A FULLY BAYESIAN APPROACH TO SAMPLE SIZE DETERMINATION FOR VERIFYING PROCESS IMPROVEMENT

A Dissertation by

Nasser Safaie

Master of Science, Amirkabir University of Technology, Tehran, Iran, 1996

Bachelor of Science, Ferdowsi University of Mashhad, Mashhad, Iran, 1990

Submitted to the Department of Industrial and Manufacturing Engineering
and the faculty of the Graduate School of
Wichita State University
in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy

December 2010

A FULLY BAYESIAN APPROACH TO SAMPLE SIZE DETERMINATION
FOR VERIFYING PROCESS IMPROVEMENT

The following faculty members have examined the final copy of this dissertation for form and content, and recommend that it be accepted in partial fulfillment of the requirement for the degree of Doctor of Philosophy with a major in Industrial Engineering.

_____
Gamal S. Weheba, Committee Chair

_____
S. Hossein Cheraghi, Committee Member

_____
Abu S. M. Masud, Committee Member

_____
M. Edwin Sawan, Committee Member

_____
Mehmet Bayram Yildirim, Committee Member

_____
Pingfeng Wang, Committee Member

Accepted for the College of Engineering

_____
Zulma Toro-Ramos, Dean

Accepted for the Graduate School

_____
J. David McDonald, Dean

DEDICATION

To Hidden Imam Mahdi (A.S.), One who Rises,
may God hasten his reappearance.

ACKNOWLEDGEMENTS

I would like to acknowledge my advisor, Dr. Gamal Weheba, for his dedicated support, sustained guidance, and continuous supervision. Obviously, without his generosity of knowledge sharing, this dissertation would not have been completed.

I also acknowledge specifically Dr. Cheraghi for supporting me every step of the way during my academic study.

I thank my other committee members Dr. Masud, Dr. Yildirim, Dr. Sawan, and Dr. Wang for their helpful comments during all stages of my work. Their valuable suggestions made this study a valuable experience.

My deepest gratitude goes to my parents; my late father, Mr. Abbas Safaie, and my mother, Mrs. Tahereh Hosseini whom I owe my life, for their encouragement and support.

Finally, I would greatly appreciate my wife, Mansoureh, and my children, Alireza and Zahra, for their invaluable encouragement, patience, and cooperation throughout my academic studies. They have sacrificed much so that I could pursue my goals.

ABSTRACT

There has been significant growth in the development and application of Bayesian methods in industry. The Bayes' theorem describes the process of learning from experience and shows how knowledge about the state of nature is continually modified as new data become available. This research is an effort to introduce the Bayesian approach as an effective tool for evaluating process adjustments aimed at causing a change in a process parameter. This is usually encountered in scenarios where the process is found to be stable but operating away from the desired level. In these scenarios, a number of changes are proposed and tested as part of the improvement efforts. Typically, it is desired to evaluate the effect of these changes as soon as possible and take appropriate actions. Despite considerable research efforts to utilize the Bayesian approach, there are few guidelines for loss computation and sample size determination. This research proposed a fully Bayesian approach for determining the maximum economic number of measurements required to evaluate and verify such efforts. Mathematical models were derived and used to establish implementation boundaries from economic and technical viewpoints. In addition, numerical examples were used to illustrate the steps involved and highlight the economic advantages of the proposed procedures.

TABLE OF CONTENTS

TABLE OF CONTENTS (continued)

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ACC	Average Coverage Criterion
AIAG	Automotive Industry Action Group
ALC	Average Length Criterion
ARL	Average Run Length
ARL_0	Average Run Length in Control
ARL_1	Average Run Length Out of Control
BSSD	Bayesian Sample Size Determination
CLT	Central Limit Theorem
CUSUM	Cumulative Sum
EWMA	Exponentially Weighted Moving Average
HPD	Highest Posterior Density
i.i.d.	independent and identically distributed
LSL	Lower Specification Limit
MEU	Maximum Expected Utility
SPC	Statistical Process Control
SSD	Sample size Determination
USL	Upper Specification Limit
WOC	Worst Outcome Criterion

CHAPTER 1

INTRODUCTION

One of the most important uses of control charts is to distinguish assignable causes from chance causes of process variation and verify whether the process is in control or not. Chance (Common) causes are inherent in the process all the time. A process that is operating with only chance causes of variation is stable, i.e., the process performance is not a function of time and follows the random law. In this case, the process is predictable and in a state of statistical control. On the other hand, a process that is operating in the presence of assignable causes is out of control. The assignable causes create a portion of the variability in a set of samples that need to be investigated.

Chakraborti et al. (2009) indicated that control chart applications involve two phases. In phase I, a set of historical data is analyzed to assess stability and identify special causes. If no special causes are present, then the in control process parameters are estimated. In phase II, the data is sequentially collected over time to assess whether the performance has been changed from the estimated level. However, there are some scenarios where the process is found to be centered away from the desired level, and changes are sought to improve its performance. In this case, the proposed changes must be tested to verify their effect (Wadsworth et al., 2002). The duration of this activity can be referred to as a transition period. Under these conditions, users of the traditional control charts are limited to two options: assume complete ignorance of the old process level and repeat phase I, or utilize the desired level as the standard value for phase II. The former requires accumulating a new set of data, while the latter may result in an accept/reject decision without providing an updated estimate of the process parameter.

The main purpose of this dissertation is to explore applications of the Bayesian approach as a third alternative to update prior knowledge about the process parameter. This includes the development of a mathematical model for determining the maximum economic number of measurements. The Bayes' theorem is applied to combine prior beliefs with evidence from the sample to make an inference about the parameter of interest. Mathematical models have been derived and used to establish implementation boundaries from economic and technical viewpoints. In addition, numerical examples have been used to illustrate the steps involved and highlight the economic advantages of the procedures developed.

The following chapter presents a review of the relevant literature and highlights some of the merits of the Bayesian techniques for sample size determination (SSD). Chapter 3 is devoted to a discussion regarding differences and similarities between the traditional and Bayesian methods. In Chapter 4, the research gap is identified, and objectives of this research are stated. Chapter 5 presents the procedures followed in deriving mathematical models for determining sample size. Models are developed to include three cases: the first two involve a shift in the process average, while the third involves a reduction in the process variance. Concluding remarks and recommendations for future research are presented in Chapter 6.

CHAPTER 2

BACKGROUND AND LITERATURE REVIEW

This chapter includes a review of the literature pertaining to estimation and test of hypothesis, statistical process control (SPC), different control charts, and determining the sample size using a variety of methods. This literature review is divided into five main sections. The first section discusses the attributes and applications of traditional and Bayesian methods concerning parameter estimation and hypothesis testing. The second section discusses the importance of SPC and different control charts including Shewhart, Cumulative Sum (CUSUM), and Exponentially Weighted Moving Average (EWMA). The fourth section reviews various methods for determining sample size and compares the two popular methods: traditional and Bayesian. The fifth section addresses sequential sampling methods and the sequential Bayesian analysis. In the last section, different kinds of Taguchi's loss/utility functions and their properties are presented.

2.1 Estimation and Test of Hypothesis

The Bayesian and traditional methods deal differently with estimation and test of hypothesis for a parameter of interest. The main reason for this is related to the role of the posterior distribution in the Bayesian approach. The posterior distribution contains both sample and prior information, and any reference about the parameter should embrace the characteristics of its distribution. Here, new information is combined with previous information. When new information is obtained, the probability (uncertainty) is revised so that it may represent all available information (Congdon, 2007). In the Bayesian approach, parameters are random variables of prior distributions reflecting an accumulated state of knowledge. However, in the traditional methods, parameters are fixed values and are estimated based on the sample

information. Hence, it is necessary to calculate a statistic that is a function of the current sample data. The practitioner must do calculations for a specific problem. There is some evidence that the results of these two approaches tend to be the same. Broemeling (2009) indicated that in the presence of a non-informative prior, there are similarities between these two methods. That is why, in this situation, the posterior distribution is identical to the likelihood distribution. Hence, any reference from the posterior distribution depends entirely on the sample information, and as mentioned before, all information of the sample is summarized in the likelihood distribution. Another similarity between these two methods is that the power of the test in the traditional method, $1 - \beta$, is equivalent to Bayes function or expected loss in the Bayesian view (Bolstad, 2010). The differences between these two methods in estimation and test of hypothesis are discussed in the next section.

2.1.1 Point Estimation

Point estimation identifies a single number, calculated from a random sample data, as being reasonably close to the unknown parameter. According to Congdon (2005), an estimator is a function of the sample (i.e., any statistic), while an estimate is the realized value of an estimator (i.e., a number). In traditional methods, probabilities are to be interpreted as relative frequencies. In addition, the parameter is unknown but fixed and it is estimated based on the observed values in the sample (Montgomery and Runger, 2007). Sometimes probabilities are interpreted as individual prior beliefs about parameter values, which are translated to the Bayesian approach. In this approach, a parameter is considered to be a quantity, whose variation can be described by a subjective prior distribution. Prior distribution is one's belief and formulated before sampling. The updated prior based on the sample is called posterior distribution. Briefly, posterior distribution is a conditional distribution based upon observing the

sample. It is still a random quantity, not a fixed number, so considering different possible values of a parameter is possible. This flexibility is dealt with by calculating a different posterior distribution, given the different prior distributions. The mean of the posterior distribution is used as a point estimate of the parameter. This is because the posterior mean has the smallest standard error and minimizes the posterior variance or mean squared error (Reichert et al., 2002). This standard error is less than σ because the Bayesian method is using prior information as well as sample information to estimate the parameter. As an example of a normal Bayesian estimator, assume that θ is a parameter of interest, $x \sim N(\mu, \sigma_g^2)$, and the prior distribution of θ is $N(\mu, \sigma^2)$. The posterior distribution of θ is also normal with the mean and variance expressed by

$$\mathrm{E}(\theta|x) = \frac{\sigma^2}{\sigma^2 + \sigma_g^2}\, x + \frac{\sigma_g^2 \sigma^2}{\sigma_g^2 + \sigma^2}\, \mu \tag{2.1}$$

$$\mathrm{Var}(\theta|x) = \frac{\sigma_g^2 \sigma^2}{\sigma_g^2 + \sigma^2} \tag{2.2}$$

As can be seen, the posterior mean is an estimator of μ with a linear combination of the prior information and sample information. As the prior information becomes more non informative, the Bayes estimator gives more weight to the sample knowledge. On the other hand, if the prior information is informative so that $\sigma_g^2 > \sigma^2$, then the weight of prior is increased. Berger (1985) indicated that the calculated standard error of a point estimate in the Bayesian approach is less than that in traditional methods, unless a non informative prior is utilized.

2.1.2 Interval Estimation

The interval estimation provides an interval of values, instead of a single estimated value, bounded by lower and upper limits, within which the true value of the unknown parameter is asserted to lie. In traditional methods, a sample of pre specified size *n*, a sampling distribution of the statistic computed, and a confidence level $(1-\alpha)$ should be specified. Then the confidence

interval is calculated as an interval between two values $\widehat{\theta_L}$ and $\widehat{\theta_u}$, which has enclosed the parameter θ with a specified probability. The Bayesian analog of a traditional confidence interval is called a credible set (Congdon, 2003). In Bayesian interval estimation methods, independent and identically distributed (i.i.d.) observations (x_1, x_2 ,…, x_n) are drawn; then, based on the prior and the sample, the maximum likelihood function is calculated. Applying Bayes' theorem, the posterior is computed and utilized to construct the credible interval. A credible interval is the Bayesian version of a confidence interval, except that it is easier to be interpreted. Pham-Gia and Turkkan (1992) defined credibility interval as a probability statement attached to the interval as a measure of the strength of one's belief, taking into account one's prior knowledge and the observations. The notion of Highest Posterior Density (HPD) is often used to determine an appropriate credible interval. A specific credible interval (HPD) has the feature that the density for every point inside the interval is greater than that for every point outside the interval. In other words, every point included has higher probability density than every point excluded. A $(1-\alpha)100\%$ HPD credible set includes only those points with the largest posterior density (Carlin and Louis, 2009). For example in a traditional confidence interval, assume $\alpha = 5\%$; then the interval is calculated as $\mu \epsilon (\mu_1, \mu_2)$. It is tempting to interpret that "the probability is 95% that μ is in the interval (μ_1, μ_2)." However, since the unknown parameter is a fixed value, this interpretation is wrong. Indeed, the probability of μ lying in the interval is 0 or 1. In other words, the above statement is valid before sampling, not after (please refer to section 2.3.1). In contrast, the calculated interval credibility by the Bayesian setup provides an actual probability statement, based only on the sampling information and prior opinion. Hence, it is valid to state that μ lies in the interval with some probability, not 0 or 1. This is because μ is a random variable with a probability distribution, not a fixed value. From an intuitive viewpoint, it

is notable that traditional statistics place the emphasis on only one predetermined significance level at a time, while in Bayesian statistics various credible intervals are not limited to only one interval. The practitioner could use composite intervals, such as 90% HPD, 95% HPD, or 99% HPD, simultaneously. Nevertheless, nothing is free. The ease of application and interpretation of the results is because the Bayesian paradigm requires more input than the traditional model.

2.1.3 Hypothesis Testing

The goal of a hypothesis test is to decide, based on a sample, which of two complementary alternatives is true. Because the parameter is an unknown fixed number in the traditional setup, the result of a hypothesis test is either true or false. That is, $a_0 = p(H_0\ is\ true\ |x)$ or $a_1 = p(H_1\ is\ true\ |x)$ is meaningless, and each of the alternatives is either to be accepted or rejected. Casella and Berger (2002) indicated that in traditional statistics, these two probabilities are unknown because the parameter is unknown, and these probabilities do not depend on the sample *x*. On the other hand, from the Bayesian viewpoint, all inferences about the parameter of interest are based on the posterior distribution, which is a probability distribution for a random variable. Despite traditional statistics, the Bayesian approach does not need to specify an arbitrary level of significance or predetermine the outcome of the test according to the selection of the sample size (Press, 1989). Thus, the above two probabilities depend on sample *x*, are meaningful, and could be calculated, and the decision between H_0 and H_1 is made based on the posterior probabilities of a_0 and a_1. Indeed, a_0 and a_1 are the probabilities of the hypotheses in light of data and prior beliefs. Hence, posterior probabilities of hypotheses are the Bayesian measures in testing. Meyer and Collier (1970) showed that for only one prior belief as a null hypothesis, as many as 31 alternatives are possible to be tested in one

problem. As a result, since it can directly test a composite hypothesis, Bayesian statistics in terms of test of hypothesis is more practical compared to traditional statistics.

Carlin and Louis (2000) discussed four specific properties of the test procedure and explained how the two traditional and Bayesian approaches deal with them. First, in traditional hypothesis testing, the two alternatives are nested, one within the other, while in practice, many cases involve a choice between more than two models that are not nested. On the other hand, in the Bayesian paradigm, there is no restriction on the number of hypotheses that may be concurrently considered, and there is no need for them to be nested within any of the others. Second, assume in testing a point null hypothesis: $H_0: \theta = \theta_0$ and $H_1: \theta \neq \theta_0$. One of the differences between these two methods for test of hypotheses is that sometimes H_0 is being rejected even though the true mean is negligibly different from θ_0. In other words, it is said that in hypothesis testing, the loss of distance from the true hypothesis is not accurately measured by the traditional approach. It is common that the loss from an incorrect decision increases as a function of the "distance" of the parameter θ from the true value of θ_0. However, in the Bayesian methods, anyone can seek to utilize the information by simply processing the likelihood function with their own prior. It is possible to determine the posterior and calculate the exact amount of loss corresponding to the personal belief as a prior. Third, in the traditional approach, it is common to say "fail to reject the null hypothesis" instead of "accept the null hypothesis." The reason for this is that the result can only offer evidence against the null. A small *P*-value indicates that the alternative has more power. However a large *P*-value does not result in the equivalency of two alternatives; it only indicates the lack of evidence that they are not equivalent. Fourth, the *P*-value is only computable for data arising from the original design in the traditional method. However, it is customary for unforeseen situations to arise during the

experiment, resulting in data that are not exactly what was intended in the experiment design. The Bayesian techniques avoid this limitation. Next section discusses the role of SPC in process improvement and application of different control charts for process monitoring.

2.2 Statistical Process Control

Statistical process control has been used to control quality for several decades and has proven to be extensively effective in helping factories meet competitive challenges. SPC, as a passive statistical method, is a powerful collection of problem-solving techniques, which is useful in achieving process stability and increasing capability through the reduction in variability (Woodall and Montgomery, 1999). According to Stoumbos et al. (2000), SPC is applied on four categories: process monitoring, planning, evaluating customer satisfaction, and forecasting. Among these, process monitoring is used to control production processes over time to detect changes in process performance as soon as possible. Responding to needs of the customer, the research community continuously attempts to provide more effective tools.

The control charts, as the most useful process monitoring tools in SPC, detect changes in a process that may affect the quality of the output (MacCarthy and Wasusri, 2002). Shewhart (1931) stated that "The control chart may serve, first, to define the goal or standard for a process that management strives to attain; second, it may be used as an instrument for achieving that goal, and third, it may serve as a means of judging whether the goal has been reached." Duncan (1956) defined a control chart as "a statistical device principally used for the study and control of repetitive processes."

Montgomery (2009) discussed five advantages of using the control chart. First, the control chart helps judge the past process performance, facilitates process supervision, and the product inspection. It also enables workers to detect and remove special causes. Second, the

control chart helps keep the process in state of control which is in agreement with this statement, "do it right the first time", and to have an effective process control. Third, it helps avoid two mistakes; over-adjust (unnecessary adjustment) and under-adjust (do nothing), both lead to deterioration performance. Forth, the pattern of points on the control chart provides diagnostic information regarding some changes in the process that may improve the process performance. The effects of changes also can be measured quickly and reliably. Fifth, the control chart provides information about the process capability which is tremendously useful to the process designers.

Woodall (2006) noted that the eventual goal of SPC is the reduction or elimination of variability in the process by identifying the assignable causes. Control charts are more effective when integrated with a comprehensive SPC program. Figure 2.1 shows elements of a control chart in which θ and α denote, respectively, the sample statistic and type I error probability.

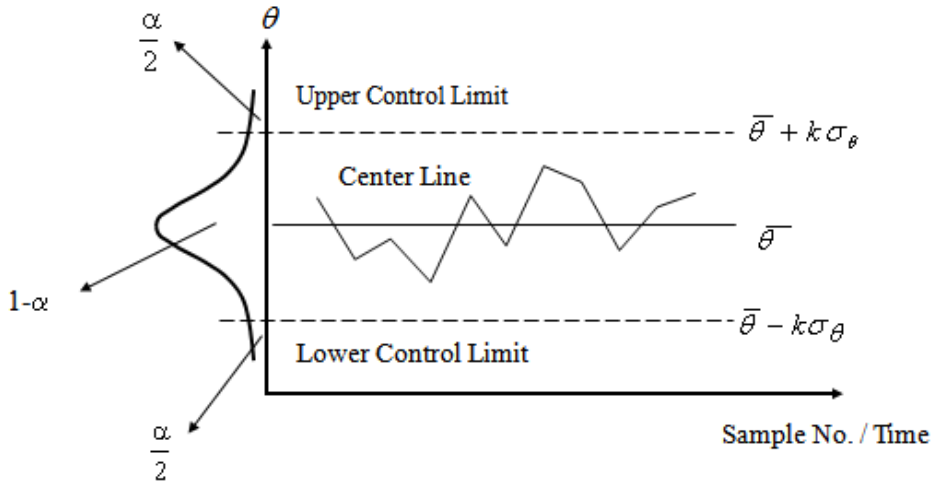


Figure 2.1 Elements of a control chart (Montgomery, 2009)

As mentioned previously, one of the main purposes of control charting is to distinguish between common and assignable causes of variation in order to prevent overreaction and under reaction to variability in process performance over time. When uncommon sources of variability

exist, sample statistics have been plot outside the control limits. This is a signal that some investigation of the process should be made and corrective actions taken to remove these unusual sources of variability. Researchers have pointed out a substantial difference between the control charts. The manner in which past information is weighted is the point to be emphasized in this difference. Tagaras (1998) pointed out that Shewhart charts do not assign any weight to earlier sample statistics, whereas CUSUM charts assign equal weights, and EWMA chart assigns more weight to recent information.

