatives, setting each derivative equal to zero, and solving the resulting system of equations yields the following MLEs:

$$\hat{p}_{a1} = \frac{n_{a1}\left(n_{a2}+n_{b2}+m_{a2}+m_{b2}\right)}{(n+m)\left(n_{a2}+n_{b2}\right)} \qquad \hat{p}_{a2} = \frac{n_{a2}\left(n_{a2}+n_{b2}+m_{a2}+m_{b2}\right)}{(n+m)\left(n_{a2}+n_{b2}\right)}$$

$$\hat{p}_{b1} = \frac{m_{b1}\left(n_{a2}+n_{b2}+m_{a2}+m_{b2}\right)}{(n+m)\left(m_{a2}+m_{b2}\right)} \qquad \hat{p}_{b2} = \frac{m_{b2}\left(n_{a2}+n_{b2}+m_{a2}+m_{b2}\right)}{(n+m)\left(m_{a2}+m_{b2}\right)} \qquad (4.11)$$

$$p^{*}_{a2\cap b2} = \frac{(m_{b2}n_{a2} - n_{b2}m_{a2})(n_{a2} + n_{b2} + m_{a2} + m_{b2})}{(n+m)(m_{a2} + m_{b2})(n_{a2} + n_{b2})}$$

All the MLEs given in equation (4.11), and the subsequent MLEs of Procedure IV equations (4.14), are unbiased. See Appendix C for details.

Based on these five MLEs estimation of all five bivariate normal parameters is possible. Denote the mean and standard deviation of the strength in mode A and B respectively as: μ_a, σ_a, μ_b and σ_b, and let ρ_{ab} be the correlation. Then, since $\Pr(X \le \mathrm{PL}_{a1}) = p_{a1}$ and $\Pr(X \le \mathrm{PL}_{a2}) = p_{a2}$ where $X \sim N(\mu_a, \sigma_a^2)$, using simple algebra gives the MLEs:

$$\mu^{*}_{a} = \frac{\mathrm{PL}_{a1}\Phi^{-1}(p^{*}_{a2}) - \mathrm{PL}_{a2}\Phi^{-1}(p^{*}_{a1})}{\Phi^{-1}(p^{*}_{a2}) - \Phi^{-1}(p^{*}_{a1})} \qquad \sigma^{*}_{a} = \frac{\mathrm{PL}_{a2} - \mathrm{PL}_{a1}}{\Phi^{-1}(p^{*}_{a2}) - \Phi^{-1}(p^{*}_{a1})}$$

$$\mu^{*}_{b} = \frac{\mathrm{PL}_{b1}\Phi^{-1}(p^{*}_{b2}) - \mathrm{PL}_{b2}\Phi^{-1}(p^{*}_{b1})}{\Phi^{-1}(p^{*}_{b2}) - \Phi^{-1}(p^{*}_{b1})} \qquad \sigma^{*}_{b} = \frac{\mathrm{PL}_{b2} - \mathrm{PL}_{b1}}{\Phi^{-1}(p^{*}_{b2}) - \Phi^{-1}(p^{*}_{b1})} \qquad (4.12)$$

Estimating the correlation ρ_{ab} is more difficult. But, since $p_{a2\cap b2} = f(p_{a2}, p_{b2}, \rho_{ab})$, the MLEs p^{*}_{a2}, p^{*}_{b2}, and $p^{*}_{a2\cap b2}$ can be used to estimate ρ^{*}_{ab}. For the purpose of estimating the correlation, Procedure III is identical to Procedure II. The MLEs p^{*}_{a2}, p^{*}_{b2}, and $p^{*}_{a2\cap b2}$ given in (4.10) are identical to the MLEs p^{*}_{a}, p^{*}_{b} and $p^{*}_{a\cap b}$ given in equation (4.7). Due to this similarity between procedures II and III, the results from Section 4.1.2 can be used to analyze the effect of different p_{a2} and p_{b2} values on the efficiency of the correlation estimate. Since all the bivariate normal parameters are unknown, it is not possible to set proof-load levels to give the desired p_{a2} and p_{b2} values. As previously recommended it is best to aim for p_{a2} and p_{b2} values around 0.6. Using results from Chapter 3, it is also possible to estimate the effect of different p_{a1}, p_{b1}, p_{a2} and p_{b2} values on the accuracy of the MLEs for mean and standard deviation. The

equations (3.9) and (3.10) give good estimates for the standard deviation of the MLEs μ_a^*, σ_a^*, μ_b^* and σ_b^*. The standardized gauge limits t_1 and t_2 can be estimated from the MLEs p_{a1}^*, p_{b1}^*, p_{a2}^* and p_{b2}^* by inverting the normal probability density function. For example, $p_{a1}^* = 0.3$ and $p_{a2}^* = 0.6$ gives $t_1 = -0.52$ and $t_2 = 0.25$. The sample size used to adjust equations (3.9) and (3.10) should be set equal to the number of units tested in a particular mode. This includes all the units that are subject to proof-loads in the mode of interest first, i.e. n units for mode A and m units for mode B, plus the number of units that survive testing in the other strength mode first and then are subject to proof-loads in the mode of interest. The best values for the probabilities of failure p_{a1}, p_{b1}, p_{a2} and p_{b2} to estimate the mean and/or the standard deviation of a normal distribution are given in Tables 5.3-5.6.

4.2.2 Procedure IV

Procedure IV is similar to III but includes an additional proof-load. Procedure IV is performed as follows:

1) For each of the first n units perform the proof-loads PL_{a1}, PL_{a2}, PL_{b1}, PL_{b2} continuing until the unit either fails one of the proof-loads or all the tests are completed. Let n_{a1}, $n_{a2} - n_{a1}$, n_{b1}, $n_{b2} - n_{b1}$ equal the number of the n units tested that failed the proof-loads PL_{a1}, PL_{a2}, PL_{b1} and PL_{b2} respectively.

2) For the second m units do the same tests, but change the order to PL_{b1}, PL_{b2}, PL_{a1}, PL_{a2} and let m_{b1}, $m_{b2} - m_{b1}$, m_{a1}, $m_{a2} - m_{a1}$ equal the number of the second m units that fail the proof-loads in the given order.

The resultant classification of units into groups by Procedure IV is shown graphically in Figure 4.12.

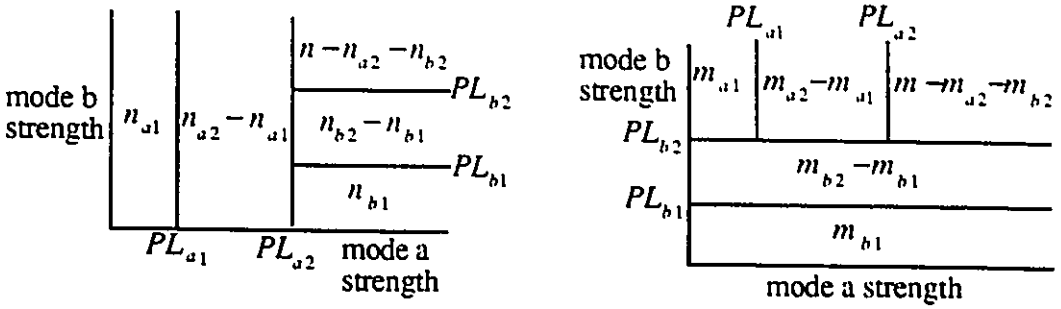

Figure 4.12: Grouping for Procedure IV
classification of the n and m units is shown in the left and right figure respectively

Based on this testing procedure, and the observed numbers of failures of each proof-load the likelihood function can be written:

$$\begin{aligned} L_{IV} = {} & p_{a1}^{n_{a1}}\left(p_{a2}-p_{a1}\right)^{n_{a2}-n_{a1}}\left(p_{b1}-p_{a2\cap b1}\right)^{n_{b1}}\left(p_{b2}-p_{b1}-p_{a2\cap b2}+p_{a2\cap b1}\right)^{n_{b2}-n_{b1}} \\ & p_{b1}^{m_{b1}}\left(p_{b2}-p_{b1}\right)^{m_{b2}-m_{b1}}\left(p_{a1}-p_{a1\cap b2}\right)^{m_{a1}}\left(p_{a2}-p_{a1}-p_{a2\cap b2}+p_{a1\cap b2}\right)^{m_{a2}-m_{a1}} \\ & \left(1-p_{a2}-p_{b2}+p_{a2\cap b2}\right)^{n+m-n_{a2}-n_{b2}-m_{a2}-m_{b2}} \end{aligned} \tag{4.13}$$

This likelihood function is similar to equation (4.10), but slightly more complicated since it has seven variables. Solving for the MLEs of these seven variables gives the same MLEs for p_{a1}^{*}, p_{a2}^{*}, p_{b1}^{*}, p_{b2}^{*} and $p_{a2\cap b2}^{*}$ as in Procedure III (i.e. the MLEs given as equations 4.11), with the following additions:

$$\begin{aligned} p_{a2\cap b1}^{*} &= \frac{\left(n_{a2}+n_{b2}+m_{a2}+m_{b2}\right)\left(m_{b1}n_{b2}+n_{a2}m_{b1}-m_{a1}n_{a2}-m_{a1}n_{b2}\right)}{\left(n+m\right)\left(n_{a2}+n_{b2}\right)\left(m_{a2}+m_{b2}\right)} \\ p_{a1\cap b2}^{*} &= \frac{\left(n_{a2}+n_{b2}+m_{a2}+m_{b2}\right)\left(m_{b2}n_{a1}+n_{a1}m_{a2}-m_{a1}n_{a2}-m_{a1}n_{b2}\right)}{\left(n+m\right)\left(n_{a2}+n_{b2}\right)\left(m_{a2}+m_{b2}\right)} \end{aligned} \tag{4.14}$$

Again, the MLE estimates for the means and standard deviations of the strengths in modes A and B are obtained using the equations (4.12). It should be noted that the estimates μ_a^{*}, σ_a^{*}, μ_b^{*} and σ_b^{*} will be identical to those obtained through Procedure III

since the expressions for p_{a1}, p_{b1}, p_{a2} and p_{b2} do not depend on either of the new observed variables n_{b1} or m_{a1}. However, since $p_{a2\cap b2} = f(p_{a2}, p_{b2}, \rho_{ab})$, $p_{a2\cap b1} = f(p_{a2}, p_{b1}, \rho_{ab})$, and $p_{a1\cap b2} = f(p_{a1}, p_{b2}, \rho_{ab})$ all three intersection MLEs obtained (i.e. $\hat{p}_{a2\cap b2}$, $\hat{p}_{a2\cap b1}$ and $\hat{p}_{a1\cap b2}$) can be used to give three different estimates of ρ_{ab}. Let $\hat{\rho}^1_{ab}$, $\hat{\rho}^2_{ab}$, and $\hat{\rho}^3_{ab}$ denote these three correlation MLEs respectively. Since $\hat{p}_{a2\cap b2}$ is always larger than either $\hat{p}_{a2\cap b1}$ or $\hat{p}_{a1\cap b2}$, $\hat{\rho}^1_{ab}$ offers the best individual estimate of ρ_{ab}. However, since using the weighted average (4.7) reduces the standard deviation of the correlation estimate over a wide range of p_{a1}, p_{b1}, p_{a2} and p_{b2} values by between 5-30%, usually the improvement is around 15%.

$$\hat{\rho}_{ab} = \frac{\hat{p}_{a2}\hat{p}_{b2}\hat{\rho}^1_{ab} + \hat{p}_{a2}\hat{p}_{b1}\hat{\rho}^2_{ab} + \hat{p}_{a1}\hat{p}_{b2}\hat{\rho}^3_{ab}}{\hat{p}_{a2}\hat{p}_{b2} + \hat{p}_{a2}\hat{p}_{b1} + \hat{p}_{a1}\hat{p}_{b2}} \qquad (4.14)$$

Note that the correlation estimate $\hat{\rho}_{ab}$, obtained from the weighted sum (4.14), is no longer a MLE, but rather represents a reasonable way of combining the three individual MLEs into one measure.

