PART ONE

Chapter I

Introduction

1.1 Background

Nepal is primarily recognized as an agrarian country in the south Asian region, since 67% of the country's population work on land and produce agricultural products (CBS, 2002). The whole population depends on agriculture for sustenance and even those people who could have had any job are also indulged in agriculture. It is the most important source of foodstuff, which avails food not only to human beings but also to animals. In Nepal, 50% of export products come from agriculture (Adhikari, 2002). The major source of raw materials needed for Nepalese industries is also the agricultural product.

Establishment and enhancement of agriculture-based industries is supposed to be vital for the sustainable development of the nation (K.C., 2002). Agro-industry, which plays linking role in chaining the agro products from producers to ultimate consumers, is the essential module to uplift the *life standards* of farmers, and to ensure the *quality of life* of the consumer (Bourlakis and Weightman, 2004). It is considered as an important concern for the national planning too.

Agro-food production system is based on different components, such as facilities of irrigation, better seed, fertilizer, agricultural equipments, agricultural loan and transportation, food processing industries, product market and product quality assurance. Farmers, food processors, retailers, and consumers are considered as the main stakeholders of agro-food production system. Governmental departments, research institutions and non-governmental organizations are engaged as essential supplementary institutions of this system. Agro-food production system can be recognized in either physical aspects or technological, economic, biological, sociological, political, health and environmental aspects.

Agricultural products are generally consumed in two ways: straight after the post harvest treatment of the product from the produced field or product supplied by the processing industries. The quality and safety attributes of the foodstuff may be affected far and wide *farm to fork* (FAO and WHO, 2002).

Today's global concern, the subject of consumer concern in food production and quality control system, possibly has direct or indirect relation between / amongst the attributes of the sectors of farmer, food processor, distributor, retailer and consumer. Till date, producers are not having pressure-feel from large food processors, and the processors not from retailers and consumers, who require better practices in relation to food hygiene and safety, environmentally managed and naturally sustained resources. In this context, the study on modelling uncertainty related to the food production and quality problems is considered to be of great importance. The statistical methods concentrated on the research problem are supposed to be essential to explore, establish, verify and to model the relations of the attributes of concerns of food production and quality control system.

1.2 Statement of the Research Problem

Food quality system may be taken as a vital question for both, consumers and producers nevertheless, food quality crises often occurred all over the world. The inherent uncertainty and variability could evidently be shown for food products and food processing procedures. As a result, statistical control systems are needed, in order to effectively control food quality, predict potential problems and give suggestions for realistic control.

The expert knowledge is often considered as the primary source of information for the food production and quality system. The development in information systems in food industries also provides us with possibilities to discover valuable information about quality problems from recorded data. However, due to the complexity of food production system, experimental data are not always considered sufficient to deal with new quality problems in a direct way. These problems can be dealt with the

help of Bayesian approach, combining experimental data with prior information about the parameters from expert knowledge or from past data.

Bayesian methods are found to be successfully applied in many areas, such as biology, medicine, finance, economics, and marketing. So far, Bayesian approaches on modelling uncertainty in food production and quality control system are not being used widely due to lack of documentation system and supply networks. However, in recent years, some food industries have initiated to build information systems to collect data about various stages of food production, quality monitoring and supply networks. These information systems will be helpful to provide us with opportunities to employ Bayesian methods for discovering interesting relations of food quality problems. Bayesian analysis and inferential method are used in this research, expecting that, it can motivate us to recognize new methods in food quality control, aid in discovering possible causes for these problems, and monitor those causal factors to predict potential food quality problems.

1.3 Rationale of the Research

Nepal has quite a few traditional small food-processing industries and some modern industries (K.C., 2002). According to Nepal Standard Industrial Classification (CBS, 1998), there are 661 such industries. Essentially, in pre-harvest sectors of agriculture, various studies have been made in cropping pattern, agricultural practices, breed development, insecticides use, and advancement of the indigenous food. However, so far, no sufficient studies have been made in post-harvest systems of food production and quality control in methodical and statistical modelling approaches. Statistical modelling approaches have not found been used in agro-food production system, post harvest processing, food industrial environment, and consumer risk assessment. A little number of researches on specific areas is sponsored by some international non-governmental organizations such as WFP and WHO.

Food materials are prepared usually as the mixture of different constituents; as a result, they have complex nature. The variability in food attributes is found due to the causes of raw materials, biological diversity, resources, processing method, and vendor system. Because of the wide range of variability, food quality problems have high degree of uncertainty, and constraints (parameter of interest) related to food are uncertain (Martens, 1983).

The theoretical understanding of uncertainty and variability is found rare in food quality problems. Therefore, we have to rely on other sources of knowledge to model the uncertainty. Probabilistic approach is considered the only suitable method to deal with uncertainty. To obtain reliable information about system the prior knowledge is incorporated with the data and this approach provides the basis for applying Bayesian methodology on food quality control problems.

The problems concerning the consumers, food science activities concentrate to the food quality and safety (Hills, 2001). The fundamental issues of food science may be nutritional value, shelf life, safety, uses property and packaging. Unknown quantities of interest in quality control experiments might be industrial safety, utility factor, product component, future outcome etc.

Research activity on industrial sector is almost non-existence in Nepal, (Bajracharya, 2001). In food science activities, there is a lack of making inference statistically. To deal with such problems, mathematical model is much needed, which can help the food scientists to predict and to control the food quality. Such models are very useful to model hazard rate, life-testing, failure to specification, shelf life, reliability estimation, capability measure, and confidence bounds. Models based on probability are essential when variability and uncertainty characterized such food quality attributes.

The classical statistics is used to obtain point and interval estimates of the population parameter and to test the hypothesis without combining prior information. While, Bayesian statistics is capable to use the posterior predictive level keeping informed the priors; and model discrimination is more applied than it has currently done in the

food science (Van Boekel, 2003). In this approach, uncertainties about parameters of food quality system are quantified according to probabilities, and then they are updated as information gathered from the experiment.

Prior information is often available for food quality control from the earlier studies. These studies can be used as prior information because the mechanism of action of quality control devices is typically physical, making the effects local (limited) and not systemic (general). Local effects are often predictable from prior information when adjustments to a mechanism are minor.

Product inspections and compositional analyses of the product in a manufacturing system are done either by destructive method or by non-destructive method of analysis. A small size of sample is preferred for application of destructive method of analysis. Sample having small size is not sufficient to give precise conclusion in classical method. In this context, we seek to employ the method, using which a small sized sample can give a precise conclusion. In case of Bayesian method with hierarchical priors, it is possible to obtain precise result with small sample size. It is expected that it could save the loss due to inspection using destructive method.

1.4 Research Questions

The main question of this research is how quality concerns in food production and quality system can be modelled; the relations are set up and they are validated using statistical method. It involves making inference, modelling in probabilistic or stochastic means and checking the proposed model fit.

The study will attempt to explore the queries about:

1. What are the distributions of the variables selected for the study related to quality problems in food production process?
2. What are the statistical models applicable to express relations between such variables?
3. How application of modelling can be done for the selected factors / variables using Bayesian methodologies?

1.5 Review Practicalities

Hawthorn *et al.* (1984) discussed different statistical methods that can be used in agro industry sectors. Hills (2001) stated how the modelling approach is most exiting and influencing in food process industries and how it provides the foundation for improvement in food quality. Mitasova and Milas (1998) explained how process modelling is aimed at improving our understanding and predicting the impact of natural and socio-economic processes and their interactions. Duncan (1970) remarked that the estimation of lot and process characteristics could be successfully obtained if a prior probability of outcome in a phenomenon is known.

Lunning (2002) focused on consumer-driven quality management in food production systems using a product-based approach, which integrates organizational and technological aspects of food product quality into one techno-managerial concept and presented an integrated view of quality management. Hubbard (2003) emphasized on the use of fundamentals of Statistics, design of experiments, statistical quality control, six sigma techniques and net control of products and processes. Bourlakis and Weightman (2004) explained in detail the management of the food supply system from food produced on the farm to food industries, wholesale and retail markets and consumers.

For many attributes of food product, a certain effect is considered for each factor influencing quality. In order to model the effect of those factors, quantitative models are needed, when variability and uncertainty characterize the food quality attributes. Van der Vorst (2005) has provided the ways to quantify quality attribute of food products on performance measurement systems.

In advanced food industries, the data about food production and quality control are recorded for the statistical point of view. These data are applicable to maintain modern quality trends, like, Total Quality Management (TQM) (Barendsz, 1998), Hazard Analysis and Critical Control Point (HACCP) (Horchner et al., 2006). These data are also put into operation in drawing statistical control charts, obtaining process capability and exploring the causal relations on the subject of process and

product characteristics. Knowledge about these relations provides a possibility to prevent problems by monitoring and practically controlling the corresponding factors. Therefore, it is meaningful to employ most powerful statistical methods to identify new problems and discover causal relations.

Presence of an uncertainty in the real world observation is considered as the heart of the statistics and modelling it in probabilistic manner. In statistics, there are two main ways of dealing with an uncertainty and expressing the reliability of the result in precise way. The classical approach is used for modelling the phenomena solely based upon the data and other approach is used for modelling it combining uncertainty concerning population parameter as the prior information with data. The posterior probability estimate based on prior information is said to be the Bayesian estimation.

O'Hagan (2003) has suggested the method of Bayesian inference to create a model to link data to parameters, formulate prior information about parameters, combine the two sources of information using Bayes' theorem, and use the resulting posterior distribution to derive inferences about parameters. The approaches of the Bayesian inference are found in several books and papers; its theoretical utility is taken in Lindley (1971).

Prior information needs to be elicited to conduct Bayesian analysis; if there is prior information about the hypothesis, clearly, we need to incorporate it in the analysis (Berger, 1985). Fully Bayesian approach covers Empirical Bayes' Method, Hierarchical Bayes, MCMC (Markov Chain Monte Carlo), and Hierarchical Bays' Model (ISBA, 1992). Priors such as reference priors, default priors or objective Bayes priors, reference prior algorithm are used for modelling (Bernardo & Smith, 1994).

1.6 Structure of the Thesis

The thesis is primarily divided into two parts. The first part has four basic chapters. The first chapter deals with the introduction of the study. The second chapter contains the objectives, major issues, sources of sample and data and the delimitations of the study. The third chapter is the literature review with reference to the study. A review of the fundamental principles and scholarly work concerning quality, Bayesian methods and computational methods are presented in this chapter. The first section of the third chapter is review of the quality and quality concerning works and the second section to the review of Bayesian method. The third section of this chapter is related to the brief introduction of the computational programs. The fourth chapter deals with the methodology concerning the research. The first section of it is related to the Bayesian methodology for the inferential procedure and second section is in relation to different models concerning the quality control aspects.

The second part of the thesis is compiled with the five contributory research articles corresponding to the objectives of the study. It consists of fifth chapter to ninth relating to the major contribution about research data and model assessment. The last chapter of the thesis is for the conclusions and recommendations for the further areas of the research.

Chapter II

Objectives

2.1 Objectives of the Study

The broad objective of this study is to select some consumer concerns of food products and quality, and to set up and validate the relations between the variables using statistical methods. It involves making inference by prediction, modelling in probabilistic means, obtaining the advances on inferences and quality control method and checking proposed model on some aspects of consumer concern quality.

The following are the specific objectives of the study:

1. To explore statistical models applicable to express the relations between variables relating to some concerns of food production and quality system.
2. Modelling the factors selected for the study in Bayesian point of view
3. To express the predictive values of the parameters in updated posterior density and
4. To check the fitted model using Bayesian approaches

2.2 Major issues and Variables

One of the major analytical issues is identification of prior distribution using past data or the expert knowledge related to food quality problems. If nothing is known about the prior, a prior (non-informative) is selected which has a negligible effect on the data. Another issue is selection of conditional cases and likelihood function of the factors to be studied for the data borrowed from experimental results and to express it in Bayesian framework.

The practical issues of the study are consumer concerns of food quality. The major characteristics studied in this work are: weight of the pouch product, the proportion of non-conforming items, modelling the uncertainty related to failure in operation

points (identification of CCP), predicting the product quality of rice by the evaluation of porosity, and estimation of gluten content for the better product quality of wheat flour in terms of protein content.

The variables of the study are consumer concerning characteristics of the food production and quality control system. For the statistical analysis and modelling purpose, the physical characteristics of food, quality attribute of the product, failure points of quality, compositional quality variables are selected as main variables.

2.3 Sources of Sample and Data

The main sources of data for this research are documented data, the data from the evaluation of processing units; system maintained data and practical field/ lab data. The data have been collected from different sources for the purpose of analysis, drawing inferences and fitting model. The procedure of data collection is based on the industrial visit, literature survey and appraising documents of the quality management. No specific sampling procedures have been adopted in data collection. The food industries were selected in convenience, and data were gathered in personal contact with the management and quality personnel. Some of the data have been taken from the works of the researchers. Collected data have been used to set up Bayesian modelling and Bayesian inference.

Dairy Development Corporation, Balaju, Kathmandu; Himalayan Snacks, Banepa, Kavre; Rijal Tasi Industry, Itahari Sunsari; Mahalaxmi Flour Mill, Sonapur, Sunsari, are the major contributors of data. The dissertations, seminar papers and project reports of the students of Central Campus of Technology are the other sources of data.

2.4 Delimitations of the Study

Food product and its quality has a large number of the subjects, such as food values, preservation periods, safety, uses property, packaging, adulteration, pesticide used, hazard rates and reliability estimation etc. Also, there are hundreds of food items used in our daily life. In this study, an attempt is made only to use Bayesian methodologies to analyze new data of some concerns food quality. The study is focused only with application of Bayesian modelling on some issues of consumers primary concerns related to agro-food production and quality control data.

The study is delimited to some aspects of quality such as weight of the product, the proportion of non-conforming items, identification of critical control point (CPP) in canning process, quality of rice in terms of porosity, and gluten content of wheat flour in terms of protein for the better product quality.

Chapter III

Literature Review

3.1 Quality

3.1.1 Introduction

Quality, especially the food quality, is known as an opinion term, based on consumer's perception concerning a product or a service. It does not solely depend on maintaining the written standards or specification indicated. The judgment of the quality of a product decidedly depends on the response of the consumer. It is defined in terms of meeting consumers' requirement over the conformance to the specifications.

Based on this viewpoint, Crosby (1979) defined quality as 'conformance to requirements'. By using customer loss function, Taguchi (1986) emphasized on customers' requirements. Deming (1986) refused to accept everything outside the specification always wrong and inside the specification entirely correct. For a given acceptance quality level, Mendenhall and Sincich (1995) described both the consumer's risk and producer's risk using operating characteristics curve. English (1999) defined quality as the 'meeting customer's expectation consistently and not necessarily exceeding them'.

The quality issues in food sector have been gaining importance over the past few years. Food industries are able to exercise great power and influence over society. They can act as an agent of changing social practice and cultural food habits. With increasing competitiveness in the food market, they are trying to use development of science and technology to create new food products, to maintain food quality, to revolutionize consumption pattern and many other issues arisen in terms of production and quality control. Food producers, food quality personnel and food industries are continually searching for methods and means, which permit production of goods that meet as much as possible the characteristics demanded by the market.

Quality assurance methods and techniques can provide a useful tool for approaching evolving markets in the correct way. Application of statistical quality control analysis to monitor analytical methods in laboratory help a manufacturer improve the reliability of reporting and identifying the source of problems more quickly.

3.1.2 Principles of Quality Control / Management

Various established quality principles have been invented by different quality pioneers. Crosby (1979), Shewhart (1986), Deming (1986), Juran (1988), and Ishikawa (1988) presented some of the renowned quality philosophies. In their view, quality is not intangible, it is controllable or manageable. Taguchi (1987) advocated the use of designed experiments to improve measurement and calibration systems. Deming (1986), Juran (1988), Crosby (1979), Ishikawa (1986), Shewhart (1986), Imai (1989 & 1997), English (1999) have given the methodical way of quality management/control. The majority of their efforts are on the customer focus, continuous process improvement and scientific methods. Meeker and Escobar (1998) and Condra (1993) focused on the reliability modelling of the quality issues. Meeker and Escobar define the reliability in terms of quality as the probability that a product or subject will perform its intended function under operating conditions, whereas Condra emphasizes the reliability to be quality over time. They comprehend that the measurement of the quality based on the massive amount of data or information need to be simultaneously collected and stored as the production process is running. This leads every body to be acquainted with statistical procedure with sufficient information in data form while managing quality.