2.2.1 Shewhart Control Charts

The first control chart was introduced by Shewhart (1931) in the 1920s to monitor the mean of independently distributed data from a univariate normal process. Since then, control schemes have been extensively applied in improving the quality of the process. The statistical model considered by Shewhart assumes constant process parameters and independent identically distributed (i.i.d.) process measurements. This is expressed as

$$x_i = \mu + \varepsilon_i \tag{2.3}$$

where x_i is a sample data, μ is the process average, and ε_i is the i.i.d. normal random error with mean zero and variance σ^2. Shewhart control charts are more effective when the in control process data is stationary and uncorrelated. According to equation (2.3), data tend to fluctuate around a constant average in a predictable manner. Since their establishment, the control schemes have found extensive applications in improving the quality of the processes to date. However, control charts can only be effective when the underlying assumptions (independency and normality) are not violated. These charts lack robustness when these conditions are not followed (Woodall and Mahmoud, 2005).

A Shewhart control chart is modeled as follows: observations are made over time on a process, and control charts are used to detect departures from statistical control. The center line is μ_w, and the general model for calculating upper and lower control limits is

$$\mu_w \pm L\sigma_w \tag{2.4}$$

where L is the distance of the control limit from the center line (control limit coefficient), μ_w is the mean of the sample statistic w, and σ_w is standard deviation of the statistic. Deming (1986) indicated that the three-sigma Shewhart chart works for a variety of broad situations, and no one has been able to create an enhancement to it. He pointed out that under the assumption of known and constant variance σ^2, if the distribution of the quality characteristic is reasonably well approximated by the normal distribution, then the use of three-sigma limits (L = 3) is applicable.

As can be seen from this model, the control statistic computed is a function of the last sample only and the cumulative or weight of the prior observations is ignored. This feature makes Shewhart control charts less efficient for detecting small and moderate-sized changes. The increasing emphasis today on high-quality products increases the importance of detecting small changes in the process average.

2.2.2 CUSUM Control Charts

Montgomery (2009) stated that "The Shewhart control chart uses only the information about the process contained in the last sample observation and it ignores any information given by the entire sequence of points. This feature makes the Shewhart control chart relatively insensitive to small process shifts, say, on the order of about 1.5 σ or less". In the Cumulative Sum (CUSUM), proposed by Page (1954), information is accumulated over time. Because of the combination of information obtained from several samples, the CUSUM is known to detect small

shifts in process parameters well. This control chart can be applied to individual observations. CUSUM statistics are computed as

$$C_i^+ = max[0, x_i - (\mu_0 + K) + C_{i-1}^+] \tag{2.5}$$

$$C_i^- = max\left[0, (\mu_0 - K) - x_i + C_{i-1}^-\right] \tag{2.6}$$

where the initial values are $C_0^+ = C_0^- = 0$, μ_0 is the target value, *K* is the reference value (often selected halfway between target μ_0 and the out of control value of the mean μ_1), and $C_i^+$and C_i^- are one-sided upper and lower CUSUMs, respectively. The process turns out to be out of control if C_i^+ or C_i^- exceed the decision interval *H*. The recommended values for *H* and *K* are 5σ and 0.5, respectively. CUSUM has been studied by many authors. Hawkins and Olwell (1998) and Zantek (2006) have widely discussed its properties and compared this chart with other charts in particular. They concluded that the CUSUM could be an effective alternative to the Shewhart control chart in phase II process monitoring to detect small shifts.

2.2.3 EWMA Control Charts

The Exponentially Weighted Moving Average (EWMA), introduced by Roberts (1959), is a weighted average of all past and current observations, and reveals shifts in a previously stable process. The statistic of this control chart is expressed by

$$Z_i = \lambda X_i + (1 - \lambda) Z_{i-1} \qquad 0 < \lambda \leq 1 \tag{2.7}$$

where X_i is the sequence of observations, λ is the smoothing constant factor, and Z_i is a predicted value of process average at time i. The starting value for $i = 1$ is the process target and defined as the center line, so that $Z_0 = \mu_0$. The control limits for the control scheme are

$$\mu_0 \pm L\sigma \sqrt{\frac{\lambda}{(2 - \lambda)}} \tag{2.8}$$

Lucas and Saccucci (1990) pointed out that EWMA and CUSUM are excellent alternatives to the Shewhart control chart for phase II process monitoring situations. EWMA differs from CUSUM by applying the additional weighting factor, which causes the sensitivity of shift identification. They indicated that EWMA control schemes have average run length properties similar to those for CUSUM control schemes.

The probabilities of type I and type II errors are frequently used to express the statistical properties of control charts. In other words, the statistical effectiveness of a control scheme is determined by the speed with which it detects the occurrence of assignable causes and by the frequency of false alarms. The average run length (ARL) has been employed as a performance indicator to evaluate the effectiveness of various control schemes. Borror et al. (1999) indicated that EWMA charts are more robust (insensitive) to the normality assumption than other control charts. Under the normality assumption with $L = 3$, the in control average run length (ARL_0) should be located at 370.4 samples on the average. The value of the chart parameters L, width of the control limit, and λ can be chosen such that the desired ARL_0 is maintained at 370.4. EWMA differs from CUSUM by applying the additional weighted factor, which increases the sensitivity of the shift detection. They indicated that EWMA control schemes have average run length properties similar to those for CUSUM control schemes. They proved that the EWMA can be viewed as a weighted average insensitive to the normality assumption. Therefore, it is an ideal control chart to use with individual observations. It is not appropriate to look for non random patterns in the EWMA chart because the EWMA statistic is auto correlated. Crowder (1989) and Hawkins et al. (2007) widely discussed the properties of the EWMA chart. They approved the EWMA can be viewed as a weighted average insensitive to the normality assumption. Therefore, it is an ideal control chart to use with individual observations

In the latter two methods, knowledge is accumulated over time. From the above citations, it can be seen that the EWMA scheme is a valuable tool and can compete with CUSUM control charts in detecting small shifts in the process mean.

Woodall (2000) indicated that EWMA and CUSUM charts are very efficient for certain cases and they have not replaced the Shewhart charts, which can be easily used in extensive applications for detecting special causes.

Colosimo and Castillo (2007) pointed out that all properties of the Shewhart and CUSUM control charts are well known in ideal circumstances where all process observations follow a common normal distribution with known in-control independent parameters. Moving away from the independence assumption justifies the use of an EWMA chart, the calibration of which still involves known process parameters and so is implicitly best suitable to processes with long histories. However, there are some scenarios in which process parameters are hardly ever known to a sufficient precision for these theoretical calculations to be credible. When the parameters are unknown, Bayesian methods in statistical process control are applied where only partial information of the process is available. This approach provides a good bridge between the extremes of total knowledge and total ignorance of process parameters (Bayes, 1783).

2.2.4 Economic Design of Control Charts

The economic design of control chart was first developed by Duncan (1956). The main concern of Duncan's model is to minimize total cost in the presence of only a single special cause. In this model, the practitioner estimates a number of process parameters and cost parameters, which leads to an optimal economic design. Duncan's model deals with applying control chart to maintain a stable level of the process. Duncan's work was the stimulus for a great amount of research. Lorenzen and Vance (1986) generalized Duncan's model and

presented a unified approach for determining an optimal parameter for univariate control charts. Most researchers developed economical models in this area to maintain the current level of the process performance. Limited efforts were made to improve the process average or variance. In practice, there are scenarios that require decreasing variability or moving the average closer to the target design. Weheba and Nickerson (2004) developed an economical model for improving the current process level. This model includes two cost functions. First, a reactive function provides the related costs incurred while the process is stabilized at a given level. Second, a proactive function considers costs while the process attains an improved level. By utilizing incremental economics, these two functions are assembled as a combined function in order to evaluate the process improvement. Furthermore, this model is sensitive to a number of process parameters that were left out in other models. The model is applicable to the Deming cycle for continuous improvement. Their model can be applied to reduce variability and change the process mean in phase II of the Shewhart control chart.

2.3 Sample Size Determination

Sample size determination plays a significant role in the design and analysis of any data sampling project. It is a key component of the planning of studies in many applied sciences. Directly or indirectly, sample size affects all steps of the study from design to analysis and results. There are a number of sampling techniques that can be utilized to quantify the impact of engineering efforts and verify their effect on performance of a process. These include both traditional as well as Bayesian techniques, and only some of them consider the economic aspect of sampling. According to Adcock (1997), SSD methods can be divided into two major groups: traditional and Bayesian.

2.3.1 Traditional and Bayesian Methods for SSD

In the traditional method, before sampling, two values are pre selected: the interval half width *H* and confidence level of the interval (1–α)100%. If a random sample of n observations, $x_1, x_2, \ldots, x_n$, are to be drawn from a normal distribution with finite mean μ and variance σ^2, and sample mean $\bar{x}$ that estimates the population mean, the requirement that the unknown population mean μ lies within $\pm$ *H* of the sample mean with probability (1– α), that is

$$\bar{x} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \tag{2.9}$$

Leading to

$$H = Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \tag{2.10}$$

The statistic $Z_{\alpha/2}$ is the upper (α/2)100% point of a standard normal distribution and obtained from the table (Mendenhall and Sincich, 2007). If σ is known for a pre specified *H*, then the sample size is expressed by

$$n \geq \left(\frac{Z_{\alpha/2}}{H}\right)^2 \sigma^2 \tag{2.11}$$

In this inequality, the standard deviation σ is assumed to be known, and any uncertainty implicit in σ^2 is ignored. Adcock (1988) criticized this method in three ways. First, since it is assumed that σ is known, the hidden uncertainty in σ^2 is ignored. Second, a random sample is assumed to be collected from a normal distribution, while the normality assumption is not appropriate everywhere. Third, it is assumed that the sample means are normally distributed, but in the case of sample mean, an appeal to the Central Limit Theorem (CLT) suggests that equation (2.9) should provide a satisfactory sample size.

For hypothesis testing, assume a null hypothesis is $H_0: \mu = \mu_0$ versus an alternative $H_1: \mu = \mu_1 \neq \mu_0$. If β is defined as a probability of type II error, $1 - \beta$ is a power of

the test, and $|\mu_1 - \mu_0| = H$ while H is specified, then Desu and Raghavarao (1990) showed that n^* can be calculated by

$$n^* = \left(\frac{Z_{\alpha/2} + Z_{\beta/2}}{H}\right)^2 \sigma^2 \tag{2.12}$$

Both H_0 and H_1 are specified for μ, and $n^*$is calculated by controlling the n^* and $1 - \beta$. In other words, the sample size and a rejection region for the H_0 are selected so that the following two inequalities are concurrently satisfied:

$$P(reject\ H_0 | \mu = \mu_0) \leq \alpha \tag{2.13}$$

$$P(reject\ H_0 | \mu \neq \mu_0) \geq 1 - \beta \tag{2.14}$$

Joseph et al. (1997) argued that calculating the sample size without considering the prior information and uncertainty may lead to a larger sample size than is necessary. In addition, the normality assumption may not be applicable in many practical scenarios. On the other hand, the Bayesian sample size determination (BSSD) approach accommodates with uncertainty of prior knowledge of the process, which is discussed in the next sections.

Further research for traditional methods can be found in two monographs: Kraemer and Thiemann (1987) and Desu and Raghavarap (1990). These two studies explicitly overview and present the traditional methods for solving a variety of sample size determination problems.

The SSD is essentially Bayesian in nature, according to Berger and Sellke (1987), since it absolutely requires prior information. As Adcock (1995) explained, the Bayesian approach is accompanied by uncertainty of prior knowledge and is an efficient tool for SSD, which is applicable to any univariate distribution. A Bayesian treatment of the SSD involves the introduction of either a credibility interval or a loss/utility function. Lindley (1997) divided the

BSSD approaches into two techniques. The first group of techniques involves applying the Highest Posterior Density (HPD) credible interval and its length, and statistical criteria, while the second group involves utilizing a loss/utility function, sample cost constraints, and economical criteria. He compared these two methods on practical aspects and concluded that the main advantage of the first group is its simplicity, comprehension, and applications. In addition, identification of desired coverage probability and an expected length (the average probability of the parameter being contained in a given length) by the user is easy for HPD criteria. In contrast, although the concept of loss/utility is applied in the second group, it is difficult to balance the three quantities, i.e., coverage, length, and sampling cost, within a single loss/utility function (Pham-Gia, 1997). In the next sections some literature about these methods is reviewed.

2.3.2 HPD Criteria

Adcock (1988) introduced a number of criteria based on HPD interval and its length. HPD regions result in the shortest intervals for any given coverage and provide a convenient summary of the posterior knowledge about the parameter of interest. That is, since the posterior contains the most critical information, these criteria are linked to either the posterior variance or posterior coverage probabilities for intervals with a given length. According to Joseph et al. (1995b), HPD has two equivalent properties. First, the density for every point inside this interval is greater than that for every point excluded. Second, the interval, for a given probability content, is as short as possible. Pham-Gia and Turkkan (2003) noted that the posterior variance of distribution of the parameter of interest is better than any other criterion for SSD to consider since it has the most valuable information. Hence, many of the criteria suggested by researchers are related to either the variance of the posterior distribution or the posterior coverage probabilities for intervals of a given length.

Let μ be the parameter of interest, $f(\mu)$ the prior distribution, $x = (x_1,\ldots,x_n)$ the sample with size n, $f(x)$ the marginal pre-posterior or predictive marginal distribution, and $f(\mu|x)$ the posterior distribution for μ given x. These distributions are denoted by

$$f(x) = f(x|\mu)f(\mu)d\mu \tag{2.15}$$

$$f(\mu|x) = f(x|\mu)f(\mu)/f(x) \tag{2.16}$$

As can be seen, equation (2.11) depends on H, α, and σ, but these three values are not sufficient for the calculation of equations (2.15) and (2.16); other values are needed. Specifically, the sample size and the values regarding the data to be collected are needed. Since the sample is not yet drawn, these values are unknown. For solving this issue using the predictive distribution in equation (2.15), three HPD criteria were developed. The main idea behind the HPD criteria is to find an optimal sample size satisfying $\int T(x)f(x)dx < \rho$, where $f(x)$ is the pre posterior marginal distribution of x, $T(x)$ is a statistic based on an individual sample x, and ρ is a fixed probability or a fixed length of the interval. This idea is the stimulus for developing the following three HPD criteria by Joseph et al. (1995a): average coverage criterion (ACC), average length criterion (ALC), and worst outcome criterion (WOC).

Assume that an individual observation x, a fixed HPD interval of length l, a parameter of interest μ, and a calculation of minimum sample size n are defined. ACC guarantees that the expected coverage probability would be greater than or equal to $1-\alpha$. In other words, ACC is based on the idea that the average probability of μ is contained in a fixed posterior interval, with length l, with a fixed probability that holds, on average, over all possible samples, i.e.,

$$\int \left\{ \int_a^{a+l} f(\mu|x,n)d\mu \right\} f(x)dx \geq 1-\alpha \tag{2.17}$$

Adcock (1988) first applied this concept to estimate normal means, with a symmetric interval of $(a, a+l)$ as a tolerance region for the mean. A more general case, for non symmetric posterior distribution, is discussed by Joseph et al. (1995a). They considered $a,$ such that $(a, a+l)$ to be the HPD interval. Hence, an interval of minimum length is obtained.

In the ALC the roles of coverage probability and length of interval are reversed. The coverage probability of the HPD interval, or average posterior credible interval, is fixed at $1-\alpha,$ whereas in ACC the length was fixed. Let *x* be a sample of size *n* and a length of $l'(x,n)$. The sample size is selected such that the expectation of these lengths is less than or equal to *l*. This is mathematically expressed by

$$\int l'(x,n) f(x) dx \leq l \tag{2.18}$$

where $l'(x,n)$ is the length of $(1-\alpha)100\%$ posterior credible interval for data *x*. The ALC ensures that the average length of $(1-\alpha)100\%$ posterior credible intervals weighted by $f(x)$ is less than or equal to *l*.

In WOC, the most conservative criterion, it is assumed that the worst possible measurements are taken with the largest variability. This criterion ensures that the minimum coverage probability of a fixed length HPD interval is at least equal to a confidence level of $(1-\alpha)100\%$. In other words, both the length and coverage probability of the HPD interval hold concurrently.

Joseph and Belisle (1997) applied the above criterion to determine sample size for normal means and differences between normal means with known variances. They concluded that because the posterior variance depends on *n* and does not change with the different measurements, ACC, ALC, and WOC resulted in the same sample size *n*.

BSSD has received considerable attention in the literature (Goldstein, 1981; Gould, 1993; Adcock, 1995; Joseph et al., 1997; and Lee and Zelen, 2000). Recalling that the HPD criterion do not include any loss/utility aspects, the next section deals with the Bayesian methods of sample size determination utilizing loss/utility functions to minimize the cost of sampling.

2.3.3 Fully Bayesian Methods for SSD

The whole idea of fully BSSD is to minimize overall cost, including decision loss and the cost of conducting the sampling. Pham-Gia and Turkkan (2005) indicated that fully BSSD pays attention to the cost corresponding to an optimal decision or to the average cost when the prior information is taken into account in determining the posterior risk. It is often the case that the lower the decision loss, the larger the sample size, and consequently the larger the cost of sampling. Thus, these two quantities are in opposition to each other. Hence, there should be a trade-off between sample size and loss cost.

The early work about BSSD was introduced by Raiffa and Schlaifer (1961), which dealt with the sequence (n, x, d, θ) denoting, respectively, sample size, collected data, terminal decision, and value of the unknown parameter. They identified a utility function of U (n, x, d, θ) to quantify the sample size n. They also assumed that θ has a density p (θ) known before sampling. Their solution procedure is followed in reverse order (θ, d, x, n), taking the expected value of utility over θ and x, and maximizing over θ and n. Let x_1, x_2, …, x_n represent n measurements of a random variable with density of $p(x|\theta)$ and $\theta \sim p(\theta)$. After observing x, it is desired to make a decision d about θ. The optimal sample size n^* is calculated by maximizing the expected utility function (MEU) as

$$n^* = \max_n \int [\max_d \{U(n, x, d, \theta) p(\theta|d, x, n) d\theta\} p(x|n)] dx \tag{2.19}$$

where $p(x|n)$ is the pre posterior distribution of x before sampling, while no information is available about θ, and $p(\theta|x,n)$ is the posterior distribution after sampling.

Following Raiffa and Schlaifer's work, Lindley (1997) considered the utility function and included a sampling cost of C for each observation as

$$U(n,x,d,\theta) = U(d,\theta) - Cn \tag{2.20}$$

As can be seen, the right side of equation (2.20) does not depend on *x*. Lindley did not consider the setup cost in his utility function. Berger (1985) expressed the loss function as

$$L(n,x,d,\theta) = -U(d,\theta) - Cn \tag{2.21}$$

to determine an upper bound for a sequential sampling approach. He defined two fully BSSD methods; namely, optimal fixed sample size and sequential sampling. The Berger's methods for BSSD are discussed in the next section.