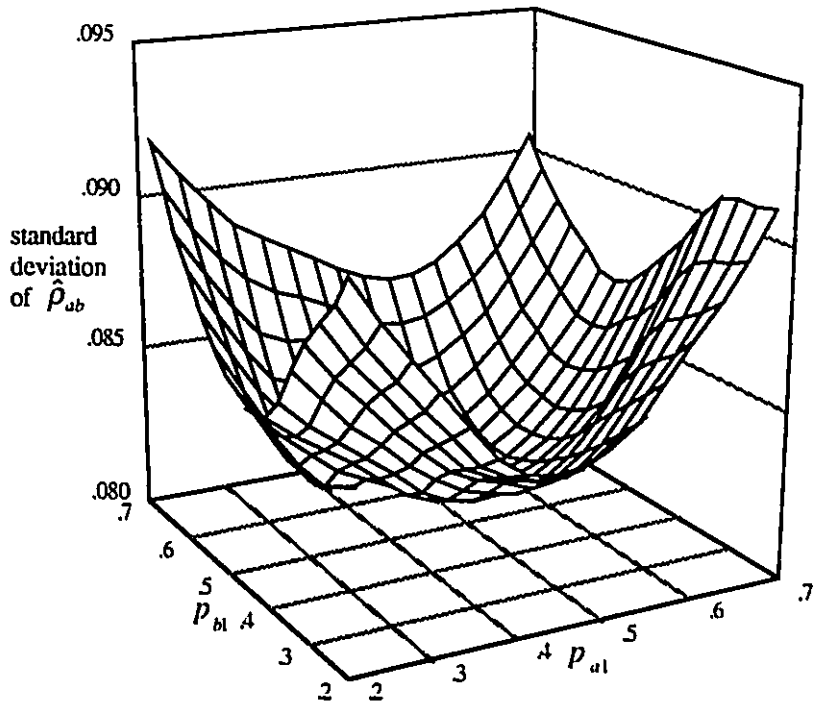

Figure 4.13: Standard Deviation of Simulated ρ^*_{ab} I - Procedure IV
$n+m = 300$, $\rho_{ab} = 0.6$, and in all cases $p_{a2} = 0.8$ and $p_{b2} = 0.8$

Procedures III and IV produce identical MLEs for the means and standard deviations of a bivariate normal distribution. See the discussion in Section 4.2.1 for an analysis of how changes in failure probabilities p_{a1}, p_{b1}, p_{a2} and p_{b2} effect those MLEs. For Procedure IV, unlike Procedure III, the values of p_{a1} and p_{b1}, in addition to p_{a2} and p_{b2} values, affect the correlation estimate. The results obtained for Procedure IV are presented graphically in Figures 4.13 and 4.14. These 3D surface plots, and numerous other simulations, suggest that the best results will be obtained if we have p_{a2} and p_{b2} somewhere between 0.5 and 0.8 and p_{a1} and p_{b1} between 0.25 and 0.35 lower. Note that the standard deviation of the correlation estimates obtained with procedures II, III and IV is better if the MLE of the probabilities are used rather than the true values even if they are known. This is due to a self-correcting feature in the estimation procedures II, III and IV.

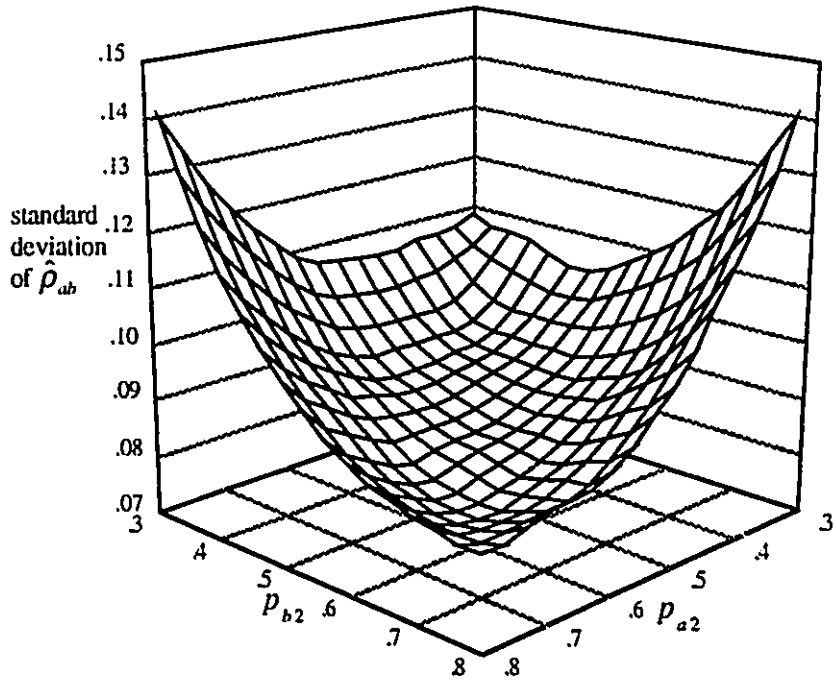

Figure 4.14: Standard Deviation of Simulated ρ^*_{ab} II - Procedure IV
$n+m = 300$, $\rho_{ab} = 0.6$, and in all cases $p_{a1} = p_{a2}-0.2$ and $p_{b1} = p_{b2}-0.2$

4.3 Comparison of Procedures I, II, III and IV

The various correlation estimation procedures are presented in order of increasing complexity. Procedure I is the simplest; it requires a single proof-load in each strength mode, and the proof-loads are performed in the same sequence for all the units. Procedure II is slightly more complex with the order of the single proof-load in each strength mode varying for different units. Procedure III extends to two proof-loads in one strength mode followed by a single proof-load in the second strength mode. Finally, Procedure IV suggests two proof-loads in each of the strength modes. However, the procedures vary in other aspects as well. The major points of comparison, other than the ease of implementation, are summarized in Table 4.2

Table 4.2: Comparison of Procedures I, II, III and IV

Procedure #	Quality of ρ_{ab} estimate	Other parameters estimated	prior knowledge required
I	excellent	none	p_a and p_b
II	good	p_a and p_b	none
III	good	μ_a, μ_b, σ_a, σ_b	none
IV	excellent	μ_a, μ_b, σ_a, σ_b	none

To a large extent the choice between the proposed procedures depends on the prior information available. If the probabilities of failure on each of the proof-loads p_a and p_b is known with certainty Procedure I is a logical choice. However, this prior knowledge is usually not available.

Procedures II and III yield identical correlation estimates as shown in Sections 4.1.2 and 4.2.1. The advantage of Procedure III is that it allows the estimation of the individual strength means and standard deviations, not only the probability of failure under the proof-loads. This is a significant advantage in many cases where this additional information is required. The correlation estimates obtained through Procedure IV are comparable in quality to those obtained by Procedure I, and significantly better than the estimates obtained through Procedures II or III. In addition, like Procedure III, Procedure IV allows the estimation of all five bivariate normal parameters and requires no prior knowledge. In conclusion, the best procedure depends on how many proof-load tests are feasible, what prior knowledge is available, and what bivariate normal estimates are required.

CHAPTER 5
Optimal Grouping Criteria

This chapter addresses the question of optimal grouping, and quantifies the amount of efficiency lost through the use of grouped data rather than variables data. As mentioned, in many circumstances the group limits used to classify units are predetermined. For example at Eaton Yale Inc. (Hamilton, Ontario) creating step-gauges specifically for an application is not practical due to the large number of different products produced. As a result, at Eaton, they use existing generic step-gauges that are incremented in thousands of an inch. However, in some situations, it is possible and practical to set application specific group limits. Clearly, all group limit designs will not be equally informative. As a result, it is possible to determine the optimal grouping criteria for particular applications.

Section 5.1 discusses gauge limit design for one-sided acceptance sampling plans to detect mean shifts. The results agree with past work (Beja and Ladany, 1974), and extend the analysis to the more general multiple group case. The results giving optimal group limits to detect one-sided mean shifts in a normal distribution presented in Tables 5.1 and 5.2 also appear in Steiner et al. (1993A). Section 5.2 considers the further extension to two-sided tests. The analysis from Section 5.1 can also be used to determine optimal group limits in the two-sided test case. The resulting optimal group limits are applicable for both the two sets of weights approach of Section 3.2.1 and the MLE approach of Section 3.2.2. Group limit design for Shewhart control charts is addressed in Section 5.3. The optimal group limits found are applicable in all four control chart design approaches from Section 3.3. Optimal group limits to detect mean and standard deviation

shifts in a normal process are determined, as well as optimal limits to detect shape and scale shifts of a Weibull process. Note that Table 5.5 gives the optimal group limits to detect mean shifts from a normal process and also appears in Steiner et al. (1993B). For the case of Shewhart control charts, we also find optimal limits for simultaneous detection of parameter shifts. This is of interest since often a process may become "out of control" through either a mean or standard deviation shift. It is very interesting to note that the results in Section 5.3 appear elsewhere in the literature in a totally different context. In Balakrishnan and Cohen (1991), Chapter 7, the results in Table 5.10 appear in a slightly different format. The problem considered by Balakrishnan and Cohen (1991) is finding the best linear unbiased estimate (BLUE) by selecting specific order statistics. For more details and additional references see their book. In the problem to find the BLUE the objective is to minimize the estimate's variance, whereas in the optimal grouping problem, considered here, the objective is to maximize the Fisher information available about a parameter. The equivalence of the two problem exists because by the Rao-Cramer inequality (Kendall et al. 1978) the inverse of the Fisher information is a lower bound for the variance of an unbiased estimate, and a best estimate would attain this bound. The solution to find the BLUE is given in terms of percentage points which when translated by the inverse cumulative density function give the optimal group limits. In Section 5.4 optimal group probabilities for the correlation estimation Procedure I from Chapter 4 are determined. Procedure I is the only correlation estimation procedure presented where group probabilities can be set a priori, however, the results also provide a guide for the other procedures from Chapter 4.

5.1 One-Sided Acceptance Sampling Plans

Acceptance sampling problems are often given in terms of acceptable and rejectable proportions of non-conforming or out-of-specification units. As a result, we

shall restrict our attention in this section to the classical case where we are interested in detecting mean shifts when the standard deviation is known. For one-sided mean shift acceptance sampling plans with single-step gauges, Beja and Ladany (1974), Sykes (1981), and Evans and Thyregod (1985) have shown that when the error risks are equal, the optimal group limit to detect mean shifts from μ_a to μ_r should be placed at $(\mu_a + \mu_r)/2$. Beja and Ladany (1974) also suggested that the optimal group limits for a two-step gauge should be placed symmetrically about $(\mu_a + \mu_r)/2$. Using this rule of thumb, a one dimensional search for the optimal group limits is possible. Note however that this solution will only be optimal if the error risks are equal. Our results for optimal group limits to detect one-sided mean shifts agree with these past results, and extend the analysis to the general k-step gauge and to the case when error rates are not necessarily equal.

Suppose we wish to design an acceptance sampling plan that will detect a one-sided mean shift of a certain magnitude from a normal distribution with specified error rates. In this situation, the weight based methods presented in Section 3.1 will give the optimal testing procedure. If the error rates are small and/or the magnitude of the shift to be detected is small, then the required sample size will be large enough so that $\bar{z}$ (the average weight from Section 3.1) is approximately normally distributed. In this case, the optimal group limits are determined by minimizing expression (3.4), the required sample size, subject to the constraint that the group limits remain ordered. Formally, let $\mathbf{t}$ be the k dimensional vector of standardized group limits. Then the multi-dimensional minimization problem is

$$\text{minimize } \{n(\mathbf{t}) + m(\mathbf{t})\}$$

where

$$m(\mathbf{t}) = \begin{cases} M & \text{if the } t\text{'s are not ordered,} \\ & \text{i.e., if } t_j > t_{j+1}, \text{ for any } j = 1,\ldots k \\ 0 & \text{otherwise} \end{cases}$$

M is a large number and $n(\mathbf{t})$ is expression (3.4). Note that the solution to the above minimization problem depends on the size of mean shift we wish to detect, and on the chart's error rates. This optimization problem can be easily solved by the Nelder Mead multi-dimensional Simplex Algorithm (Press et al., 1988). The Nelder Mead algorithm is quite efficient in most circumstances, however if efficiency is of great concern, it would be better to use the Fletcher-Reeves algorithm (Press et al., 1988). The Fletcher-Reeves algorithm is more efficient because it uses not only the function, but also its gradient. The calculation of the gradient of equation (3.4) with respect to the unknown t_j's is not unduly difficult, but looks rather complex. See Appendix D for the calculation of the gradient. The amount of work necessary to find the optimal group limits in the special case when the error rates α and β are equal is reduced since the optimal group limits must be symmetric about $(\mu_a + \mu_r)/2$. Using this fact reduces in half the number of variables to consider.

The optimal standardized group limits for selected error rates and mean shifts of a half, one, and one and a half sigma unit for a standard normal process are given in Tables 5.1 and 5.2. To determine the actual group limits to use in a specific example, the standardized group limits presented in the tables must be translated. In the case of a $N(\mu,\sigma)$ process, simply multiple each of the group limits given in the tables by σ and add μ.