Young and Guess (2002) explained how such data are stored and used in a real time database with regression modelling to predict strength. They emphasized that carefully devised data preparation can guard the analyst against miss-specified model assumptions and consequently incorrect estimates. Meeker and Escobar (1998) stated that there are often many practical cases where a better fit of the data are from non-normal distributions, however the normal distribution is often assumed for many applications during the quality improvement process.

The technique of quantitative risk analysis has been developed to more accurately represent risk in systems composed of a network of interacting factors. Vose (1996) remarked that the key difference in quantitative risk assessment as 'it attempts to take into account every possible value for each variable and weights each possible state of affairs by the probability of its occurrence'. Notermans & Mead (1996) proposed that techniques of quantitative risk assessment should be incorporated into HACCP systems. Vose (2000) explained a number of mathematical techniques such as Monte Carlo simulation and other numerical methods that have been developed for the purpose of quantitative risk analysis. The view of Vose (1996) on the subject 'to take into account every possible value for each variable and weights each possible state of affairs by the probability of its occurrence' leads us to the use of Bayesian approach in the quality control system.

3.1.3 Review of Bayesian Approaches in Food Quality Control

In Bayesian method, Corney (2000) explained the method of designing food with Bayesian Belief Networks. Van Boekel (2003) gave an overview of some typical food-science problems. He emphasized on the use of Bayesian approaches in food quality modelling, modelling food quality change and food safety, product design and classification of foods. Fearn (2003) persuaded to apply Bayesian methods on many unexploited opportunities in the agro-food production chain. Stein (2003) provided Bayesian network technique, which serves as a model for networks occurring in food security, and the relations in the model reflect causal impact between events. For the design and food supply system, Van Beek (2003) suggested to use quantitative modelling. Van der Voet and Paulo (2003) discussed on some explorations into Bayesian modelling of risks due to pesticide intake from food. FAO / WHO (2000) provided the regulations for the food contamination monitoring and food borne disease surveillance. According to Campbell (2000), FDA has embarked on an initiative to investigate how Bayesian design and analysis can be used effectively. Campbell referred to follow www.fda.gov/cdrh/pdf/p970015b.pdf, and www.fda.gov/cdrh/pdf/p970033b.pdf. Pennello (2006), [on behalf of FDA], presented a guidance paper for using Bayesian methods in planning clinical trials.

FDA (2006), itself, proposed a manual for using Bayesian analysis in clinical trials. Barker (2003) explained on application of Bayesian Belief Network models to food safety science. Kennedy *et al.* (2009) demonstrated some of the potential uses of Bayesian approaches in quantitative microbiological risk assessment to integrate information sources for food.

3.2 Bayesian Statistics

3.2.1 An Introduction to Bayesian Statistics

Bayesian statistics is an approach to data analysis that provides a rational method for learning from evidence as it accumulates. Classical methods use the prior information only in the design stage; while Bayesian method uses the prior information not only in the design stage, but utilizes it as the vital part of the analysis as well. The basic idea in Bayesian statistics is that one's uncertainty about an unknown quantity of interest is represented by probabilities for possible values of that quantity.

The uncertainties concerned in the statistical problems are dealt comfortably using *probabilistic models*. In statistical analysis, probabilistic models are i) based on sampling re-sampling empirical data, ii) based on parametric modelling of data, and iii) based on parametric modelling of data with prior information. The third type of modelling, '*parametric modelling of data with prior information'* is known as the Bayesian modelling approach.

The term 'Bayesian' is linked to the emergence of statistical thinking some 250 years ago. The name Bayesian is going to be used in statistical procedure from the mid 20^{th} century (1950s- 60s). The name is given after the name of Reverend Thomas Bayes (1702-1761); in whose name, we have the Bayes' theorem in probability, posthumously published (in 1763) a paper *'An essay towards solving a problem in the doctrine of chances'*.

Bayesian paradigm states that probability is the only measure of one's uncertainty about an unknown quantity. Within this paradigm, parameters are treated as random variables. It is just a description of their uncertainty, but not variability. Bayesian approach to statistics firmly based on axiomatic foundation, which provides a logical structure and assure the mutual consistency of the methods proposed (Bernardo, 2003). International Society for Bayesian Statistics declares that:

> Bayesian inference provides a logical, quantitative framework for iterative process of integrating, accumulating information, assess the current state of knowledge, gather new data to address remaining questions, and then update and refine understanding to incorporate both new and old data. …… It has been applied in a multitude of scientific, technological, and policy settings. …… This approach can provide flexible methods to conduct conditional analysis and other amendment to experiments in normal line. It can be useful in complex modelling condition where a conventional analysis is difficult to implement or does not exist (ISBA, 1992).

Bayesian analyses are often computationally intense. However, recent advances in computational algorithms and many-fold increases in computing speed have made it possible to carry out calculations for almost any Bayesian analysis. These advances have resulted in a remarkable increase in the use of Bayesian methods (Malakoff, 1999). The basic tool that enabled the advances is a method called Markov Chain Monte Carlo (MCMC), (Gamerman, 1997; Gilks, Richardson and Spiegelhalter, 1996). Windows version of Bayesian Inference Using Gibbs Sampling (WinBUGS) is commonly available computer program dedicated to making Bayesian calculations (Congdon, 2003).

3.2.2 Underlying Principles of the Bayesian Approach

The basic principles of the Bayesian approach are: (i) uncertainties in the parameter of interest are expressed through the specification of probability distribution, (ii) the probability statement about a hypothesis is updated using Bayes' rule, and (iii) the updated probability is used for making decision and modelling of complex system.

3.2.3 Steps in Bayesian Methods

The following steps are used in implementing Bayesian methods.

- Formulation of prior information; selection of the prior distribution of parameter of interest
- Generating data creating a statistical model, linking data to parameters using model
- Combining the information from two sources using Bayes' theorem
- Use of the posterior distribution to derive inferences about parameters

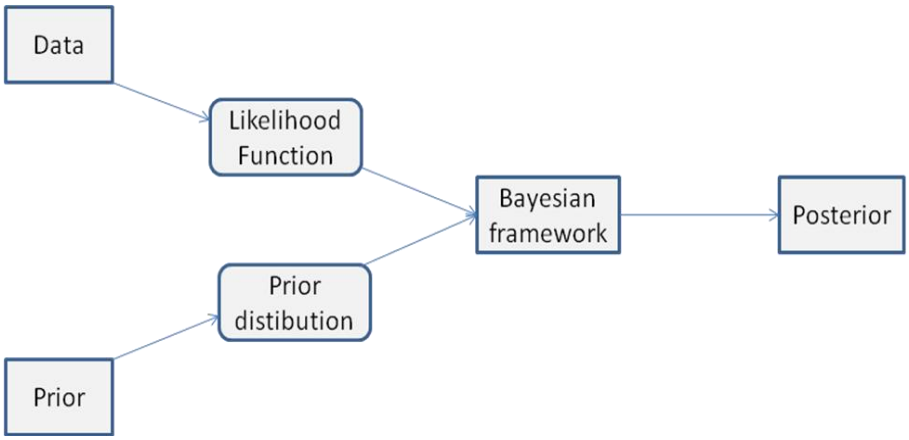


Figure 3.1: An Outline of Bayesian method (O'Hagan, 2003)

3.2.4 Mathematical Notations and Foundation

The probability of an event *E* given the set of possibilities with the sum total of data available, ***Ω,*** is denoted by $P(E \mid \Omega)$ or simply $P(E)$. Usually, ***Ω*** is used for unobservable random vectors, typically parameters.

The probability of an unknown quantity (an observable random vector) θ_i, $i = 1,2,.....,k$; given the information available *'H'* (the states of nature or hypothesis) is denoted by $P(\theta_i|H)$ or $P(\theta_i)$. Similarly, the probability of data *(X)* relevant to the values of unknown quantity under hypothesis *'H'* is denoted by $P(X|H)$ or $P(X)$; where $X = \{x_1, x_2, ...x_n\}$.

$P(\theta|C)$ stands for the general probability density of random vector $\theta \in \Theta$ under condition 'C' and $P(x|C)$ stands for the general probability density of random vector $x \in X$ under condition 'C'. So that,

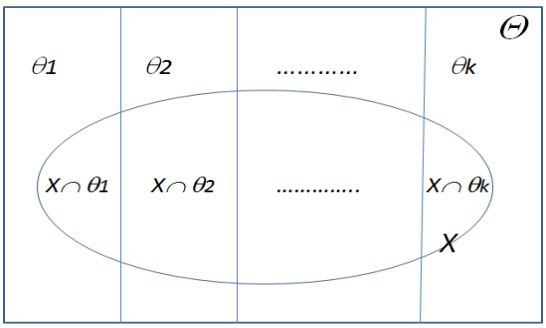
$$P(\theta|C) \geq 0 \quad \text{and} \quad \int_{\Theta} P(\theta|C)\, d\theta = 1$$

$$P(x|C) \geq 0 \quad \text{and} \quad \int_{X} P(x|C)\, d\theta = 1$$

Bayes' formula:

Let, the set of parameter of interest θ with possible outcomes $\theta_1, \theta_2, ..., \theta_k$ form a partition of parameter space **Θ,** if

$$\bigcup_{i=1}^{k} \theta_i = \Theta.$$



If $\theta_1, \theta_2, ..., \theta_k$ are mutually exclusive, and **X** is any other random event in ***Θ*** comes from the data ***X,*** then the events $\theta_1 \cap X$, $\theta_2 \cap X, \ldots\ldots, \theta_k \cap X$ form a partition of data **X**; thus,

$$X = \bigcup_{i=1}^{k} (\theta_i \cap X).$$

The probability of X is given by

$$\Pr(X) = \sum_{i=1}^{k} \Pr(\theta_i \cap X)$$

By multiplicative rule of probability,

$$\Pr(\theta_i \cap X) = \Pr(\theta_i) \times \Pr(X|\theta_i)$$

Finally, if $\Pr(\theta_i) \geq 0$ for all i, then by the theorem of total probability

$$\Pr(X) = \sum_{i=1}^{k} \Pr(\theta_i) \times \Pr(X|\theta_i)$$

The update of the prior probability $\Pr(\theta_i)$ to the posterior probability $\Pr(\theta_i|X)$ after observing data X is

$$\Pr(\theta_i|X) = \frac{\Pr(\theta_i \cap X)}{\Pr(X)}$$

or,
$$\Pr(\theta_i|X) = \frac{\Pr(\theta_i) \times \Pr(X|\theta_i)}{\sum_{i=1}^{k} \Pr(\theta_i) \times \Pr(X|\theta_i)},$$

which is known as *Bayes' Formula.*

$\Pr(\theta)$ is commonly known as the *prior probability* of a random vector $\theta \in \Theta$

$\Pr(\theta_i|X)$ is the *posterior probability* of θ_i given the data X. It represents the probability of observable random vector θ_i after the data have been observed.

$\Pr(X|\theta_i)$ is the *conditional probability* of data given θ_i. It summarizes the likelihood of X given θ_i.

$\sum_{i=1}^{k} \Pr(\theta_i) \times \Pr(X|\theta_i)$ is the *marginal probability* of X.

Generalization of Bayes' Formula :

The *posterior distribution function* (well known as *posterior density*) of θ_i given data *X*, denoted by $p(\theta_i | x)$, represents the probability of parameter θ_i after the data (*X*) have been observed. If *g(x)* denotes the marginal probability of *X*, the generalized Bayes' theorem to derive posterior distribution is

$$p(\theta_i|X) = \frac{\pi(\theta_i) \times g(X|\theta_i)}{g(X)}.$$

Therefore, the posterior density is

$$p(\theta_i|X) = \frac{\pi(\theta_i) \times g(X|\theta_i)}{\sum_{i=1}^{k} \pi(\theta_i) \times g(X|\theta_i)}, \text{ for the discrete model and}$$

$$p(\theta_i|X) = \frac{\pi(\theta_i) \times g(X|\theta_i)}{\int_{\Theta} \pi(\theta_i) \times g(X|\theta_i)\, d\theta}, \text{ for the continuous data and all } \theta \in \Theta.$$

Here, $\pi(\theta_i)$ denotes the *prior density* (pmf for discrete and / or pdf for continuous $\theta \in \Theta$). It summarizes one's belief about the probability of parameter θ_i before data (*D* or *X*) have been observed.

$g(X|\theta_i)$ denotes the *conditional probability* of *X* given θ_i. It summarizes the likelihood of data *X* given θ_i.

$g(X)$ denotes the *marginal probability* of *X* (also called *unconditional probability,* occasionally). This is equal to the sum of the quantities in the numerator for all events θ_k. It is also known as *predictive distribution* of *X*, since it represents our existing predictions of the value of *X* taking into account both the uncertainty about the value of θ and the residual uncertainty about *X* when θ is known (Lee, 1997).

Prior Distribution:

Before starting an experiment and obtaining data, we assign the probability distribution to the possible value of the unknown quantity, known as the prior distribution. In principle, the prior is based on the investigator's personal knowledge of the quantities of interest or on expert's opinion. If absolutely nothing is known

about that quantity, something called a *non-informative* prior distribution is specified. In experiments undergoing fixed review, however, the prior distribution is usually based on data from relevant previous experiment.

Posterior Distribution:

After data gathering and making use of information about parameter of interest, the prior probabilities are mathematically updated according to Bayes' theorem. The updated probabilities, known as posterior probabilities, are the probabilities for the values of unknown quantity after data observed. This approach is a scientifically convincing way of combining previous information with current data. The approach regulates to changing levels of evidence: *today's posterior probabilities become tomorrow's prior probabilities* (ibid.).

3.2.5 Likelihood Function and Bayes' Theorem

Let, X be the random vector such that $X = (X_1, X_2, \ldots\ldots, X_n)$ and x be the numerical realization (observation vector) such that $x = (x_1, x_2, \ldots\ldots \ldots, x_n)$. Let, the pdf (or pmf) of X as the realization of θ is denoted by $g(x|\theta)$ for all $x \in S$ and $\theta \in \Theta$. If we represent the prior distribution of θ by $\pi(\theta)$, then the posterior density function is

$$p(\theta|X) = \frac{\pi(\theta)\ g(X|\theta)}{g(X)}.$$

The *joint probability distribution* of the data (X) and the parameter of interest θ is given by $g(X,\theta)$. For the data point $X = x_i$ and given value θ the joint probability is $g(x_i,\theta)$. The product of the densities $g(x_i,\theta)$ for all $i = 1, 2, \ldots, n,$ is said to be likelihood function and is given by

$$L(X,\theta) = L(x_1, x_2, \ldots\ldots\ldots, x_n; \theta) = g(x_1,\theta)\cdot g(x_2,\theta)\cdot \ldots\ldots\ldots \cdot g(x_n,\theta)$$

$$= \prod_{i=1}^{n} g(x_i,\theta).$$

The conditional probability of $X = x_i$ for given value θ is denoted by $g(x_i|\theta)$. The probability function $g(X|\theta) = \prod_{i=1}^{n} g(x_i|\theta)$ is the likelihood of θ and written

as $L(\theta)$ or $L(\theta|X)$. It gives the predictions at to what the data should look like if the parameter takes place the particular value θ (ibid.).

The posterior density $p(\theta|X) = \frac{\pi(\theta)\ g(X|\theta)}{g(X)}$ can be written as

$$p(\theta|X) = \frac{1}{g(X)} \cdot \pi(\theta) \cdot g(X|\theta)$$

or, $$p(\theta|X) \propto \pi(\theta) \cdot L(\theta)$$

or, $$Posterior \propto Prior \times Likelihood$$

Taking *log* on both sides,

$$log(posterior) = log(prior) + log(likelihood)$$

The constant of proportionality is $\frac{1}{g(X)}$, known as *normalizing constant*.