2.4 Sequential Analysis Methods

In order to define the sequential sampling clearly, it is necessary to define the random sampling first. The *n* elements are said to be a random sample if they are chosen from a population in such a way that every set of *n* elements in the population has an equal probability of being selected. Unlike in sequential sampling, the elements are selected one by one. After each sample, a decision is made to accept or reject the null hypothesis and stop sampling. Otherwise sampling will be continued. In this sampling technique, the decision may occur as early as the first sample or any subsequence sample. It is also possible that the decision would require a large sample. Thus, it is notable that in the sequential sampling, the sample size *n* is a random variable (Mendenhall and Sincich, 2007).

The objective of sequential sampling as developed by Wald (1947) is "avoidance of oversampling." Sequential analysis is concerned with methods of analyzing the samples in which

the final number of the sample size is not predetermined and depends on the observations as they are obtained. According to Wetherill (1961), this method includes two elements: first, a stopping rule, which determines whether the sampling should be stopped or continued, and second, a decision rule, which states what action to be taken after sampling has stopped. The goal of sequential analysis is to determine the optimum of the above two elements that meet specific criteria in a given problem. The two most common methods of fully BSSD are optimal fixed sample size and sequential analysis. If a stopping rule implies that the sample size would not exceed a specified number, then the corresponding procedure is called optimal fixed sample size. Otherwise, the procedure is sequential.

For more declaration of the definition of the two methods, assume that the observations x_i ,i=1,2,… represent independent sequential observations from the density $p\ (x \mid \theta)$ with the unknown parameter θ, with which some decision is to be made. The sampling cost depends on the number of observations taken and the method by which the observations are taken (fixed or sequential). Berger (1985) pointed out that in the optimal fixed sample size technique, one selects a sample size n in advance, makes the observations, and ultimately, based on amount of loss, makes a decision. The distinguishing characteristic of the sequential analysis method is that the observations are taken one at a time, and a decision is made after each observation either to stop sampling and choose an action or to take another observation. In other words, after each sample, a new sequential problem is created. The advantage of sequential analysis is that one can collect the exact amount of observations for making a decision with desired accuracy. However, this method involves a considerable amount of notation and theory to obtain a reasonable answer. The current Bayes risk should be compared to the expected Bayes risk that is calculated

if more observations are taken. This comparison implies that the decision maker knows what needs to be done (Lai, 2001).

The bulk of the literature is concerned with sequential sampling. In a comprehensive review, Ghosh (1991) traces the applications of the sequential analysis concept to the control charts, sequential sampling inspection, and two-stage designs. Introduced by Shewhart in 1924, the control chart, a powerful tool for statistical process control, plots the averages of observations of a quality characteristic in a sequential order. Toward the end of the 1920s, Dodge and Romig (1929) developed acceptance sampling plans as an alternative for 100% inspection, which is concerned with incoming or receiving inspection. A sample is drawn from the lot, some quality characteristic of the units in the sample is inspected, and a decision is made to accept or reject the lot. It happens in sequential sampling that the measurements would be taken not all at one time, but in two or more stages, which was introduced by Thompson (1933). The design and application of a number of extensions of acceptance sampling plans, including single, double, multiple, and sequential sampling plans, are discussed in the literature. As mentioned previously, choosing a loss/utility function plays a crucial role in fully BSSD methods.

2.5 Loss and Utility Functions

Loss and utility functions are key elements of Bayesian decision analysis in every quality engineering method. As Roshan (2004) defined, loss functions are mathematical functions that describe the relationship between the estimation error and the economic loss associated with that error. The components of loss include not only waste and expense, but also costs during production such as inspection, repair, scrap, and rework. Loss and utility functions play a key role in evaluating a decision for Bayesian methods. Loss functions depict the economic impact of

choosing a specific action. If action "*a*" is performed and θ results to be the true state of nature, then a loss $L(\theta, a)$ is incurred.

Berger (1985) discussed three loss functions, the first being the squared-error loss expressed as $(\theta - a)^2$. The popularity of this loss function is due to the simplicity of its application. Second, the absolute error loss, denoted by $|\theta - a|$, is applied when the loss function is linear. Third, "0-1" loss is used in two-action decision problems; for instance, this loss function can be used to evaluate a test of hypothesis. The loss is zero for a correct decision and is 1 for a wrong decision.

Taguchi (1986) quantified the quality concept using Gauss's quadratic loss function. According to him "quality is the loss imparted to society from the time a product is shipped." He argued that the loss due to a product's variation in performance is proportional to the squared deviation of a performance characteristic from its target. He concluded how the variability is the key concept that relates quality to money. This idea led him to express the variability through developing a loss function. Taguchi expressed the loss functions based on three possible situations: target is the best, smaller is better, and larger is better.

2.5.1 Target is the Best

This situation considers quality characteristics for which a target value is more appropriate. In this case, Taguchi's loss function is expressed by

$$L(x) = k(x - m)^2 \tag{2.22}$$

where k is a positive constant and is influenced by the financial importance of the quality characteristic, x is the value of the quality characteristic, and m is the target value or nominal. If the quality characteristic x is at the target value m, then the loss $L(x)$ is zero. As x deviates from m, the loss increases dramatically in a quadratic manner. Some examples for x in this category

include the dimension of a part as thickness, diameter, length; viscosity of a specific type of oil; and so on. Figure 2.2 displays the Taguchi's view of loss function.

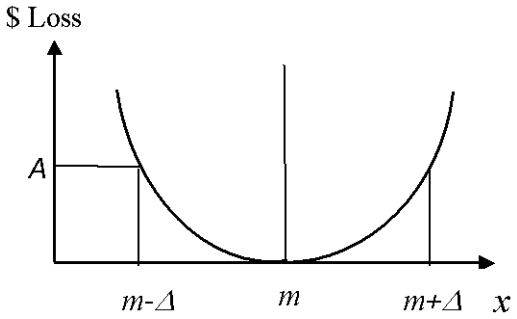


Figure 2.2 Taguchi's loss function (Target the Best)

Taguchi's loss function can take into account both the loss due to deviation from the target and loss to society as repair cost, scrapping, or replacement of the product. The society includes all elements that are directly or indirectly in contact with the product as customer, environment, and manufacturer. To determine the value of *k*, suppose the specification limits of the engineering design for the product are ($m \pm \Delta$). This tolerance range, Δ, corresponds to the maximum allowable variation beyond which the average performance is not adequate. That is, if the quality characteristic *x* is at Δ units from the target design (*m*), then the manufacturer should spend an amount equal to *A*, by average, to repair a product before shipping it to the customer. Otherwise, the manufacturer has to tolerate the loss due to the product not meeting customer requirements. Indeed, this tolerance is the maximum limit for shipping the product.

If *A* stands for the cost of repair or replacement per unit, then by utilizing equation (2.22), *k* is calculated as $k = A/\Delta^2$ (Mitra, 1998). The loss function can be rewritten as

$$L(x) = (A/\Delta^2)(x - m)^2 \qquad (2.23)$$

Equation (2.23) is normalized and can be applied for a variety of quality control problems on the same scale. By taking the expected value of the loss function, the uncertainty that is hidden in the

distribution could be recovered. The expected loss is the average over many instances of the measurements. (Maghsoodloo, 1992) showed that the expectation of loss with respect to distribution of *x* is given by

$$E[L(x)] = k[\sigma^2 + (\mu - m)^2] \tag{2.24}$$

where μ and σ^2 represent the mean and variance of *x*, respectively. From an intuitive viewpoint, it is important to note that equation (2.24) can be utilized to represent the uncertainties associated with the estimation of both $\sigma^2$and μ. These variables are treated as random variables in application of the Bayesian theorem.

2.5.2 Smaller is Better

Some quality characteristics are always non negative, and the desired value of them is zero. For instance, fuel consumption of an aircraft, the wear of a bearing, or shrinkage of a gasket could be classified in this category. Figure 2.3 (a) illustrates such a case. The loss function is expressed by

$$L(x) = (A/\Delta^2)x^2 \tag{2.25}$$

The expected, or average, loss over many instances of the product is given by

$$E[L(x)] = k[\sigma^2 + \mu^2] \tag{2.26}$$

2.5.3 Larger is Better

In some scenarios, the target of a non-negative quality characteristic is infinite or as large as possible. One example of this case can be strength in bridge beams. Here, no specific target is identified in advance. Figure 2.3 (b) illustrates this case. The loss function of this case is expressed by

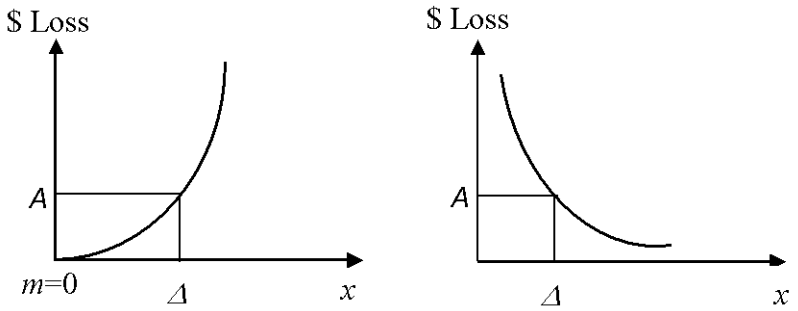
$$L(x) = A\,\Delta^2\,(\frac{1}{x^2}) \tag{2.27}$$

The expected loss is also expressed by

$$E\,[\,L(x)\,] = A\,\Delta^2 E\,(\frac{1}{x^2}) \tag{2.28}$$

where

$$E\,(\frac{1}{x^2}) \approx \frac{\sum_{i=1}^{n}(\frac{1}{x_i^2})}{n} \tag{2.29}$$



(a) Smaller is Better (b) Larger is Better

Figure 2.3 Taguchi's loss function

Spiring (1993) modeled loss with a finite maximum by introducing an inverted normal density function. A review of further work is available from Sun et al. (1996), Spiring and Yeung (1998), Wu and Shu, (2007), and Cudney and Drain (2007).

CHAPTER 3

DISCUSSION

The main goal of this chapter is to distinguish between traditional and Bayesian frameworks in applications and limitations. To achieve this goal, the first section discusses the attributes and applications of two methods. Moreover, this section introduces the different kinds of conjugate priors that have a particular rule in calculation of posterior parameters. The next section compares and contrasts these two frameworks. At the end of the chapter, limitations of each method are presented. The question of why Bayesian is more appropriate for parameter estimation in a transition period is also answered.

3.1 Traditional and Bayesian Methods

The typical questions regarding traditional and Bayesian methods involve the applications, strengths, and weaknesses of each approach. The strength of traditional statistical methods is inherent in simplifying the computation, which stems from having been established before the advent of modern technology. Jaynes (2003) discussed three important conditions in which the traditional approach works well. First, the variables of a study follow a simple distribution that is determined by a small number of parameters. Second, prior knowledge about the variables of the study is not available. Third, a large number of observations are available. In some contexts where traditional statistical methods are not very effective, Bayesian statistics are more applicable. For example, when there is a small data set, traditional estimators can be inappropriate since the likelihood may not be well-approximated by a normal distribution. The Bayesian inference rather than traditional methods can be performed much better in this situation. In the last decade, a “Short Run SPC” problem has received considerable attention in the literature. This problem occurs more often in small production runs. Colosimo and Castillo

(2007) pointed out the difficulties that may be encountered for parameter estimation using the small data set. One of the appropriate approaches for this problem is applying a Bayesian perspective. This obviously makes sense when a small data set is available and prior beliefs can be incorporated to help obtain posterior distribution from which inferences could be made.

On the other hand, when the sample size is large enough (n > 30) the results of traditional inferences are always considerable (Peck et al., 2005). Another instance is the incorporation of prior knowledge and the use of a data set from multiple studies. Industries usually have sufficient prior information available, which is useful for accurate estimation of parameters. Applying the Bayesian approach is a way to incorporate this information in estimation (Marin and Robert, 2007). Regardless, industries are to provide information about the uncertain quantities. The language of uncertainty is probability, and only through the Bayesian approach could the uncertainty be addressed (Berger, 2000).

To carry out the Bayesian approach, three steps are necessary. First, before sampling, the prior distribution $f(\theta)$ should be specified that expresses the prior beliefs and current state of knowledge regarding the parameter. This method allows for a subjective prior distribution. If the prior is carefully selected, then the results of Bayesian methods are very effective. Second, a method of experimental measurement should be applied to describe the distribution of the data given the parameter of interest, i.e., $f(x|\theta)$. The likelihood function $l(\theta)$ can be thought of as any function proportional to $f(x|\theta)$, that is, $L(\mu) \propto p(x|\theta)$. It not only ensures that the data are collected randomly but that the data can only affect the posterior through this function. It can be treated as representing the information about θ coming from data. The third step is based on the combination of information from prior and likelihood to update the knowledge about θ by

calculating the posterior distribution $f(\theta|x)$ after sampling. This is found through the Bayes' theorem which is expressed by

$$Posterior\ Distribution\ =\ \frac{Likelihood * Prior\ Distribution}{\int Likelihood * Prior} \tag{3.1}$$

Thus, Bayes' theorem describes the process of learning from experience and shows how knowledge about the parameter represented by prior information is continually modified as new data becomes available.

The effectiveness of the Bayesian and traditional methods depends upon the types of problems considered. It can be concluded from an extensive literature review (for instance, see Lee, 2004; Bolstad, 2004; and Gill, 2008), that the Bayesian approach is more applicable than the traditional approach in the following situations: first, when it is desirable to accumulate numerical data of the parameter of interest; second, when decision-making is desired in terms of loss associated with every scenario; and third, when informative prior knowledge is available. However, Bayesian analysis is not recommended when there is not a good understanding of the procedure, since the objectivity of the prior information is crucial in determining the sample size and decision making.

Lindley (1983) summarized the kind of problems that fit better within the Bayesian framework. He described the strength of the Bayesian methods as follows: Assume the estimation of an unknown parameter of interest is desired. Some data about the value of this quantity and some background information relevant to the underlying problem are available. The merit of the Bayesian approach is that it provides an explicit framework to combine these three elements and to obtain a revised value of the unknown parameter. According to Albert (2009), by offering an uncertainty and explicitly allowing the incorporation of external or prior information, the Bayesian approach leads to more effective procedures in many instances. In

other words, as more data is obtained, the posterior inferences are updated, so the Bayesian method is more flexible compared to other methods.

3.1.1 Choice of Prior Distribution

Colosimo and Castillo (2007) introduced three common choices of prior distribution. In the class of conjugate priors, the prior and posterior distributions have the same form of distribution with different parameters. A conjugate prior ensures that the posterior has a closed form in the sense that it simplifies calculations, especially in sequential applications of Bayes' theorem. Robert (2001) tabulated the posterior distribution of each kind of conjugate prior distribution. For instance, the conjugate prior for the mean of a normal distribution is the normal distribution, and the corresponding conjugate prior for the variance of a normal distribution is the scaled inverse-χ^2 distribution. The class of non-conjugate priors results in a posterior with a different form than the prior distribution. The non-informative priors, so called objective priors, include a lack of information about the parameter.

3.2 Other Differences Between Traditional and Bayesian Methods

There are three more differences between traditional and Bayesian methods (Jeffreys, 1961). First, in traditional methods, a number of techniques are used for developing statistics: Maximum likelihood, Unbiasedness, Minimum variance, and least squares principles. In Bayesian methods, the three most important techniques are the Bayes risk, Minimax Principle, and Invariance Principle. Second, traditional methods do not take into account prior and loss information. The loss function used in Bayesian analysis would effectively help the researcher in the decision making process. The greatest need for applying Bayesian analysis is in decision problems for which prior and loss information are a crucial part of the problem. Third, the Bayesian methods use a *P*-value instead of the probability of type I and type II error for

evaluation of a test procedure. For example, if $H_0: \theta \leq \theta_0$ and $H_1: \theta > \theta_0$ and $X \sim N(\theta, \sigma^2)$, then $\alpha_0 = P(\theta \leq \theta_0 | x) = \Phi((x - \theta_0)/\sigma)$ and $P-value = P(X \geq x) = 1 - \Phi((x - \theta_0)/\sigma)$, where the *P*-value is the smallest level of significance that would lead to rejection of null hypothesis, i.e., the chance of rejecting the null when it is true, and H_0 is rejected when the *P*-value is less than the probability of type I error (Tan et al., 2010).

Berry (1997) indicated some important differences between Bayesian and traditional methods. The Bayesian approach is subjective; the basis for all Bayesian inferences is the current distribution of the various unknowns. This is usually called a prior distribution when it applies at the start of a sample, and a posterior distribution when it is conditional on the experimental results. Each individual has the personal probability distribution of unknowns. Whether this distribution is prior or posterior, it is subjective. Two persons' posterior distributions are usually closer than their prior distributions (Berry, 2006). But there is no guarantee that people's opinions are in agreement, even in the face of what some consider compelling evidence. In this regard, more generally the Bayesian approach mirrors science. The subjectivity requires assessing prior probability. This requirement forces users to relate current information to other available information. In addition, there is no agreement on definition of a single, all purpose prior distribution to be adapted to the type of data in a sample, which is drawn from the population of study. Adcock (1988) argued that due to existence of so many users of the observations of the sample, it is very difficult to present a generally applicable loss/utility function, i.e., a kind of general purpose of a loss/utility function by which a problem could be justified for all situations. He mentioned that this property distinguishes Bayesian from traditional methods because Bayesians evaluate information that is available separate from the sample of interest. This information is rarely identical in two different settings. The Bayesian

methods consider all available evidence for making an inference, while the traditional methods use only the information of the last sample. According to Tsiamyrtizis and Hawkins (2005), one of the essential properties of the Bayesian approach is to allow for better use of partial information, making a good bridge between the extremes of total knowledge and total ignorance of process parameters. The important conclusions that they derived from this property is that the Bayesian framework converges rapidly to the true process average, regardless of the original different beliefs.

3.3 Limitations of Bayesian Methods

This section is concerned with the limitations of the Bayesian approach for applications. Press (2003) extensively discussed subjective and objective Bayesian statistics and criticized the complexity of Bayesian calculations and the hidden arbitrariness of the prior distribution. One of the main reasons of criticism is that a subjective concept of prior probability exists in the Bayesian approach rather than objectivity (which means a fact has been discussed for many years). The question of how to select a prior distribution is one of the most common criticisms in the Bayesian approach because in practice any prior distribution should be justified. He argued that even though a subjective prior distribution has the worth of additional replications of the data, it is not always easy to translate the prior information to a meaningful distribution, and the parameter of prior distribution must be assessed. In addition, the subjective prior may not be applicable in a general situation because it belongs to a particular practitioner. Marcellus (2008a) summarized a number of limitations of Bayesian framework. A practical difficulty is that more information is necessary about the process than that of the traditional methods, even though acquiring this knowledge can yield real benefits. Another difficulty is that the probabilities used in Bayes' theorem must sometimes be assessed. This is an activity that requires special training

and cleverness. Furthermore, according to Marcellus (2008b), Bayesian monitoring is not robust to the structure of the process. The calculations are changed by changing the process distribution. Yet another difficulty is that valuable visual information might be lost if traditional methods were not used in conjunction with Bayes' theorem. Typically, the Bayesian approach is not recommended when there is no good understanding of the procedure, since the objectivity of prior information is crucial in SSD and decision making. (Gelman et al., 2004).

3.4 Limitations of Traditional Methods

This section is concerned with summarizing the limitations of traditional methods, which was discussed in previous chapters. As mentioned in Chapter 1, there are some scenarios in the end of phase I of traditional control charts where the process is found to be stable, but the performance is not at the desired level. In these scenarios, a number of changes must be made to improve the performance of the process. Typically, the effect of these changes should be evaluated in order to take appropriate actions. The duration of these activities, as shown in Figure 3.1, has been referred to as a transition period. At this point, an effective model is desired to calculate the loss accompanied with each action taken. In addition, this model should address an appropriate sampling plan for data collection to get to the result as soon as possible.