Table 5.1: Optimal Group Limits and Weights
Standard Normal Distribution, $\alpha = \beta$.
The calculation of n assumes $\alpha = \beta = 0.001$

μ_1	k	n	λ	i =	1	2	3	4	5	6	7
0.5	1	235.5	0	t_i	0.25						
				z_i	-0.4001	0.4001					
	2	186.0	0	t_i	-0.3417	0.8417					
				z_i	-0.6052	0	0.6052				
	3	179.6	0	t_i	-0.6925	0.2500	0.925				
				z_i	-0.74	-0.2188	0.2188	0.74			
	4	164.6	0	t_i	-0.9384	0.1139	0.6139	1.4384			
				z_i	-0.8395	-0.3667	0	0.3667	0.8395		
	5	161.1	0	t_i	-1.1254	-0.3743	0.2500	0.8743	1.6254		
				z_i	-0.9172	-0.4771	-0.1511	0.1511	0.4771	0.9172	
	6	158.9	0	t_i	-1.2749	-0.5751	-0.0142	0.5142	1.0751	1.7749	
				z_i	-0.9804	-0.5642	-0.2653	0	0.2653	0.5642	0.9804
1	1	55.6	0	t_i	0.5000						
				z_i	-0.8070	0.8070					
	2	44.4	0	t_i	-0.0424	1.0424					
				z_i	-1.1789	0	1.1789				
	3	41.2	0	t_i	-0.3428	0.5000	1.3428				
				z_i	-1.4062	-0.3972	0.3972	1.4062			
	4	0.0	0	t_i	-0.5373	0.1813	0.8187	1.5373			
				z_i	-1.5600	-0.6495	0	0.6495	1.5600		
	5	9.2	0	t_i	-0.6723	-0.0357	0.5000	1.0357	1.6723		
				z_i	-1.6692	-0.8257	-0.2615	0.2615	0.8257	1.6692	
	6	8.8	0	t_i	-0.7697	-0.1941	0.2767	0.7233	1.1941	1.7697	
				z_i	-1.7492	-0.9553	-0.4503	0	0.4503	0.9553	1.7492
1.5	1	22.4	0	t_i	0.7500						
				z_i	-1.2275	1.2275					
	2	18.1	0	t_i	0.2661	1.2339					
				z_i	-1.7172	0	1.7172				
	3	17.0	0	t_i	0.0273	0.7500	1.4727				
				z_i	-1.9817	-0.5190	0.5190	1.9817			
	4	16.6	0	t_i	-0.1068	0.4829	1.0171	1.6068			
				z_i	-2.1358	-0.8189	0	0.8189	2.1358		
	5	16.4	0	t_i	-0.1867	0.3132	0.7500	1.1868	1.6867		
				z_i	-2.2293	-1.0090	-0.3225	0.3225	1.0090	2.2293	
	6	16.3	0	t_i	-0.2365	0.1971	0.5706	0.9294	1.3029	1.7365	
				z_i	-2.2882	-1.1366	-0.5429	0	0.5429	1.1366	2.2882

Table 5.2: Optimal Group Limits and Weights
Standard Normal Distribution, $\alpha = 0.001$, $\beta = 0.005$

μ_1	k	n	λ	i =	1	2	3	4	5	6	7
0.5	1	198.1	0.0070	t_i	0.2889						
				z_i	-0.3878	0.4125					
	2	156.3	0.0090	t_i	-0.2954	0.8870					
				z_i	-0.5880	0.0204	0.6220				
	3	144.0	0.0099	t_i	-0.6405	0.3009	1.2428				
				z_i	-0.7197	-0.1949	0.2423	0.7603			
	4	138.4	0.0103	t_i	-0.8812	0.0587	0.6685	1.4929			
				z_i	-0.8162	-0.3402	0.0263	0.3925	0.8620		
	5	135.4	0.0106	t_i	-1..0634	-0.3151	0.3079	0.9321	1.6839		
				z_i	-0.8913	-0.4483	-0.1227	0.1791	0.5048	0.9418	
	6	133.6	0.0107	t_i	-1.2084	-0.5121	-0.0468	0.5746	1.1359	1.8372	
				z_i	-0.9521	-0.5333	-0.2351	0.0297	0.2948	0.5936	1.0070
1	1	46.6	0.0256	t_i	0.5725						
				z_i	-0.7618	0.8533					
	2	37.3	0.0333	t_i	0.0459	1.1288					
				z_i	-1.1147	0.0792	1.2429				
	3	34.6	0.0367	t_i	-0.2387	0.5968	1.4438				
				z_i	-1.3259	-0.3028	0.4901	1.4854			
	4	33.5	0.0385	t_i	-0.4178	0.2891	0.9254	1.6528			
				z_i	-1.4649	-0.5414	0.1037	0.7552	1.6533		
	5	32.9	0.0395	t_i	-0.5384	0.0827	0.6141	1.1525	1.8020		
				z_i	-1.5608	-0.7049	-0.1481	0.3741	0.9436	1.7760	
	6	32.6	0.0402	t_i	-0.6230	-0.0658	0.3983	0.8443	1.3207	1.9127	
				z_i	-1.6290	-0.8229	-0.3278	0.1193	0.5716	1.0847	1.8684
1.5	1	18.8	0.0498	t_i	0.8471						
				z_i	-1.1378	1.3202					
	2	15.2	0.0654	t_i	0.3034	1.3495					
				z_i	-1.5924	0.1616	1.8456				
	3	14.3	0.0717	t_i	0.1636	0.8762	1.6076				
				z_i	-1.8291	-0.3309	0.7057	2.1367			
	4	13.9	0.1082	t_i	0.0443	0.6198	1.1540	1.7579			
				z_i	-1.9625	-0.6099	0.2005	1.0273	2..3134		
	5	13.7	0.0762	t_i	-0.0250	0.4587	0.8917	1.3331	1.8502		
				z_i	-2.0414	-0.7843	-0.1104	0.5349	1.2347	2.4245	
	6	13.6	0.0771	t_i	-0.0676	0.3493	0.7170	1.0763	1.4567	1.9090	
				z_i	-2.0903	-0.9006	-0.3216	0.2176	0.7654	1.3757	2.4958

Table 5.3: Optimal Group Limits and Weights
Standard Weibull Distribution, $\alpha = \beta$
The calculation of n assumes $\alpha = \beta = 0.001$

μ_1	k	n	λ	i =	1	2	3	4	5	6	7
0.5	1	110.1	0.0068	t_i	0.5934						
				z_i	−0.8764	0.3872					
	2	90.5	0.0104	t_i	0.3044	0.9060					
				z_i	−1.2995	−0.2719	0.5120				
	3	85.0	0.0122	t_i	0.1964	0.5372	1.1057				
				z_i	−1.5657	−0.6410	−0.0069	0.5714			
	4	82.7	0.0132	t_i	0.1442	0.3743	0.7126	1.2454			
				z_i	−1.7478	−0.8848	−0.3300	0.1464	0.6044		
	5	81.5	0.0138	t_i	0.1152	0.2855	0.5218	0.8492	1.3484		
				z_i	−1.8772	−1.0579	−0.5563	−0.1383	0.2464	0.6245	
	6	80.8	0.0143	t_i	0.0978	0.2313	0.4101	0.6441	0.9587	1.4273	
				z_i	−1.9711	−1.1859	−0.7246	−0.3467	−0.0069	0.3164	0.6376
1	1	1.4	0.0369	t_i	0.8591						
				z_i	−1.5449	0.7278					
	2	6.2	0.0508	t_i	0.5466	1.2247					
				z_i	−2.1412	−0.3949	0.9482				
	3	4.9	0.0569	t_i	0.4308	0.8507	1.4481				
				z_i	−2.4468	−0.9031	0.0622	1.0550			
	4	4.4	0.0599	t_i	0.3770	0.6815	1.0614	1.6015			
				z_i	−2.6151	−1.1947	−0.3962	0.3244	1.1157		
	5	4.1	0.0616	t_i	0.3485	0.5874	0.8677	1.2176	1.7137		
				z_i	−2.7131	−1.3779	−0.6907	0.0877	0.4948	1.1536	
	6	4.0	0.0627	t_i	0.3320	0.5285	0.7508	1.0125	1.3386	1.7992	
				z_i	−2.7735	−1.5007	−0.8960	−0.3704	0.1222	0.6141	1.1788
1.5	1	14.9	0.089	t_i	1.1408						
				z_i	−2.0899	1.0529					
	2	12.6	0.1127	t_i	0.8224	1.5417					
				z_i	−2.7530	−0.4487	1.3437				
	3	12.0	0.1205	t_i	0.7106	1.1712	1.7656				
				z_i	−3.0436	−1.0135	0.1300	1.4802			
	4	11.8	0.1235	t_i	0.6616	1.0014	1.3934	1.9076			
				z_i	−3.1843	−1.3137	−0.3990	0.4497	1.5561		
	5	11.7	0.1249	t_i	0.6368	0.9059	1.2014	1.5481	2.0046		
				z_i	−3.2592	−1.4934	−0.7263	−0.0325	0.6510	1.6028	
	6	11.7	0.1255	t_i	0.6228	0.8454	1.0832	1.3488	1.6619	2.0742	
				z_i	−3.3027	−1.6105	−0.9493	−0.3549	0.2124	0.7884	1.6337

Table 5.4: Optimal Group Limits and Weights
Standard Weibull Distribution, $\alpha = 0.001$, $\beta = 0.005$

μ_1	k	n	λ	i =	1	2	3	4	5	6	7
0.5	1	94.5	0.0208	t_i	0.6224						
				z_i	−0.8456	0.4006					
	2	78.1	0.0277	t_i	0.3281	0.9425					
				z_i	−1.2530	−0.2324	0.5240				
	3	73.6	0..0308	t_i	0.2179	0.5704	1.1441				
				z_i	−1.5034	−0.5868	0.0291	0.5811			
	4	71.7	0.0324	t_i	0.1648	0.4056	0.7505	1.2834			
				z_i	−1.6695	−0.8169	−0.2803	0.1786	0.6122		
	5	70.7	0.0333	t_i	0.1355	0.3157	0.5586	0.8894	1.3850		
				z_i	−1.7837	−0.9769	−0.4951	−0.0930	0.2752	0.6308	
	6	70.2	0.0340	t_i	0.1180	0.2608	0.4460	0.6840	0.9998	1.4618	
				z_i	−1.8638	−1.0927	−0.6531	−0.2911	0.0345	0.3422	0.6427
1	1	27.3	0.0758	t_i	0.9155						
				z_i	−1.4598	0.7655					
	2	23.0	0.0994	t_i	0.5991	1.2957					
				z_i	−2.0218	−0.2825	0.9845				
	3	21.8	0.1090	t_i	0.4838	0.9166	1.5263				
				z_i	−2.2987	−0.7617	0.1671	1.0872			
	4	21.4	0.1137	t_i	0.4311	0.7457	1.1341	1.6825			
				z_i	−2.4459	−1.0315	−0.2685	0.4213	1.1436		
	5	21.2	0.1162	t_i	0.4035	0.6507	0.9376	1.2938	1.7948		
				z_i	−2.5295	−1.1986	−0.5467	0.0294	0.5841	1.1776	
	6	21.1	0.1177	t_i	0.3877	0.5912	0.8190	1.0857	1.4164	1.8789	
				z_i	−2.5800	−1.3096	−0.7395	−0.2394	0.2305	0.6965	1.1994
1.5	1	13.1	0.1512	t_i	1.2196						
				z_i	−1.9534	1.1143					
	2	11.1	0.1913	t_i	0.8994	1.6413					
				z_i	−2.5730	−0.2637	1.4069				
	3	10.6	0.2051	t_i	0.7897	1.2643	1.8807				
				z_i	−2.8341	−0.7956	0.3119	1.5424			
	4	10.4	0.2110	t_i	0.7422	1.0926	1.4980	2.0361			
				z_i	−2.9576	−1.0745	−0.1931	0.6312	1.6170		
	5	10.3	0.2139	t_i	0.7182	0.9961	1.3017	1.6625	2.1451		
				z_i	−3.0226	−1.2402	−0.5041	0.1685	0.8341	1.6628	
	6	10.3	0.2155	t_i	0.7047	0.9349	1.1810	1.4571	1.7854	2.2260	
				z_i	−3.0601	−1.3480	−0.7150	−0.1402	0.4119	0.9742	1.6931

Tables 5.3 and 5.4 present the optimal group limits and weights for a standard Weibull process where the mean and standard deviation are equal to unity. Comparing the optimal limits for normal and Weibull group limits shows that the Weibull limits are not symmetric even when the error rates are equal. To compare the optimal limit in Tables 5.1 and 5.2 to the optimal Weibull process limits it is necessary to add unity to the normal process limits. In general, the optimal Weibull limits are shifted to lower values when compared to the optimal limits for a normal process.

5.2 Two-Sided Acceptance Sampling Plans

This section derives optimal group limits for the two-sided acceptance sampling plans of Section 3.2. For two-sided mean shift acceptance sampling plans or acceptance control charts I know of no previous work on optimal gauge design. Section 3.2 considers the hypothesis tests where the difference between μ_a^+ and μ_a^- is large in terms of sigma units, and thus the two-sided hypothesis test can be thought of as equivalent to two one-sided tests. As a result, the optimal group limits for two-sided mean shift detection will be in two clusters, half near μ_a^+ and the other half near μ_a^-. Since the two clusters of group limits are so far apart, the lower cluster of limits has very little effect on the ability to detect mean shifts to μ_r^+, and vice versa. As a result, the optimal group limits for the two sided test can be determined accurately from the analysis done for the one-sided tests in Section 5.1. Since adding a group limit near $(\mu_a^+ + \mu_a^-)/2$ will also have little effect, it is recommended that an even number of group limits be chosen. Half of the group limits will be determined by considering the lower hypothesis test, and the other half determined based on the upper hypothesis test. In other words, for the normal process use Tables 5.1 and 5.2, and for a Weibull process use Tables 5.3 and 5.4.

For example, say we wish to design an acceptance sampling plan or acceptance control chart based on 6 group limits that is to detect mean shifts of one sigma unit when

$\mu_a^+ = 10$, $\mu_a^- = 5$ $\sigma = 0.6$, (thus $\mu_r^+ = 10.6$, $\mu_r^- = 4.4$), $\alpha = 0.001$, $\beta = 0.005$, and the process is approximately normal. Then, from Table 5.2, the optimal standardized group limits for 3 step-gauge are –0.2387, 0.5968, 1.4438. Thus the optimal group limits for this two-sided problem are –0.2387*0.6+5 = 4.86, 0.5968*0.6+5 = 5.36, 5.87, –0.2387*0.6+10 = 9.86, 10.36, 10.87. Notice that there are two distinct groups of gauge limits.

5.3 Shewhart Control Charts

The design of step-gauges for Shewhart type control charts is motivated by Stevens (1948). Stevens proposed designing two-step gauges so that they maximize the expected Fisher information about the null hypothesis. Shewhart charts attempt to detect whenever the process is no longer stable at the target value (or null hypothesis). As a result, the problem of determining the best group limits for control, may be thought of as equivalent to designing a step-gauge to best estimate the parameter of interest when the null hypothesis holds. This implies that the grouping criteria that maximizes the expected Fisher information at the null hypothesis should be used. Stevens considers only the case of detecting mean and standard deviation shifts of a normal process with a two-step gauge. This section extends the methodology, first to the general multiple group case, and second to the Weibull process. In addition, the process is often monitored for both mean and standard deviation shifts using data from the same step-gauge. However, the optimal step-group limits are not the same for these two purposes, so the issue of a compromise gauge design for simultaneous parameter shift detection is also considered.