$g(X)$ is *marginal (or unconditional) distribution* of data; in most of the cases it does not have the closed form, depends only on the data (*X*, and not on θ).

This shows that, posterior density summarizes the total information after viewing the data and provides a base for posterior inference regarding θ. Consequently, to arrive at the probabilistic inference about the consistency of the model with the data the prior distribution and the likelihood are essential, which is the basis of the Bayesian analysis.

3.2.6 Priors: An Overview

The Bayesian approach to learning starts with some prior knowledge or assumptions about the model structure. This initial knowledge is represented in the form of a prior probability distribution over model structures. The probability of a hypothesis, unconditional, set before observing data, is the prior distribution of the parameter of interest.

Bayesian inference relies on the marginal likelihood and prior. Prior specification is critical feature of any Bayesian study, the primary assignment, which reflects

knowledge about uncertainty. If experimental data contains sufficient signal, a prior does not greatly influence the posterior. If the posterior is highly dependent on the likelihood function (data), it may not contain sufficient prior information.

Priors are treated as an adaptive allocation device, governed by both soliciting expert opinions and by occurring efficacy data with some hypothesized statistical assumption. Different types of priors are specified for the Bayesian analysis.

Three interpretations can be given to prior distributions: (i) as frequency distributions based perhaps on previous data, (ii) objective representations of what it is rationale to believe about a parameter, or (iii) a subjective measure of what a particular individual actually believes. Categorically, the priors cannot be sorted in exclusive partitions. In general, we outline priors in broad three classes: i) Objective priors, ii) Subjective priors, and iii) Empirical-Hierarchical priors (Beal, 2003).

Objective Priors:

The prior, which has negligible effect on the posterior, and if the results are entirely based on the data then, it is an objective prior. Objective Bayesians prefer such priors. They try to suggest as little information as possible in attempt to allow the data to carry as much as possible in the posterior distribution. Often it is called "letting the data speak themselves" or "prior ignorance". Such priors generate *non-informative priors*. Non-informative priors are a bit fast for the analysis even if they have no consistent information. Non-informative priors are commonly used if we (i) do not want to influence the inference, in some particular direction, (ii) have a little expertise, (iii) have intricacy to elicit or translate expert opinions or counsels in mathematical form (iv) want the inference to be robust to misspecification of the prior.

Suppose, we have the distribution of the parameter of interest before data have been observed is $\pi(\theta)$. If it does not favour one θ value over another, then $\pi(\theta)$ is a non-informative prior.

Non-informative priors are well-known in different names such as; *diffuse prior, vague prior, flat prior, default prior, improper objective prior, natural objective*

prior, reference prior. In most of the cases, these priors are interchangeably used to denote prior distributions representing very weak prior.

Diffuse prior / Vague prior / Flat prior

When we have no clear and concentrated prior knowledge about the parameter of interest, so that we have no particular reason to believe that $\theta = a_1$ rather than $\theta = a_2$, in the neighbourhood of '*a*', then, the prior that we choose is the *diffuse prior* or *vague prior*. The distribution of such prior is flat relative to the likelihood function, so they are also known as *flat prior*.

Suppose, the prior distribution is from the uniform density defined as:

$$\pi(\theta) = \frac{1}{b-a}, \qquad \textit{for all } a < \theta < b, \quad a \text{ and } b \text{ are some constants,}$$

then, $\pi(\theta)$ is a non-informative flat prior.

Uniform prior:

When it is assumed that θ has an unknown probability between 0 and 1, which is the case somewhat against 'know nothing about θ', then the prior, which we select, is a *uniform prior*. Suppose, $\pi(\theta) = c$, $c > 0$ a constant,then $\pi(\theta)$ is a uniform prior. A uniform prior of the form,

$$\pi(\theta) = \begin{cases} 1 & \text{for } 0 \le \theta \le 1 \\ 0 & \text{otherwise} \end{cases}$$

is used as the completely ignorance of the prior. It is sometimes known (criticized) as the Bayes' postulate, but different from Bayes' theorem.

Default prior / Improper prior:

Default priors or *improper priors* are the priors, which are not suitable for computing model posterior probabilities and inappropriate to analytical point of view. If the parameter of interest *'θ'* is taken as the random variable, expresses an uncertainty, ranges over 0 to ∞ or -∞ to ∞, then the flat prior does not exist, such a prior is called an *improper prior* (Carlin and Louis, 1996).

$\pi(\theta)$ is improper if

$$\int_{\Theta} \pi(\theta)\ d\theta = \infty$$

Improper priors usually yield of non-informative priors but they are frequently used if they turn out to proper posterior distribution.

Reference prior:

If a prior distribution of the parameter of interest is not clearly recognized, and, one does not want to use a vague prior, then a prior relative to the data and having minimal effect on the posterior inference is referred, known as *reference prior*. Reference priors are non-informative; since they depend upon data and model with compared to prior beliefs, and they have trivial effect on the posterior. An essential element of the reference prior is that reference distribution only depends on the asymptotic behaviour of the assumed probability model (Bernardo, 2003).

The reference priors are widely used in the conditions: (i) when no pertinent prior information is freely available, (ii) when information is subjective (just a belief) and analysis desires an objective one, (iii) when two or more values of an unknown parameter equally stronger but not agreed arguments on prior beliefs exist. In such cases, the reference prior, a neutral prior function, is used even if it is not a probability distribution. It is as the limiting form of the posterior, a technical device to obtain a proper posterior distribution (ibid.). Some times the reference prior is called indifference prior. The reference analysis is information theoretic concept to derive appropriate reference posterior, which is based on supposed model and observed data.

Let, for some parameter of interest $\theta \in \Theta$, $g(X|\theta)$ is some probability mechanism that generates the data *(X)*, and a real valued function $g = g(\theta)$ is the quantity of interest of the model parameter θ. To obtain posterior probability distribution of the quantity of interest $g(\theta|X)$, it is necessary to specify a joint prior $\pi(\theta, \lambda)$; where $\lambda \in \Lambda$, is a nuisance parameter vector. It is required to identify the form of $\pi(\theta, \lambda)$, which has a minimum effect on the posterior distribution, so that

$p(\theta|X) \propto \int_{\Lambda} g(X|\theta,\lambda) \cdot \pi(\theta,\lambda)\, d\lambda$, is a θ reference prior.

There are different types of reference priors suggested by various research experts such as: Haar prior, Haldan's prior (Lee, 1997), exchangeable prior, dilution prior, intrinsic prior, criteria based automatic prior, spike and slab prior, (Ghosh, 2006), Zellner prior (Zellner, 1971) and K-L divergence (Kullback and Leibler, 1968). A frequently used reference prior in most of the Bayesian articles is Jeffreys' prior (Jeffreys, 1961).

Jeffreys' prior:

Jeffryes' prior is a form of reference prior, which offers an alternative way of prior computation that is invariant under transformation. The reference prior distribution of a real valued parameter that exhibit asymptotic normality in their posterior coincides with Jeffreys' prior. It is given by

$\pi(\theta) = \sqrt{I(\theta)}$, where $I(\theta)$ is the expected Fisher transformation in the model.

$$I(\theta) = -E_{x|\theta}\left[\frac{\partial^2}{\partial\theta^2}\log g(x|\theta)\right] = -E_{x|\theta}\left[\frac{\partial^2}{\partial\theta^2}\log L(\theta)\right]$$

If we transform the unknown parameter θ to $\gamma = \gamma(\theta)$, then

$$\frac{\partial}{\partial\gamma}\log L(\gamma|x) = \frac{\partial}{\partial\theta}\log L(\theta)\frac{\partial\theta}{\partial\gamma}$$

Jeffreys' prior is invariant to one to one transformation,

i,e, $\sqrt{I(\gamma)} = \sqrt{I(\theta)}\left|\frac{d\theta}{d\gamma}\right|$

Hence, computing the Jeffreys' prior for γ directly produces the same answer, instead in computing the Jeffreys' prior for θ and subsequently performing the usual Jacobian transformation to the γ scale. In the multi parameter case, the Jeffreys' prior is given by

$\pi(\theta) = \sqrt{I(\theta)}$ and

$$I_{ij}(\theta) = -E_{X|\theta}\left[\frac{\partial^2}{\partial\theta_i^{\,2}\partial\theta_j^{\,2}}\log g(x|\theta)\right] = -E_{X|\theta}\left[\frac{\partial^2}{\partial\theta_i^{\,2}\partial\theta_j^{\,2}}\log L(\theta)\right]$$

Jeffreys has suggested that arbitrariness in the choice of parameters could have no difference in the result. Jeffreys priors are improper and non-informative, but the posterior distribution obtained using this priors are proper.

Subjective Priors:

The prior that is set up in the form of previous experimental data or expert knowledge is called a subjective prior. Such prior knowledge is described by an *informative prior* distribution, since it is based upon individual expose. Informative priors are the values of some parameters or the density of the parameters that follow. Informative priors are not dominated by likelihood and have an impact on the posterior distribution. The uses of informative priors can decrease the sample size in experiment and make ease the computation. Experiments conducted in abroad, old registries / records, experimental data on very similar products, and pilot studies are possible sources of such prior information. The priors based on data from other studies are also known as *Quantitative priors*, which are easy to evaluate. The prior studies are expected to be similar to the current study in, as many as possible, the following aspects: procedure, parameters, objective, population, sites, and time-frame. There are some subjective believes, which are difficult to express in mathematical forms. In choosing a subjective prior, an analytical approach frequently used in the Bayesian inference is called the *conjugate prior* (Lee, 1997).

Conjugate prior:

For a family of distribution, a prior is said to be conjugate to the likelihood, if resulting posterior distribution is, as well, a family of that distribution. Suppose, the prior distribution is a member of distributional family $D(\alpha)$; the likelihood or the distribution of experimental data $f(y|\theta)$, and if the resulting posterior distribution $f(\theta \mid y)$ is also a member of the same distributional family $D(\tilde{\alpha})$, then the prior distribution is said to be a conjugate.

Definition:

Let, the likelihood function is $L(\theta \mid x)$ or simply $L(\theta)$ for experimental data, a class of prior distribution is said to from a conjugate family if the posterior density

$$p(\theta|x) \propto \quad \pi(\theta) \quad \cdot L(\theta)$$

is the class of Π for all x whenever the prior density is π.

To use conjugate prior, the prior is assumed as a member of some parametric family of distributions, and for the experimental data an appropriate member of the distributional family that resulting into posterior distribution of the family matching to prior distribution is identified.

Bayesian analysis with conjugate prior distribution provides a simple convenient method for combining expert judgment with observation. Conjugate priors are used mostly for analytical tractability and for sequential use of previous posterior as the new prior to the next model (Bernardo and Smith 1994). The exponential families from which likelihood function of data have a conjugate is drawn. This approach is widely available in practice.

Table 3.1 Some conjugate priors for common likelihood functions

Likelihood (Family)	**Conjugate prior**
Binomial (n,θ)	$\theta \sim$ *Beta (a, b)*
Poisson (λ)	$\lambda \sim$ *Gamma(α, β)*
Normal (μ,σ²), σ² known	$\mu \sim$ *Normal (μ_0, σ_0^2)*
Normal (μ,σ²), μ known	$\frac{1}{\sigma^2} \sim$ *Gamma(α_0, β_0)*
Gamma (α,β), α known	$\beta \sim$ *Gamma(α_0, β_0)*
Beta (a, b), b known	$b \sim$ *Gamma(α_0, β_0)*
Multinomial	*Dirichlet*
μ unknown, V known	*Multinomial Normal*
μ known, V unknown	*Inverse Dirichlet*

Hierarchical Priors (Hyper-priors):

A prior on the hyper-parameter is termed as *hierarchical prior* or *hyper-prior*. The Bayesian models with hierarchical priors can be expressed naturally using probabilistic model. Let, $\theta \in \Theta$ be a random vector representing a parameter of interest, where, $\theta = \{\theta_1, \theta_2, ..., \theta_n\}$. If it is assumed that each θ has been drawn from the same prior distribution, makes a logic in the Bayesian inference, then θ is a hierarchical prior (Bernardo, 2003).

Hierarchical parameters are useful, even, when applied only to single parameter often offering a more intuitive interpretation for the parameter's role. Hierarchical priors are often designed using conjugate prior both for analytical ease and in order that previous knowledge can be readily expressed. Hierarchical priors can be expressed naturally using probabilistic graphical model.
Let, the prior density is in the form

$$p(\theta|\gamma) = \int \left[\prod_{k=1}^{K} p(\theta_k|\gamma) \right] p(\gamma)\, d\gamma,$$

where, γ represents the hyper-parameter, θ_k is hyper-prior and $p(\theta|\gamma)$ hierarchical prior.

Each parameter θ_k is independent of given hyper-parameter, though they are dependent marginally. *Empirical Bayes* method refers the practice of optimizing hyper-parameter γ of the priors so as to maximize the marginal likelihood of the set $p(x|\gamma)$, corresponding to the shape and scale of the prior.

3.2.7 Some useful Terminologies in Bayesian Analysis

Kernel:

Let, a function $g(X)$ be a density function such that $g(X)=C\cdot k(X)$, for some *C*, which does not involve *x*, then $k(X)$ is said to be *kernel* of the function $g(X)$. The *kernel* of the conditional density function X given θ is $k(X)$, if $g(X|\theta)=C(\theta)\cdot k(X)$, for some $C(\theta)$ which does not involve X. The purpose of $C(\theta)$ is to make the density function integrate to one. In deriving posterior density, by omitting the constants, the use of kernel makes computation much easier to follow. Direct computation is used if recognizing kernel is difficult.

Sufficiency:

An estimator is said to be sufficient for a parameter $\theta \in \Theta$ when it contains all the information in the sample about the parameter. A sufficient estimator ensures that all information of sample can furnish with respect to the estimation of a parameter is being unlisted. If the likelihood function can be expressed as the product of two function such that one of them does not contain θ, then the estimator is said to be sufficient.

Let $X = x_1, x_2, \ldots\ldots, x_n$, be a set of sample from family of distribution $G_\theta : \theta \in \Theta$. A statistics $T = t(x)$ is sufficient for θ or for the family of the distribution $G_\theta : \theta \in \Theta$, if and only if the conditional distribution X given $T = t(x)$ does not depend on θ.

Let, the likelihood function

$$L(\theta)=\prod_{i=1}^{n} g(x_i,\theta) \text{ is possible to write in the form}$$

$$L(x_1, x_2,.....x_n;\theta)= L_1(t_n,\theta)\ L_2(x_1, x_2,.....x_n),$$

where, the second function of the right hand side does not contain θ, then the estimator $T = t_n$ is said to be sufficient for θ. T may be a real valued or a vector valued function.

In other form, if

$$g(x|\theta) = g(t|\theta)\ g(x|t)$$

then, t is a sufficient statistics for θ given x.

The statistics $t = t(x)$ is sufficient for θ given x if there exist non-negative functions h and f such that the likelihood function may be factorized in the form

$$g(x|\theta) = h(t,\theta)\ f(x).$$

It follows that t is sufficient for θ if and only if

$$L(\theta|x) \propto L(\theta|t)$$

whenever, $t = t(x)$ and constant of proportionality does not depend on θ.

In case of model fit in generalized exponential family, a sufficient statistics always exists (Lindley, 1970).

The role of sufficiency is important in the Bayesian computation while conjugate prior distribution is used. If t is sufficient, the posterior distribution of θ depends only on the data X through $t(x)$, and may be directly computed in terms of $p(t|\theta)$, so that,

$$p(\theta|x) = p(\theta|t)\ \propto\ p(t|\theta)\,p(\theta).$$

Maximum Likelihood Estimator:

The maximum likelihood estimate (MLE) of the unknown parameter, $\hat{\theta}$, is the value of Θ corresponding to the maximum of likelihood, $l(\theta|x)$, where x is a given vector of observation. MLE is the value of Θ that the 'most likely' to produce the data x. The likelihood of θ given x, denoted by $l(\theta|x)$, is equivalent to a function $g(x|\theta)$.