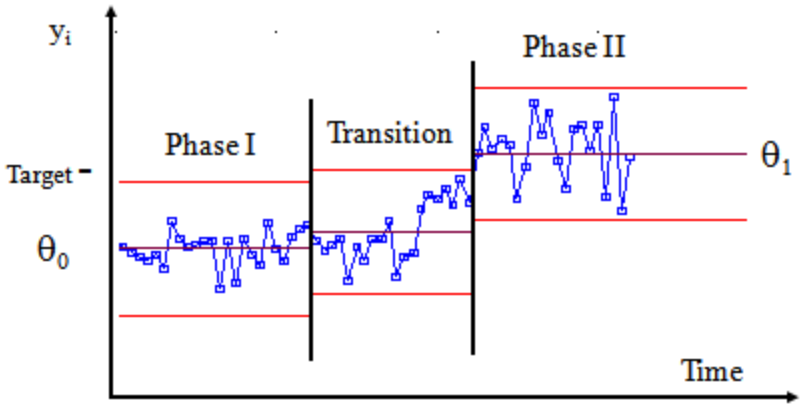


Figure 3.1 Phases of chart application and transition period

The traditional approach is not applicable in these scenarios for two reasons. First, traditional methods do not determine monetary loss or gain corresponding to each action. Second, it is usually discovered in sampling that data are not stationary or independent which violates the underlying assumptions of traditional control charts. On the other hand, the Bayesian approach is more effective in these situations. In Bayesian methods, all accumulated numerical data from phase I are used as informative prior knowledge to evaluate the effect of changes. Table 3.1 shows a tabulated comparison between traditional and Bayesian methods. One of the main purposes of this research is to use the advantages of traditional control charts to overcome the limitations of Bayesian methods and vice versa. In meeting this purpose, the modeling approach used in this research establishes values for prior parameters based on the information obtained during phase I to overcome the subjectivity of the prior. On the other hand, the modeling approach utilizes the Bayes' theorem to calculate and update the posterior parameters during the transition period to overcome the limitation of traditional method for calculating the posterior information.

TABLE 3.1

A COMPARISON BETWEEN TRADITIONAL AND BAYESIAN METHODS

Traditional Methods	**Bayesian Methods**
Use only information of the last sample and ignore any information given by the entire sequence of points.	Use all available evidence for making an inference.
Are applicable only when the in control process data is stationary and uncorrelated.	Are also applicable when data are non stationary or auto correlated in other situations.
Are not applicable when the assumptions (independency, normality, and constant parameter) are violated.	Are free of these assumptions.
Do not provide an updated estimate for the process parameter during the transition period.	Are able to update process parameters during the transition period utilizing prior information obtained from control chart.
Are not applicable to calculate loss in SSD.	Calculate loss in SSD.
Are not applicable for using partial information.	Are applied for optimal use of partial information on the process.
Are easy to calculate.	Are complicated to calculate.
Need less information of the process.	Need more information of the process.
Use graphical tools to show results.	Do not show results graphically.
Are robust to the structure of the process.	Are not robust to the structure of the process.
Do not use the prior information.	Are subjective and require assessing prior.

CHAPTER 4

RESEARCH PROCEDURE

The cost of maintaining the process at the level at which it is stabilized may result in poor quality. Responding to the needs of customers, the research community continuously attempts to provide more effective tools. Bayesian methods have recently received significant attention due to their ability to increase quality and reduce the total cost in scenarios where traditional methods are not applicable. This chapter includes the problem statement and declaration of literature gaps regarding the problem. In addition, the research objective is presented.

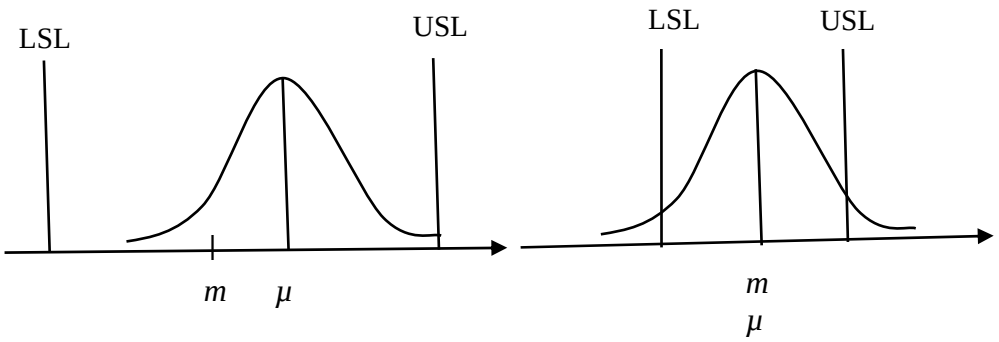
4.1 Research Gap

Establishing stability during phase I of the control chart application is one of the most important aspects of statistical process control. However, the argument has been made that changes must be introduced to the process in order to improve its performance. These changes may target a shift in the process average, a reduction of the total variation, or both. Figure 4.1 depicts the three possible scenarios commonly encountered upon the conclusion of phase I.

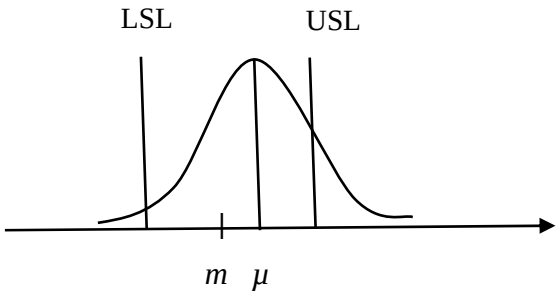
In the first case, the process is found to be off center with low variability, and the process average should be changed to the designed level. The second case involves the scenario in which the process is centered but with high variability. In this case, some changes are to be made to reduce the process variance. In the third case, the process is off center with high variability. Both average and variance are to be moved to the designed level. As such, efforts should be made to move the process average closer to the target and reduce the process variance to the designed level.

In all cases, changes must be made and tested to verify that the desired level of performance has been attained. The duration of these activities has been referred to as a

transition period. Typically, Shewhart's model has not held true in representing the measurements collected. These are more likely to be non-stationary and auto-correlated. The Bayesian approach is more suited to these scenarios.



(a) Off Center (b) Centered with High Variability



(c) Off Center with High Variability

Figure 4.1 Possible scenarios for process performance in conclusion of phase I

The gain associated with efforts to shift the process mean closer to the target or reduce its variability requires the determination of an optimal sampling plan. Traditional SPC literature provides strong evidence supporting the ability of control charts to detect a shift in the process level during phase II. However, none of the available techniques allow for an effective utilization of the prior knowledge gained during phase I. The case where the process is found to be in control at the desired level would be the only exception. In all other cases, this practice may result in increasing the cost of improvement. That is, some changes need to be made on the process to move the average or reduce the variability or both. However, the effect of these

changes should be tested using a sampling procedure. Practitioners need an effective sampling plan that can be used to evaluate process changes and quantify their effect. These concerns are the focus of this dissertation.

While the cited Bayesian methods draw considerable attention to optimization models for sample size determination, limited attempts have been made to incorporate a utility or loss function in determining the sample size. Moreover, there are no guidelines for addressing the consequent loss of making a wrong decision during the transition period.

All research efforts cited for the economic design of control chart were aimed at phase II applications. The same was concluded for the statistical as well as the economic-statistical design models. Indeed, prior knowledge is utilized under the assumption that the process parameters have been reliably estimated at the conclusion of phase I.

In this research, efforts are be made to develop a theoretical model for determining the sample size required to guide engineering efforts aimed at changing the process level.

4.2 Problem Statement

In the conclusion of phase I in Shewhart control charts, one of the three scenarios represented in Figure 4.1 may be discovered. Hence, it is necessary to provide updated estimates of the process level after attempting some changes to the process. These issues have a critical effect on the assessment of process performance during the transition period. Calculation of the maximum sample size based on the economic worth of improvement accompanied with the changes is the main concern of this research.

4.3 Research Objectives

The objective of this research is twofold. The first is to develop a mathematical model for determining the sample size required to verify changes in the process level. The second objective

is to demonstrate operation of the sampling procedure. Applications include a shift in the process average and reduction of process variability. After developing a mathematical model for each case, a sampling procedure with guidelines is developed to secure effective implementation of the model. Simulated data are used in the numerical examples for each case to verify the effect of changes on the process. The research procedure in order to accomplish the research objectives is as follows:

1. Determine the loss.
2. Identify the economic criterion.
3. Develop mathematical models for determining the max economic sample size and the sampling procedure to follow in verifying process improvement.
4. Provide numerical examples to demonstrate performance of the sampling procedures.

CHAPTER 5

MATHEMATICAL MODELING

This chapter presents the theoretical model for determining the maximum economic sample size required to estimate the actual process parameter. The model development includes two steps. In the first step, a loss matrix is constructed for a process parameter that is away from the target level in the end of phase I of the Shewhart control chart. This loss matrix corresponds to one of two actions. The Bayes function and the Bayes risk are formulated for each possible action, either making some changes in the process or leaving the process as it is. Regarding the loss function identified for each situation, the first condition as an economical criterion by which the practitioner could make a decision is introduced. If the practitioner makes a decision to attempt some change in order to improve the process, the formulation of the procedure is continued.

For verification that these changes resulted in moving the process parameter to the target level, the maximum economic sample size is determined by two methods, namely approximation and exact. In the exact method, an equation for calculating the maximum sample size n^* is derived. The second criterion concerning the technical feasibility of the process improvement is also introduced. In the second step, the sampling procedure is demonstrated and performance assessment is conducted to implement the model. Furthermore, the theoretical model is applied in three cases, and the simulated data is used to verify the model. The research procedure is shown in Figure 5.1.

The organization of this chapter is as follows. Notations are detailed first. Then the mathematical models accompanying sample size determination formulation and sampling procedure are developed for three cases: sudden process average, gradual shift for process

average, and reduction of process variance. Each case is applied with a numerical example to illustrate the steps involved and highlight the economic advantages of the research procedure.

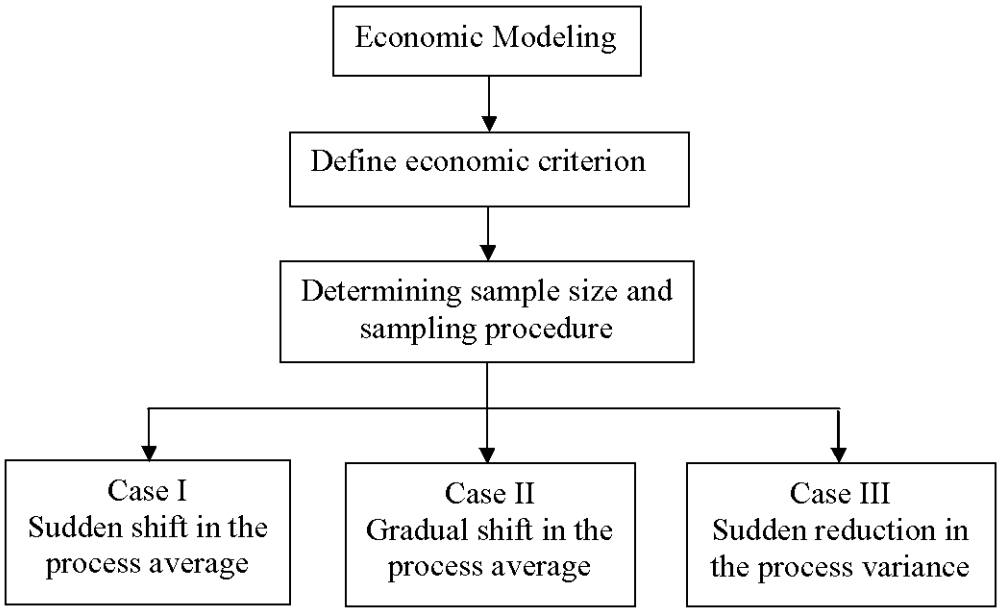


Figure 5.1 Research procedure

5.1 Notations

The mathematical model and sampling procedure are developed using the following notations:

x	Single quality characteristic for univariate manufacturing process
L	Loss function
m	Design target value for process parameter μ
n	Sample size
n^*	Maximum economic sample size
A	Cost of repairing or replacing one unit of product ($)
Δ	One half of tolerance spread
k	Proportionality constant in Taguchi's loss function, which is equal to A/Δ^2

σ_n^2	Posterior variance after collecting sample of *n* measurements
σ_g	Standard deviation of measurement error (gauge capability)
r	Bayes risk
C	Cost of each measurement ($ / unit))
CLT	Central limit theorem
W	Setup cost ($)
a	Bayes action
a_1	Taking no action to improve process
a_2	Action of improving process
$R(a_1)$	Loss function, associated with a_1 action
$R(a_2)$	Loss function, associated with a_2 action
$r(a_2)$	Bayes risk associated with a_2 action
p	Probability that unknown parameter will change
1-p	Probability that unknown parameter will not change
N	Number of future units produced
$\overline{X}$	Average of *n* measurements
S^2	Variance of *n* measurements
(μ_1, σ_1^2)	Process parameters after phase I
(μ_2, σ_2^2)	Desired target values for process parameters
(μ_n, ρ_n)	Parameters of posterior distribution after *n* measurements

5.2 Loss in Accordance with Possible Actions

The process considered is one of discrete manufacturing of a single quality characteristic *x*. The specification limits for *x* are expressed as $m \pm \Delta$. It is also assumed that a state of

statistical control has been established and the quality characteristic was shown to follow a normal distribution with mean $\mu_1$and variance σ_1^2 in which one or both parameters are off the desired performance level. Production is expected to continue over future periods to supply a total of *N* units.

Two options are considered for process management. The first option (a_1) involves taking no action and continuing operation at an estimated per unit loss of L_1, which according to equation (2.24) is given by

$$L_1 = k[\sigma_1^2 + (\mu_1 - m)^2] \tag{5.1}$$

where $k = A/\Delta^2$, and A represents the repair or replacement cost for one unit of product. The second option (a_2) involves making changes to the process to reduce the expected loss to $\$L_2$. As such, L_2 can be expressed as

$$L_2 = k[\sigma_2^2 + (\mu_2 - m)^2] \tag{5.2}$$

The second option will also require an initial investment of $\$W$ for tooling and setup. The probability that the changes will successfully result in improving the process level is estimated a priori as *p*. This information can be summarized in a loss matrix, as shown in Table 5.1.

TABLE 5.1

LOSS MATRIX

	Maintain Current Level (a_1)	**Improve to New Level (a_2)**
Probability of Success (*p*)	$N(L_1)$	$N(L_2 - L_1) + W$
Probability of Failure (1-*p*)	$N(L_1)$	W

The expected value of the loss, namely the loss function, for the two options can be calculated as

$$R(a_1) = p[N(L_1)] + (1-p)[N(L_1)] = N(L_1)$$

$$R(a_2) = p[N(L_2 - L_1) + W] + (1-p)W = pN(L_2 - L_1) + W \quad (5.3)$$

The criterion for economic improvement requires that the average expected value of the loss under the second option be less than that under the first. That is,

$$R(a_2) < R(a_1)$$

$$\text{i.e.,} \quad pN(L_2 - L_1) + W < N(L_1) \quad (5.4)$$

When this criterion is satisfied, a_2 would represent the optimal Bayes rule in which the expected loss is minimized.

Assuming that the economic criterion holds true, some changes have to be made to improve the process. The changes may be aimed at causing a shift in the process average closer to the design target or a reduction of the process variability. The mathematical derivation depends on the anticipated results and the way the process will react to the changes made. In Case I, the process average is assumed to experience a sudden shift in response to the changes. In Case II, the average is assumed to trend during the transition period to the new level. Case III involves changes that cause a sudden reduction in the process variance.

5.3 Case I: Sudden Shift in the Process Average

This case is concerned with verifying the effect of changes made to shift the process average. In this case, the process has been shown to be in a state of statistical control with mean μ_1 (off target, i.e., $\mu_1 \neq m$) and variance σ_1^2. Hence, some changes should be made to re-center the process closer to the design target while maintaining the same process variance σ_1^2. In other words, the process variance will be monitored during the transition period using the same traditional control chart from phase I. As a first step, the economic criterion of equation (5.4)

should be satisfied. Under this condition, a_2 represents the optimal Bayes rule in which the expected loss is minimized.

In addition, to verify the effect of the proposed changes a sample of size *n* measurements is required at the rate of $\$C$ per unit. As such, the average Bayes risk associated with a_2 can be calculated as

$$r(a_2) = E[R(a_2)] = E[\, pN(L_2 - L_1) + W + Cn]$$
$$= pNE[(L_2 - L_1)] + W + Cn \tag{5.5}$$

Now suppose that an unbiased gauge with variance $\sigma_g^{\,2}\,(\sigma_g^{\,2} < \sigma_1^2)$ is available to collect a total of *n* measurements and that the process average will experience a sudden shift in response to the changes made. The resulting sequence of measurements $x_1, x_2,\ldots, x_n$ can be used to update the prior belief regarding the process average. Utilizing the Bayesian theorem, the likelihood function of *x* given μ is expressed by

$$f(x|\mu) = \frac{1}{\sqrt{2\pi\sigma_g^2}} \exp -\frac{1}{2}\left(\frac{x-\mu}{\sigma_g}\right)^2 \tag{5.6}$$

As was shown by Box and Tiao (1973), the posterior density after *n* measurements is also a normal function in the form

$$f(\mu|x_1, x_2, \ldots, x_n) = \frac{1}{\sqrt{2\pi\rho_n}} \exp -\frac{1}{2}\left(\frac{\mu-\mu_n}{\rho_n}\right)^2 \tag{5.7}$$

where the posterior mean μ_n and posterior variance ρ_n are given by

$$\mu_n = \frac{n\sigma_1^2\bar{X} + \sigma_g^2\mu_1}{\sigma_g^2 + n\sigma_1^2}, and \quad \rho_n = \frac{\sigma_1^2\sigma_g^2}{\sigma_g^2 + n\sigma_1^2} \tag{5.8}$$

where $\overline{X}$ is the sample average, which is a sufficient statistic for calculating the posterior distribution and, based on the Central Limit Theorem (*CLT*), follows the normal distribution (Carlin and Louis, 2009). This is especially true under the assumption of sudden shift.