As mentioned above, we are interested in assessing the expected Fisher information in a sample of size n. The information about the parameter θ in the sample of data from a k-step gauge is given by:

$$I(\theta \mid \mathbf{Q}) = \left(\frac{\partial \ln(L(\theta \mid \mathbf{Q}))}{\partial \theta}\right)^2$$

$$= \left(\sum_{j=1}^{k+1} \frac{Q_j}{\pi_j(\theta)} \frac{d\pi_j(\theta)}{d\theta}\right)^2. \qquad (5.1)$$

However, the log-likelihood for a sample of size n will be formed by the sum of n log-likelihoods, each of identical expectation (Edwards, 1972). As a result, it is equivalent, for our purposes, to consider the expected information in a single observation. The expected information in a sample of size one at θ, $E(I(\theta))$, may be obtained by conditioning on the group into which the observation is classified. In particular, if $\mathbf{e}_1$, $\mathbf{e}_2$, ..., $\mathbf{e}_{k+1}$ denote the unit vectors of length $(k+1)$, then

$$E(I(\theta)) = \sum_{j=1}^{k+1} I(\theta \mid \mathbf{e}_j)\pi_j(\theta)$$

$$= \sum_{j=1}^{k+1} \frac{1}{\pi_j(\theta)}\left(\frac{d\pi_j(\theta)}{d\theta}\right)^2. \qquad (5.2)$$

The parameter θ can represent any parameter of interest. The optimal group limits to detect mean or standard deviation shifts of a normal distribution are determined in Section 5.3.1. In Section 5.3.2 the optimal group limits for Weibull shape and scale parameter shifts are given. In each case, the group probability function $\pi_j(\theta)$, is adapted to the particular parameter and distribution of interest.

5.3.1 Normal Process

For the normal distribution the group probabilities are

$$\pi_j(\mu,\sigma) = \int_{t_{j-1}}^{t_j} \phi(y)dy = \int_{t_{j-1}}^{t_j} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right)dy.$$

The standardized group limits t_j are utilized since ultimately interest lies in finding the information about the standard normal. Assume, without loss of generality, $\sigma = 1$. Then,

$$\phi(y) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{(y-\mu)^2}{2}\right),$$

and $$\frac{d\pi_j(\mu)}{d\mu} = \phi(t_{j-1}) - \phi(t_j),$$

and the expected information about μ from a single observation is

$$E(I(\mu)) = \frac{\phi(t_1;\mu)^2}{\pi_1(\mu)} + \sum_{j=2}^{k}\frac{\left(\phi(t_{j-1};\mu)-\phi(t_j;\mu)\right)^2}{\pi_j(\mu)} + \frac{\phi(t_k;\mu)^2}{\pi_{k+1}(\mu)}. \qquad (5.3)$$

The group limit design problem is to find the standardized group limits t_j's, that maximize this expected information. Without loss of generality, the calculations in the table assume $\mu = 0$. Finding the best location for the physical group limits x_j's, from the optimal standardized limits is straightforward. If the process produces units that match a normal distribution with mean μ and standard deviation σ then $x_j = \sigma t_j + \mu$.

The function $E(I(\mu))$ is not concave. If the first $(k-1)$ group limits are fixed and the k^{th} group limit is allowed to become arbitrarily large, the expected information asymptotically approaches a minimum. Extensive experimentation suggests, however, that the expected information function is unimodal. As a result, this non-linear optimization problem may be solved using either the Nelder-Mead multi-dimensional simplex method or the Fletcher Reeves algorithm (Press, et al., 1988). The Fletcher Reeves algorithm is more efficient but requires the gradient of (5.3) with respect to t. The gradient of (5.3) is given in Appendix E. Moreover, the optimal group limits are

symmetric about zero, thus the number of variables in the problem can be reduced by one half.

Table 5.5 (see also Kulldorff, 1961) gives the optimal group limits and the efficiency of a k-step gauge relative to exact measurement. Efficiency is defined as the ratio of the statistical information available using groups to the information available using variables. Clearly, the use of more than two or three groups significantly increases the efficiency of an observation. Table 5.5 shows that more information about μ is available in ten five-group optimally gauged observations, than is available in nine exact measurements. If exact measurement is uneconomical, then a properly designed gauge is an excellent alternative.

Table 5.5: Optimal Group Limits to Detect Mean Shifts
assume that when process is "in control" $\mu = 0$

k	Efficiency	t_1	t_2	t_3	t_4	t_5	t_6
1	0.6366	0.0					
2	0.8098	–0.6120	0.6120				
3	0.8825	–0.9817	0.0	0.9817			
4	0.9201	–1.244	–0.3824	0.3824	1.244		
5	0.9420	–1.4468	–0.6589	0.0	0.6589	1.4468	
6	0.9560	–1.6108	–0.8744	–0.2803	0.2803	0.8744	1.6108

The expected Fisher information approach can also be used to design group limits when the objective is to detect shifts in the standard deviation of a process. Assuming Y to be normally distributed with $\mu = 0$ gives

$$\phi(y;\sigma) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(\frac{-y^2}{2\sigma^2}\right)$$

$$\frac{d\pi_j(\sigma)}{d\sigma} = \frac{1}{\sigma}\left(t_{j-1}\phi\left(t_{j-1};\sigma\right) - t_j\phi\left(t_j;\sigma\right)\right).$$

Thus the expected information about σ, for a single observation, is written:

$$E(I(\sigma)) = \frac{t_1^2\phi(t_1;\sigma)^2}{\sigma\,\pi_1(\sigma)} + \sum_{j=2}^{k}\frac{\left(t_{j-1}\phi(t_{j-1};\sigma) - t_j\phi(t_j;\sigma)\right)^2}{\sigma\,\pi_j(\sigma)} + \frac{t_k^2\phi(t_k;\sigma)^2}{\sigma\,\pi_{k+1}(\sigma)}. \qquad (5.4)$$

Again we wish to find the group limits t_j's, that maximize this expected information. Without loss of generality, assume $\sigma = 1$. If the expected information is to be maximized when the process standard deviation is σ, multiple the group limits found for the case where $\sigma = 1$ by the desired σ value. Since extensive empirical study suggests this function is unimodal, the maximization problem can also be efficiently solved by the Nelder Mead or Fletcher Reeves algorithm. The gradient of $E(I(\sigma))$ is given in Appendix E. The results are shown in Table 5.6. Note that for an even number of groups, the middle step gauge placement has arbitrary sign, and is not zero as in the mean shift case. This is because, to detect standard deviation shifts, a group limit placed at $t = 0$ will provide no additional information.

Table 5.6: Optimal Group Limits to Detect Sigma Shifts
assume that when process is "in control" $\sigma = 1$

k	Efficiency	t_1	t_2	t_3	t_4	t_5	t_6
1	0.3042	±1.5758					
2	0.6522	–1.4825	1.4825				
3	0.7074	–1.4520	1.1855	2.0249			
or 3	0.7074	–2.0249	–1.1855	1.4520			
4	0.8244	–1.9956	–1.1401	1.1401	1.9956		
5	0.8588	–1.9827	–1.1193	0.9837	1.6189	2.3267	
or 5	0.8588	–2.3269	–1.6190	–0.9837	1.1190	1.9821	
6	0.8943	–2.3130	–1.6002	–0.9558	0.9558	1.6002	2.3130

The group limits that maximize the expected information about σ are not the same as those that maximize the expected information about μ. As a result, since charts

are often used to monitor for mean shifts and standard deviation shifts simultaneously a compromise gauge design is considered. Often the detection of mean shifts is given priority. For this reason, the proposed methodology for the grouping design allows some flexibility in the amount of emphasis given to detecting mean and standard deviation shifts. The analysis proceeds by using the weighted sum of efficiency ratings for mean and standard deviation estimation as the optimization criteria. In other words, maximize

$$Eff(\mu,\sigma;d) \quad = \quad d\,Eff(\mu)+(1-d)Eff(\sigma),$$

where d is the weight, $Eff(\mu)$ is the efficiency of mean estimation, and $Eff(\sigma)$ is the efficiency of standard deviation estimation. Unfortunately, this new combined efficiency criteria is a bimodal function, and therefore the optimization procedure may yield a local maximum. There is a boundary at around $d = 0.35$ where the improved mean estimate yielded by using a group limit at $t = 0$ is outweighed by the better standard deviation estimate obtained by staggering the group limits around $t = 0$. Empirical results suggest that the global maximum can be found using two different specific starting guesses for the group limits. The best group limits with the middle limit at $t = 0$ and the best group limits staggered about $t = 0$ are found and the solution with the largest $Eff(\mu,\sigma;d)$ value is chosen. Tables 5.7 presents the optimal 3-group gauge limits for different weights found in this manner. Table 5.8 gives the optimal compromise group designs for various number of group limits when the mean estimation is given greater weight ($d = 0.7$).

Table 5.7: Suggested 3-Group Limits to Detect Mean and Sigma Shifts
assume that if a process is "in control" $\mu = 0$ and $\sigma = 1$

weight d	Efficiency μ	Efficiency σ	t_1	t_2	t_3
0.1	0.6543	0.7335	-1.3974	1.0861	1.9584
0.2	0.7030	0.7248	-1.3384	0.9559	1.8742
0.3	0.7587	0.7060	-1.2835	0.7752	1.7625
0.4	0.8523	0.6481	-1.3906	0	1.3906
0.5	0.8569	0.6446	-1.3577	0	1.3577
0.6	0.8623	0.6379	-1.3117	0	1.3117
0.7	0.8685	0.6262	-1.2529	0	1.2529
0.8	0.8749	0.6066	-1.1779	0	1.1779
0.9	0.8803	0.5757	-1.0859	0	1.0859

Table 5.8: Optimal Group Limits to Detect Mean and Sigma Shifts
$d = 0.7$, assume that if process is "in control" $\mu = 0$ and $\sigma = 1$

k	Efficiency μ	Efficiency σ	t_1	t_2	t_3	t_4	t_5	t_6
1	0.6366	0	0					
2	0.7822	0.4664	-0.8487	0.8487				
3	0.8685	0.6262	-1.2529	0	1.2529			
4	0.9082	0.7384	-1.5500	-0.5295	0.5295	1.5500		
5	0.9333	0.8039	-1.7703	-0.8768	0	0.8768	1.7703	
6	0.9489	0.8486	-1.9481	-1.1366	-0.3889	0.3889	1.1366	1.9481

5.3.2 Weibull Process

When designing a step-gauge for Shewhart control charts when the underlying distribution is Weibull the methodology presented in Section 5.3.1 is used. However, since the Weibull is defined in terms of shape and scale parameters a and b, the optimal step-gauge design to detect shifts in a and b independently are first considered. Later the optimal group design for detecting shifts in both parameters simultaneously are discussed. The group limit designs for simultaneous parameter detection would be

appropriate if, for example, we are interested in detecting mean and/or standard deviation shifts of a Weibull process.

If the process produces parts that are best modeled by a Weibull distribution, the group probabilities can be written:

$$\pi_j(a,b) = \exp\left(-\left(\frac{t_{j-1}}{b}\right)^a\right) - \exp\left(-\left(\frac{t_j}{b}\right)^a\right)$$

The standard Weibull is defined to have unit shape and scale parameters, $a = b = 1$.

With interest focused on the shape parameter of the Weibull distribution, and assuming, without loss of generality, that $b = 1$, we can write:

$$\pi_j(a) = \exp\left(-t_{j-1}^a\right) - \exp\left(-t_j^a\right),$$

and thus, $$\frac{d\pi_j(a)}{da} = t_j^a \ln\left(t_j\right)\exp\left(-t_j^a\right) - t_{j-1}^a \ln\left(t_{j-1}\right)\exp\left(-t_{j-1}^a\right).$$

We can assume $b = 1$ because, as will be explained later, the optimal limits found for the standard Weibull are easily translated to any value of b. From the above expressions, the expected information about the shape parameter, from a single observation, is:

$$E\left(I(a)\right) = \sum_{j=1}^{k+1} \frac{\left(t_j^a \ln\left(t_j\right)\exp\left(-t_j^a\right) - t_{j-1}^a \ln\left(t_{j-1}\right)\exp\left(-t_{j-1}^a\right)\right)^2}{\exp\left(-t_{j-1}^a\right) - \exp\left(-t_j^a\right)} \qquad (5.5)$$

This equation can be maximized using either the Nelder Mead multi-dimensional simplex algorithm, or more efficiently by Fletcher Reeves algorithm (see Appendix D for details on the gradients of $E\left(I(a)\right)$ and $E\left(I(b)\right)$). Table 5.9 presents the results of the maximization problem. The efficiency rating is calculated relative to the amount of information available in variables data.