Some times, it is assumed that Bayesian analysis is natural extension of maximum likelihood (Carlin and Louis, 1996). Many of the desirable features of maximum likelihood are large sample properties, which Bayesian analysis is exact for small samples. In Bayesian analysis, the MLE is used as a point estimate. The mean of the posterior distribution is an estimator that, in classical terms, is equivalent to MLE.

The method of maximum likelihood is widely used for estimating and hypothesis testing. When the sample size is very large, the posterior distribution in Bayesian analysis will be concentrated around the MLE. In case of large sample size and weak prior information, MLE can be used to approximate posterior mean.

Exchangeability:

Exchangeability is considered as a key idea in statistical inference, in general, but it is particularly important in the Bayesian approach. Two observations are exchangeable if they provide equivalent statistical information. Thus, two items randomly selected from a particular population can be considered exchangeable, if the items in a trial are exchangeable with the items in the population for which the device is intended. If the items are ready to act as exchangeable, then the trial can be used to make inferences about the entire population. Otherwise, the trial tells us very little about the larger population. The concept of a *representative sample* can thus be expressed in terms of exchangeability.

Exchangeability may depend on the statistical model used. If for example, the undesirable event rate for a particular output depends on the any index then the outputs are exchangeable *conditional on* that index (Bernardo, 2003). That is, two items will provide equivalent statistical information, but only after, we account for differences in index. Therefore, any discussion of exchangeability should also include a discussion of the statistical models used.

Exchangeability is also visualized in terms of trials. Two trials are exchangeable if they provide equivalent statistical information about some *super-population*. Again, the trials may be exchangeable, but only after the justification the other factors with the appropriate statistical model (ibid.).

The use of Bayesian hierarchical models enables us to combine information from different sources that may be exchangeable on some levels but not on others. If trials are exchangeable, then Bayesian hierarchical models enable us to make full use of the information from all the trials (Bernardo & Smith 1994).

Predictive Distribution:

The Bayesian approach is applied for the derivation of a special type of posterior probability; namely, the probability of future events given outcomes that have already been observed, called the *predictive probability*. The probability distributions for all possible values of future outcomes are called the *predictive distribution*. Predictive distributions have many uses such as: i) determining when to stop a trial, ii) helping a investigator make decisions by predicting the outcome, given the observed outcomes of past trial, iii) predicting an outcome from a validated substitute, iv) adjusting trial results for missing data and v) model checking (Lee, 1997).

Predictive distribution is marginal likelihood for the next observation when posterior distribution is used to marginalize, elsewhere, θ. Let, $X = \{x_1, x_2, ..., x_n\}$, $x_i \in X$ be a set of exchangeable observation and it is desired to predict the value of a future observation $x_{n+1} \in X$, generated by the same random mechanism that has generated the data; then $p\left(x_{n+1} \middle| X = x_1, ..., X = x_n; \Theta\right)$ or simply $p\left(x \middle| X, \Theta\right)$ is the predictive distribution describing uncertainty on the value that x will take, given the information provided by X and other knowledge Θ.

If $\pi(\theta)$be a prior density on the value of θ, the probability model (data)

$$g\left(x \middle| \theta, X\right) = g\left(x \middle| \theta\right)$$

then, the *predictive density* of the future observation given the post observation is

$$h\left(x \middle| X\right) = \int_{\Theta} g\left(x \middle| \theta\right)\ p\left(\theta \middle| X\right)\ d\theta .$$

Also,

$$h\left(x_{n+1} \middle| x\right) = \int_{\Theta} h\left(x_{n+1} \middle| \theta\right)\ p\left(\theta \middle| x\right)\ d\theta$$

is the predictive distribution for x_{n+1}, which summarize the information concerning the new observation given the likelihood, the prior and the data.

Intrinsic Discrepancy:

Intrinsic discrepancy is the very general measure of the divergence between the two probability distributions. It is practical to define the useful convergence property of the distributions. Intrinsic discrepancy is defined as the measures of fundamental inconsistency in between two model selected. In Bayesian analysis, the selection of model is considered as a creditable job as the selection of prior; and intrinsic discrepancy provides a measure the disagreement between the models (Bernando and Smith, 1994). A common measure of the intrinsic discrepancy between two density functions $f_1(x)$ and $f_2(x)$ for some data $x \in X$ is defined as

$$\delta\{f_1, f_2\} = \min\left\{ \int_X f_1(x)\log\left(\frac{f_1(x)}{f_2(x)}\right)dx, \int_X f_2(x)\log\left(\frac{f_2(x)}{f_1(x)}\right)dx \right\}$$

Let, $g = g(X|\theta_1),\ \theta_1 \in \Theta_1$ and $h = h(X|\theta_2),\ \theta_2 \in \Theta_2$

are two models given the family Θ_1 and Θ_2 respectively, then the intrinsic discrepancy between models g and h is given by

$$\delta\{g, h\} = \min_{\theta_1 \in \Theta_1, \theta_2 \in \Theta_2} \delta\{g(x|\theta_1), h(x|\theta_2)\}$$

3.3 Computational Methods in Bayesian Paradigm

3.3.1 Introduction

Simulation is known as the technique to observing a real system in operation. It allows us to collect pertinent information about the behaviour of the system by exceeding a computerized method. Collected data are then used to design the system simulation. Simulation is not viewed only as an optimization procedure, rather, a technique for the estimation, and a measure of performance of the modelled system. The simulation methods, frequently used in Bayesian computations are Markov Chain Monte Carlo, Gibbs Sampler, Monte Carlo method, Kalman filter and Bootstrap and algorithms commonly used are Metropolis, and Metropolis-Hastings algorithms.

3.3.2 Markov Chain Monte Carlo

Markov Chain Monte Carlo (MCMC) is a general method for the simulation of stochastic processes. MCMC is a class of algorithm for sampling from probability distribution constructing Markov chain that has the desired distribution as its equilibrium distribution. Basic idea of Markov chain Monte Carlo is to construct a *Markov chain* that will converge to the target distribution, and draw samples from that chain. The simulated Markov chain is such, whose invariant state follows a given probability in a high dimensional state space.

In MCMC method, samples are generated from posterior distribution to compute the desired estimates. MCMC method fits with the simulated annealing procedure and generates fair samples from probability. The generated samples are then used for system learning and verification, scientific computation and optimization, and Bayesian inference.

The MCMC method developed all the way through Markov chain principle, Monte Carlo simulation, Metropolis algorithm, Metropolis-Hastings algorithm and Gibbs sampling. The essentials of the MCMC method are Metropolis, Metropolis-Hastings algorithms and Gibbs sampling. The most popular basic MCMC method is called

Gibbs sampling or successive substitution sampling. It is the easiest method for simulation and fundamental engine of the WinBUGS software.

The foundation of the Markov chain Monte Carlo method is appeared in conjunction with the distinguished paper of Metropolis *et al.* (1953). In MCMC simulation, to draw the random samples from a specific probability distribution a Markov chain is designed whose long run equilibrium is that distribution. A computer language is written to simulate Markov chain and run it sufficiently enough to attain the approximate equilibrium. The state of the Markov chain is recorded and then an approximate draw from equilibrium is taken.

Initially, Metropolis *et al.* (1953) used a symmetric Markov chain; later the developments by Hastings (1970) included an adaptation to the case of non-symmetric Markov chains. The famous paper by Geman and Geman (1984) on image restoration and Gelfand and Smith (1990) on computation of marginal densities provided landmark to the statistics community showing that MCMC can be applied effectively to the Bayesian inferences.

The MCMC using Metropolis-Hastings algorithm was commonly accepted as the general method for the simulation of stochastic processes having probability densities known up to a constant of proportionality (Geyer, 1992). In his article, Geyer proposed the way of estimating Monte Carlo error using standard non parametric method on one long run of the Markov chain. Chib and Greenberg (1995) provided a detail introductory exposition of the Metropolis- Hastings algorithm, a powerful Markov chain method to simulate multivariate distribution. They have proclaimed the Gibbs sampler as the special case of the Metropolis-Hastings algorithm (ibid.). Green (1995) worked on dimension varying problems and presented the use of reversible jump Markov chain Monte Carlo Method for the time series model.

Chen and Shao (1998) developed an alternative simulation based Monte Carlo method for Bayesian analysis of constrained parameter problems to determine the properties of the desired posterior distribution and to solve the problems with

normalizing constants that naturally arises in hierarchical modelling. Chen and Shao (1999) described how to estimate credible and HPD intervals in Bayesian inference form the marginal densities using MCMC sampling algorithm. They also developed a method for HPD and credible interval estimation using importance sampling and used in Bayesian hierarchical model (ibid.).

Liechty and Roberts (2001) developed MCMC methods for analyzing both Markov and non-Markov versions of continuous-time latent models or the state space models or the hidden Markov models.

The Monte Carlo method of simulation and Metropolis-Hastings algorithm are found most attentively used for the Bayesian applications in Besag and Clifford (1989, 1991); Besag, York and Mollie (1991); Carlin and Gelfand (1991); Besag and Green (1993); Smith and Roberts (1993); Tierney (1994); Gilks, Richardson and Spiegelhalter (1996).

The detailed discussion of the Gibbs sampler, MCMC and their applications in the Bayesian forecasting process are found in Gilks (1992), Gilks and Wild (1992), York (1992), Casella and George (1992), Fishman (1995), Cowles and Carlin (1996), Carlin (1996), Tanner (1996), Gammerman (1997), Drapper (2000), Robert and Casella (2004), Berg (2004).

3.3.3 Monte Carlo Method

Monte Carlo method is recognized as a method of statistical trials. It is based on simulating statistical experiments by means of computational techniques and recording numerical characteristics obtained from these experiments. It is used to find the simple way of higher integration concerning the marginal and joint distribution. Monte Carlo simulation refers to the use of random sample to estimate the output of an experiment. The error in this method is estimated by finding the standard deviation of the quantities being simulated. Monte Carlo method is regarded as the ancestor to the present day simulation.

Initially, the Monte Carlo Method was introduced by Metropolis and Ulam (1949) dealing with a class of problems in mathematical physics. They proposed it as a statistical approach to the study of the integro-differential equation. The mathematical description given for this method was the study of a flow, which consists of deterministic and stochastic processes.

Metropolis *et al.* (1953) proposed a method of sampling (calculations) by fact computing machines used extensively for numerical problems in statistical mechanics. Hammersley and Handscomb (1964) presented the use of the proposed Metropolis method in statistical mechanics. Hastings (1970) generalized the sampling method introduced by Metropolis *et al.* (1953), with an exposition of the relevant theory, techniques of application and Monte Carlo error estimation. Using Monte Carlo method, Shao (1989) studied the approximation of the Bayesian action and its posterior expected loss. He proposed two accuracy measures of the Monte Carlo approximation.

Modern Monte Carlo method is the sampling scheme for distributions with large state spaces known up to a multiplicative constant. The theoretical overview of the method of Monte Carlo simulation is given in the Appendix C-1. Modern Monte Carlo method has two approaches: Importance sampling and Gibbs sampling.

3.3.4 Importance Sampling

Importance Sampling is a general scheme for sampling from complex distributions. It is simple method for sampling from posterior distributions in some cases. It can be more efficient than simple Monte Carlo, particularly for tail probabilities. Importance sampling also provides a solution to the question of how one can update beliefs as data in particle filtering (Griffiths, Tenenbaum & Kemp 2008).

Importance sampling is a perfect sampling in the context of Monte Carlo Sampling. Monte Carlo importance sampling has contributed to extending the Bayesian computational toolkit (Gelman & Rubin, 1992). The significance and method of

effective use of importance sampling is found in Owen and Zhou (2000). The method of importance sampling is given in the Appendix C-2.

3.3.5 Gibbs Sampling

Gibbs sampling is an algorithm to generate a sequence of samples from the joint probability distribution of two or more random variables. Such a sequence of samples is used to approximate the joint distribution or to compute complex integral. Gibbs sampling is a special case of Metropolis-Hasting algorithm, and thus an example of Markov chain Monte Carlo algorithm. The algorithm is named after the physicist J. W. Gibbs, in reference to an analogy between the sampling algorithm and statistical physics. The algorithm was formulated by Geman and Geman (1984), and called the Gibbs sampler.

The roots of Gibbs sampler found in Metropolis *et al.* (1953), further developed by Hastings (1970). The Gibbs sampling is a Markovian updating scheme. Geman and Geman (1984) developed a systematic form of application of it. It is originally introduced in the context of image processing. With the paper of Geman and Geman(1984), Gibbs sampler started to gain popularity. By revealing its potentiality in wide variety of conventional statistical problem, Gelfand and Smith (1990) generated new interest in the Gibbs Sampler. They demonstrated its applicability to complex statistical modelling.

Gilks *et al*. (1989) suggested its applications in Bayesian cluster analysis. Mack *et al*. (1990) used it for genetic linkage analysis. Carlin, Gelfand and Smith (1991) appreciated it for change-point problems. Schervish and Carlin (1992) explored in detail the general convergence conditions needed for the Gibbs sampler and other algorithms and Roberts and Polson (1990) discussed the rates of convergence. Carlin, Polson and Stoffer (1992) applied it for model selection from non-normal scale mixture densities and non-linear state space modelling.

The application of Gibbs sampling is found simple for fully-conjugate Bayesian models. Gelfand *et al*. (1990) illustrated its use in missing data, ordered means and

growth curve models, and Gelfand and Smith (1990) for hierarchical models, variance components and errors in variables.

For the non-conjugate models, computational difficulties are found while using Gibbs sampler. For the estimation of generalized linear models with random effects, Zeger and Karim (1991) dealt with non-conjugacy by rejection sampling from a normal envelope centred at the mode of the sampling density.

With many examples, Tanner (1991) presented the application of the Gibbs sampler in classical statistical inference. Gelfand, Smith and Lee (1992) illustrated the use of the Gibbs sampler for the complicated statistical calculation having impact on theory. Casella and George (1992) explained the computer-intensive algorithms of Gibbs sampler for simple and complicated cases, with examples, analytically.

Gilks and Wild (1992) proposed adaptive rejection sampling, a method for rejection sampling from any univariate log-concave probability density function, applying to a Gibbs sampling analysis of monoclonal antibody reactivity. This technique was preferred to use in particular for application of Gibbs sampling to Bayesian models with non-conjugacy.

Gelman and Rubin (1992) showed the iterative simulation methods like the Gibbs sampler and the Metropolis algorithm potentially helpful for the summarizing multivariate distributions. They focused these methods for the application in inference of the Bayesian posterior distributions in real problems. They derived their results as normal theory approximation and illustrated on a random effects mixture model for clinical data.

Thomas *et al.* (1992) initiated a program to perform Bayesian inference using Gibbs sampling called BUGS. Smith and Roberts (1993) reviewed the use of the Gibbs sampler for Bayesian computation and described the application of MCMC simulation methods in the sample-based approaches for the Bayesian inference.

Tierney (1994) outlined the practical use of Markov chain methods in simulation methodology, especially for Gibbs samples and Metropolis algorithm, for the

variance reduction techniques, for the guidelines on the choice of sample size and allocation and in exploring posterior distribution. The method of Gibbs Sampling (briefly) is given in the Appendix C-3.

3.3.6 The Metropolis Algorithm (MA)

The Metropolis algorithm is an algorithm for the simulation by means of generating Markov chain. It creates a random walk in the space of θ, which converges to a target distribution. Given a target distribution, the Markov process converges to a stationary distribution p(θ|x) that can be computed up to a normalizing constant.

MA simulates a sequence of random points θ^1, θ^2, ……. ; whose distribution converges to the target distribution. Each sequence is considered a random walk whose stationary distribution is $p(\theta|x)$. The expression of the Metropolis algorithm is given in the Appendix C-4.