Substituting values of L_1 and L_2 into equation (5.5), the expected risk $r(a_2)$ can be expressed as

$$r(a_2) = pNkE[\sigma_1^2 + (\mu - m)^2) - \sigma_1^2 + (\mu_1 - m)^2)] + W + Cn$$

$$= pNk([E(\mu - m)^2 - E(\mu_1 - m)^2] + W + Cn \quad (5.9)$$

Since $E(\mu - m)^2 = \rho_n$, then

$$r(a_2) = pNk(\rho_n - (\mu_1 - m)^2) + W + Cn \quad (5.10)$$

Substituting the value of ρ_n from equation (5.8), the expected risk *r* (a_2) is given by

$$r(a_2) = pNk\left[\frac{\sigma_1^2\sigma_g^2}{\sigma_g^2 + n\sigma_1^2} - (\mu_1 - m)^2\right] + W + Cn \quad (5.11)$$

Equation (5.11) represents the average risk as a function of sample size *n*. To explore its characteristics, sample size *n* is treated as a continuous variable as an approximation. With this assumption, the first and second derivatives of *r* (a_2) with respect to *n* can be obtained as

$$\frac{dr(a_2)}{dn} = \frac{-pNk\sigma_1^4\sigma_g^2}{(\sigma_g^2 + n\sigma_1^2)^2} + C \quad (5.12)$$

$$\frac{d^2r(a_2)}{dn^2} = \frac{2pNk\sigma_1^6\sigma_g^2}{(\sigma_g^2 + n\sigma_1^2)^3} \quad (5.13)$$

An examination of equation (5.13) reveals that *r* (a_2) is convex in *n*, since all parameters are essentially positive. Consequently, setting equation (5.12) equal to zero and solving for *n* results in the maximum economic sample size *n**, expressed as

$$n^* = \sqrt{pNk}\,C^{-1/2}\sigma_g - \frac{\sigma_g^2}{\sigma_1^2} \quad (5.14)$$

It is worth noting here that equation (5.14) is in agreement with that given by Berger (1985), except for the term under the square root. This scaling factor includes important parameters that reflect the economic value of the quality characteristic *x*, and the level of

certainty in prior beliefs. Due to the nature of n^* and the ratio of the gauge variance to the process variance, the last term in equation (5.14) can be ignored in most practical scenarios. Now, substituting the ratio A/Δ^2 for the proportionality constant k, results in a more simplified expression in the form

$$n^* = \sqrt{pN\frac{A}{C}\frac{\sigma_g^2}{\Delta^2}} \tag{5.15}$$

As can be seen, equation (5.15) represents all relevant parameters required to calculate an approximate value of n^*. It incorporates the ratio of the product replacement cost (A) to the measurement cost (C). Large values of this ratio would allow larger samples to be taken economically. The same can be said regarding the effect of the probability of success p. Larger values of p allow for larger values of n^* from an economic viewpoint, whereas the term represented by σ_g^2/Δ^2 tends to have an inverse effect on the sample size. This ratio represents a measure of the gauge capability and should be known before collecting measurements. Some industrial standards, e.g., those by the Automotive Industry Action Group (2002) require that the ratio of $(6\sigma_g / 2\Delta)$ be 10% or less for the gauge to be capable. The number of remaining units N is of special importance. The larger the values of N, the larger the corresponding values of n^*, and vice versa. This would translate into a second criterion regarding the technical feasibility of improvement. That is, the value of N should be large enough to allow for n^* to be produced. Situations where this criterion is not satisfied require re-examination of the changes proposed, and/or the cost of measurement. Otherwise, improvement efforts should proceed by implementing the changes and collecting a set of n^* measurements. The average of these measurements $\overline{X}$ would be used to obtain the posterior parameters μ_n and ρ_n using equation (5.8).

To verify the shift in the process average, an HPD credible interval in the form $\mu_2 \pm z_{\alpha/2}\sqrt{\rho_n}$ can be constructed at a given level of α. If such an interval encloses the posterior average μ_n, then it can be concluded that the changes have resulted in the desired shift. Otherwise, improvement efforts have to continue following the same procedure while testing different changes. Figure 5.2 outlines the sampling procedure to be performed in Case I.

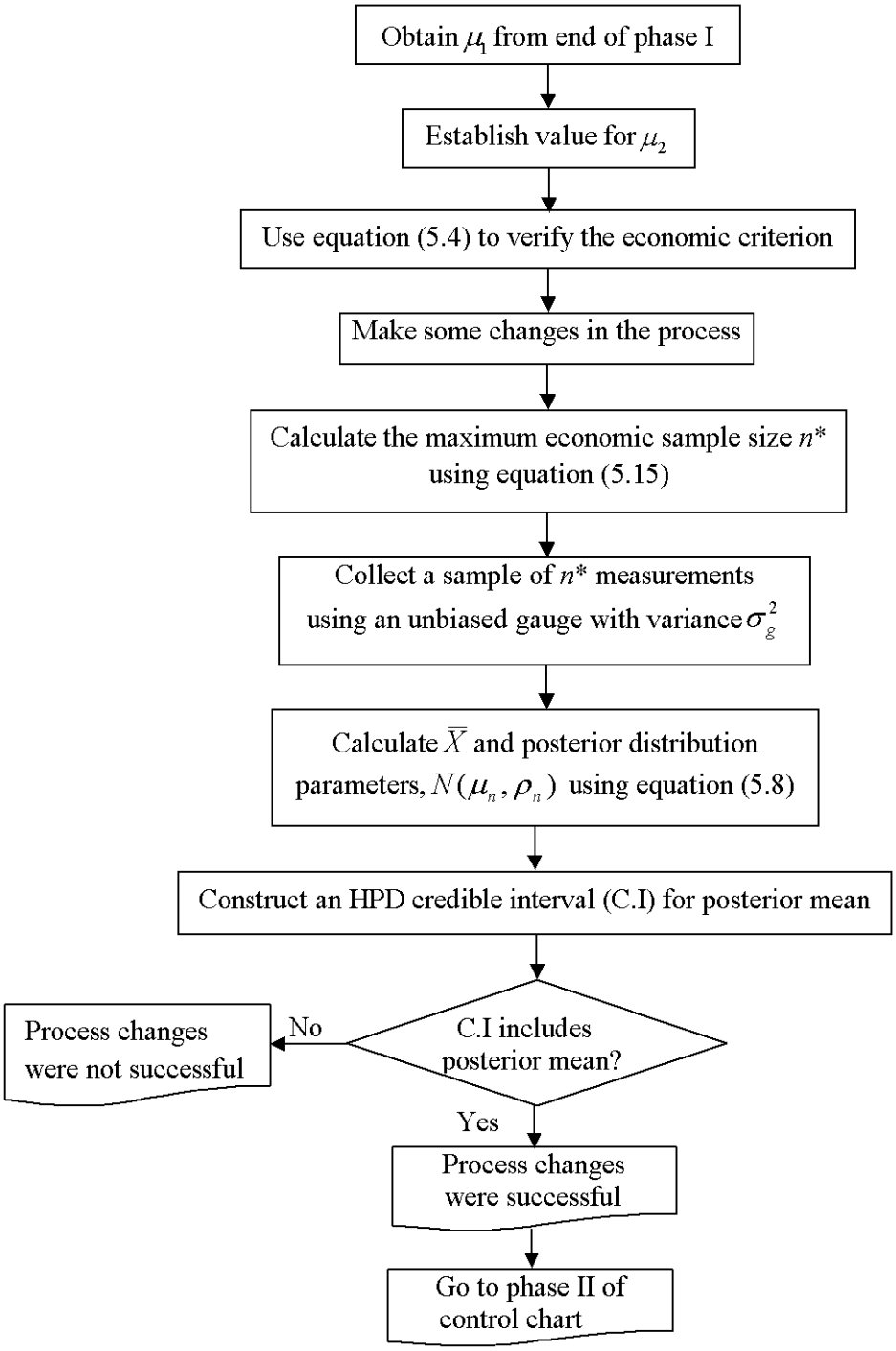


Figure 5.2 Sampling procedure for Case I

5.3.1 Numerical Example for Case I

Control charts for the mean and variance indicate that the current process for producing a component with critical dimension *x* is in a state of statistical control. The design specifications

for this dimension are given as 8.10 ± 0.05. The control charts show that the process average μ_1 = 8.06 and the variance $\sigma_1^2 = (0.01)^2$. Production is to be continued to produce a lot of N = 2,000 items. Pertaining cost parameters are estimated as follows: cost of scrapping or reworking one unit of product A = \$100, setup cost W = \$1,000, and cost of measuring one unit is C = \$10.

Two options for the process were considered. The first is to continue production without any change. Utilizing equation (5.1), the estimated loss per unit associated with choosing the first option is expected to be L_1 = \$68. In the second option, a number of changes are to be made to cause a shift in the process average closer to the target while maintaining the same process variance σ_1^2. It is expected that these changes would cause an abrupt shift from 8.06 to 8.10. Some measurements are to be taken to investigate the effect of the changes. Let an unbiased gauge with variance of $\sigma_g^2 = (0.002)^2$ be available to take measurements. The probability of success has been estimated to be p = 0.7. Figure 5.3 shows the various values for these parameters with the Bayes risk plot.

Utilizing equations (5.2), the estimated per unit loss by making changes L_2 would be \$4. Substituting the values of L_1 and L_2 into equation (5.4) gives the expected value of the loss for the two options as -88,600 < 136,000, which satisfies the economic criterion. Therefore, the second option of making changes to the process is selected as a Bayes rule.

Now using equation (5.11), the Bayes risk plot is constructed at various values of sample size n. Figure 5.3 indicates a Bayes risk plot with n^* = 5. That is, the average risk r (a_2) associated with re-centering the process is minimized at maximum economic sample size 5 which results in the value of -8,505.56 as a Bayes risk. This represents a gain as indicated by a negative result.

Alternately, equation (5.15) would result in 4.73, which should be rounded up to 5. As can be seen, this value of n^* as a maximum economic sample size is much smaller than the given value of N. This satisfies the technical feasibility of sampling. It is important to note that the calculated value of n^* is a maximum value of the number of measurements needed to verify a shift of the given magnitude.

Parameter	Value
N	2,000
A	100
Δ	0.05
$k=A/\Delta^2$	40,000
m	8.1
μ_1	8.06
σ_1	0.01
σ_g	0.002
C	10
W	1,000
p	0.7

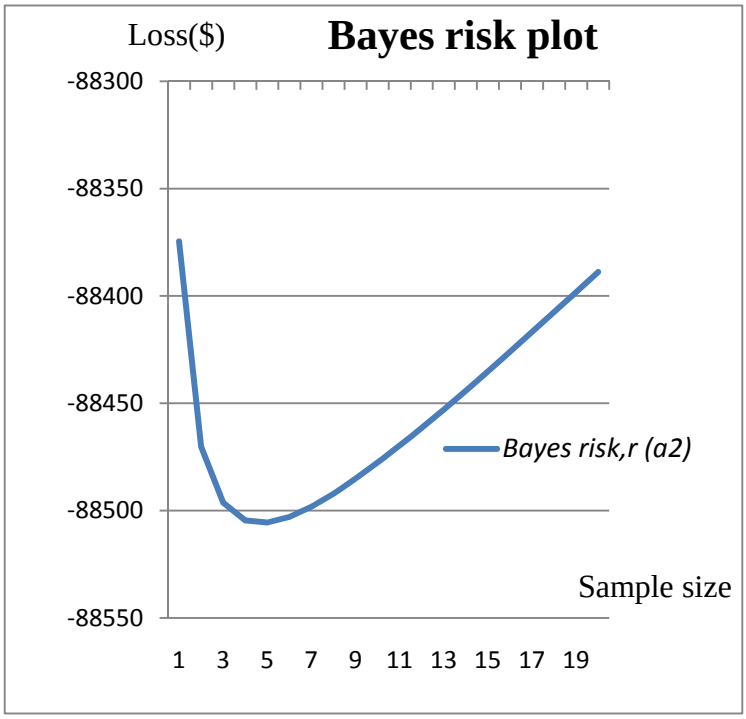


Figure 5.3 Process and cost parameters with Bayes risk plot

The five observations are simulated by the Excel Random Number Generator from a normal distribution of $N(\mu_2, \sigma_g)$. The simulate numbers are 8.10253, 8.10123, 8.09907, 8.09987, and 8.10077. A plot based on μ_2 and ρ_n is constructed to show how the prior process average is moved to the designed target. Figure 5.4 shows the credible interval with limits 8.0983 and 8.1017. This figure shows that the HPD credible interval encloses the posterior mean in the sense that the changes for re-centering the process average were successful. This confirmed that the values of μ_2 and σ_1^2 could be considered as process parameters in phase II of control chart.

Parameter	Value
n^*	5
$\bar{X}$	8.100,694
$\mu_2 = m$	8.1
ρ_n	0.000,000,79
$X_{(0.025)}$	8.098,26
μ_n	8.100,371
$X_{(0.975)}$	8.101,74
$Z_{\alpha/2}$	1.96
α	0.05

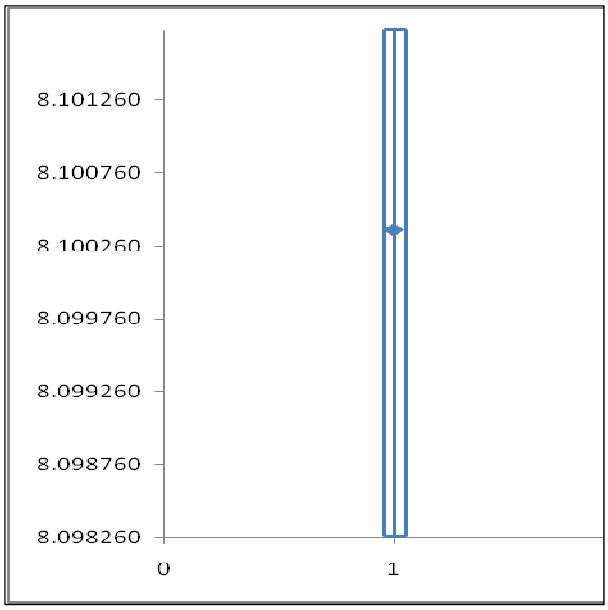


Figure 5.4 HPD credible interval of numerical example for Case I

5.4 Case II: Gradual Shift in the Process Average

In this case, the process is assumed to experience a gradual shift during the transition period. Under this assumption, $\overline{X}$ may not be a sufficient statistic. Figure 5.5 depicts the difference between Case I and Case II relative to the sudden and gradual change of process average.

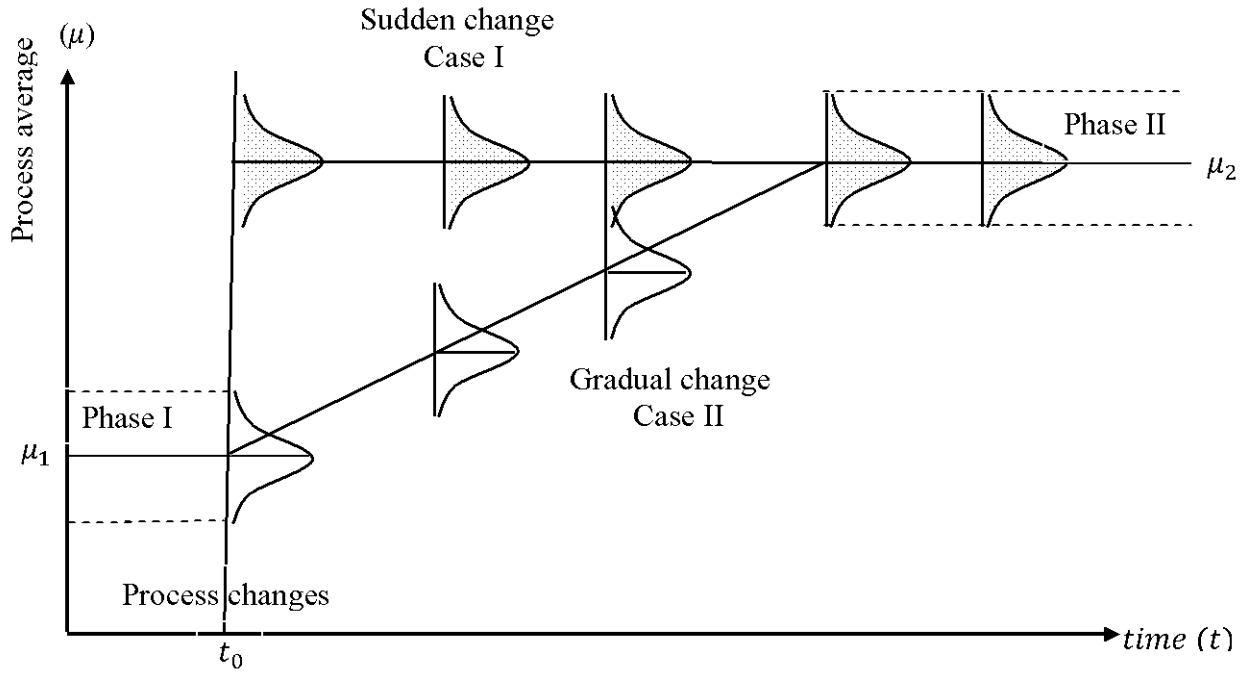


Figure 5.5 Sudden and gradual change of the process average

Since the likelihood function and the resulting posterior density in equations (5.6) and (5.7) would not change, equation (5.15) would still be utilized in calculating the maximum economical sample size *n**. However, it would be beneficial to update the prior average sequentially. That is, values of the posterior mean would be updated after each measurement (x_i) using

$$\mu_{i+1} = \frac{\sigma_1^2 x_{i+1} + \sigma_g^2 \mu_i}{\sigma_g^2 + \sigma_1^2} \tag{5.16}$$

whereas the posterior variance would be obtained as

$$\rho = \frac{\sigma_1^2 \sigma_g^2}{\sigma_g^2 + \sigma_1^2} \tag{5.17}$$

In equation (5.16), the prior average is denoted as μ_1, which is updated to the posterior mean after collecting the first measurement x_1. This process should repeat following each measurement until enough statistical evidence of the shift could be obtained or a total of *n** measurements have been collected. Once again, an HPD credible interval in the form $\mu_2 \pm z_{\alpha/2}\sqrt{\rho}$ can be

constructed based on a specified value of α, and used to verify the shift. When such an interval encloses the posterior mean μ_n, sampling should stop, since no additional measurements need to be collected. It is important to note here that in this case, a decision may be reached before collecting the maximum economical number of measurements *n**. This sequential updating may be aided by a graph such as the one shown in Figure 5.7. It is important to note that this sequential procedure would apply in both Case I and Case II. However, the credible interval obtained in this case is based on a larger value of the posterior variance. In other words, the procedure represented in Case I provides higher accuracy in estimating the process average. The following section represents the same numerical example used in Case I with the assumption that the process experiences a gradual shift. This would help illustrate the difference between the two cases.

Figure 5.6 illustrates a flowchart of the sampling procedure for this case. The sampling plan for case I is different from that of case II. In the first case, a set sample of n^* measurements is collected once and the credible interval is calculated for that sample. In the second case, based on posterior variance ρ and designed target for process average μ_2, an HPD credible interval is constructed. Then one measurement is collected at a time. One of three decision s is made following each measurement. If the credible interval encloses the updated posterior mean the sampling is stopped. Otherwise, the sampling should be continued until the number of units meets the value of *n**. After collecting *n** measurements, if the credible interval still does not enclose the posterior mean, the process changes were not successful.