When the scale parameter of the Weibull is of interest a similar analysis is possible. Assuming, without loss of generality, that $a = 1$, gives the group probabilities

$$\pi_j(b) = \exp\left(-\frac{t_{j-1}}{b}\right) - \exp\left(-\frac{t_j}{b}\right),$$

and therefore, $$\frac{d\pi_j(b)}{db} = \frac{t_{j-1}}{b^2}\exp\left(-\frac{t_{j-1}}{b}\right) - \frac{t_j}{b^2}\exp\left(-\frac{t_j}{b}\right).$$

So $$E(I(b)) = \sum_{j=1}^{k+1} \frac{\left(\frac{t_{j-1}}{b^2}\exp\left(-\frac{t_{j-1}}{b}\right) - \frac{t_j}{b^2}\exp\left(-\frac{t_j}{b}\right)\right)^2}{\exp\left(-t_{j-1}/b\right) - \exp\left(-t_j/b\right)}. \qquad (5.6)$$

The group limits that maximize expression (5.6) above are given in Table 5.10.

Table 5.9: Optimal Group Limits to Detect Shape Shifts
assume that if the process is "in control" $a = 1$

k	Efficiency	t_1	t_2	t_3	t_4	t_5	t_6
1	0.2801	0.1189					
2	0.6557	0.1418	3.2891				
3	0.7527	0.1505	2.6936	4.5643			
4	0.8285	0.0516	0.2486	2.6173	4.4970		
5	0.8692	0.0534	0.2580	2.3339	3.6005	5.3934	
6	0.8990	0.0245	0.1154	0.3257	2.2921	3.5641	5.3593

Table 5.10: Optimal Group Limits to Detect Scale Shifts
assume that if the process is "in control" $b = 1$

k	Efficiency	t_1	t_2	t_3	t_4	t_5	t_6
1	0.6476	1.5936					
2	0.8203	1.0176	2.6112				
3	0.8910	0.7540	1.7716	3.3652			
4	0.9269	0.6004	1.3545	2.3720	3.9656		
5	0.9476	0.4999	1.0998	1.8538	2.8714	4.4650	
6	0.9606	0.4276	0.9269	1.5273	2.2813	3.2989	4.8925

Often we are interested in both parameters simultaneously. This is the case when interest lies in the mean and/or the standard deviation of a Weibull process. For that reason, the optimal step-gauge design is determined for shifts in both parameters. As the relationship between shape and scale parameters and the corresponding mean and standard deviation is complex, an optimization criteria based on the average efficiency is proposed. Table 5.11 presents the results, showing the step-gauge design that has the highest average efficiency to detect scale and shape parameter shifts. For this compromise solution, the non-linear optimization procedure may give only a local maximum. Extensive empirical study suggests that we get different local maxima of the average of (5.5) and (5.6) depending on how many of the initial gauge limits start on either side of the mean value of the standardize Weibull, i.e. $t = 1$. Consequently, for a k-step gauge there are $(k+1)$ local maximums.

Table 5.11: Optimal Group Limits to Detect Shape and Scale Shifts
assume that when the process is "in control" $a = b = 1$

k	Average Efficiency	t_1	t_2	t_3	t_4	t_5	t_6
1	0.4096	2.4552					
2	0.6351	0.2591	2.5473				
3	0.7613	0.2148	1.7140	3.6038			
4	0.8257	0.1181	0.6377	2.0668	3.9147		
5	0.8736	0.0814	0.4054	1.4950	2.7639	4.5538	
6	0.9021	0.0626	0.3027	1.0291	2.0483	3.2591	5.0264

The empirical results also suggest that to obtain the global optimal, as presented in Table 5.11, a starting solution must have an equal number of group limits on either side of $t = 1$. For the case of a odd number of group limits, the extra group limit is better placed on the upper side of the mean. So, for example, the best gauge placement using

only $k = 3$ gauge limits, would be at 0.215, 1.715, and 3.6, and this grouping criterion has an average efficiency of 76% for detecting scale and shape parameter shifts.

The optimal group limits presented in Tables 5.9-5.11 all show the best group limits for detecting shifts in the standard Weibull when $a = b = 1$. Fortunately, these optimal limits can easily be rescaled for the general Weibull. To translate the optimal group limits from the case $a = b = 1$, use the formula:

$$x_i = bt_i^a \qquad i = 1, \ldots, k+1$$

where t_i represent the standardized group limits presented in Tables 5.9-5.11, and x_i are the rescaled group limits.

5.4 Destructive Testing Procedure I

This section addresses the question of optimal group probabilities to estimate the correlation using Procedure I of Section 4.1. Note that in Chapter 4 the group divisions were defined in terms of group probabilities rather than group limits. However, group probabilities can easily be translated to group limits through the use of the inverse cumulative density function of the normal distribution, i.e. $t_i = \Phi^{-1}(p)$. Given an actual correlation ρ_{ab} and the sample size n, the values of p_a and p_b that minimize the predicted standard deviation given in equation (4.6) can be found. Due to the well behaved nature of the function that approximates the standard deviation of ρ^*_{ab} Nelder Mead multidimensional simplex method (Press et al. 1988) is appropriate. The results showing the pair of p_a and p_b values that minimize the approximate standard deviation for ρ^*_{ab} for actual ρ_{ab} between -0.95 and 0.95 are plotted in Figure 5.1.

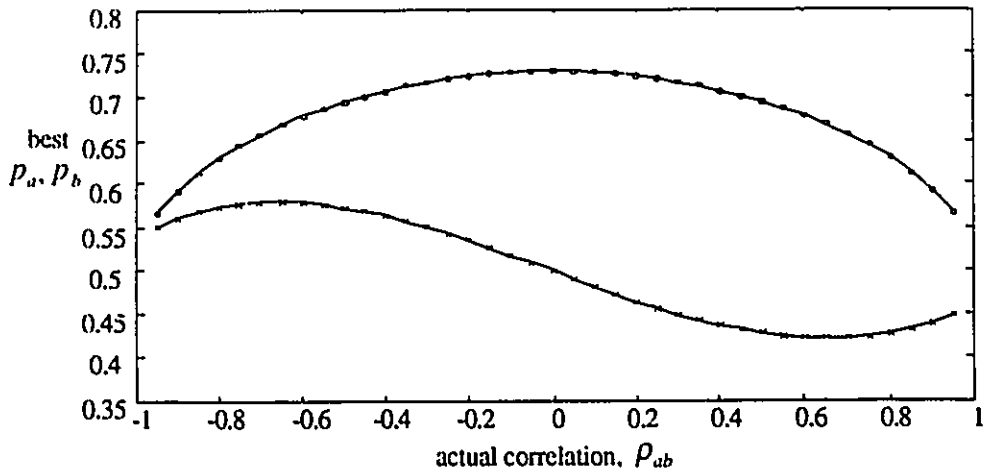

Figure 5.1: Optimal p_a and p_b Values to Estimate ρ_{ab}
'o' = best p_a, 'x' = best p_b

The plots for optimal p_a and p_b values appear approximately quadratic and cubic in nature respectively, and can be very closely approximated by fitting polynomial regression lines. This gives

$$\text{best } p_a = 0.733 - 0.165\rho_{ab}^2$$
$$\text{best } p_b = 0.499 - 0.184\rho_{ab} + 0.146\rho_{ab}^3 \qquad (5.7)$$

Determining the optimal values for p_a and p_b requires a prior estimate for the correlation ρ_{ab}. However, due to the relative insensitivity of the standard deviation of ρ_{ab}^* near the optimal p_a and p_b values (explored in the Section 4.1.1), choosing good p_a and p_b values can be done even with little idea of the actual ρ_{ab}. With little prior information regarding the correlation level, proof-load levels close to $p_a = 0.65$ and $p_b = 0.45$ are recommended. These proof-load levels provide correlation estimates with close to optimal standard deviation values for any true correlation level. Procedure I utilizes a single proof-load in each mode, thus in each dimension there is a single group limit. Based on the above suggested group probabilities, the best group limits are –0.126 and

0.385 for strength mode A and B respectively. These group limits are given in terms of a standardized normal process.

CHAPTER 6
Summary, Conclusion and Possible Extensions

This thesis develops quality control and improvement techniques based on grouped data. Grouped data commonly occur in industry when exact measurements are either prohibitively expensive or impossible. The methodology presented allows the creation of acceptance sampling plans, acceptance control charts, and Shewhart control charts based on grouped data. In addition, a number of correlation estimation procedures are derived that are applicable when data is grouped due to destructive testing.

In Chapter 1 the concept of grouped data is introduced, and an outline of the three major areas of application is presented. The three application areas are acceptance sampling plans, control charts and correlation estimation under destructive tests. Chapter 1 also provides a detailed literature survey of previous research relating to the use of grouped data in quality control and improvement problems. In Chapter 2, much of the notation used in the thesis is defined, and algorithms, derived from existing work in the literature, for the calculation of maximum likelihood parameter estimates from grouped data are presented. In addition, Chapter 2 shows that existing *ad hoc* quality control techniques often used in industry for grouped data are inadequate. Chapter 3 turns to the derivation of one-sided and two-sided acceptance sampling plans, acceptance control charts and Shewhart control charts based on grouped data. The solution methodology is based on the asymptotic properties of the chosen test statistic. In each case, different solution strategies are compared and contrasted. In addition, Chapter 3 discusses in detail the design of such plans and charts when utilizing small sample sizes. Chapter 4 presents four different procedures that use proof-loading to estimate the correlation between

destructively measured strength properties. Unlike existing techniques, all the procedures involves grouping units in two modes and require no precise measurements. The first two procedures are adaptations of existing techniques that use a single proof-load in each mode and give only estimates of the correlation. The second two procedures are further extensions that utilize two proof-loads in each mode. The resulting additional information allows the estimation of the two individual means and standard deviations as well as the correlation. Chapter 5 addresses the issue of optimal grouping criteria. The best way to group observations depends on the application, but the optimal groupings for acceptance sampling plans, Shewhart control charts for normal or Weibull processes, and correlation estimation under destructive testing are derived through optimization techniques.

The ultimate goal of Statistical Process Control (SPC) is increased quality in manufactured products or services provided. For the most part, SPC techniques have been developed for two types of data: variables data and dichotomous data. However, existing quality control tools will not work well if they are inappropriate for the situation. For example, variables-based SPC techniques are commonly applied in an *ad hoc* way to grouped data and may lead to misleading results and incorrect decisions. Grouping data into two or more groups is a natural compromise between variables data and dichotomous data. Grouped data are common in industry and occur when precise measurement is expensive but gauging articles into groups is feasible. In this thesis the SPC methodology presented is developed and designed for grouped data. The derived sampling plans, control charts and correlation estimation procedures for grouped data are quite competitive, in terms of required sample size, to variables based methods. The slight loss in efficiency is often more than compensated by lower data collection costs, since grouping data may be easier and cheaper. In addition, the resulting charts and plans are easily implemented in a shop floor environment.

In a more general context, the presented methodology provides a framework to deal with grouped data in the areas of parameter estimation and hypothesis testing. The methods have been described in the context of quality improvement, but they are more widely applicable. The solution approaches suggested are very adaptable. The likelihood approach is appropriate for grouped data from any underlying distribution, and with any parameter of interest, since, due to the data grouping, the appropriate distribution is always multinomial. Thus, for any underlying distribution of the quality characteristic only the group probabilities change leaving the proposed design methodology unchanged. As a result, the techniques are also applicable when the underlying distribution of the quality characteristic is non-normal. For example, the proposed techniques have application in the service industry where service times are often modeled as exponentials. The weights-based methods are also very appealing due to their simplicity. In the weights-based methods each unit is assigned a weight based on the group into which it is classified, and the average weight of a sample is used as the test statistic. Thus, the weights-based methods are easily implemented since these calculations can be done without sophisticated measuring devices or computers.

A number of interesting extensions to this work are possible and are currently being pursued. Likelihood methods can be extended to sequential sampling methods through the sequential probability ratio test. Sequential procedures have the advantage of requiring, on average, smaller sample sizes to achieve the same operating characteristics as the fixed sample size solutions. Sequential testing procedures, however, have several drawbacks. The sample size required to reach a decision is not known a priori, and units must be considered one at a time. This extension would be of particular interest since it would allow the design of cumulative sum (CUSUM) charts. CUSUM charts are currently very popular in the literature since they are easy to use and are better at detecting small parameter shifts than Shewhart charts. Another possible extension is to

develop hypothesis testing procedures for the correlation under destructive testing. A methodology similar to the one presented in Chapter 3 would probably also be applicable in this case. A third extension involves the application of grouped data to experimental design. The ultimate goal of both experimental design and control charts is quality improvement. However, control charts passively monitor the output of a process until an "out of control" signal is obtained, and then investigate and remove the cause of the problem. Designed experiments, in contrast, provide a more active statistical tool for achieving quality improvement. With designed experiments a number of possible input conditions are tried, and through a statistical analysis the probable optimal combination of process inputs is determined. Inputs consist of such things as raw materials, temperature and machine settings. Often the correct combination of inputs results in a reduction in the "natural variation" of a process. This thus leads directly to more consistent and higher quality outputs.

These proposed extensions, together with the work from this thesis, would provide practitioners with a fairly complete quality control and improvement system designed for grouped data. The thesis material makes the first important steps in this direction, and provides a methodology that can be extended to the other areas. In conclusion, the presented SPC techniques based on grouped data will be a valuable addition to a quality practitioner's repertoire of quality control and improvement techniques, and provide a methodology for parameter estimation and hypothesis testing based on grouped data in general.

Bibliography

Abramowitz, M., and Stegun, I.A. (eds.) (1970), *Handbook of Mathematical Functions* (Applied Mathematics Series 55), Washington, DC: National Bureau of Standards.

American National Standards Institute (ANSI) and the American Society for Quality Control (ASQC), (1978), "Quality Systems Terminology," Milwaukee, WI, ASQC.