3.3.7 The Metropolis-Hastings Algorithm (MHA)

The Metropolis-Hasting algorithm was widely used in physics and image restoration for the integration of complex function by random sampling. Its use in Statistics is exposed in 90's (Chib and Greenberg, 1995). Suppose, $\theta \in \Theta$ has the density p(θ),we are interested to draw samples from the true joint posterior density p(θ|x)for parameter Θ. It is useful to construct Markov chain for the sample points with some state space and equilibrium distribution $p(\theta)$. Metropolis–Hasting algorithm constructs the transition probability from $\theta^t = \theta$ to the next realized state θ^{t+1}.

Let, $p(\theta \mid x) = g(\theta)/k$, where normalizing constant k may be known difficult to compute. The Metropolis algorithm generates a sequence of draws from distribution. In MA the simulated sequence $\theta^1, \theta^2, \ldots\ldots$ is a Markov chain with a unique distribution and the target distribution is equal to the stationary distribution (Chib and Greenberg, 1995). MHA generalizes MA and the jumping distribution need to be symmetric. For the correction a changed importance ratio is computed. The expression for the Metropolis-Hastings Algorithm is given in the Appendix C-5.

3.3.8 Computer Programs used for Analysis

First Bayes:

First Bayes is a simple program intended to help the beginners with teaching and learning elementary Bayesian Statistics. It deals with quite simple and standard statistical models, with an emphasis on obtaining some understanding of how the Bayesian approach works. It is not a package for doing serious statistical analysis of practical data. Tony O'Hagan is the author of it and it might be freely copied and distributed. It can be freely downloaded from http://www.shef.ac.uk/~st1ao/. First Bayes has its own website http://www.firstbayes.co.uk/, where all shorts of information about it are given.

WinBUGS:

WinBUGS, the acronym of the Widows version of Bayesian Analysis Using Gibbs Sampling, is a powerful computer package for carrying out MCMC computations. It is computing–language oriented software compatible with windows, in which the users only need to specify the structure of the model. WinBUGS uses the MCMC methods to generate samples from posterior distribution of the specified model. The development of WinBUGS had proved valuable for the implementation of Bayesian models in a wide variety of scientific discipline and it has become key factor in the growing popularity of Bayesian methods in science (Ntzoufras, 2009). It is in widespread use for serious Bayesian analysis and has been a major contributory factor to the growth of Bayesian applications (O'Hagan, 2003). It can be freely downloaded from http://www.mrc-bsu.cam.ac.uk/bugs/ .

Chapter IV

Methodology

4.1 Bayesian Inference

4.1.1 Preliminaries

Statistical procedure are found to be used, generally, either to assist in infer some causal system or to reach effective decision. In inference, the processes are dealt for finding an approximate value of population parameter and testing hypothesis regarding the parameter observing random samples. In decision process, the appropriate course of action is selected, which has minimum loss or the risk in the preference selection.

In quality control problems, the parameter of interest is often estimated as the process of drawing conclusions. Though, the accuracy of any particular approximation is not known precisely, probabilistic statements are constructed concerning the accuracy of such numbers as found over many experiments.

In Bayesian data analysis and estimation methods, all the uncertain quantities are modelled in terms of their joint probability distribution. The key principle is found to be used to construct the joint posterior distribution for all the unknown quantities in a model given the data (sample). In parametric models, the posterior distribution contains all the relevant information on the parameters to be estimated. For the Bayesian inference procedures, Bayesian methods of point estimation, interval estimation and significance testing are found to be used. For the Bayesian decision theory Bayesian advancement in the use of utility function is found to be applied.

In this chapter, some selected estimation procedures and testing techniques have been presented for drawing inference regarding the parameter of interest. Discussions have been made about how forecast is done on the predictive distribution and how the predictive distribution is used for model checking purpose.

4.1.2 Bayesian Estimation

The statistical estimation procedure to Bayesian approach is such prime approach of inference, which incorporates reasonable expectations or prior judgments. The central idea of Bayesian estimation is to update prior information on the distribution of parameters by taking measured data into account. Bayesian estimation fully depends on the posterior distribution of the parameter of interest. Bernardo (2003) stated that the final outcome of a problem of inference about any unknown quantity is nothing but the corresponding posterior distribution.

In most of the problems, a dilemma with Bayesian estimation is the computationally complicated form of the posterior distribution. The most complicated cases are parameterization, the probability distribution of the parameters and the conditional probability distribution of the measurements for given parameter values. For the computational simplicity, MCMC method is used for generating sample from the posterior distribution to compute the desired estimates (Birkes and Dodge, 1993).

Method of Bayesian Estimation:

Given a random variable *X* with the probability density of the form $f(X,\theta)$, where θ is an element of well-defined set, having $x_1, x_2, ..., x_n$ a set of possible values of *X* and $\theta_i, (i = 1,2,.....)$, the possible values of Θ.

We assume, *X* have been generated by a conditional probability model $f(X \mid \theta)$ and the probability of the unknown but observable and random θ_i as $p(\theta_i)$.

The joint pdf of *X* and Θ is

$$g(x_1, x_2, ..., x_n; \theta) = f(x_1 \mid \theta) f(x_2 \mid \theta) ..... f(x_n \mid \theta).$$

The joint marginal pdf of Θ given $X_1, X_2, ..., X_n$ is

$$g_1(x_1, x_2, ..., x_n) = \int_{\Theta} g(x_1, x_2, ..., x_n; \theta)\, d\theta; \text{ and}$$

The conditional pdf of θ given $(X_1 = x_1, X_2 = x_2, ..., X_n = x_n)$ is

$$p(\theta \mid x_1, x_2, ..., x_n) = \frac{g(x_1, x_2, ..., x_n; \theta)}{g_1(x_1, x_2, ..., x_n)} = \frac{f(x_1 \mid \theta) f(x_2 \mid \theta) ..... f(x_n \mid \theta)}{\int_{\Theta} g(x_1, x_2, ..., x_n; \theta)\, d\theta}.$$

Use of Utility Function in Bayesian Estimation:

In decision making, an unknown quantity θ, subset of parameter space Θ, is assumed as the states of nature, which affects the decision process. A function based on probability, called a *utility function,* is used to describe decision maker's preferences in decision making process. It is expected in maximizing the expected utility. Another function, called loss function in the decision making process, is opposite of the utility function, which is desired to minimize statistically. In Bayesian decision procedure, sample information, prior information and the utility functions are commonly used for the decision.

Let, '*a'* be an action or a particular decision and '*A'* the set of all possible actions. If, θ turns out to be true the states of nature for the particular action *a,* then *loss function* $l(\theta, a)$ is sustained. The utility function is given by

$$U(\theta, a) = -l(\theta, a)$$

The most widely used family of loss function is of the form

$$l(\theta, a) = c\,|\,\theta - a\,|^b, \quad \text{where, } b, c > 0.$$

When b *=1,* the loss function is proportional to the absolute error loss (AEL) and when b *=2,* it is proportional to squared error loss (SEL). Both of these loss functions increase as the distance between θ and a increases. For large gap between θ and a, square error loss gives more penalty relatively the absolute error loss. The decision rule is to know what function $a = \delta(x)$ is taken when $X_n = x_n$ is observed, given that, for all possible set of decision rules, *D:* $\delta \in D$ and $\delta : \chi \to \mathrm{A}$.

The quality of a decision rule is measured by the risk function known as expected loss,

$$R(\theta, \delta) = E\left[L(\theta, \delta(x))\right] = \int_{\chi} L(\theta, \delta(x)) f(x\,|\,\theta)\,dx.$$

For the prior distribution $\pi(\theta_n)$ and the observed value $X_n = x_n$, the method to order the possible decision rules, Bayes' risk is given by

$$B(\pi, \delta) = E_{\pi}[R(\theta_n, \delta)] = \int_{\Theta} R(\theta_n, \delta)\pi(\theta_n)\,d\theta_n$$

Suppose, a decision rule δ^* (called the Bayes' rule) which minimizes the Bayes' Risk with respect to the given prior is

$$B(\pi, \delta^*) = \inf_{\delta \in D} B(\pi, \delta)$$

For absolute error loss

$$\delta^* = median[\pi(\theta \mid x)] \text{ , i.e. } \delta^* = m,$$

where m is such number which satisfies

$$\int_{-\infty}^{m} \pi(\theta_n \mid x_n) d\theta_n = \frac{1}{2}.$$

For squared error loss, $\delta^* = E(\theta \mid x)$. It is convenient to use square error loss, because of its closed form and suitability of numerical computation (Ferguson, 1973).

Point Estimation:

In the Bayesian inference, the process of point estimation is also treated as decision problem. A point estimate of θ is some function, $\hat{\theta} = \theta(X)$, which is assumed to be an appropriate value for unknown θ. Estimation of θ is a decision problem, and to solve this, a loss function $l(\theta, \hat{\theta})$ is specified according to decision theory. The common choice for loss function called square loss, is given by $l(\theta, \hat{\theta}) = (\theta - \hat{\theta})^2$.

The expected posterior loss function, if $\hat{\theta}$ is estimated value of θ, is a risk function associated with conjunction to $\hat{\theta}$ is given by $E[l(\theta, \hat{\theta})] = R(\theta, \delta)$.

So, $$R(\theta, \delta) = E[l(\theta, \hat{\theta})] = l\left(\hat{\theta} \mid X\right)$$

$$= \int_{\Theta} l\left(\theta, \hat{\theta}\right) p(\theta \mid X) \mathrm{d}\theta \text{ , where, } p(\theta \mid X) \text{ is posterior}$$

density of θ given data X.

The corresponding Bayes estimator θ^* is that function of data which minimizes risk function, such that $\theta^* = \theta^*(X) = E[\theta \mid X]$.

If the loss function $l(\theta,\hat{\theta})$ is the square error loss, $(\theta-\hat{\theta})^2$, then the Bayes estimator (BE) of θ, say $\hat{\theta}$, is the mean of the posterior density for θ (Robert, 2004). For the large sample size, BE for parameter and Maximum Likelihood Estimator (MLE) of the same parameter are nearly equivalent. If sample results are inconsistent with prior, the BE may be considerably different from MLE. The precision function of a Bayesian approach is the multiplicative inverse of the posterior variance. The standard error (SE) of the estimator $\hat{\theta}$ of θ is the just standard deviation of $\hat{\theta}$, $\sigma_\theta = \sqrt{V(\hat{\theta})}$. The precision of the estimator is considered more if the SE is the less. For the Bayesian estimation, the full estimator is the entire posterior density itself. The following are some point estimators based on the method of maximum likelihood related to posterior density.

Mode of the distribution given posterior: $\hat{\theta} = \max[\, p(\theta \mid X)]$

Expected value of θ given posterior: $\hat{\theta} = E(\theta \mid X) = \int_{\Theta} p(\theta \mid X)\ d\theta$

Median of the posterior distribution: the point candidate is the median of the posterior distribution, where the estimator satisfies $\Pr(\theta > \hat{\theta} \mid x) = \Pr(\theta < \hat{\theta} \mid x) = 0.5$ and hence,

$$\int_{-\infty}^{\hat{\theta}} p(\theta \mid x)\ d\theta = \int_{\hat{\theta}}^{\infty} p(\theta \mid x)\ \mathrm{d}\theta = 0.5.$$

Bayesian Interval Estimation:

By Interval estimation of an unknown parameter θ in the Bayesian viewpoint, mean to construct an interval that contains $100(1-\alpha)\%$ of the posterior probability for θ. The Bayesian interval estimation is also termed as Bayesian Confidence interval (Lindley, 1965) or Credible set (Edward et. al. 1963). Bayesian interpretation of a credible set *C* is natural; the probability of a parameter belonging to the set *C* is 1- α. A formal definition is given below.

Definition:

Assume, the set C is a subset of Θ, then C is credible set with credibility (1-α)100% if

$$P_{\theta|X}(\theta \in C) = E_{\theta|X} I(\theta \in C) = \int_C \pi(\theta \mid x) d\theta \geq 1 - \alpha$$

If the posterior is discrete, then the integral becomes sum (counting measure) and

$$P_{\theta|X}(\theta \in C) = \sum_{\theta_i \in C} \pi(\theta_i \mid x) \geq 1 - \alpha .$$

This is the definition of (1- α)100% credible set, and of course for a given posterior function such set is not unique.

Highest Probability Density (HPD) Region:

For a given credibility level (1-α)100%, the shortest credible set /region is of interest. For minimized size of the region, it should correspond to the highest posterior probability (density) area.

Definition:

The (1 -α)100% HPD credible set for parameter θ is a set C, subset of Θ of the form

$$C = \{\theta \in \Theta \mid \pi(\theta \mid x) \geq k(\alpha)\},$$

where $k(\alpha)$ is the largest constant for which

$$P_{\theta|X}(\theta \in C) \geq 1 - \alpha .$$

Geometrically, if the posterior density is cut by a horizontal line at the height $k(\alpha)$, the set C is projection on the θ axis of the part of line inside the density, i.e., the part that lies below the density.

The HPD set C minimizes the size among all sets $D \in \Theta$ for which

$$P_{\theta|X}(\theta \in D) = 1 - \alpha$$

Highest Probability Density (HPD) region is the minimum sized credible region where all the points in the region have larger probability density than all points outside. HPD regions are not invariant under re-parameterization (Bernardo, 2003). The method of intrinsic estimation and intrinsic credible region given by Bernardo (2003) are given in the Appendix C-6.

4.1.3 Hypothesis Testing and Bayes Factor

Hypothesis Testing:

In Bayesian approach, strictly speaking, hypothesis-testing procedure is not found to be applied in rigid accept/reject fashion. The posterior distributions are summarized for all parameters of interest to draw inferences. The probabilities of regions in the parameter space with respect to prior and posterior measures are evaluated and models are selected and compared.

In statistical process control, it is expected to detect drifts of the parameter of interest when process sifts to out of control situation from in control situation. For the adjustment, a corrective action is made and the hypothesis H_1 against H_0 is tested, such that:

$$H_0 : \theta \in [a, b]$$

$$H_1 : \theta \notin [a, b]$$

In case, when H_0 is not rejected the process continues to operate and if H_0 is rejected some action is taken.

In Bayesian framework, the hypothesis testing is performed using the posterior distribution of $\theta_n \mid X_n$. If the data X_n is available and c is pre-specified cut-off value, the posterior probability $P_n = P(\theta_n \in [a, b] \mid X_n)$ is calculated and the null hypothesis is accepted, if $P_n \geq c$, otherwise rejected.

Bayes Factor:

The posterior distribution $p(\theta \mid x)$ of the quantity of interest, θ, conveys a full information on θ (Lee, 1997). In Bayesian approach, the testing of the models under null and alternative hypothesis is done by using Bayes factor. Bayes factor is used as the correct way to carry out model comparison. It is the ratio of the prior predictive densities to under the compared model (Kass and Raftery, 1995).

Assuming that a simple null hypothesis $H_0 : \theta \in \Theta_0$ against the alternative hypothesis $H_1 : \theta \in \Theta_1$, then, $\Theta_0 \cap \Theta_1 = \phi$ *and* $\Theta_0 \cup \Theta_1 = \Theta$ obviously.

Thus, $p(H_0) + p(H_1) = 1.$

In Bayesian framework

$$\Pr(\theta > \theta_0) = \int_{\theta_0} p(\theta \mid x)\ d\theta \quad \text{and}$$
$$\Pr(\theta_0 < \theta < \theta_1) = \int_{\theta_0}^{\theta_1} p(\theta \subset \Theta_1 \mid x)\ dx$$

Let p_0 and p_1 denote the posterior probability that θ is in the null and the alternative hypothesis sets,

$$p_0 = \Pr(\theta \subset \Theta_0 \mid x) \quad \text{and}$$
$$p_1 = \Pr(\theta \subset \Theta_1 \mid x)$$

for prior probabilities $\pi_0 = \Pr(\theta \in \Theta_0)$ and $\pi_1 = \Pr(\theta \in \Theta_1)$.