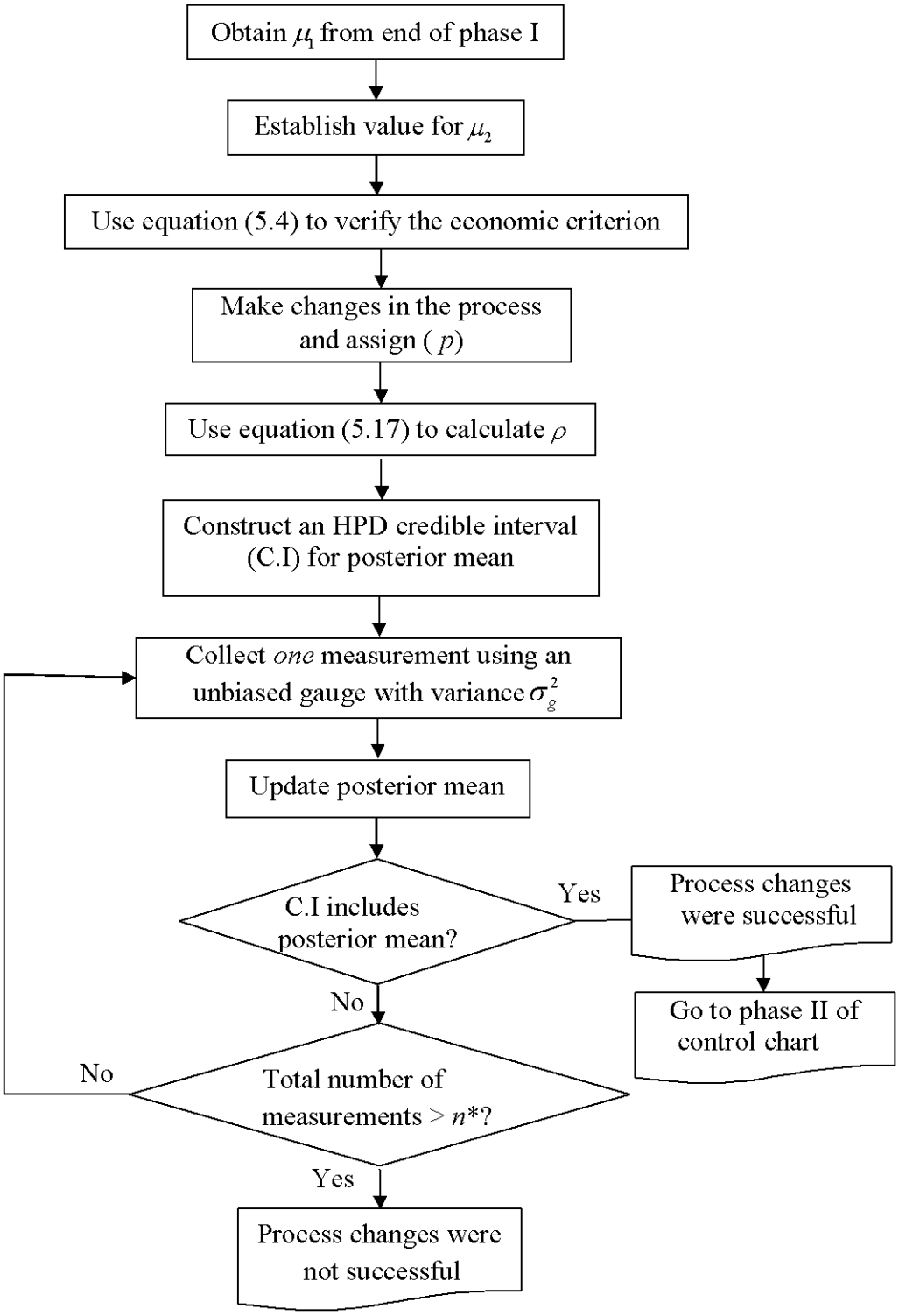


Figure 5.6 Sampling procedure for Case II

5.4.1 Numerical Example for Case II

The same example in Case I is considered with the simulated data modified to indicate a gradual shift in the process average. Figure 5.3 is still valid for the cost parameters and Bayes risk plot in this case. The sampling procedure assessment for Case II is illustrated in Figure 5.6. As can be seen, this procedure is different from that of Case I. In Case I, the posterior parameters are calculated only one time, after sample of n^* observations are collected. On the other hand, in Case II, after each observation, the posterior parameters are sequentially updated, and the effect of the changes is examined. Therefore, it is obviously possible that the procedure is stopped before a total of n^*. This procedure shows how the Bayesian method has the ability to utilize information obtained from each step of sampling.

Table 5.2 shows the simulated observations when the process average follows a gradual shift after making changes. In this case, the HPD credible interval for all the simulated data has the same value.

TABLE 5.2

HPD CREDIBLE INTERVAL FOR POSTERIOR PROCESS AVERAGE

n	**Simulate Observations with Gradual Change**	μ_{i+1}	ρ	$NORMINV(0.025, \mu_2, SQRT(\rho))$ $X_{(0.025)}$	$NORMINV(0.975, \mu_2, SQRT(\rho))$ $X_{(0.975)}$
		8.06			
1	8.07012	8.06947	0.0000038	8.096180	8.103821
2	8.07666	8.07638	0.0000038	8.096180	8.103821
3	8.08294	8.08269	0.0000038	8.096180	8.103821
4	8.09110	8.09078	0.0000038	8.096180	8.103821
5	8.10086	8.10047	0.0000038	8.096180	8.103821

Figure 5.7 shows graphically the HPD credible interval. This figure shows that the posterior mean is increasing upon a gradual trend. A unique credible interval is plotted for all observations. In this numerical example, the credible interval included the posterior mean after collecting the fifth observation. In other cases, this may occur before the maximum economical sample size, n^*.

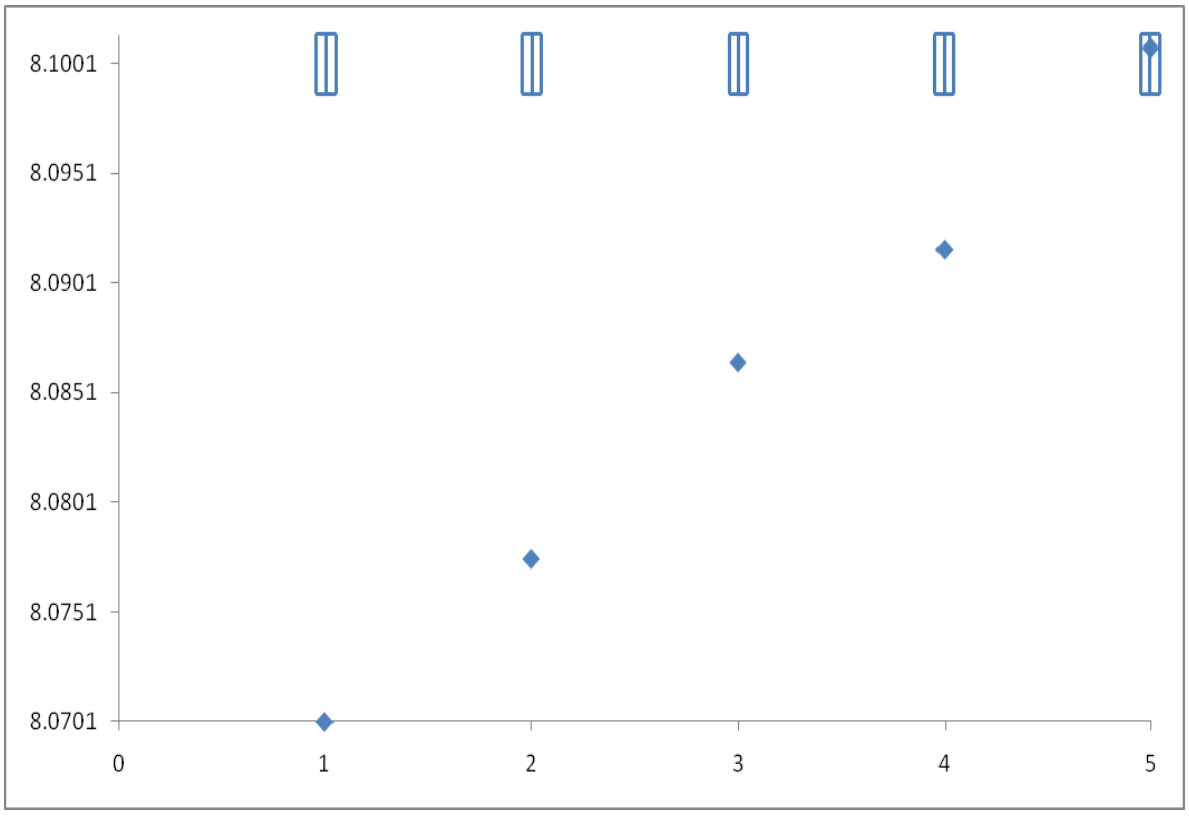


Figure 5.7 HPD credible interval of numerical example for Case II

5.5 Case III: Sudden Reduction in the Process Variance

Excessive variation in a quality characteristic may lead to undesirable effects such as rework, scrap costs, customer dissatisfaction, and so on. As such, reducing the variance plays a significant role in process quality improvement. In this case, a Bayesian method is applied to help the practitioner verify process improvement for reducing the process variance after the transition period.

Case III involves estimating the process variance in a normal distribution with a known mean but unknown variance. It is assumed that the process was in a state of statistical control, and the quality characteristic was shown to follow a normal distribution with parameters $N(\mu_1, \sigma_1^2)$. The process was found to be centered but with high variability (Figure 4.1b). Some efforts are to be made to reduce the process variance. This section presents a fully Bayesian approach for determining the sample size to verify such reduction.

According to Taguchi (1986), equation (2.24), and the loss matrix shown in Table 5.1, the corresponding expected loss associated with taking no action L_1 would be given by equation (5.1), and the loss accompanied by the improved level L_2 is calculated by equation (5.2).

In this case, as shown in Figure 5.8, the process average will not change during the transition period ($\mu_1 = \mu_2$). Assuming that equation (5.4) holds true for the economic criterion, the action a_2 is selected as a Bayes rule so that some changes should be made to reduce process variance from σ_1^2 to σ_2^2. To verify the effect of changes, the corresponding Bayes risk for a_2 is calculated by

$$r(a_2) = pNE[L_2 - L_1] + W + Cn = pNKE(\sigma_2^2 - \sigma_1^2) + W + Cn \qquad (5.18)$$

In equation (5.18), the conjugate prior of the posterior variance should be introduced. Accordingly, the posterior parameters will be calculated by combining the likelihood of the data and the prior distribution. The posterior parameters include all information about the process.

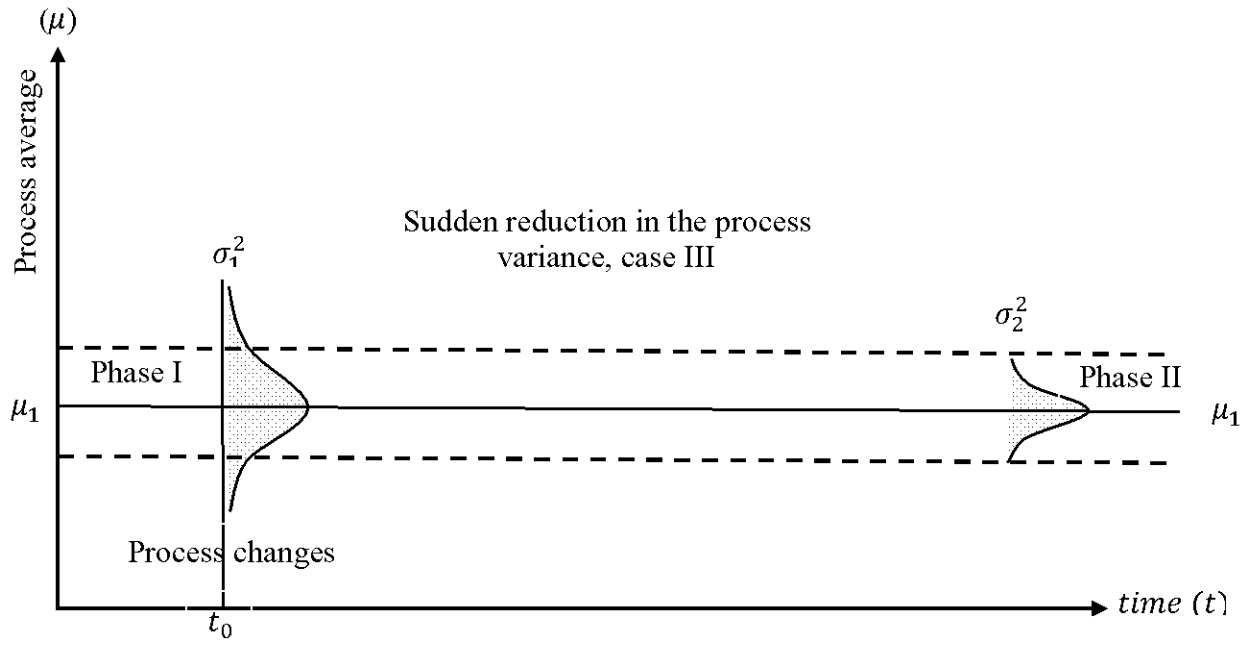


Figure 5.8 Case III: Reduction of variability after making changes in process variance

Using the prior and likelihood functions, the posterior parameters can be calculated. Robert (2001) pointed out that the corresponding conjugate prior for the variance of a normal distribution is the inverse-gamma distribution.

A special case of the inverse-gamma distribution is the scaled inverse chi-squared, which is a distribution of $(\nu_1\sigma_1^2)/\chi_{\nu_1}^2$ in which ν_1 is the number of degrees of freedom and σ_1^2 is the prior process variance (Colosimo and Castilo, 2007). The prior distribution provides the information about ν_1 observations with average squared deviation σ_1^2 , both of which can be quantified at the conclusion of phase I of Shewhart control chart. The likelihood of *n* i.i.d. observations follows a normal distribution

$$f(x_1, x_2, \dots, x_n \,|\sigma^2) = (\sigma^2)^{-n/2}\exp\,(-\frac{n}{2\sigma^2}s^2) \tag{5.19}$$

where s^2, the variance of *n* measurements, is a sufficient statistic, which contains all of the required information about the variance and is expressed as

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \mu_1)^2}{n} \tag{5.20}$$

According to Bayes' theorem, combining the likelihood function and prior distribution, the posterior distribution is calculated as

$$f(\sigma^2|x_1, x_2, \dots, x_n) = f(\sigma^2) * f(x_1, x_2, \dots, x_n\,|\sigma^2) \tag{5.21}$$

$$f(\sigma^2|x_1, x_2, \dots, x_n) \propto (\sigma^2)^{-(\frac{n+\nu_1}{2}+1)} \exp\left(-\frac{1}{2\sigma^2}(\nu_1\sigma_1^2 + ns^2)\right) \tag{5.22}$$

which is a scaled inverse-χ^2. The number of degrees of freedom, $n + \nu_1$, in this distribution is the summation of the number of measurements and prior number of degrees of freedom. The mean and variance of this distribution are given by

$$E(\sigma_n^2) = \frac{n}{n-2}\, s^2 \qquad for\ n > 2\ , and\ s > 0 \tag{5.23}$$

$$var(\sigma_n^2) = \frac{2n^2}{(n-2)^2(n-4)} s^4 \qquad for\ n > 4\ , and\ s > 0 \tag{5.24}$$

where σ_n^2 is the posterior variance for a sample of *n* measurements. The posterior variance depends on the data in the sense that when measurements become available, s^2 could be calculated. That is, the posterior is a function of the measurements only through sufficient statistics s^2.

Given the expected value of scaled inverse-χ^2, the Bayes risk can be calculated. Substituting the value of $E(\sigma_n^2)$ from equation (5.23) into equation (5.18) results in

$$r(a_2) = pNk[E(\sigma^2) - \sigma_1^2] + W + Cn = pNk\left[\frac{n}{n-2}s^2 - \sigma_1^2\right] + W + Cn \tag{5.25}$$

Assuming s^2 to be a constant, the first derivative with respect to *n* is

$$\frac{dr(a_2)}{dn} = -2pNks^2\left(\frac{1}{(n-2)^2}\right) + C \tag{5.26}$$

which should be negative for values of $n > 2$. And the second derivative is

$$\frac{d^2r(a_2)}{dn^2} = pNks^2\left[\frac{4}{(n-2)^3}\right] \tag{5.27}$$

which should be positive for all values of $n > 2$. If these conditions hold true, then setting the first derivative equal to zero results in the maximum economic sample size *n**. That is,

$$n^* = \sqrt{\frac{2pNks^2}{C}} + 2 \tag{5.28}$$

For calculating *n**, it can be assumed that $s^2 = \sigma_2^2$ the target variance of the process. Consequently, equation (5.28) will result in the maximum economic sample size, given the target has been achieved.

The sampling procedure would require the selection of *n** units from which the sample size s^2 is calculated using equation (5.20). This value is used to obtain the posterior variance σ_n^2 , given by equation (5.23). The value of s^2 should be enclosed in an HPD credible interval given by

$$\frac{\sigma_2^2}{n^*}\ \chi^2_{1-\alpha/2,\,n^*} \leq s^2 \leq \frac{\sigma_2^2}{n^*}\ \chi^2_{\alpha/2,n^*} \tag{5.29}$$

Otherwise, the changes made were not successful. Figure 5.9 illustrates the sampling procedure for Case III.

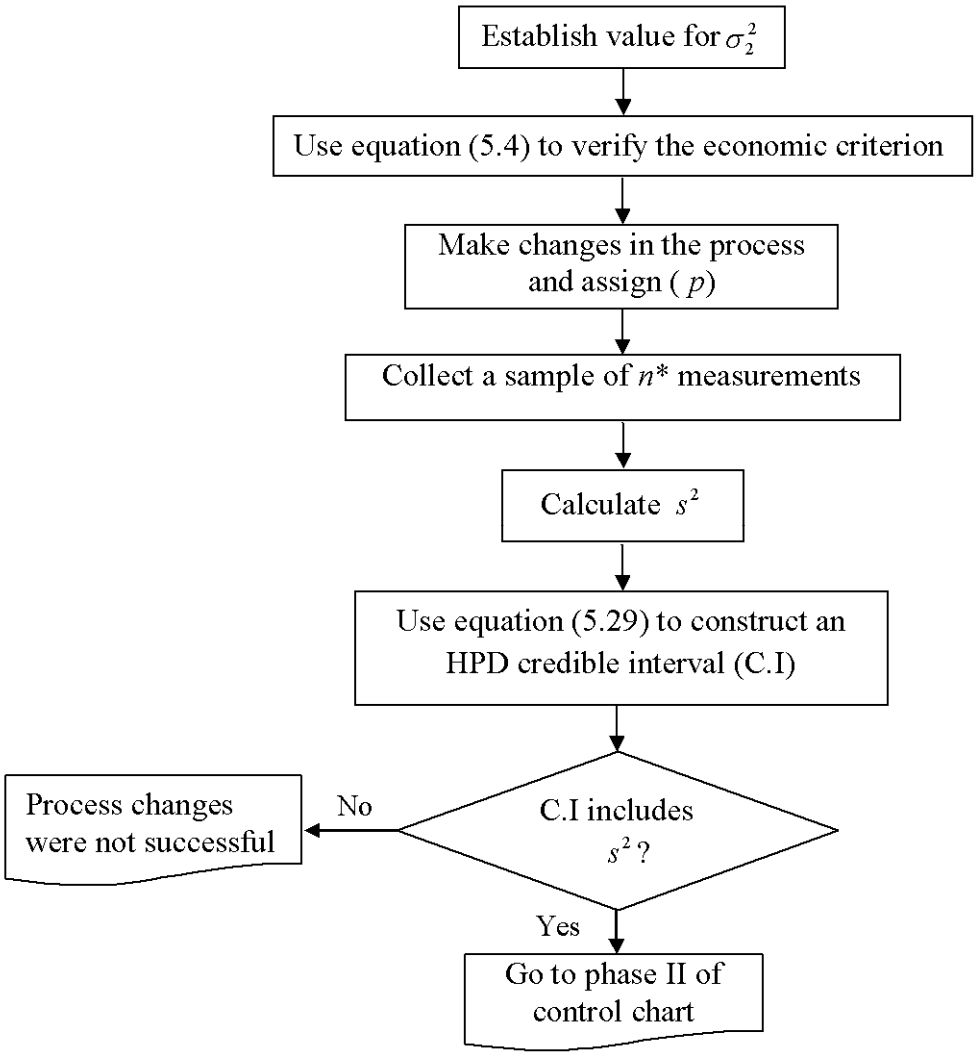


Figure 5.9 Sampling procedure for Case III

5.5.1 Numerical Example for Case III

Assume that the numerical example for Case III is the same example as in the first two cases. The control chart indicates that the dimension of a component follows a normal distribution with μ_1 =m= 8.1 and variance of σ_1^2 = (0.03) ^{2}. A target value for the process variance (σ_2^2) is established at (0.01) ^{2}. At this level of performance, the process capability ($2\Delta \quad /6\sigma$) would improve from 0.56 to 1.67. The economic criterion is satisfied because inequality (5.4) results in - 43,800 < 72,000. Some changes is made in the process and the

probability that the changes will be successful in reduction of process variance has been estimated a priori as p=0.7. Utilizing equation (5.28), the maximum economic sample size is 35.5 or 36. This result can be obtained by plotting equation (5.18) for various values of n as shown in Figure 5.10. An examination of the plot reveals that n^*=36. In addition, the n^* is much less than the value of N which make the feasibility technical criterion hold true. The corresponding loss accompanied by this value of sample size is calculated as $ - 43,110.59 which represents the obtained gain associated with the process improvement.