Amorim, S. (1982), "Experimental Designs for Estimating the Correlation Between Two Destructively Tested Variables," unpublished Ph.D. Thesis, University of Wisconsin-Madison, Dept. of Statistics.

Amorim, S. and Johnson, R.A. (1986), "Experimental Designs for Estimating the Correlation Between Two Destructively Tested Variables," *Journal of the American Statistical Society*, Vol. 81, No. 395, pp.807-811.

Balakrishnan, N., Cohen, A.C., (1991), *Order Statistics and Inference - Estimation Methods*, Academic Press, Inc. Boston.

Bartlett, N.R. and Lwin, T. (1984), "Estimating a Relationship between Different Destructive Tests on Timber," *Applied Statistics*, Vol. 33, No. 1, pp. 65-72.

Beja, A., Ladany, S.P., (1974), "Efficient Sampling by Artificial Attributes," *Technometrics*, Vol. 16, No. 4, pp. 601–611.

Box, G., Kramer, T. (1992), "Statistical Process Monitoring and Feedback Adjustment - A Discussion, *Technometrics*, Vol. 34, No. 3., pp. 251-267.

Bray, D.F., Lyon, D.A., Burr, I.W., (1973), "Three Class Attributes Plans in Acceptance Sampling," *Technometrics*, Vol. 15, No. 3, pp. 575–585.

Deming, W.E. (1986), *Out of the Crisis*, Cambridge, Massachusetts, MIT Center for Advanced Engineering Study.

Dempster, A.P., Laird, N.M., Rubin, D.B., (1977), "Maximum Likelihood from Incomplete Data via the EM Algorithm" (with discussion), *Journal of the Royal Statistical Society, Series B*, Vol. 39, pp. 1-38.

Dodge, H.F., Romig, H.G. (1929), "A Method of Sampling Inspection," *The Bell System Technical Journal*, October, pp. 613-631.

Dodge, H.F., Romig, H.G. (1959), Sampling Inspection Tables, Single and Double Sampling, Second Ed., Wiley, New York.

Drezner, Z. and Wesolowsky, G.O. (1990), "On the Computation of the Bivariate Normal Integral," *Journal of Statistical Computation and Simulation*, Vol. 35, pp. 101-107.

Dudding, B.P. and Jennett, W.J. (1944), *Control Chart Techniques when Manufacturing to a Specification*, London: British Standards Institute.

Duncan, A.J. (1950), "A Chi-Square Chart for Controlling a Set of Percentages," *Industrial Quality Control*, Vol. 7, pp. 11-15.

Duncan, A.J. (1986), *Quality Control and Industrial Statistics* (5th ed.), Homewood, IL: Richard D. Irwin.

Edwards, A.W.F., (1972), *Likelihood*, Cambridge University Press, Cambridge.

Efron, B. (1981), The Jackknife, the Bootstrap, and Other Resampling Plans. CBMS Monogram, No. 38, SIAM, Philadelphia.

Efron, B., Gong, G. (1983), "A Leisurely Look at the Bootstrap, The Jackknife, and Cross-Validation," *The American Statistician*, Vol. 37, No. 1, pp. 36-48.

Ermer, D.S., Roepke, J.R., (1991), "An Analytical Analysis of Pre-Control," ASQC Quality Congress Transactions - Milwaukee, pp. 522-527.

Evans, I.G., Thyregod, P., (1985), "Approximately Optimal Narrow Limit Gauges," *Journal of Quality Technology* , Vol. 17, pp. 63-66.

Evans, J.R., Lindsay, W.M., (1992), *The Management and Control of Quality*, 2nd edition, West Publishing Company, St. Paul, MN.

Evans, J.W., Johnson, R.A., Green D.W. (1984), "Estimating the Correlations Between Variables Under Destructive Testing, or How to Break the Same Board Twice," *Techometrics*, Vol. 26, No. 3, August, pp. 285-290.

Freund, R.A., (1957) "Acceptance Control Charts," *Industrial Quality Control*, Vol. 14, No. 4, pp. 13-22.

Galligan, W.L., Johnson, R.A., Taylor, J.R. (1979), "Examination of the Concomitant Properties of Lumber, "Proceedings *of the Metal Plate Wood Truss Conference*, St. Louis, MO, Forest Products Research Society, Madison, WI.

Gerhards, C.C. (1979), "Time-Related Effects of Loading on Wood Strength: A Linear Cumulative Damage Theory," *Wood Science*, Vol. 11, No. 3, pp. 139-144.

Ghosh B.K., Sen P.K., (eds.), (1991), Handbook of Sequential Analysis, Marcel Dekker Inc., New York.

Grab, E.L., Savage, I.R. (1954), "Tables of the expected value of 1/X for positive Bernoulli and Poisson variables," *Journal of the American Statistical Association*, Vol.49, pp. 169-177.

Green, D.W., Evans, J.W. (1983), "Estimating Correlation Between Strength Properties," *Proceedings of the Fourth Engineering Mechanics Division Specialty Conference: Recent*

Advances in Engineering Mechanics and Their Impact on Civil Engineering Practice, Vol. 2, pp. 936-939.

Green, D.W., Evans, J.W., Johnson, R.A. (1984), "Investigation of the Procedure for Estimating the Concomitance of Lumber Strength Properties," *Wood and Fiber Science*, Vol. 16, No. 3, pp. 427-440.

Haitovsky, Y., (1989), "Grouped Data," in the Encyclopedia of Statistical Sciences, Johnson N.L. and Kotz S. (eds.) 1989.

Hamaker, H.C., (1979), "Acceptance Sampling for Percent Defective by Variables and by Attributes," *Journal of Quality Technology*, Vol. 11, No. 3, pp. 139-148.

Harlow, D.G. (1989), "The effect of proof-testing on the Weibull distribution," *Journal of Materials Science*, Vol. 24, pp. 1467-1473.

Hill, D., (1956), "Modified Control Limits," *Applied Statistics*, Vol. 5, No.2, pp. 12-19.

Ingram, M., Bray, D.F., Clark, D.S., Dolman, C.E., Elliot, R.P., Thather F.S. (eds.) (1978), Sampling for Micro-biological Analysis: Principles and Specific Applications, University of Toronto Press, Toronto.

Jennett, W.J., Welch, B.L., (1939), "The Control of Proportion Defective as Judged by a Single Quality Characteristic Varying on a Continuous Scale," *Journal of the Royal Statistical Society Series B*, Vol. 6, No. 1, pp. 80-88.

Johnson, N.L., Kotz, S. (1969), *Distributions in Statistics: Discrete Distributions*, Houghtom Mifflin Company, Boston.

Johnson, N.L., Kotz, S., (1970), *Continuous Univaiate Distributions, Vol. 2*, Wiley, New York.

Johnson, N.L., Kotz, S., (eds.) (1989), *Encyclopedia of Statistical Sciences*, Wiley, New York.

Johnson, R.A. (1980), "Current Statistical Methods for Estimating Lumber Properties by Proof-loading," *Forest Products Journal*, Vol. 30, no. 1, pp. 14-22.

Johnson, R.A., Galligan, W.L., (1983), "Estimating the Concomitance of Lumber Strength Properties," *Wood and Fiber Science*, Vol. 15, No. 3, pp. 235-244.

Juran, J.M., Gryna, F.M., Bingham, R.S., (eds.) (1979), Quality Control Handbook, Third Edition, McGraw Hill New York.

Kendall, M.G., (1938), "The Conditions Under Which Sheppard's Corrections are Valid", *Journal of the Royal Statistical Society*, Vol. 101. pp. 592-605.

Kendall, M., Stuart, D., Ord, J.K., (1978), Kendall's Advanced Theory of Statistics, Volume two, Fifth Edition, Oxford University Press, New York.

Kulldorff, G. (1961), Contributions to the Theory of Estimation from Grouped and Partially Grouped Samples, John Wiley and Sons, New York.

Ladany, S.P., (1976), "Determination of Optimal Compressed Limit Gaging Sampling Plans," *Journal of Quality Technology* , Vol. 8, No. 4, pp. 225–231.

Ladany, S.P., Sinuary–Stern, Z., (1985), "Curtailment of Artificial Attribute Sampling Plans," *International Journal of Production Research*, Vol. 23, No. 5, pp. 847–856.

Lindley D.V., (1950), "Grouping Correction and Maximum Likelihood Equations," *Proceedings of the Cambridge Philosophical Society*, Vol. 46, pp. 106-110.

Macdonald, P.D.M., (1993), Private Communication.

Mace, A.E., (1952), "The Use of Limit Gages in Process Control," *Industrial Quality Control*, January, pp. 24–31.

Madsen, B. (1976), "In-Grade Testing: Degree of Damage Due to Proofloading of Lumber in Bending," Report 17, Structural Research Series, Vancouver, B.C., Canada: Dept. of Civil Engineering, University of British Columbia.

Marcucci, M., (1985), "Monitoring Multinomial Processes," *Journal of Quality Technology*, Vol. 17, No. 2, pp. 86-91.

Marymont, I., (1975), "Corrections and Inferences for Grouped Samples," M.A. Thesis, Dept. of Statistics, Tel-Aviv University.

Nelson L.S., (1987), "A Chi-Squared Control Chart for Several Proportions," *Journal of Quality Technology*, Vol. 19, No. 4, pp. 229-231.

Newcombe P.A., Allen, O.B., (1988), "A Three-Class Procedure for Acceptance Sampling by Variables," *Technometrics*, Vol. 30, No. 4, pp. 415-421.

Neyman, J., Pearson, E.S., (1928), "On the Use and Interpretation of Certain Test Criteria for the Purpose of Statistical Inference, Parts I and II" *Biometrika*, Vol. 20A, pp. 175-240 & 263-294.

Ott, E.R., Mundel, A.B., (1954), "Narrow–Limit Gaging," *Industrial Quality Control*, March, pp. 21–28.

Press, W.H., Flannery, B.P., Teukolsky, S.A., Vetterling, W.T. (1988), *Numerical Recipes*, Cambridge University Press, Cambridge.

Pyzdek, T. (1993), "Process Control for Short and Small Runs," *Quality Progress*, April, pp. 51-60.

Rao, C.R. (1973), *Linear Statistical Inference and its applications*, 2nd edition, Wiley, New York.

Ryan, T.P., (1989), *Statistical Methods for Quality Improvement*, John Wiley and Sons, New York.

Schilling, E.G., (1981), *Acceptance Sampling in Quality Control*, Marcel Dekker, New York.

Shainin, D., Shainin, P., (1989), "Pre-Control Versus $\overline{X}$ & R Charting: Continuous or Immediate Quality Improvement?", *Quality Engineering*, Vol. 1, No. 4, pp. 419-429.

Sheppard, W.F., (1898), "On the Calculations of the most probable Values of Frequency Constants for Data Arranged According to Equidistant divisions," *Proceedings of the London Mathematical Society*, Vol. 29, pp. 353-380.

Shewhart, W.A., (1931), *Economic Control of Quality of Manufactured Product*, D. Van Nostrand, New York.

Silvia, A.A., (1988), "Stoplight Control," *Quality Progress*, September, pp. 39-42.

Steiner, S.H., Geyer, P.L., Wesolowsky, G.O., (1994A), "Control Charts based on Grouped Data," *International Journal of Production Research*., Vol. 32, No. 1, pp. 75-91.

Steiner, S.H., Geyer, P.L., Wesolowsky, G.O., (1994B), "Shewhart Control Charts to Detect Mean Shifts based on Grouped Data," second revision submitted to *Technometrics*.

Steiner, S.H., Wesolowsky, G.O., (1994A), "Estimating the Correlation between Destructive Measured Variables using Proof-loading," revision submitted to *Technometrics*.

Steiner, S.H., Wesolowsky, G.O., (1994B), "Quality Control based on Grouped Data," submitted to *Journal of Quality Technology*.

Stevens, W.L., (1948), "Control By Gauging," *Journal of the Royal Statistical Society: Series B*, No. 1, pp. 54-108.

Strickler, M.D., Pellerin, R.F., Talbot, J.W. (1970), "Experiments in Proofloading Structural End Joints Lumber," *Forest Products Journal*, Vol. 20, No. 2, pp. 29-35.

Suddarth, S.K., Woeste, F.E., Galligan, W.L. (1978), "Probabilistic Engineering Applied to Wood Members in Bending/Tension," Research Paper FPL 302, U.S. Dept. of Agriculture, Forest Products Laboratory, Madison, Wisconsin.

Sullivan, L.P., (1984), "Reducing variability: A new approach to Quality," *Quality Progress*, Vol. 17, No. 7 July, pp. 15-21.

Sykes, J., (1981), "A Nomogram to Simplify the Choice of a Sampling Plan Using a Single Gauge," *Journal of Quality Technology*, Vol. 13, pp. 36-41.

Taguchi, G., Wu, Y. (1979), Introduction to off-line quality control. Central Japan Control Association, (available from American Supplier Institute, 32100 Detroit Industrial Expressway, Romulus, MI 48174.

Tiku, M.L., (1965), "Laguerre Series forms of Non-central χ^2 and F Distributions," *Biometrika*, Vol. 52, pp. 415-427.

Tippett, L.H.C., (1944), "The Efficient Use of Gauges in Quality Control," *Engineer*, Vol. 177, pp. 481-483.

Traver, R.W., (1985), "Pre-Control: A Good Alternative to $\overline{X}$-R Charts," *Quality Progress*, September, pp. 11-14.