The prior odds of H_0 verses H_1 are π_0 / π_1 and the posterior odds are p_0 / p_1

The ratio of posterior odds to the prior odds in favour H_0 verses H_1 is known as Bayes Factor (B_0).

$$B_0 = \frac{p_0 / p_1}{\pi_0 / \pi_1} = \frac{p_0}{\pi_0} \cdot \frac{\pi_1}{p_1}$$

Since, $p_0 + p_1 = 1$ *and* $\pi_0 + \pi_1 = 1$

$$B_0 = \frac{p_0(1-\pi_0)}{\pi_0(1-p_0)}$$

The decision rule is to accept H_0, if the Bayes factor is greater than 1.

The Bayes factor B_1 in favour of H_1 verses H_0 is the multiplicative inverse of B_0, $B_1 = 1 / B_0$.

Bayes factors are used as the correct way to carry out model comparison. The important theorems related to Bayes factor are given in the Appendix C-7.

4.2 Bayesian Modelling

4.2.1 Introduction to Statistical Models

In statistical procedures, a series of decisions is summarized to take actions on the results of inferences of quantities what we expect to observe. There exists high degree of uncertainty because of wide range of variability. Uncertainty under consideration is dealt comfortably by the modelling approach. It is essential to construct parametric models for represent or for sufficiently approximate the true generating mechanism of a phenomenon under study.

Models are the designed statements to predicting future events, capturing summarized trends and regularities in the observed data. A statistical model is considered as a collection of probabilistic statements that describes and interprets present behaviour or predicts future performance. Statistical models are cheaply used to describe real life problems under uncertainty (Ntzoufras, 2009). It consist of three components: the response variable Y, the explanatory variables $X_1, X_2, ..., X_p$, and a linking mechanism between the two sets of variables. The response variable Y is a stochastic part of the model because the outcome is uncertain before it is observed. The procedure is concentrated to a certain outcomes of Y and predicts a future outcome of Y. Since, Y is a stochastic variable, so,

$(Y| X_1, X_2, \ldots \ldots \ldots, X_p) \sim D(\theta)$, where $D(\theta)$ is a distribution with parameter θ.

The advantage of the models is that they impose to arrange and organize all information available in a logical way, which helps to define precisely the problem under study and facilitates exchange of knowledge. Models may be used for prediction when verified and validated, which may require data from both observation and experiments.

To describe significant dependencies among variables dependency modelling is used, whereas, to describe the causal relations between determinant factors and performance measures causation models are used (Fayyad, Piatetsky and Smyth, 1996). Probabilistic models require extensive datasets. Therefore, the strength of Bayesian approach is that they can make use of information that might not pertain

exactly to the issue at hand. The information can be weighted according to relevance or quality, and sensitivity analysis can be used to assess the priority to be given to collecting more directly relevant data. Bayesian variants of Monte Carlo integration procedures have been devised to address these objections using Gaussian process models (O'Hagan, 1995; Rasmussen and Ghahramani, 2003).

4.2.2 Modelling in Bayesian Paradigm

If the underlying processes are not enough understood, models are designed based only on the observed data. Instead, models are constructed with existing expertise, by beginning with a flexible model specified by a set of parameters, and combined it with the statistical model of the generated data set. The former is the modelling technique in standard classical approach and the latter is the Bayesian modelling approach. Bayesian modelling is the method of parametric modelling of data with prior information.

Let y be a random variable called response variable, which follows a probabilistic rule with density or probability function $f(y \mid \theta)$, where θ is the parameter vector. If, the iid sample of size 'n' of variable $y = [y_1, y_2, \ldots\ldots, y_n]^T$, then the joint distribution

$$f(y \mid \theta) = \prod_{i=1}^{n} f(y_i \mid \theta)$$

is called the likelihood of the model and contains the available information provided by the observed sample.

Usually, models are constructed in order to asses or interpret causal relationship between the response variable Y and various characteristics expressed as a variable $X_j, j \in \upsilon$, called explanatory variables; j indicates a model term (or covariate) and υ, the set of all terms under consideration. The explanatory variable is linked with the response variable via a deterministic function and a part of the original parameter vectors is substituted by an alternative set of the parameters

(denoted by β) that usually summarizes the effect of each covariate on the response variable.

In a Bayesian model selection task, the posterior distribution is obtained over a set of models given some a priori knowledge and some new observations (data). The knowledge is represented in the form of a prior over model structures *P(M)*, and their parameters *P(T|M)*, which define the probabilistic dependencies between the variables in the model (Beal, 2003).

By Bayes' rule, the posterior over models M observing data y is given by:

$$P(M \mid y) = \frac{P(M)\ P(y \mid M)}{P(y)}$$

The term $P(y \mid M)$ in the numerator is the *marginal likelihood* or *evidence* for a model M, which integrates over model parameters, and is the key quantity for Bayesian model selection. Also,

$$P(y \mid M) = \int P(\theta \mid M)\ P(y \mid \theta, M)\, d\theta$$

For model structure, the posterior distribution is computed over parameters as:

$$P(\theta \mid y, M) = \frac{P(\theta \mid M) P(y \mid \theta, M)}{P(y \mid M)}$$

The predictive density of a new response y′ given the responses $y = \{y_1, y_2, \ldots\ldots, y_n\}$ is obtained as

$$P(y' \mid y, M) = \int P(\theta \mid y, M) P(y' \mid \theta, y, M) d\theta \text{ , or simply}$$

$$P(y' \mid y, M) = \int P(\theta \mid y, M) P(y' \mid \theta, M) d\theta$$

If y' is conditionally independent of $y|T$, posterior distribution of x' associated with the new response value y' is obtained as

$$P(x' \mid y', y, M) \propto \int P(\theta \mid y, M) P(x', y' \mid \theta, M) d\theta$$

The process of assembling information into a Bayesian model is a multi-stage one, using data and information of many types. It is important to note that, even though these models provide a structure into which the available data can be incorporated

and use expert opinion; where there are no data, this does not mean that the models are a substitute for experimental data. The greatest advantage of Bayesian models is that they can be used to facilitate decision analysis despite inadequate data; this is especially important as some types of data are not likely to be readily collected at all (ibid.).

4.2.3 Graphical Models

Graphical models are an intuitive tool for visualising conditional independency relationships between variables. A graphical model expresses a family of probability distributions on sets of variables in a model. Graphical models provide a backbone upon which it has been possible to derive efficient message-propagating algorithms for conditioning and marginalising variables in the model given observation (data), (Heckerman, 1996; Cowell et al., 1999; Jordan and Weiss, 2002).

Each arc between two nodes in the graphical model represents a probabilistic connection between two variables. The terms ‘node’ and ‘variable’ are used interchangeably. Depending on the pattern of arcs in the graph and their type, different independence relations can be represented between variables. The pattern of arcs is commonly referred to as the *structure* of the model.

The graphical model specification can be done via directed acyclic graphs (DAG) from the doodle menu of WinBUGS. To construct DAG from model, specification of the nodes is needed to represent the variables of the model and the edges to represent dependencies between the variables induced by the model. Node and variables are depicted by rectangular or oval boxes depending on their type, while edges are depicted using unidirectional arrows (Ntzoufras, 2009).

4.2.4 Normal Regression Model

Normal regression models are most accepted models, in which the response variable *Y* is considered to be a continuous random variable distributed with the normal distribution with the parameters μ (mean) and σ^2 (variance). The normal regression model is summarized as:

$$Y \mid X_1, X_2, ..., X_p \sim N\left(\mu(\beta, X_1, X_2, ... X_p),\ \sigma^2\right)$$

with, $\mu(\beta, X_1, X_2, ..., X_p) = \beta_0 + \sum_{j=1}^{p} \beta_j X_j$

where, $(\beta_0, \beta_1, ..., \beta_p)^T$ *and* σ^2 are the regression parameters.

An alternative formulation of the regression model is, representing response variable directly as a function of the explanatory variable plus a random normal error with mean 0 and variance σ^2 .

$$Y = \beta_0 + \beta_1 X_1 + ... + \beta_p X_p + \varepsilon \ ; \qquad \varepsilon \sim N(0, \sigma_\varepsilon^{\ 2}).$$

Likelihood Specification in Normal Regression Model:

To simplify computational notation, we denote the response variable given explanatory variable $Y \mid X_1, X_2, ..., X_p$ simply by *Y,* and the expected value $E(Y|X_1, X_2, \ldots\ldots, X_p)$ by $E(Y)$ or μ.

Let, $x_{i1}, x_{i2}, ..., x_{ip}$ be the values of the explanatory variable X_1, $X_2, \ldots\ldots, X_p$ and with a sample size *n* corresponding response values $y = (y_1, y_2, ..., y_n)^T$ for individuals $i = 1, 2, \ldots, n$; then the model is expressed as

$$Y_i \sim N\left(\mu_i, \sigma^2\right)$$

$$\mu_i = \beta_0 + \beta_1 x_{i1} + ... + \beta_p x_{ip} \qquad \text{for } i = 1, 2, \ldots\ldots, n.$$

Independent Prior Specification:

The basic way of assuming a priori regarding the parameters in the normal regression model is the use of independent distributions.

$$f(\beta,\sigma^2)=\prod_{j=0}^{p} f(\beta_j)\cdot f(\sigma^2)$$

$$\beta_j \sim N(\mu_{\beta_j},\varsigma_j^2) \qquad for, j=0,1,\ldots\ldots,p$$

$$\sigma^2 \sim inv\ \text{gamma}\ (a, b)$$

Most of the computational software for Bayesian analysis prefer to use precision (τ) instead of variance σ^2. So, the specification is expressed as

$$f(\beta,\tau)=\prod_{j=0}^{p} f(\beta_j)\cdot f(\tau)$$

and $\tau \sim gamma(a,b)$

Conjugate Prior specification:

The normal distribution is assigned as conjugate prior for the $\beta|\sigma^2$ and an inverse gamma distribution for σ^2 for the normal regression model. The priori for the joint distribution of $[\beta,\sigma^2]$ follows *normal-inverse gamma* distribution. It is symbolized as

$$\beta \mid \sigma^2 \sim N_p(\mu_\beta,\ c^2V\sigma^2) \qquad \text{and} \qquad \sigma^2 \sim IG(a,b)$$

where, $V=(X^TX)^{-1}$ and

c^2 is a parameter controlling overall magnitude of the prior variance (Zellner 1986); the default choice of $c^2 = n$ (Kass and Washerman, 1995).

4.2.5 Random Effect Model

If Y_{ij} is the response variable for the jth individual of the ith goup, μ is the general mean, a_i is the effect of response variable when observed in the group i, then the model equation for Y_{ij} is

$$Y_{ij}=\mu_i+\varepsilon_{ij} \quad j=1,2,\ldots\ldots\ldots\ldots,n_i; \qquad i=1,2,\ldots\ldots\ldots,N$$

Or, $Y_{ij}=\mu_i+a_i+\varepsilon_{ij}$

where, ε_{ij} is the residual and a_i is the random variable, so the model is said to be random effect model.

In classical approach the model is written as,

a_i are iid and $a_i \sim N\left(0, {\sigma_a}^2\right)$ and

ε_{ij} are iid and $\varepsilon_{ij} \sim N\left(0, {\sigma_\varepsilon}^2\right)$

where, the random variables a_i and ε_{ij} are mutually independent.

In Bayesian paradigm, the model is summarised as:

$$Y_{ij} \mid a_i, {\sigma_\varepsilon}^2 \sim N\left(a_i, {\sigma_\varepsilon}^2\right)$$

$$a_i \mid \mu, {\sigma_a}^2 \sim N\left(\mu, {\sigma_a}^2\right)$$

The priors of the parameters μ, ${\sigma_\varepsilon}^2$ and ${\sigma_a}^2$ are assumed as:

$$\mu \sim N\left(\mu_0, {\sigma_0}^2\right)$$

$$\frac{1}{{\sigma_\varepsilon}^2} \sim gamma(a_1, \lambda_1)$$

$$\frac{1}{{\sigma_a}^2} \sim gamma(a_2, \lambda_2)$$

If we consider the following non-informative priors

$$\pi(\mu) \propto 1$$

$$\pi\left({\sigma_\varepsilon}^2\right) \propto \frac{1}{{\sigma_\varepsilon}^2}$$

$$\pi\left({\sigma_a}^2\right) \propto \frac{1}{{\sigma_a}^2}$$

then, we have the classical random effect model.

If we consider,

$$Y_{ij} \sim N\left(\mu_{ij},\ \sigma^2\right)$$

$$\mu_{ij} = \mu + a_i$$

$$\pi(a_i) \propto 1, \qquad \pi(\mu) \propto 1$$

$$\pi\left({\sigma_\varepsilon}^2\right) \propto \frac{1}{{\sigma_\varepsilon}^2},$$

then, we have the classical fixed effect model

4. 2.6 Generalized Linear Model

Generalized Linear Model (GLM) is a wide class of model used for analysis of both quantitative and qualitative response variable (McCullagh & Nelder 1989). The GLM can be used to relate a discrete random variable to two continuous predictor variables and one categorical predictor variable (Richardson and Best, 2003).

GLMs are regarded as the natural extension of the normal linear regression model. They are based on exponential family of distribution, which include the common distributions such as normal, binomial and Poisson. In many applications, in which response variable has a non normal distribution rather it is used to knowing how common probability functions can be expressed in exponential family form (Ntzoufras, 2009).

The stochastic component of this model consists of response variable $Y_i \sim D(\theta)$. The systematic component is the function of the explanatory variables. The linear combination of the explanatory variables is used in GLM, so it is called a linear predictor. The linear function $g(\theta)$ is the mathematical expression which connects the parameter of the response variable Y with linear predictor and the covariates. A location parameter is linked with the linear predictor. GLM is summarized in terms of stochastic component as:

$$Y_i \sim \exp \quad f\{\vartheta_i, \phi, a(), b(), c()\}$$

$Y_i \sim \exp \quad f\{\vartheta_i, \phi, a(), b(), c()\}$ denotes the exponential family with location parameter ϑ_i, scale parameter ϕ and $a(), b(), c()$ are the functions needed to specify the structure of the specific distribution (Ntzoufras, 2009).

Where, the systematic component $\eta_i = X_i\beta = \beta_0 + \sum_{j=1}^{n} x_{ij} \cdot \beta_j$

a canonical distribution parameter $\vartheta_i = R(\theta_i)$

a link function $g(\theta_i) = g\left(R^{-1}(\vartheta_i)\right) = g_\vartheta(\vartheta_i) = \eta_i$ and

the model parameters $\theta_{\mathrm{m}} = \left(\beta^T, \phi^T\right)$.

4.2.7 Bayesian Models with Hierarchical Priors

Naturally, Bayesian models are hierarchical in nature. The name hierarchical model derives from the hierarchy in which observations and parameters are structured. Bayesian hierarchical modelling is the methodology to combine prior results with a current study to obtain estimates of the parameters. This approach is known as the approach of borrowing strengths from other studies (FDA, 2006).

If knowledge is insufficient to specify the parameters of interest, the Bayesian model with more than two levels is used, which is known as Bayesian hierarchical modelling. In Bayesian hierarchical modelling, the prior distribution $f(\theta \mid a)$ of a model parameter θ with prior parameter *a* can be considered one level hierarchy. The likelihood as the final stage of Bayesian model is the final stage of the hierarchy (Ntzoufras, 2009). In some structures, the priors frequently use a series of conditional distribution called hierarchical stages of prior distribution.

The posterior distribution is written as:

$$f(\theta \mid y) \propto f(y \mid \theta)\ f(\theta, a)\ \ f(a, b)$$
$$\propto f(y \mid \theta)\ f(\theta \mid a)\ f(a \mid b)$$

The prior distribution is characterized by two levels of hierarchy, $(\theta \mid a)$ as the first level and $f(a \mid b)$ as the second level. The prior of the upper level of hierarchy is called the hyper-prior and its parameter as the hyper-parameter. In above hierarchical priors, $f(a \mid b)$ is the hyper prior and *b* is the hyper parameter of the prior *a.* George et al. (1996) discussed Bayesian analysis of hierarchical models where the conjugate prior is adopted at first level.