Parameter	**Value**
N	2,000
A	100
Δ	0.05
$k = A/\Delta^2$	40,000
$m = \mu_2 = \mu_1$	8.1
σ_1	0.03
σ_2	0.01
C	10
W	1,000
p	0.7

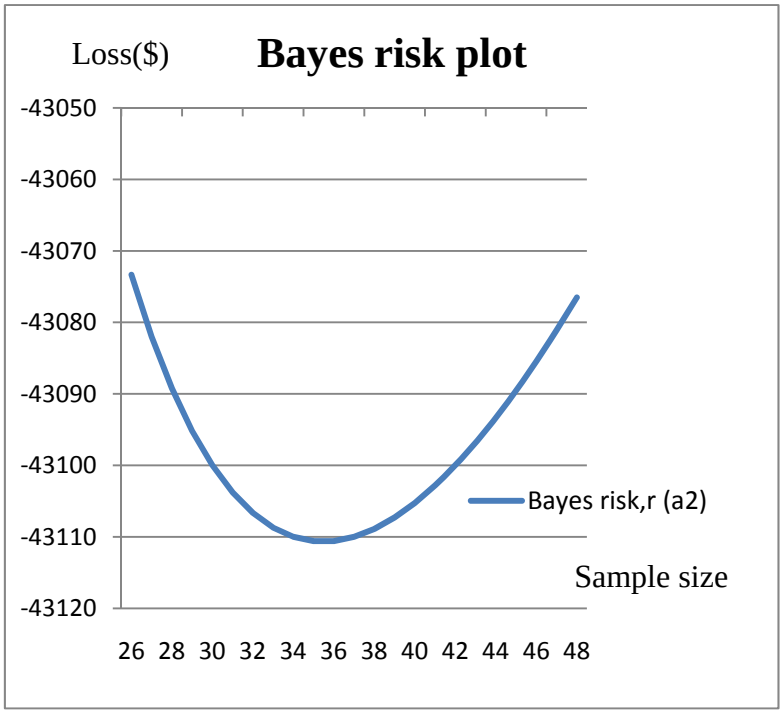


Figure 5.10 Process and cost parameters with Bayes risk plot for Case III

Utilizing equation (5.29) an HPD credible interval is constructed with a specified α. Figure 5.11 shows that s^2 lies within the credible interval.

Parameter	Value
σ_2^2	0.000,100
n^*	36
$\chi^2_{\alpha/2,n^*}$	16.012,764
α	0.05
$\chi^2_{1-\alpha/2,\,n^*}$	1.689,869
LCI	0.000,058
s^2	0.000,142
UCI	0.000,150

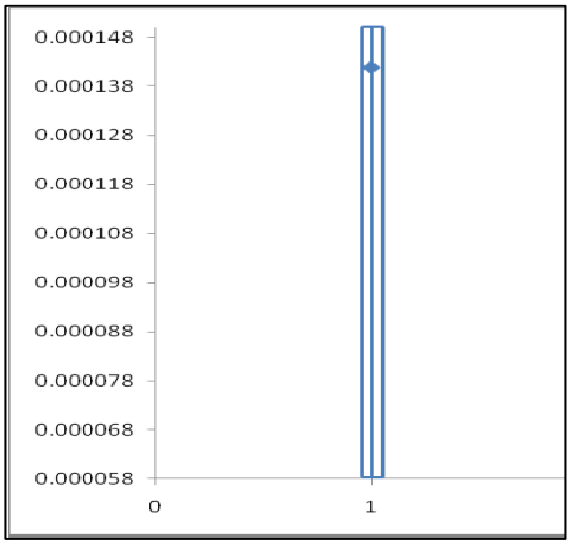


Figure 5.11 HPD credible interval of numerical example for Case III

Another way for showing the credible interval for the process variance is shown in Figure 5.12.

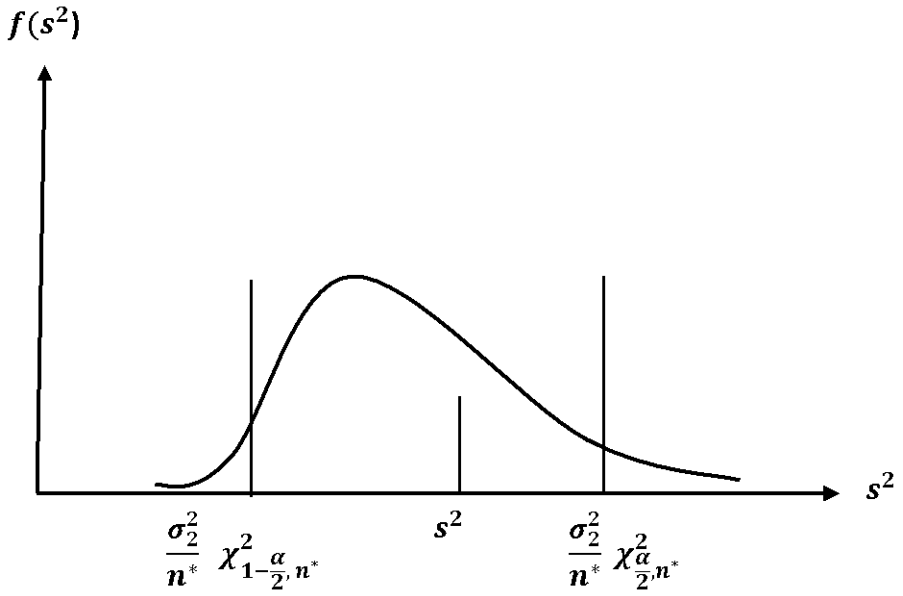


Figure 5.12 HPD credible interval for s^2 in scaled inverse-χ^2 for Case III

Both Figures (5.11) and (5.12) illustrate that the changes for reduction of variance were successful. This confirms that the new process parameters are μ_1 and σ_2^2. The two parameters could be considered as process parameters in phase II of control chart.

CHAPTER 6

CONCLUSIONS AND FUTURE RESEARCH

One of the main purposes of this research was to show the application of Bayesian methods where the traditional methods are not applicable in calculating the sample size and determining monetary loss. It also attempted to clarify advantages and limitations of each approach.

There are some scenarios where the process level is found to be away from the desired target in the conclusion of phase I, and corrective changes must be made to improve the performance. A number of sampling techniques can be utilized to quantify the impact of engineering efforts and verify their effect of changes on the process performance. The duration of this activity is called a transition period. In the transition period, some assumptions of Shewhart control charts are violated because observations for verification of process improvement may not be non-stationary or auto-correlated. In addition, traditional methods are not applicable for determining the associated loss with each corrective option. On the other hand, the fully Bayesian methods can be applied in this period.

The modeling approach used in this research was built on the ability of the Bayesian methods to update process parameters during the transition period free of any assumption. The applied approach also utilized the knowledge gained during phase I to overcome the limitations of the Bayesian methods.

The first objective of this research was to develop a mathematical model for sample size determination for verification of process improvement accompanied by loss computation during the transition period after completion of phase I of Shewhart control chart. In achieving the first objective, a loss function was defined and used to calculate the expected Bayes risk associated

with each improvement alternative. Two criteria, economical and technical, were defined, and equations for calculating the maximum economic sample size were derived to verify the process improvement. The equations for sample size determination were developed in two ways, namely approximated and accurate. Due considerations were given to the gauge capability, the cost of sampling, and the cost of process setup. Methods by which the resulting risk function can be optimized with respect to sample size were considered. Three applications of the mathematical model were explored to verify a sudden or gradual shift in the process average or a reduction in the process variance. The assumptions and characteristics of three applications were distinguished. In addition, applying the developed model in each case by illustrative examples verified the model. In case of changes to the process average, the sample size n cannot be determined without knowledge of the type action taken and the process reaction. Therefore, this case was discussed in two separated Case I and Case II with two different process reactions. The third case involved a sudden reduction in the process variance.

The second objective of this research was to represent a sampling procedure with guidelines to secure effective implementation of the model. In achieving the second objective, efforts were made to introduce a sampling plan for each case separately. The sampling procedures depend only on knowledge of the process reaction to the changes made. Utilizing a flowchart, all necessary steps for an appropriate sampling were demonstrated. Each case was followed by a numerical example to show the implementation of the corresponding model. The values of prior distribution were established, and parameters of the posterior distribution were calculated utilizing the simulated data. The properties of the Bayes' theorem were applied to update the posterior parameters by combining the information of prior and likelihood function of

data. An HPD credible interval was constructed for each numerical example to show graphically how the process parameters were changed.

6.1 Contributions

This dissertation served the following purposes:

- Developing a Bayesian method to determine the maximum economic sample size required to test the effects of changes to a process. Results of this method verified that the process mean or variance has really changed closer to the target level.
- Deriving the mathematical models to establish implementation boundaries from economic and technical viewpoints.
- Applying Taguchi's quadratic loss function to quantify the loss incurred when an action for changing the process is taken.
- Calculating the maximum economic sample size based on the economic worth of the improvement process reaction to the changes.
- Distinguishing between the possible scenarios related to parameters. Three cases were identified. The first two cases involved a normal distribution with unknown mean but known variance in which the process average had a sudden shift or a gradual shift to the desired target. The third case involved a normal distribution with known mean and unknown variance in which the process variance should be reduced during the transition period.
- Demonstrating the examples to clarify the step-by-step implementation of the model and highlighting the economic advantages of the developed procedures.
- Developing a sampling procedure for each case to secure implementation of the model properly.

- Overcoming the criticism of no visibility of the Bayesian approach by illustrating the results with an HPD credible interval and other graphs.
- Overcoming the criticism of subjectivity for prior in the Bayesian approach by obtaining the prior information from the control chart in conclusion of phase I.

6.2 Concluding Remarks

The strength of the developed fully Bayesian approach as compared to the traditional approach can be viewed as follows:

- The Bayesian approach is more applicable than traditional methods in the following situations: First, it is desirable to accumulate the numerical data of the parameter of interest; second, the decision making is desirable in terms of loss associated with every scenario; third, informative prior knowledge is available.
- Considering the uncertainty of prior information in the Bayesian approach leads to more effective procedures. In other words, as more data is obtained, the posterior inferences are updated so that the Bayesian method is more flexible when compared to other methods.
- Traditional methods do not take into account prior or loss information. The only approach to deal with the calculation of loss that utilizes the prior information is the Bayesian approach. The greatest need to apply Bayesian analysis is in decision-making problems for which prior and loss information is a crucial part of the problem.
- Bayesian methods consider all available evidence for making an inference, while traditional methods use only information of the last sample.
- In many cases involving industrial projects, process parameters are unknown. Bayesian methods allow for optimal use of partial information on the process. This approach

provides a good bridge between the extremes of total knowledge and total ignorance of process parameters.

- The Bayesian paradigm has the ability to make inferences where sufficient statistics are not available. In other words, Bayesian statistics along with prior information are applicable through limited samples that may violate the assumptions of random sampling.
- By utilizing Bayesian posterior distribution, estimates are often calculated with smaller variance than that of traditional estimates.

6.3 Future Research

This research could be extended for processes having other conditions. The following sections outline suggested areas for future research.

6.3.1 Both Parameters Unknown

In this dissertation, three scenarios were discussed. It may be important to consider a scenario in which both the mean and variance are targets for improvement. Then another conjugate prior should be introduced, and this situation affects the mathematical model and procedure sampling. In addition, in the developed model, it was assumed that the quality characteristic follows a normal distribution. This assumption for Case I, based on the CLT, is unavoidable. However, if this assumption is removed from Case II and Case III, the model and the results are affected.

6.3.2 Multivariate Quality Characteristic

In this research, the quality characteristic was considered as a univariate. An extension to the multivariate study is needed in some scenarios in which more than one variable is involved.

6.3.3 Shifting Pattern for Process Average

In this dissertation, the gradual shift for the process average in Case II follows only one particular pattern. Some scenarios involve a slower or a faster trend for gradual shift. The rate of trend may affect the sampling procedure or calculation of the maximum economic sample size. Special attention is needed to investigate the effect of such scenarios on the results.

6.3.4 Measurement Error

In this study, an unbiased gauge was assumed. The existence of two sources of gauge variability, namely, bias and imprecision, affects the model development. The significance of studying this situation would be to evaluate the effect of measurement error on the optimal sample size and the monetary loss.

6.3.5 Prior Probabilities

In order to construct the loss matrix, in this dissertation, only two prior probabilities were assumed, i.e., p and $(1 - p)$. In some scenarios, more than two options with different probabilities may be involved. More investigations need to be attempted for development of the mathematical models in these scenarios.

BIBLIOGRAPHY

BIBLIOGRAPHY

Adcock, C. J. (1988). A Bayesian Approach to Calculating Sample Sizes. *The Statistician, 37*, pp. 433-439.

Adcock, C. J. (1995). The Bayesian Approach to Determination of Sample Sizes: Some Comments on the Paper by Joseph,Wolfson, and du Berger. *Statistician, 44*, pp. 155-161.

Adcock, C. J. (1997). Sample Size Determination: A Review. *The Statistician, 46* (2), pp. 261-283.

Albert, J. (2009). *Bayesian Computation with R.* New York: Springer Science+Business Media, LLC.

Automotive Industry Action Group. (2002). *Measurement System Analysis* (third ed.). Detroit: AIAG.

Bayes, T. (1783). An Essay Towards Solving a Problem in the Doctrine of Chaces. *Philosophical Transactions of the Royal Society, 53*, pp. 370-418.

Berger, J. (1985). *Statistical Decision Theory and Bayesian Analysis* (second ed.). New York: Springler-Verlag.

Berger, J., and Sellke, T. (1987). Testing a Point Null Hypothesis: The Irreconcilability of P Values and Evidence. *Journal of the American Statistical Association, 397*, pp. 112-122.

Berger, J. O. (2000). Bayesian Analysis: A Look at Today and Thoughts of Tomorrow. *Journal of the American Statistical Association, 95*, pp. 1269-1276.

Berry, D. A. (1997). Teaching Elementary Bayesian Statistics with Real Applications in Science. *The American Statistician, 51* (3), pp. 241-246.

Berry, D. A. (2006). Bayesian Clinical Trials. *Nature Reviews Drug Discovery, 5*, pp. 27-36.

Bolstad, W. (2004). *Introduction to Bayesian Statistics.* Hoboken, NJ: John Wiley and Sons.

Bolstad, W. M. (2010). *Understanding Computational Bayesian Statistics.* Hoboken, NJ: John Wiley and Sons.

Borror, C. M., Montgomery, D. C. and Runger, G. C. (1999). Robustness on the EWMA Control Chart to Nonnormality. *Journal of Quality Technology, 31* (3), pp. 309-316.

Box, G. E. P., and Tiao, G. C. (1973). *Bayesian Inference in Statistical Analysis.* New York: Wiley-Hall.

Broemeling, L. D. (2009). *Bayesian Methods for Measures of Agreement.* Boca Raton: Taylor and Francis Group, LLC.

Carlin, B. P., and Louis, T. A. (2009). *Bayesian Methods for Data Analysis.* Boca Raton, FL: Chapman and Hall.

Carlin, B. P., and Louis, T. A. (2000). *Bayes and Empirical Bayes Methods for Data Analysis* (second ed.). New York: Chapman and Hall/CRC.

Casella, G., and Berger, R. L. (2002). *Statistical Inference.* Pacific Grove, CA: Thomson Learning.

Chakraborti, S., Human, S. W., and Graham, M. A. (2009). Phase I Statistical Process Control Charts: An Overview and Some Results. *Journal of Quality Engineering, 21*, pp. 52-62.

Colosimo, B. M., and Castillo, E. D. (2007). *Bayesian Process Monitoring, Control and Optimization.* Boca Raton: Taylor & Francis Group.

Congdon, P. (2003). *Applied Bayesian Modelling.* Chichester: John Wiley and Sons.

Congdon, P. (2005). *Bayesian Models for Categorical Data.* Chichester: John Wiley and Sons.

Congdon, P. (2007). *Bayesian Statistical Modelling* (second ed.). Chichester: John Wiley and Sons.

Crowder, S. V. (1989). Design of Exponentially Weighted Moving Average Schemes. *Journal of Quality Technoliogy, 21* (2), pp. 155-162.

Cudney, E. A., and Drain, D. (2007). Effective Use of Process Capability Indices for Supplier Management. Edited by G. Bayraksan, W. Lin, Y. Son, and R. Wysk. *Proceedings of Industrial Engineering Research Conference* (pp. 1-10).

Deming, W. E. (1986). *Out of the Crisis.* Cambridge, MA: Massachusetts Institute of Technology.

Desu, M. M., and Raghavarao, D. (1990). *Sample Size Methodology*. New York: Academic Press, Inc.

Dodge, H. F., and Romig, H. G. (1929). A Method of Sampling Inspection. *Bell System Tech. J.8*, pp. 613-631.

Duncan, A. J. (1956). The Economic Designn of X_Bar Charts Used to Maintain Current Control of a Process. *Journal of the American Statistical Association, 51* (274), pp. 228-242.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2004). *Bayesian Data Analysis* (second ed.). Boca Raton FL: Chapman and Hall/CRC Press.

Ghosh, B. K. (1991). A Brief History of Sequential Analysis. In *Handbook of Sequential Analysis*. Edited by B. K. Ghosh and P. K. Sen. New York: Dekker.

Gill, J. (2008). *Bayesian Methods.* New York: Chapman and Hall.

Goldstein, M. (1981). A Bayesian Criterion for Sample Size. *Annals of Statistics, 9*, pp. 670-672.

Gould, A. L. (1993). Sample Sizes for Event Rate Equivalence Trials Using Prior Information. *Statistics in Medicine, 12*, pp. 1209-1223.

Hawkins, D. M., and Olwell, D. H. (1998). *Cumulative Sum Charts and Charting for Quality Improvement.* New York, NY: Springer Verlag.

Hawkins, D. M., Choi, S., and Lee, S. (2007). A General Multivariate Exponentially Weighted Moving Average Control Chart. *Journal of Quality Technology, 39*, pp. 118-125.

Jaynes, E. T. (2003). *Probability Theory: The Logic of Science.* New York: Cambridge University Press.

Jeffreys, H. (1961). *Theory of Probability* (third ed.). London: Oxford University Press.

Joseph, L., and Belisle P. (1997). Bayesian Sample size Determination for Normal Means and Differences Between Normal Means. *The Statistician, 46*, pp. 209-226.

Joseph, L., and Wolfson, D. B. (1997). Interval-Based Versus Decision Theoretic Criteria for the Choice of Sample Size. *The Statistician, 46* (2), pp. 145-149.

Joseph, L., du Berger, R. and Belisle, P. (1997). Bayesian and Mixed Bayesian/Likelihood Criteria for Sample size Determination. *Statistics in Medicine, 16*, pp. 769-781.

Joseph, L., Wolfson, D. B., and du Berger R. (1995a). Sample Size Calculations for Binomial Proportions via Highest Posterior Density Intervals. *The Statistician, 44* (2), pp. 143-154.

Joseph, L., Wolfson, D. B., and du Berger R. (1995b). Some Comments on Bayesian Sample size Determination. *The Statistician , 44*, pp. 167-171.

Kraemer, H. C., and Thiemann, S. (1987). *How Many Subjects?: Statistical Power Analysis in Research.* Newbury Park: Sage.

Lai, T. L. (2001). Sequential Analysis: Some Classical Problems and New Challenges. *Statistica Sinica, 11*, pp. 303-408.

Lee, P. (2004). *Bayesian Statistics.* New York: Oxford University Press.

Lee, S. J., and Zelen, M. (2000). Clinical Trials and Sample Size Considerations: Another Perspective. *Statistical Science, 15*, pp. 95-110.