Vardeman, S.B., (1986), "The Legitimate Role of Inspection in Modern SPC," *The American Statistician*, November, Vol. 40, No. 4, pp. 325-328.

Wadsworth, H.M., Stephens, K.S., and Godfrey, A.B. (1986), *Modern Methods for Quality Control and Improvement*, New York: John Wiley.

Wald, A., (1943), "Tests of statistical hypotheses concerning several parameters when the number of observations is large," *Transactions of the American Mathematical Society*, Vol. 54, pp. 426-482.

Wilks, S.S., (1938), "The large-sample distribution of the likelihood ratio for testing composite hypotheses," *Annals of mathematical Statistics*, Vol. 9, pp. 60-62.

Winterhalter, A.J., (1945), "Development of Reject Limits for Measurements," *Industrial Quality Control*, Vol. 1, Nos. 4 and 5.

Wolynetz, M.S., (1979), "Maximum Likelihood Estimation from Confined and Censored Normal Data," *Applied Statistics*, Vol. 28, pp. 185-195.

APPENDICES

Appendix A: Notation

This appendix summarizes most of the notation and acronyms used in this thesis.

Latin alphabet

a	Weibull shape parameter
b	Weibull scale parameter
c	arbitrary constant
CLT	central limit theorem
f_N	probability density function (p.d.f.) of the normal distribution
F_N	cumulative density function (c.d.f) of the normal distribution
f_W	p.d.f of the Weibull distribution
F_W	c.d.f. of the Weibull distribution
GLR	generalized likelihood ratio
H_0	null hypothesis
H_1	alternate hypothesis
H_{-1}	alternative hypothesis in the downward direction
i	unit index
j	group index
LACL	lower acceptance control limit
LCL	lower control limit
m	multiple of sigma units used to set control limits for Shewhart charts
MLE	maximum likelihood estimate
n	sample size
OC	operating characteristic
p_a	probability of failure under proof-load in mode A.
p_a^*	maximum likelihood estimate of p_a
PL_a	proof-load level in mode A.
Q	vector of sample grouping
Q_i	observed number of units in group i
t_i	location of standardized gauge limits
UACL	upper acceptance control limit
UCL	upper control limit
w	weights associated with H_{-1} vs. H_1
X_i	location of physical gauge limits
Y	random variable representing value of quality characteristic
z	weights associated with H_0 vs. H_1 test

Greek Alphabet

α	type I error rate
α'	actual type I error rate
β	type II error rate
β'	actual type II error rate
γ^*	critical GLR value
ϕ	probability density function
Φ	cumulative distribution function of the standard normal distribution
λ	critical likelihood ratio value
λ_α	adjustment of λ for sample size increase based on false alarm equation
λ_β	adjustment of λ for sample size increase based on power equation
μ	mean of the normal distribution
μ_0	stable mean value
μ_1	alternative mean value greater than μ_0
μ_{-1}	alternative mean value less than μ_0
μ_a	acceptable mean value
μ_r	rejectable mean value
μ_w	expected average w weight
μ_z	expected average z weight
π_j	probability of falling into group j
ρ_{ab}	correlation between strength modes A and B
ρ^*_{ab}	MLE for ρ_{ab}
θ	parameter of interest
θ_a	acceptable parameter value
θ_r	rejectable parameter value
θ_0	target parameter value
θ_1	parameter value of interest greater than θ_0
θ_{-1}	parameter value of interest less than θ_0
σ	standard deviation of the normal distribution
σ_w	standard deviation of the w weights
σ_z	standard deviation of the z weights

Appendix B: Interpretation of Weights

The group weight for group j, as expressed by equation (3.16), can be rescaled to be approximately equal the expected value of an observation that falls into group j given $\mu = \mu_0$. Since the weights can be rescaled, it is possible, without loss of generality, to restrict attention to the case when $\mu_0 = 0$ and $\sigma = 1$. For the normalized problem, μ_1 represents the size of mean shift we wish to detect given as a multiple of σ.

First, find the expected group value given $\mu = \mu_0$.

$$E\left(y \mid y \in j^{th} \text{ group}\right) = \frac{\int_{t_{j-1}}^{t_j} x\phi(x)dx}{\int_{t_{j-1}}^{t_j} \phi(x)dx}$$

$$= \frac{\phi(t_{j-1}) - \phi(t_j)}{Q(t_{j-1}, t_j)}$$

where $\phi(x) = \frac{1}{\sqrt{2\pi}}\exp\left(-x^2/2\right)$, the p.d.f. of the standard normal

$Q(x,y) = \int_{x}^{y} \phi(s)ds$,

and the t_j's are the standardized gauge limits.

Now consider the Taylor expansion in μ_1 of the weight assigned to all units that fall into group j, see equation (3.16) (assume $\mu_1 = -\mu_{-1}$).

$$\text{weight for group } j = \ln\left[\frac{\int_{t_{j-1}}^{t_j} \exp\left(-\frac{1}{2}(x-\mu_1)^2\right)}{\int_{t_{j-1}}^{t_j} \exp\left(-\frac{1}{2}(x+\mu_1)^2\right)}\right]$$

$$= \frac{2\left(\frac{1}{\sqrt{2\pi}}\exp\left(-t_{j-1}^2/2\right)-\frac{1}{\sqrt{2\pi}}\exp\left(-t_j^2/2\right)\right)}{\frac{1}{\sqrt{2\pi}}\int_0^{t_{j-1}}\exp\left(-x^2\right)dx-\frac{1}{\sqrt{2\pi}}\int_0^{t_j}\exp\left(-x^2\right)dx}\mu_1 + O\left(\mu_1^3\right)$$

$$= 2\frac{\phi\left(t_{j-1}\right)-\phi\left(t_j\right)}{Q\left(t_{j-1},t_j\right)}\mu_1 + O\left(\mu_1^3\right)$$

$$= 2\mu_1 E\left(y \mid y \in j^{th}\text{ group}\right) + O\left(\mu_1^3\right)$$

For reasonable values of t_{j-1} and t_j ($-5 < t < 5$) the coefficients for all terms in the above expansion of order higher than three is very small and decreasing as the order becomes higher. For typical values of μ_1, such as 1 or 2, these higher terms can be ignored, and the group weight for group j is approximately equal to $2\mu_1$ times the expected value of an observation that falls into group j given $\mu = \mu_0$.

Appendix C: Expected Value of Proof-load MLEs

This appendix contains the details of the proof that the MLEs given by equations (4.7) and (4.11) are all unbiased estimates.

Showing that the MLEs given by equations (4.7) are unbiased requires the following intermediate results:

$$E(n_a) = np_a, \qquad E(n_b) = n(p_b - p_{a\cap b})$$

$$E(m_a) = m(p_a - p_{a\cap b}), \qquad E(m_b) = mp_b$$

$$E(n_a \mid n_a + n_b) = \frac{p_a(n_a + n_b)}{p_a + p_b - p_{a\cap b}}$$

$$E(m_b \mid m_a + m_b) = \frac{p_b(m_a + m_b)}{p_a + p_b - p_{a\cap b}}$$

$$E(n_a m_b) = E(n_a)\,E(m_b)$$

$$E(m_a n_b) = E(m_a)\,E(n_b)$$

Using these expressions gives

$$\mathrm{E}\left(\frac{n_a(m_a + m_b)}{n_a + n_b} \mid n_a + n_b\right) = \frac{\mathrm{E}(n_a \mid n_a + n_b)}{n_a + n_b} E(m_a + m_b)$$

$$= \frac{\dfrac{p_a}{p_a + p_b - p_{a\cap b}}(n_a + n_b)}{n_a + n_b}\left(m(p_a - p_{a\cap b}) + mp_b\right)$$

$$= mp_a$$

Therefore, since this result does not depend on $n_a + n_b$

$$\mathrm{E}\left(\frac{n_a(m_a + m_b)}{n_a + n_b}\right) = mp_a$$

Similarly, $\mathrm{E}\left(\frac{m_b(n_a+n_b)}{m_a+m_b}\right) = np_b$

Also, we have

$$\mathrm{E}\left(\frac{m_b n_a - m_a n_b}{m_a+m_b} \,\middle|\, m_a+m_b\right) = \frac{\mathrm{E}(n_a)\mathrm{E}(m_b \mid m_a+m_b) - \mathrm{E}(n_b)\mathrm{E}(m_a \mid m_a+m_b)}{m_a+m_b}$$

$$= \frac{\dfrac{np_a p_b(m_a+m_b)}{p_a+p_b-p_{a\cap b}} - \dfrac{n(p_b-p_{a\cap b})(p_a-p_{a\cap b})(m_a+m_b)}{p_a+p_b-p_{a\cap b}}}{m_a+m_b}$$

$$= np_{a\cap b}$$

Therefore, $\mathrm{E}\left(\frac{m_b n_a - m_a n_b}{m_a+m_b}\right) = np_{a\cap b}$

and $\mathrm{E}\left(\frac{m_b n_a - m_a n_b}{n_a+n_b}\right) = mp_{a\cap b}$

Thus, considering the MLEs in expressions (4.7),

$$p_a^* = \frac{n_a(n_a+n_b+m_a+m_b)}{(n+m)(n_a+n_b)} = \frac{1}{n+m}\left(n_a + \frac{n_a(m_a+m_b)}{n_a+n_b}\right)$$

Therefore, $\mathrm{E}(p_a^*) = \frac{1}{n+m}(np_a+mp_a) = p_a$

Similarly, $p_b^* = \frac{1}{n+m}\left(m_b + \frac{m_b(n_a+n_b)}{m_a+m_b}\right)$

and therefore $\mathrm{E}(p_b^*) = p_b$

Also, $p_{a\cap b}^* = \frac{(m_b n_a - m_a n_b)(n_a+n_b+m_a+m_b)}{(n+m)(n_a+n_b)(m_a+m_b)}$

$$p_{a\cap b}^* = \frac{1}{n+m}\left(\frac{(m_b n_a - m_a n_b)}{m_a+m_b} + \frac{(m_b n_a - m_a n_b)}{n_a+n_b}\right)$$

$$\therefore \quad \mathrm{E}\left(p^{\bullet}_{a\cap b}\right) = \frac{1}{n+m}\left(np_{a\cap b} + mp_{a\cap b}\right) = p_{a\cap b}$$

Therefore, the MLEs given in equations (4.7) are all unbiased.

The MLEs given for Procedure III by equations (4.11) can be shown to be unbiased in a similar manner. To illustrate the technique it is shown that $\mathrm{E}\left(p^{\bullet}_{a2}\right) = p_{a2}$ and $\mathrm{E}\left(p^{\bullet}_{a2\cap b2}\right) = p_{a2\cap b2}$, the unbiasedness of the MLEs $p^{\bullet}_{a1}$, $p^{\bullet}_{b1}$, $p^{\bullet}_{b2}$, $p^{\bullet}_{a1\cap b2}$ and $p^{\bullet}_{a2\cap b1}$ follow directly.

The following intermediate results are required:

$$\begin{aligned}
E\left(n_{a2}\right) &= np_{a2}, & E\left(n_{b2}\right) &= n\left(p_{b2} - p_{a2\cap b2}\right) \\
E\left(m_{a2}\right) &= m\left(p_{a2} - p_{a2\cap b2}\right), & E\left(m_{b2}\right) &= mp_{b2} \\
E\left(n_{a2} \mid n_{a2} + n_{b2}\right) &= \frac{p_{a2}\left(n_{a2} + n_{b2}\right)}{p_{a2} + p_{b2} - p_{a2\cap b2}} \\
E\left(m_{b2} \mid m_{a2} + m_{b2}\right) &= \frac{p_{b2}\left(m_{a2} + m_{b2}\right)}{p_{a2} + p_{b2} - p_{a2\cap b2}} \\
E\left(n_{a2}m_{b2}\right) &= E\left(n_{a2}\right)E\left(m_{b2}\right) \\
E\left(m_{a2}n_{b2}\right) &= E\left(m_{a2}\right)E\left(n_{b2}\right)
\end{aligned}$$

Using these expressions gives

$$\begin{aligned}
\mathrm{E}\left(\frac{n_{a2}\left(m_{a2} + m_{b2}\right)}{n_{a2} + n_{b2}} \middle| n_{a2} + n_{b2}\right) &= \frac{\mathrm{E}\left(n_{a2} \mid n_{a2} + n_{b2}\right)}{n_{a2} + n_{b2}} E\left(m_{a2} + m_{b2}\right) \\
&= \frac{\dfrac{p_{a2}}{p_{a2} + p_{b2} - p_{a2\cap b2}}\left(n_{a2} + n_{b2}\right)}{n_{a2} + n_{b2}}\left(m\left(p_{a2} - p_{a2\cap b2}\right) + mp_{b2}\right) \\
&= mp_{a2}
\end{aligned}$$

Therefore, since this result does not depend on $n_{a2} + n_{b2}$

$$\mathrm{E}\left(\frac{n_{a2}\left(m_{a2} + m_{b2}\right)}{n_{a2} + n_{b2}}\right) = mp_{a2}$$