The following is an example of the hierarchical model,

$$Y_{ij} = \mu + a_i + \varepsilon_{ij}$$

where, model effect a_i are iid and $a_i \sim N\left(0,\ \sigma_a{}^2\right)$ and $\in_{ij} \sim N\left(0,\ \sigma_\varepsilon{}^2\right)$

The above model can be alternatively written as: $Y_{ij} \sim N\left(\mu_{ij},\ \sigma^2\right)$

with $\mu_{ij} = \mu + a_i$ and $a_i \sim N\left(0,\ \sigma_a{}^2\right)$ $i=1,2,.....,n$ and $j=1,2,.....m$

An example of Hierarchical modelling is given in the Appendix C-8.

4.2.8 Model Checking in Bayesian Paradigm

There are several ways of checking model fit in Bayesian paradigm. Simply, the prediction within Bayesian framework is done with predictive distribution. The posterior predictive distributions are used for model evaluation and checking. For the model checking, chiefly, the following ways are adopted (Ntzoufras, 2009).

i) Comparison of actual and predictive frequencies for discrete data
ii) Comparison of cumulative frequencies for actual and predictive values for continuous data
iii) Comparison of ordered predictive and actual values for continuous data
iv) Checking individual observation using residual
v) Checking the goodness of fit for model

Using WinBUGS software, the model checking is done with the computation of marginal likelihood. The evaluation of the model is done using posterior densities and the following methods.

i) Posterior Bayes factor
ii) BIC (Bayesian Information Criteria)
iii) AIC (Akaike Information Criteria)
iv) DIC (Deviance Information Criteria)

PART TWO

Chapter V

An Application of Bayesian Method in Estimating the Weights of Packaged Food

5.1 Introduction

The processed dairy product, ghee, is filled in pouch (packet) using a very precise computer controlled machine. The pouches are generally filled in with liquid ghee in a litre or a half litre pouch, using computerised filling machine. The weight of pouch is the consumer's primary interest for the quality assurance and to measure consistency of the producer's claim. The focus of this study is on drawing inference about the weight of pouched ghee, regarding the weights of 1 litre pouches. The details of the study using Bayesian method for predicting the mean weight of pouched ghee of a lot having known process variability is given in Khatiwada and Sthapit (2008).

5.2 Model

The distribution of the characteristics being measured is assumed to be distributed normally because of the well-mechanised production system and producing batches of thousand items. Let, X be the weight of ghee in a packet, x_{ij} denotes the weight of j^{th} packet of i^{th} sample from a lot of size N; (i =1,2,……n) (j =1,2,……..k_i). The total number of samples observed is 'n'. The parameter of interest is mean weight (θ) of the pouches.

The Gaussian distribution is the model assumption for the data (X) with the parameters θ and σ, specifically,

$$X \sim N\left(\theta, \sigma^2\right) \qquad (5.1)$$

θ and σ are the process mean and standard deviation. The distribution of X for given θ is the *likelihood* of θ.

Also, the assumption for the *prior distribution* of θ is Gaussian with mean θ_0 and standard deviation σ_0,i.e.,

$$\theta \sim N\left(\theta_0,\ {\sigma_0}^2\right) \qquad (5.2)$$

Also, the *posterior distribution* of the parameter of interest (mean) given the data can be obtained as Gaussian (normal), the proof can be found in Lindley (1970), Berger (1985) and Lee (1997).

$$\theta | X \sim N\left(\theta_1, {\sigma_1}^2\right) \qquad (5.3)$$

θ_1 and ${\sigma_1}^2$ are the posterior mean and variance, where

$$\theta_1 = {\sigma_1}^2\left(\frac{\theta_0}{{\sigma_0}^2} + \frac{n\cdot\hat{\theta}}{\sigma^2}\right) \text{ and } {\sigma_1}^2 = \left(\frac{1}{{\sigma_0}^2} + \frac{n}{\sigma^2}\right)^{-1}$$

and, *n* is the total number of sample; $\hat{\theta}$ is the estimated value (from data) of θ.

The *predictive distribution* of the new sample after obtaining the posterior density of the first n samples is given by $\left(X_{n+1} \mid X_n\right) \sim N\left(\theta_p, {\sigma_p}^2\right)$.

$$\left(X_{n+1} \mid X_n\right) \sim N\left(\theta_1,\ (\sigma^2 + {\sigma_1}^2)\right) \qquad (5.4)$$

where, $\theta_p = \theta_1$, and $\sigma_p{}^2 = \sigma^2 + {\sigma_1}^2$

5.3 Sample and Data

Random samples having equal number of the pouches have been taken from the finished product to measure the weights. The detail of the data is given in Appendix A-1. From the expert's opinion, it is known that the mean weight of the pouch is assumed to be distributed normally with some target value of 930 gm with a variation scale of ± 30gm. The specified lower limit of the average weight is expected to be 920gm. Twenty five samples of size five have been selected in different time-period and the average weights measured.

The data summary (weight in gm):

900	*905*	*914*	*913*	*927*	*915*	*900*	*918*	*908*	*916*
918	*924*	*925*	*934*	*929*	*920*	*925*	*930*	*930*	*930*
924	*930*	*934*	*922*	*934*					

5.4 Analyzing Data

Initially, for the purpose of summarizing data the point estimates and the confidence interval for the estimates have been obtained using the classical method. From the mechanization of the industry, the target value (the average weight of the individual pouch) was set as 930g with tolerance limits of ± 30g from the target value. Thus, the process spread $\mu \pm 3\sigma$ has been kept as 930 ± 30, where the assumed process standard deviation (σ) for the weight of the individual pouch is found to be 10. The summary of the assumed model parameter and the actual values obtained from the data have been computed and depicted in the Table 5.1.

Table 5.1 Summary of the parameters and confidence intervals using classical method

Source	Mean	SD	Confidence limits (for, σ = 10)			lower specification limit = 920 Target value =930	
			Q1, Q3	95%	3σ	P(*X*<920)	P(*X*≥930)
Model X|θ~N(930,100)	930	10 (known)	923.25, 936.75	910.4, 949.6	900.00, 960.00	0.1587	0.5
Data, x_i x_i|θ~N(921,100)	921	10.169 (computed)	914.25, 927.75	901.4, 940.6	890.86, 951.14	0.4602	.184
The standard error of mean, SE($\bar{x}$) = 10.169 (computed)							

5.5 Model Assessment

5.5.1 Bayesian Inference with Informative Normal Prior

To estimate the average weight of the pouch, the size of a sample (k) has been taken as 5. Thus, the variance of the sample average weight of the pouch is figured out as 20, using $\sigma_{\bar{x}}^2 = \sigma^2 / k$. This information is assumed as the prior information, and the Bayesian framework is applied to obtain estimates and probabilities (Khatiwada and Sthapit, 2008).

Table 5.2 Summary of the prior density, likelihood and the posterior density with precisions

Density of X	Prior density $\pi(\theta)$	prior precision	likelihood of data $f(X\|\theta)$	Data precision	Posterior density $p(\theta\|X)$	Posterior precision
$X \sim N(\theta, \sigma^2)$	$\theta \sim N(\theta_0, \sigma_0^2)$		$x \mid \theta \sim N(\hat{\theta}, \sigma^2)$		$\theta\|X \sim N(\theta_1, \sigma_1^2)$	
$N(\theta, 10^2)$	$N(930, 20)$	0.05	N(921.00, 100), n =25	0.25	N(922.5, 3.333)	0.30

Table 5.3 Summary of the posterior distribution

Posterior density $\theta\|X$	50% HDR	95% credible interval	3σ limits for mean	P ($\theta \leq 920$)	P ($\theta > 930$)
N(922.5, 3.3333)	921.27, 923.73	918.92, 926.08	917.00, 928.00	0.0855	0.0000

Table 5.4 Summary of the predictive distribution

Predictive density $X_{i+1}\|X_i$	50% HDR	95% credible interval	3σ limits for X_{i+1}	P(X<920)	P(X≥930)	1-P(900 ≤X≤960)
N(922.5, 103.33)	915.64, 929.36	902.57, 942.43	891.87, 953.13	0.4028	0.2303	0.0136

Plots using First Bayes:

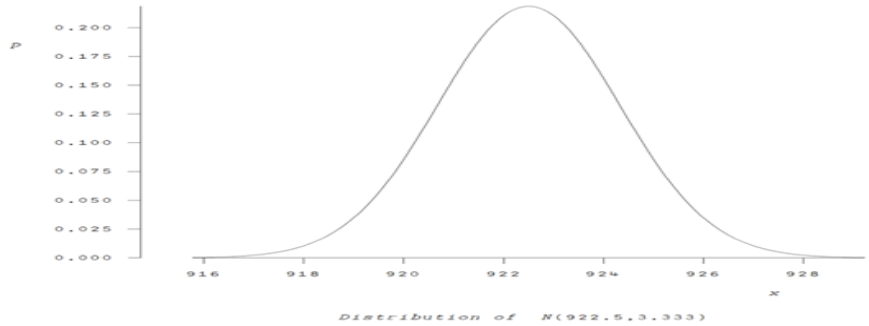


Figure 5.1 Plot of the posterior density of average weight *(θ)*

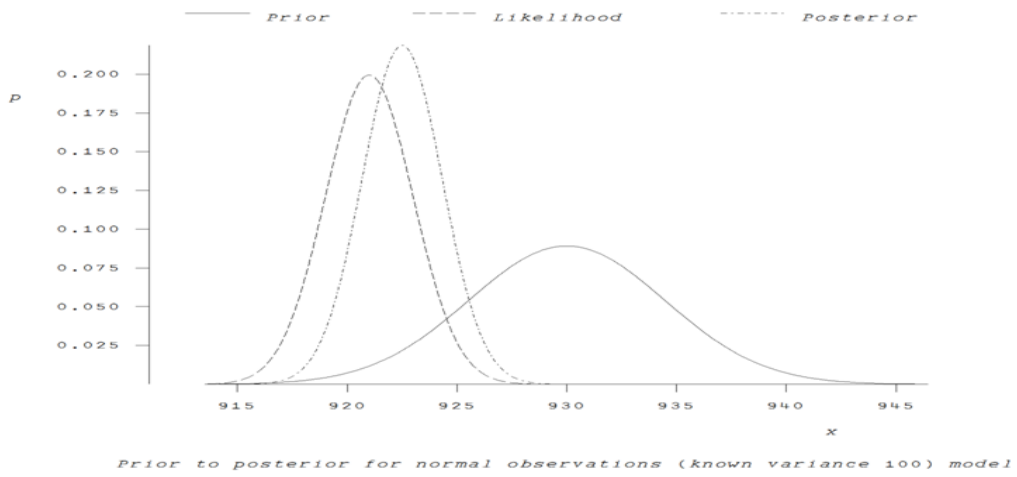


Figure 5.2 Triplot of the prior density (solid line), posterior density (dotted line) and likelihood (dashed line)

.

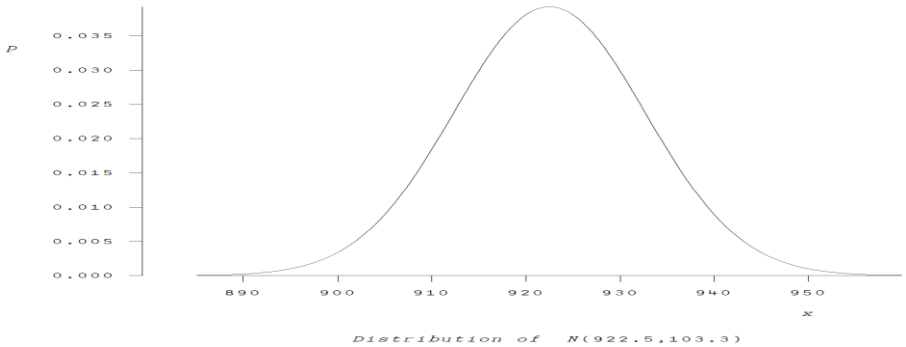


Figure 5.3 Plot of the predictive density of the weight (*X*)

The WinBUGS Result:

The use of normal prior (informative) to the normal likelihood yields a Normal-Normal Bayesian Model. The WinBUGS language for the Bayesian inference using normal (informative) prior is given in the Appendix B-1. The distribution of the posterior parameters using WinBUGS are presented in Table 5.5 and trace plot for the 5000 samples and 30000 samples from 30000 iterations discarding first 1000 values is given in Figure 5.4.

Table 5.5 Summary of the posterior density using normal informative prior

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
theta	922.5	1.825	0.02374	918.9	922.5	926.1	1001	5000
theta	922.5	1.816	0.00999	918.9	922.5	926.1	1001	30000

Kernel density plots:

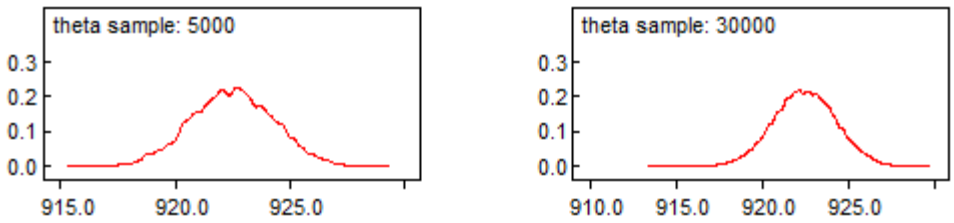


Figure 5.4 Kernel density plots of the posterior distribution through MCMC using WinBUGS

Trace:

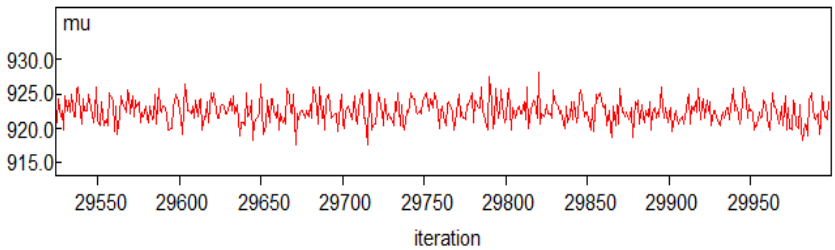


Figure 5.5 Trace of the posterior distribution through MCMC for last 500 iterations in an updating 30000 iterations.

5.5.2 Bayesian Inference using Non-informative Prior

Assuming that nothing is known about the prior distribution (it is our suspect on the information provided by the management) about the parameter of interest (θ), the model is updated by using a non-informative prior, such as

$$\theta \sim N(0,\ 1.0E6).$$

The WinBUGS language for the inference is given in the Appendix B-2. The results are summarized in Table 5.6.

Table 5.6 Summary of the posterior density using non-informative prior

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
theta	921.0	1.999	0.02600	917.1	921.0	924.9	1001	5000
theta	921.0	1.989	0.01095	917.1	921.0	924.9	1001	30000

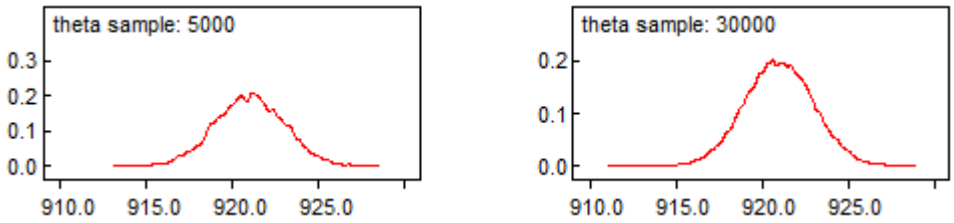


Figure 5.6 The density plots of the posterior distribution for 5000 and 30000 iterations

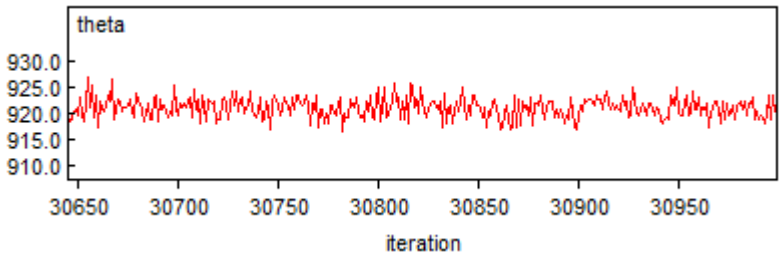


Figure 5.7 The trace of the last 400 iteration of the posterior distribution (using non-informative prior)

5.6 Results and Summary

Table 5.1 shows the estimated process mean (= 921) is less than the target value (930), with almost equal to the given process variance (100). The estimated 3σ limit for mean is not found within the natural tolerance limits (900-960). A sample falling below the lower specification limit (920) is found to be 46 out of 100 chances. The 95% classical confidence interval is obtained within the natural tolerance limit. The precision of the estimate is 0.25.