Lindley, D.V. (1997). The Choice of Sample Size. *The Statistician, 46* (2), pp. 129-138.

Lindley, D. V. (1983). Theory and Practice of Bayesian Statistics. *The Statistician, 32*, pp. 1-11.

Lorenzen, T. J., and Vance, L. C. (1986). The Economic Design of Control Charts: A Unified Approach. *Technometrics, 28* (1), pp. 3-10.

Lucas, J. M., and Saccucci, M. S. (1990). Exponentially Weighted Moving Average Control Schemes: Properties and Enhancements. *Technometrics, 32* (1), pp. 1-12.

MacCarthy, B. L., and Wasusri, T. (2002). A Review of Non-Standard Applications of Statstical Process Control (SPC) Charts. *International Journal of Quality & Reliability Management, 19* (3), pp. 295-320.

Maghsoodloo, S. (1992). *Introduction to Taguchi-Based Quality: Design and Improvement.* Auburn, AL: Engineering Learning Resources, College of Engineering, Auburn University.

Marcellus, R. L. (2008a). Bayesian Monitoring to Detect a Shift in Process Mean. *Quality and Reliability Engineering International, 24*, pp. 303-313.

Marcellus, R. L. (2008b). Bayesian Statistical Process Control. *Quality Engineering, 20* (1), pp. 113-127.

Marin, J., and Robert, C. (2007). *Bayesian Core: A Practical Approach to Computational Bayesian Statistics.* New York: Springer.

Mendenhall, W., and Sincich, T. (2007). *Statistics for Engineering and the Sciences* (fifth ed.). New Jersey: Pearson Prentice Hall.

Meyer, D. L., and Collier, R. O., Jr. (1970). *Bayesian Statistics.* Arlington Heights, IL: F. E. Peacock Publishers, Inc.

Mitra, A. (1998). *Fundamentals of Quality Control and Improvement.* Upper Saddle River, NJ: Prentice-Hall, Inc.

Montgomery, D. C. (2009). *Introduction to Statistical Quality Control* (sixth ed.). New York: Wiley.

Montgomery, D. C., and G. C. Runger. (2007). *Applied Statistics and Probability for Engineers* (fourth ed.). New York: Wiley.

Page, E. S. (1954). Continuous Inspection Schemes. *Biometrika*, pp. 100-114.

Peck, R., Devore, J., and Olsen, C. (2005). *Introduction to Statistics and Data Analysis.* Thomson Learning.

Pham-Gia, T. (1997). On Bayesian Analysis, Bayesian Decision Theory and the Sample Size Problem. *The Statistician, 46,* pp. 389-392.

Pham-Gia, T., and Turkkan, N. (1992). Sample size Determination in Bayesian Analysis. *The Statistician, 41,* pp. 389-397.

Pham-Gia, T., and Turkkan, N. (2003). Determination of Exact Sample Sizes in the Bayesian Estimation of the Difference of Two Proportions. *The Statistician, 52* (2), pp. 131-150.

Pham-Gia, T., and Turkkan, N. (2005). Bayesian Decision Criteria in the Presence of Choices under Quadratic and Absolute Value Loss Functions. *Statistical Papers, 46* (2), pp. 247-266.

Press, S. J. (1989). *Bayesian Statistics: Principles, Models, and Applications.* New York: John Wiley and Sons, Inc.

Press, S. J. (2003). Subjective and Objective Bayesian Statistics. NJ: John Wiley and Sons.

Raiffa, H., and Schlaifer, R. (1961). *Applied Statistical Decision Theory.* Boston: Harvard University Graduate School of Business Administration.

Reichert, P., Schervish, M., and Small, M. J. (2002). An Efficient Sampling Technique for Bayesian Inference with Computationally Demanding Models. *Technometrics, 44* (4), pp. 318-327.

Robert, C. P. (2001). *The Bayesian Choice* (second ed.). New York: Springer-Verlag.

Roberts, S. W. (1959). Control Charts Based on Geometric Moving Averages. *Technometrics,* pp. 239-250.

Roshan, J. V. (2004). Quality Loss Functions for Nonnegative Variables and Their Applications. *Journal of Quality of Technology, 36* (2), pp. 129-138.

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

Spiring, F. A. (1993). The Reflected Normal Loss Function. *Canadian Journal of Statistics, 21,* pp. 321-330.

Spiring, F. A., and Yeung, A. S. (1998). A General Class of Loss Functions with Industrial Applications. *Journal of Quality Technology, 30,* pp. 152-162.

Stoumbos, Z. G., Reynolds, M. R. Jr, Ryan, T. P., and Woodall, W. H. (2000). The State of Statistical Process Control as We Proceed into the 21st Century. *Journal of the American Statistical Association, 95*, pp. 992-998.

Sun, F., Laramee, J., and Ramberg, J. (1996). On Spiring's Inverted Normal Loss Function. *Canadian Journal of Statistics, 24*, pp. 241-249.

Tagaras, G. (1998). A Survey of Recent Developments in the Design of Adaptive Control Charts. *Journal of Quality Technology, 30* (3), pp. 212-230.

Taguchi, G. (1986). *Introduction to Quality Engineering.* Tokyo: Asian Productivity Organization.

Tan, M., Tian, G. L., and Ng, K. (2008). *Bayesian Missing Data Problems: EM, Data Augmentation, and Noniterative Computation.* Boca Raton, FL: Chapman and Hall/CRC Press.

Tan, M., Tian, G. L., and Ng, K. W. (2010). *Bayesian Missing Data Problems: EM, Data Augmentation, and Noniterative Computation.* Boca Raton, FL: Taylor and Francis Group, LLC.

Thompson, W. R. (1933). On the Likelihood that One Unknown Probability Exceeds Another in View of the Evidence of Two Samples. *Biometrika, 25*, pp. 285-294.

Tsiamyrtizis, P., and Hawkins, D. M. (2005). A Bayesian Scheme to Detect Changes in the Mean of a Short-Run Process. *Technometrics, 47* (4), pp. 446-456.

Wadsworth, H. M., Stephens, K. S., and Godfrey, A. B. (2002). *Modern Methods for Quality Control and Improvement* (second ed.). NewYork: Wiley.

Wald, A. (1947). *Sequential Analysis.* New York: Wiley.

Weheba, G. S., and Nickerson D. (2004). The Economic Design of X-bar Charts: A Proactive Approach. *Quality and Reliability Engineering International, 21*, pp. 1-14.

Wetherill, G. B. (1961). Bayesian Sequential Analysis. *Biometrika, 48* (3 and 4), pp. 281-292.

Woodall, W. H. (1986). Weakness of the Economic Design of Control Charts. *Tehnometrics, 28*, pp. 408-409.

Woodall, W. H. (2000). Controversies and Contradictions in Statistical Process Control. *Journal of Quality Technology, 20* (4), pp. 515-521.

Woodall, W. H. (2006). Use of Control Charts in Health Care Monitoring and Public Health Surveillance. *Journal of Quality Technology, 38*, pp. 89-104.

Woodall, W. H., and Mahmoud, M. A. (2005). The Inertial Properties of Quality Control Charts. *Technometrics, 47,* pp. 425-436.

Woodall, W. H., and Montgomery, D. C. (1999). Research Issues and Ideas in Statistical Process Control. *Journal of Quality Technology, 31* (4), pp. 376-386.

Wu, C. W., and Shu, M. H. (2007). A Bayesian Procedure for Assessing Process Performance Based on Expected Relative Loss with Asymmetric Tolerances. *Journal of Applied Statistics, 34* (9), pp. 1109-1123.

Zantek, P. F. (2006). Design of Cumulative Sum Schemes for Start-Up Processes and Short Runs. *Journal of Quality Technology, 38* (4), pp. 365-375.

APPENDICES

APPENDIX A

DEFINITIONS OF TERM S

Some of the major terms used in this dissertation are defined as follows:

Alpha Risk (α-risk): The probability that the hypothesis test will end with a rejection of the null hypothesis incorrectly. For instance, $\alpha = 5\%$ means that the null hypothesis will be wrongly rejected only 5 times out of 100 times.

Beta Risk (β-risk): The probability that the hypothesis test will end with an acceptance of the null hypothesis incorrectly.

Bayes Rule: A decision rule that minimizes incurred loss.

Bayes Risk: The amount of risk corresponding to the Bayes rule.

Bayesian and Traditional Inference: Respectively, a parameter regarded as a random variable that is calculated based on an accumulation of all earlier information, and a parameter regarded as a fixed variable that is calculated based on only the last sample.

Central Limit Theorem: If a random sample of n observation (n is sufficiently large), $x = (x_1, x_2,\ldots, x_n)$, is drawn from a population with finite mean μ and variance σ^2, then the sampling distribution of the sample mean $\bar{x}$ is approximated by a normal distribution with $\mu_{\bar{x}} = \mu,\ \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$.

Degree of Freedom: A value that can freely vary in the calculation of a statistic. In other words, if the number of parameters is deducted from the number of values, a degree of freedom would result. The degree of freedom for an estimate is defined as the number of independent observations by which the estimate is calculated.

HPD: A specific credible interval with the characteristic that every point included has higher probability density than every point excluded.

Inverse-Gamma, Chi-Square (χ^2), Inverse Chi-Square, Scaled Inverse Chi-Square: Respectively, the conjugate prior distribution for the normal variance; a special case of the gamma distribution, with $\alpha = \nu/2$ and $\beta = 1/2$; a special case of the inverse-gamma distribution, with $\alpha = \nu/2$ and $\beta = 1/2$; a conjugate prior for variance parameter in normal distributions.

Likelihood Function: A function of θ for observed data, *x*, and denoted by $f(\theta|x)$. This function contains all necessary experimental information about θ after drawing an observation and helps making any inference about the parameter of interest. In addition, two different likelihood functions, which are proportional to each other, provide the same knowledge about the desired parameter.

Mean: The point of balance or average of a data set.

Objective Prior Information: Those observations or records are available before data collection in the sense that they could be evaluated quantitatively. An objective prior distribution shows that there is very little available information about the process.

Parameter: Numerical descriptive measures of a population, usually unknown.

Posterior Distribution ($f(\theta|x)$): Calculated by combining the prior distribution, $f(\theta)$, the sample information, *x*, and the likelihood function, $f(x|\theta)$, after data collection. All inferences could be made by this distribution. The posterior distribution contains all current information about the parameter of interest. One of the most important advantages of the Bayesian approach is the flexibility with which posterior inferences can be summarized. Each summary (for example, mean, variance, and mode) has its own interpretation, which would be helpful about parameter inferences.

Prior Distribution: How the information about the underlying process is quantified prior to sampling. Prior information is an important component for carrying the Bayes' theorem. The prior probability distribution is the analyst's belief regarding the parameter prior the observation of data. If the sample size is large, then the prior information will not have an effect on inference, and the data will be left to speak for themselves. However, prior information strongly affects parameter estimation when the sample size is small. With a small group of data, the prior information can weigh heavily on the result of the posterior computation.

P-Value: The smallest level of significance that would lead to rejection of null hypothesis. That is, H_0 is rejected when the *P*-value is less than the probability of type I error (α).

Random Variable: A numerical valued function, x, defined over a sample space. Each simple event in the sample space is assigned a value of x.

Random Sample: Random variables $x_1, x_2,\ldots, x_n$, if they are mutually independent, and independent and identically distributed (i.i.d.). On the other hand, in a sequential sampling, the size is not identified in advance, and the elements are selected one by one.

Statistic: Numerical descriptive measure computed from a sample data set.

Subjective Prior Information: An attempt to bring additional information to bear on a problem. The effect of this prior information on the posterior is as if there were additional replications of the data, and then the subjective prior may be the only way to offer an acceptable solution for the problem. Unlike objective prior information, subjective prior information may provide insufficient information to solve the problem.

Sufficient Statistics: A function of sample data that includes all available sample knowledge concerning the parameter of interest. For example, $\bar{X}$and S^2 are sufficient statistics for μ and σ^2, respectively.

Unbiased Gauge Variability: The capability of a gauge to measure the true value correctly by average, while the precision refers to the inherent variability in the measurement system.

Transition Period: A period of activity at the conclusion of phase I of the Shewhart control chart where the process level is found to be away from the target design, and changes are required to improve performance. At this point, some corrective changes should be proposed and tested to verify the process improvement.

Variance: The variation of a data set in terms of the amounts by which the observations deviate from the mean.

APPENDIX B

STRENGTH OF BAYESIAN METHODS

The following example shows, with the Bayesian method, how soon two different priors of the same process will converge to the same number, unlike the Shewhart chart which cannot get close two different beliefs**.** The example of Box and Tiao (1973) was the stimulus for this section.

Example

A process improvement team is considering two levels of performance. The current level, L_0, is where the prior is approximated by N (800, 50^2). The second level, L_1, represents an improved level, which is anticipated to be N (950, 80^2). Process changes will be made during the transition period, and individual observations will be collected sequentially. The measuring instrument is known to produce data with uncertainty of $\sigma_g = 10$. Assume the true average is 950. After making some changes in the process, it is desired to verify the shift in the average. Utilizing two priors and the likelihood from simulated data, the posterior mean and variance are calculated.

The two priors are influential in deciding the posterior distribution. When the two different priors are combined with the same likelihood, different posterior distributions are produced.

Figure B.1 shows posterior distributions at both levels L_0 (the first row) and L_1 (the second row) for simulated observations. As can be seen after the first observation, two beliefs, which are represented by the posterior distribution, are much closer to each other than before.

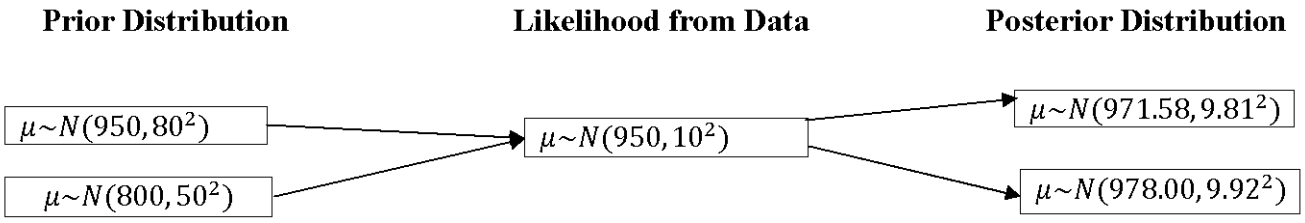


Figure B.1 Values of prior and posterior parameters for two levels L_0 and L_1

Figure B.2 shows that the two posteriors are much closer to each other than their priors. Moreover, after only eight iterations, the variations of the first prior dropped from 80 to 3.53, and those of the second prior dropped from 50 to 3.53. At this iteration, the values of the two posterior means will converge to the same number of 950.

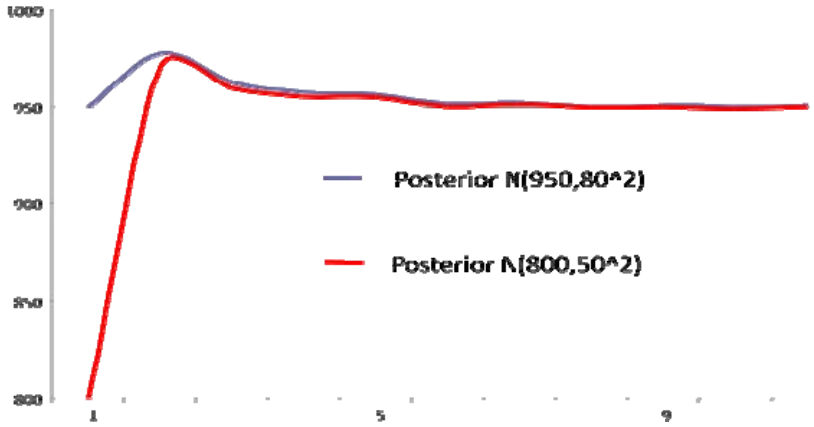


Figure B.2 Convergence of two posteriors indicating shift in process average [simulated $x_i \sim N(950, 10^2)$]

The important conclusions that can be derived from this example are first, the Bayesian framework converges rapidly to the true process average regardless of the original beliefs. Second, no specific assumptions were made regarding the process behavior during transitions. Third, the framework requires simple calculations of the posterior estimates.

APPENDIX C

PROGRAMMING CODE

For drawing an HPD credible interval, the following program code was developed using Visual Basic for Application (VBA) in Excell. This program was run to plot the corresponding graphs for three cases, as discussed in Chapter 5.

```
Dim i, j, t1, t2 As Single

 n = Val(NoData.Text)
 Range("B2:B" & (n + 1)).Select
 ActiveSheet.Shapes.AddChart(xlXYScatter, 350, 50, 500, 300).Select
 ActiveChart.ChartType = xlXYScatter
 ActiveChart.Legend.Delete
 ActiveChart.Axes(xlCategory).MaximumScale = n
 ActiveChart.Axes(xlCategory).MajorUnit = 1

 L = ActiveChart.PlotArea.InsideLeft
 T = ActiveChart.PlotArea.InsideTop
 W = ActiveChart.PlotArea.Width
 H = ActiveChart.PlotArea.Height
 Mx = Cells(2, 2)
 Mn = Mx
 For i = 2 To n + 1
    For j = 2 To 6
       If Cells(i, j) > Mx Then
          Mx = Cells(i, j)
       ElseIf Cells(i, j) < Mn Then
          Mn = Cells(i, j)
       End If
```

```
   Next j
Next i

ActiveChart.Axes(xlValue).MinimumScale = Mn
ActiveChart.Axes(xlValue).MaximumScale = Mx
ActiveChart.SetElement (msoElementPrimaryValueGridLinesNone)
Dim S1 As String
If Val(DecPlaces.Text) > 0 Then
   S1 = "0."
   For i = 1 To Val(DecPlaces.Text)
      S1 = S1 + "0"
   Next i
Else
   S1 = "0"
End If
ActiveChart.Axes(xlValue).TickLabels.NumberFormat = S1

Dim X1, X2, Y1, Y2, Lt As Single
Dim W2 As Single
W2 = ActiveChart.Axes(xlValue).Width

For i = 1 To n
   X1 = (W - W2) / n * i + W2 + 2 + (i - 1) * 0.45
   If n >= 10 Then X1 = X1 - (i - 1) * 0.45
   X2 = X1
   t1 = Cells(i + 1, 3)
   t2 = Cells(i + 1, 6)
   Y1 = (1 - (t1 - Mn) / (Mx - Mn) / 1.025) * (H - 15)
   Y2 = (1 - (t2 - Mn) / (Mx - Mn) / 1.025) * (H - 15)
```

```
 ActiveChart.Shapes.AddLine(X1, Y1, X2, Y2).Line.Weight = 1.2
   Lt = 4
   ActiveChart.Shapes.AddLine(X1 - Lt, Y1, X1 + Lt, Y1).Line.Weight = 1.2
   ActiveChart.Shapes.AddLine(X1 - Lt, Y2, X1 + Lt, Y2).Line.Weight = 1.2
   t1 = Cells(i + 1, 4)
   t2 = Cells(i + 1, 5)
   Y1 = (1 - (t1 - Mn) / (Mx - Mn) / 1.025) * (H - 15)
   Y2 = (1 - (t2 - Mn) / (Mx - Mn) / 1.025) * (H - 15)
   ActiveChart.Shapes.AddLine(X1 - Lt, Y1, X1 + Lt, Y1).Line.Weight = 1.5
   ActiveChart.Shapes.AddLine(X1 - Lt, Y2, X1 + Lt, Y2).Line.Weight = 1.5
   ActiveChart.Shapes.AddLine(X1 - Lt, Y1, X1 - Lt, Y2).Line.Weight = 1.5
   ActiveChart.Shapes.AddLine(X1 + Lt, Y1, X1 + Lt, Y2).Line.Weight = 1.5
Next i
```