Also,

$$\mathrm{E}\left(\frac{m_{b2}n_{a2}-m_{a2}n_{b2}}{m_{a2}+m_{b2}}\,\middle|\, m_{a2}+m_{b2}\right) = \frac{\mathrm{E}(n_{a2})\mathrm{E}(m_{b2}\mid m_{a2}+m_{b2})-\mathrm{E}(n_{b2})\mathrm{E}(m_{a2}\mid m_{a2}+m_{b2})}{m_{a2}+m_{b2}}$$

$$= \frac{\dfrac{np_{a2}p_{b2}(m_{a2}+m_{b2})}{p_{a2}+p_{b2}-p_{a2\cap b2}} - \dfrac{n(p_{b2}-p_{a2\cap b2})(p_{a2}-p_{a2\cap b2})(m_{a2}+m_{b2})}{p_{a2}+p_{b2}-p_{a2\cap b2}}}{m_{a2}+m_{b2}}$$

$$= np_{a2\cap b2}$$

Therefore, $\mathrm{E}\left(\dfrac{m_{b2}n_{a2}-m_{a2}n_{b2}}{m_{a2}+m_{b2}}\right) = np_{a2\cap b2}$

and $\mathrm{E}\left(\dfrac{m_{b2}n_{a2}-m_{a2}n_{b2}}{n_{a2}+n_{b2}}\right) = mp_{a\cap b}$

Thus $p^{*}_{a2} = \dfrac{n_{a2}(n_{a2}+n_{b2}+m_{a2}+m_{b2})}{(n+m)(n_{a2}+n_{b2})} = \dfrac{1}{n+m}\left(n_{a2}+\dfrac{n_{a2}(m_{a2}+m_{b2})}{n_{a2}+n_{b2}}\right)$

Therefore, $\mathrm{E}(p^{*}_{a2}) = \dfrac{1}{n+m}(np_{a2}+mp_{a2}) = p_{a2}$

Also, $p^{*}_{a2\cap b2} = \dfrac{(m_{b2}n_{a2}-m_{a2}n_{b2})(n_{a2}+n_{b2}+m_{a2}+m_{b2})}{(n+m)(n_{a2}+n_{b2})(m_{a2}+m_{b2})}$

$$p^{*}_{a2\cap b2} = \frac{1}{n+m}\left(\frac{(m_{b2}n_{a2}-m_{a2}n_{b2})}{m_{a2}+m_{b2}}+\frac{(m_{b2}n_{a2}-m_{a2}n_{b2})}{n_{a2}+n_{b2}}\right)$$

$$\therefore \quad \mathrm{E}(p^{*}_{a2\cap b2}) = \frac{1}{n+m}(np_{a2\cap b2}+mp_{a2\cap b2}) = p_{a2\cap b2}$$

Therefore, the MLEs p^{*}_{a2} and $p^{*}_{a2\cap b2}$ given in equations (4.11) are unbiased.

Appendix D: Gradient of Sample Size Formula

In addressing the question of optimal gauge limits in the case of a one-sided acceptance sampling plan it is necessary to find the gauge limits that minimize the required sample size. Although this minimization problems can be solved directly through the use of the Nelder-Mead multidimensional simplex algorithm more efficient techniques exist that utilize the gradient of the function to be minimized. As a result, this appendix shows the calculation of the gradient.

To determine the optimal gauge limits to detect shifts in the mean of a normal distribution, the sample size, as given in equation (3.4), must be minimized . For given α, β, and acceptable and rejectable mean values μ_a and μ_r, equation (3.4) is only a function of the standardized gauge limits **t**. From Section 3.1, equation (3.4) is

$$n = \left(\frac{\Phi^{-1}(1-\beta)\sigma_z(\mu_r) - \Phi^{-1}(\alpha)\sigma_z(\mu_a)}{\mu_z(\mu_r) - \mu_z(\mu_a)} \right)^2.$$

Taking the derivative with respect to the gauge limit t_i gives:

$$\frac{\partial n}{\partial t_i} = 2\left(\frac{\Phi^{-1}(\alpha)\sigma_z(\mu_a) - \Phi^{-1}(1-\beta)\sigma_z(\mu_r)}{\mu_z(\mu_a) - \mu_z(\mu_r)} \right)$$

$$\left(\frac{\Phi^{-1}(\alpha)\partial\sigma_z(\mu_a)/\partial t_i - \Phi^{-1}(1-\beta)\partial\sigma_z(\mu_r)/\partial t_i}{\mu_z(\mu_a) - \mu_z(\mu_r)} \right)$$

$$\left(\frac{\Phi^{-1}(\alpha)\sigma_z(\mu_a) - \Phi^{-1}(1-\beta)\sigma_z(\mu_r)}{(\mu_z(\mu_a) - \mu_z(\mu_r))^2} \right)\left(\frac{\partial\mu_z(\mu_a)}{\partial t_i} - \frac{\partial\mu_z(\mu_r)}{\partial t_i} \right)$$

where $\frac{\partial\mu_z(\mu)}{\partial t_j} = \phi(t_j,\mu)(z_j - z_{j+1}) + \frac{\pi_j(\mu)}{\pi_j(\mu_r)}\left(\phi(t_j,\mu_r) - \frac{\pi_j(\mu_r)}{\pi_j(\mu_a)}\phi(t_j,\mu_a) \right)$

$$-\frac{\pi_{j+1}(\mu)}{\pi_{j+1}(\mu_r)}\left(\phi(t_j,\mu_r)-\frac{\pi_{j+1}(\mu_r)}{\pi_{j+1}(\mu_a)}\phi(t_j,\mu_a)\right)$$

and $$\frac{\partial\sigma_z(\mu)}{\partial t_j} = \sigma_z(\mu)^{-1}\left(-2\mu_z(\mu)\frac{\partial\mu_z(\mu)}{\partial t_j}+\frac{\partial\pi_j(\mu)}{\partial t_j}z_j^2+\frac{\partial\pi_{j+1}(\mu)}{\partial t_j}z_{j+1}^2 \right.$$
$$\left. +2\pi_j(\mu)z_j\frac{\partial z_j}{\partial t_j}+2\pi_{j+1}(\mu)z_{j+1}\frac{\partial z_{j+1}}{\partial t_j}\right)$$

This follows since, as given in Section 3.1,

$$\mu_z(\mu) = \sum_{j+1}^{k+1}\pi_j(\mu)\ln\left(\frac{\pi_j(\mu_a)}{\pi_j(\mu_r)}\right),$$

$$\sigma_z(\mu) = \sqrt{\sum_{j=1}^{k+1}\pi_j(\mu)\ln\left(\frac{\pi_j(\mu_a)}{\pi_j(\mu_r)}\right)-\mu_z^2(\mu)}$$

and

$$\pi_j(\mu) = \int_{t_{j-1}}^{t_j}\phi(y;\mu)dy,$$

$$\phi(t,\mu) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}(t-\mu)^2\right)$$

$$\frac{\partial\pi_j(\mu)}{\partial t_j} = \phi(t_j,\mu)$$

$$\frac{\partial\pi_{j+1}(\mu)}{\partial t_j} = -\phi(t_j,\mu)$$

Appendix E: Normal Information Gradient

The optimal gauge limits for Shewhart control charts are the limits that maximize the expected statistical information available in a sample. The maximization can be done most quickly if also provided with gradients. The gradient of the expected information for the mean and standard deviation of a normal distribution is derived below. For Shewhart type charts to detect mean shifts it is necessary to maximize (5.3) which for given α, β, μ_0 and μ_1 is a function only of the gauge limits $\mathbf{t}$. The equation (5.3) is repeated below:

$$E(I(\mu)) = \frac{\phi(t_1;\mu)^2}{\pi_1(\mu)} + \sum_{j=2}^{k} \frac{\left(\phi(t_{j-1};\mu) - \phi(t_j;\mu)\right)^2}{\pi_j(\mu)} + \frac{\phi(t_k;\mu)^2}{\pi_{k+1}(\mu)}$$

Then since,

$$\pi_j(\mu) = \int_{t_{j-1}}^{t_j} \phi(y;\mu)dy,$$

$$\phi(x,\mu) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}(x-\mu)^2\right) \qquad \frac{\partial\theta}{\partial x} = \frac{\mu - x}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}(x-\mu)^2\right)$$

$$\frac{\partial\pi_j}{\partial t_j} = \phi(t_j,\mu) \qquad \frac{\partial\pi_{j+1}}{\partial t_j} = -\phi(t_j,\mu)$$

Defining $t_0 = -\infty$, and $t_{k+1} = \infty$, the derivative of the expected information about the mean with respect to the group limits is:

$$\frac{\partial E(I(\mu))}{\partial t_j} = \frac{2\left(\phi(t_j) - \phi(t_{j+1})\right)}{\pi_{j+1}}\frac{\partial\phi(t_j)}{\partial t_j} + \frac{\left(\phi(t_j) - \phi(t_{j+1})\right)^2}{\pi_{j+1}^2}\phi(t_j)$$
$$- \frac{2\left(\phi(t_{j-1}) - \phi(t_j)\right)}{\pi_j}\frac{\partial\phi(t_j)}{\partial t_j} - \frac{\left(\phi(t_{j-1}) - \phi(t_j)\right)^2}{\pi_j^2}\phi(t_j)$$

For Shewhart type charts to detect standard deviation shifts expression (5.4) must be maximized, which also for given α, β, μ_0 and μ_1 is a function just of the gauge limits **t**. The equation (5.4) is repeated below:

$$E(I(\sigma)) = \frac{t_1^2\phi(t_1;\sigma)^2}{\sigma\,\pi_1(\sigma)} + \sum_{j=2}^{k}\frac{(t_{j-1}-t_j)^2\left(\phi(t_{j-1};\sigma)-\phi(t_j;\sigma)\right)^2}{\sigma\,\pi_j(\sigma)} + \frac{t_k^2\phi(t_k;\sigma)^2}{\sigma\,\pi_{k+1}(\sigma)}$$

Using the above equations gives:

$$\frac{\partial E(I(\sigma))}{\partial t_j} = \frac{2\left(t_j\phi(t_j)-t_{j+1}\phi(t_{j+1})\right)}{\pi_{j+1}(\sigma)}\frac{\partial\phi(t_j)}{\partial t_j} + \frac{\left(t_j\phi(t_j)-t_{j+1}\phi(t_{j+1})\right)^2}{\pi_{j+1}^2(\sigma)}\phi(t_j)$$
$$-\frac{2\left(t_{j-1}\phi(t_{j-1})-t_j\phi(t_j)\right)}{\pi_j(\sigma)}\frac{\partial\phi(t_j)}{\partial t_j} - \frac{\left(t_{j-1}\phi(t_{j-1})-t_j\phi(t_j)\right)^2}{\pi_j^2(\sigma)}\phi(t_j)$$

Appendix F: Weibull Information Gradients

The optimal gauge limits for Shewhart control charts are the limits that maximize the expected statistical information available. This appendix provides the gradients with respect to the gauge limits for the expected information of a grouped Weibull distribution.

For Shewhart type charts to detect shape parameter shifts the expected information as expressed in equation (5.5) must be maximized. Given α, β, μ_0 and μ_1 equation (5.5) is a function only of the gauge limits **t**. Equation (5.5) is repeated below:

$$E(I(a)) = \sum_{j=1}^{k+1} \frac{\left(t_j^a \ln(t_j)\exp(-t_j^a) - t_{j-1}^a \ln(t_{j-1})\exp(-t_{j-1}^a)\right)^2}{\exp(-t_{j-1}^a) - \exp(-t_j^a)}$$

Then defining

$$h_i = t_i^a \ln(t_i)\exp(-t_i^a)$$

and

$$g_i = a h_i + t_i^a \exp(-t_i^a) - a h_i t_i^a,$$

gives

$$\frac{\partial E(I(a))}{\partial t_j} = \frac{2(h_i - h_{i-1})g_i}{t_i\left(\exp(t_{i-1}^a) - \exp(t_i^a)\right)} - \frac{a(h_i - h_{i-1})^2 t_i^a \exp(t_i^a)}{t_i\left(\exp(t_{i-1}^a) - \exp(t_i^a)\right)^2}$$
$$- \frac{2(h_{i+1} - h_i)g_i}{t_i\left(\exp(t_i^a) - \exp(t_{i+1}^a)\right)} + \frac{a(h_{i+1} - h_i)^2 t_i^a \exp(t_i^a)}{t_i\left(\exp(t_i^a) - \exp(t_{i+1}^a)\right)^2}$$

For Shewhart type charts to detect scale parameter shifts from a Weibull distribution the optimal gauge limits maximize the expected information as expressed in equation (5.6):

$$E(I(b)) = \sum_{j=1}^{k+1} \frac{\left(\frac{t_{j-1}}{b^2}\exp\left(-\frac{t_{j-1}}{b}\right) - \frac{t_j}{b^2}\exp\left(-\frac{t_j}{b}\right)\right)^2}{\exp\left(-t_{j-1}/b\right) - \exp\left(-t_j/b\right)}$$

Now define

$$d_i = \frac{t_i}{b^2}\exp\left(-\frac{t_i}{b}\right)$$

Then using the above definitions gives:

$$\frac{\partial E(I(b))}{\partial t_j} = \frac{2(d_{i-1}-d_i)\left(\frac{t_i}{b}-1\right)d_i}{t_i\left(\exp(t_{i-1}/b)-\exp(t_i/b)\right)} - \frac{bd_i(d_{i-1}-d_i)^2}{t_i\left(\exp(t_{i-1}/b)-\exp(t_i/b)\right)^2}$$
$$- \frac{2(d_i-d_{i+1})\left(\frac{t_i}{b}-1\right)d_i}{t_i\left(\exp(t_i/b)-\exp(t_{i+1}/b)\right)} + \frac{bd_i(d_i-d_{i+1})^2}{t_i\left(\exp(t_i/b)-\exp(t_{i+1}/b)\right)^2}$$