The posterior mean and variance using an informative prior are 922.5 and 3.33 respectively. The posterior precision ($1/\sigma_1^2 = 0.3$) is equal to the sum of the prior precision and data precision (Table 5.2). The probability of occurring average weight beyond the lower specification (120) is 8.5 per 1000 (Table 5.3). No part of the average weight appears more than the target value (930). There is almost sure probability that the average weight occurring between 917 and 928 (Table 5.3).

Table 5.4 shows the predictive distribution of the weight (X). Further, if a random sample is drawn, given the posterior distribution, the probability that the sample value will occur beyond the lower specification is 0.4028. The chance of being a newly drawn random sample above the lower specification is 0.5972. For all new draws, given the posterior distribution, the probability of occurring average weight exceeding 930 is 0.2303. Also, the predictive probability of the weight of a next sample drawn falling outside the natural tolerance limit, is 0.0136.

The model $X \sim N(930, 100)$ allowed that the chance of happening weight below 920gm (lower specification) not to be more than 15.9%. Data information shows the part of the weight below specification is 46% (Table 5.1). Using Bayesian method, the probability of the average weight occurring below specification is found to be 8.5% only. Moreover, if a new sample will be drawn, the chance that it will fall below specification is 40.3%. There is almost zero probability of occurring an average weight more than the target value (Table 5.3), and that of the newly drawn sample is 0.23 (Table 5.4). The capability index for the lower half (Mitra, 2001) for the posterior distribution is 0.25, that of the data is 0.1.

From the WinBUGS result using MCMC, it is observed that the mean of the distribution converges to 922.5 for a moderate and extended iteration with standard deviation 1.8. By the use of kernel density plot (Figure 5.4), it is shown that the convergence of the posterior mean and its underlying distribution and the increasing smoothness of the curve as increase in iteration. The consistency in sample values for a cross section of MCMC iteration is shown by using trace plot (Figure 5.5).

In case of updating the likelihood with a non-informative prior, the posterior estimate is found to be much closed to classical maximum likelihood estimate. The 95% credible set is same as 95% classical confidence interval. Clearly, this indicates that the non-informative prior has the negligible impact on the model. Bayesian method works only as the process of strengthening the updated inference, in this case. The software WinBUGS facilitates in generating data from the posterior distribution using MCMC. As the iteration increases the posterior density for the parameter of the interest seems to be more precise (Figures 5.5 - 5.7).

Remarks

The area under discussion of this study is related with the consumer's concern, whether the weights of the pouches are underweight. From the posterior distribution, using conjugate normal prior with known variance, it is observed that the target value is beyond the 3-sigma control limits and there is a chance of 0.085 being the sample average below specification limit. It can be concluded, therefore, that the process should be adjusted to the direction that can increase in the average weight of pouches.

Chapter VI

Analysis of Beta Prior Distribution for an Acceptance Sampling Plan of Food Product

6.1 Introduction

In an instant noodle factory, ready to use (fairly-cooked) product is packed and shipped. In the process of cutting, cooking, spraying, de-metalling and packaging conditions, the crack, break or rupture (nonconforming item) in the product are occurred and such items are screened out before packaging. In this study, prior information about the proportion of items nonconforming produced has been taken from the past record of the noodle factory. For the modelling purpose of experimental data, the samples have been taken from online product flow. The prior distribution is elicited using past data. Initially, from sample data an acceptance sampling plan is developed using classical approach. Then, the likelihood of sample data and the prior distribution are combined together for updating the information. An analysis on beta prior is carried out and different sampling plan for the quality attribute are designed from the updated distribution. The new plan so developed is compared with the sampling plan prepared by classical method.

6.2 Model

Suppose, $x_1, x_2, ...x_k$ be the number of nonconforming items observed in the sample of size $n_1, n_2, ...n_k$ respectively. The number of nonconforming items *X*, for a sample of size n_i in the i^{th} lot, with a lot nonconforming proportion *p*, is distributed with binomial probability. The probability mass function of ($X = x_i = n_i p$) is given by

$$P(X = x_i) = \binom{n_i}{x_i} p^{x_i} (1-p)^{n_i - x_i}, \qquad x_i = 0, 1, 2, ..., n_i ;\ n_i, < N_i \quad (6.1)$$

Further, it is assumed that the prior distribution of the parameter of interest *(p)* follows the beta distribution,

$p \sim Beta\ (a, b)$

$$\pi(p) = \frac{1}{B(a,b)} p^{a-1}(1-p)^{b-1}, \quad 0 \le p \le 1, \quad a,b > 0. \tag{6.2}$$

The prior mean and variance are respectively

$$\bar{p} = \frac{a}{a+b} \tag{6.3}$$

$$\sigma^2 = \frac{ab}{(a+b)^2(a+b+1)} \tag{6.4}$$

The value of $\bar{p}$ and σ^2 are estimated from past production records as the process average and spread respectively. In all situations of a factory account, it is difficult to obtain calculated σ^2. In case of beta distribution, the size of sample previously observed, n', is taken as the sum of the parameters (Lee, 1997) i.e.,

$$n' = a + b \tag{6.5}$$

The method of prior selection and the details of deriving posterior distribution are found discussed in Khatiwada (2010). The posterior density is obtained as

$$\pi(p \mid x) = \frac{1}{B(a+c,\ n+b-c)} \ p^{a+c-1}\ (1-p)^{n+b-c-1} \tag{6.6}$$

(where, $c = \sum x_i$ and $n = n_i$)

The posterior mean and variance are obtained as

$$\bar{p}^* = \frac{a+c}{n+a+b} \tag{6.7}$$

$$\sigma^{2^*} = \frac{(a+c)(n+b-c)}{(n+a+b)^2(n+a+b+1)} \tag{6.8}$$

6.3 Sample and Data

To observe the process of product control and to take sample (data) an inspection is made in online product flow. Fifty random samples of different sizes are taken from online product flow of different lots of same size as convenience. The sample proportion nonconforming items are calculated (presented in Appendix A-2) and plot of the nonconforming proportion is shown in Figure 6.1.

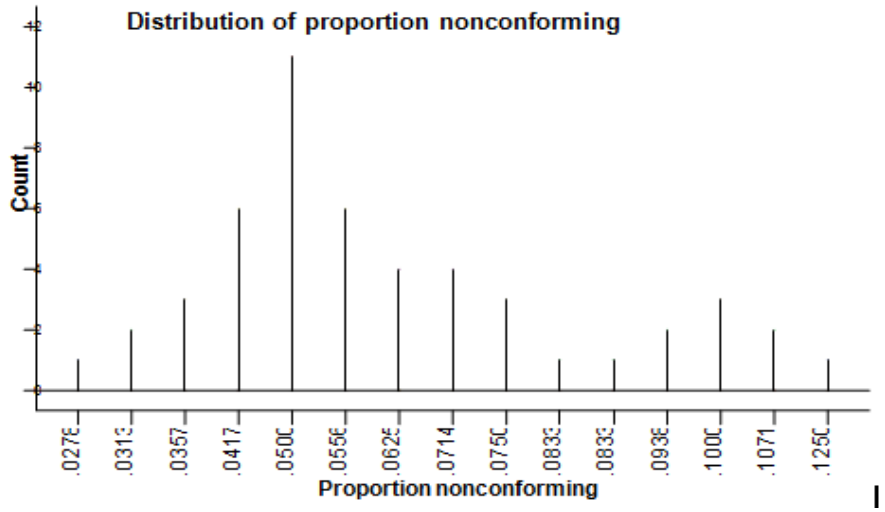


Figure 6.1 Distribution of sample nonconforming proportion

6.4 Analyzing Data

Sample mean nonconforming proportion is found to be 0.0617 with a standard deviation of 0.023. The minimum nonconforming proportion is obtained as 0.0278 and maximum of the same as 0.125. The 95% confidence interval for mean nonconforming proportion is obtained (0.00 - 0.0957). The 90^{th} percentile value of the sample nonconforming proportion is obtained as 0.0928.

In Khatiwada (2009), by taking limiting quality level (LQL) = 0.062 (the sample average point) and consumers' risk, β =0.1, different alternative sampling plans have been proposed in conventional method using binomial distribution, as given in Table 6.1.

Table 6.1 Sampling plan for $\beta = 0.1$ using binomial cdf

c	0	1	2	3	4	5	6	7	8	9	10
n	36	62	85	106	127	148	168	188	208	227	246

6.5 Model Assessment

6.5.1 Prior information and its model

According to past record of the industry, the proportion nonconforming to packaging is found to be recorded as 0.05. The size of sample inspected for this purpose, is found to be 40. The past record has not been available for calculated process variation. For an average of 0.05 nonconforming proportion, using equation (6.3) and (6.5), the parameters of beta prior are fixed as $a = 2$ and $b = 38$. The plot of prior distribution, Be (2, 38), is shown in the Figure 6.2. The statistical information obtained from the distribution beta (2, 38) is given in the Table 6.2.

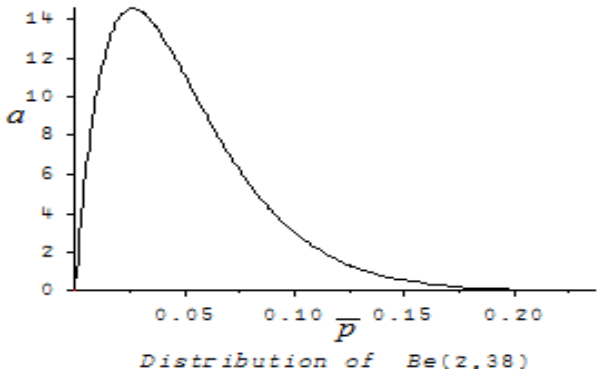


Figure 6.2 Plot of the prior distribution of average nonconforming proportion

Table 6.2 Summary of the Beta (2, 38) prior distribution

mean	median	mode	s d	$\Pr(\bar{p} \le 0.05)$	95% confidence interval		10^{th} Percentile	90^{th} Percentile	99.99^{th} Percentile
0.05	0.0427	0.0263	0.0340	0.5871	0.0013	0.1166	0.0137	0.0961	0.2639

6.5.2 Combining Prior and Sample Data

The prior information and sample data are combined to obtain posterior density (Khatiwada, 2010). Minimum sample size necessary and different maximum number of observed nonconforming items have been calculated, and presented in Table 6.3. The summary of the posterior distribution is given in Table 6.4. Plots of posterior density for four different $Be(a+c, n+b-c)$ are given in Figures 6.3[(*i*) – (*iv*)].

Table 6.3 Sampling plan for and $\beta = 0.1$ using posterior distribution

c	0	1	2	3	4	5	6	7	8	9	10
n	22	44	65	85	105	124	142	162	180	197	215
Posterior Be (a+c, n+b-c)	(2,60)	(3,81)	(4, 104)	(5,120)	(6, 139)	(7,157)	(8,174)	(9,193)	(10,210)	(11, 226)	(12, 234)

Table 6.4 Summary of posterior distribution $Be(a+c, n+b-c)$

Be(a+c, n+b-c)	mean	median	mode	s d	95% Credible Interval		10^{th} percentile	90^{th} percentile	99.99^{th} percentile
Be(2,60)	0.0323	0.0274	0.0167	0.0223	0.0008	0.0758	0.0088	0.0623	0.1773
Be(3,81)	0.0357	0.0321	0.0244	0.0201	0.0038	0.0753	0.0134	0.0629	0.1571
Be(4,101)	0.0381	0.0352	0.0291	0.0186	0.0071	0.0748	0.0169	0.0631	0.1442
Be(5,120)	0.0400	0.0376	0.0325	0.0175	0.0100	0.0746	0.0197	0.0634	0.1356
Be(6,139)	0.0414	0.0393	0.0350	0.0165	0.0125	0.0741	0.0220	0.0635	0.1288
Be(7,157)	0.0427	0.0408	0.0370	0.0157	0.0148	0.0740	0.0241	0.0637	0.1241
Be(8,174)	0.0440	0.4229	0.0389	0.0152	0.0168	0.0741	0.0259	0.0642	0.1207
Be(9,193)	0.0446	0.0431	0.0400	0.0145	0.0184	0.0734	0.0272	0.0638	0.1166
Be(10,210)	0.0455	0.0441	0.0413	0.0140	0.0200	0.0734	0.0286	0.0640	0.1141
Be(11, 226)	0.0464	0.0451	0.0426	0.0136	0.0215	0.0736	0.0299	0.0645	0.1120
Be(12, 234)	0.0488	0.0476	0.0451	0.0137	0.0236	0.0761	0.0322	0.0670	0.1144

Plots of posterior Distribution:

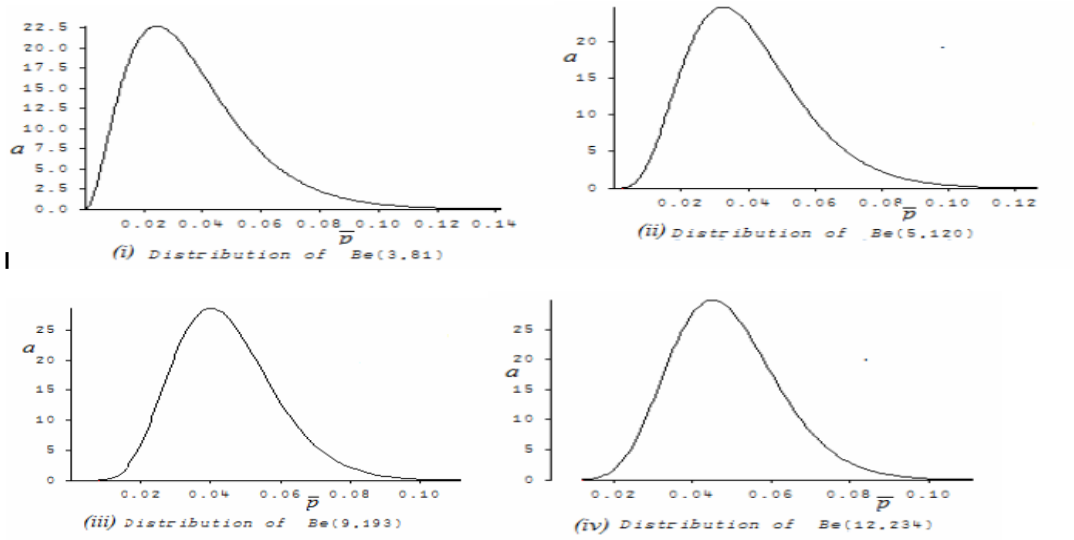


Figure 6.3 Plots of posterior distribution

6.6 Summary and Remarks

Considering Table 6.3, it is observed that the necessary sample sizes, when using the beta prior, are slightly less than that required in the traditional sample based method (Table 6.1). The 95% credible intervals for proportion nonconforming for the posterior plans, $Be(a+c, a+b-c)$ are observed shorter than the 95% confidence interval of the sample nonconforming proportion as well as the prior distribution, $Be(a,c)$. All the 90th percentile values of the nonconforming proportion for the posterior distribution are found to be less than that of the prior distribution and sample data.

The result of this study shows that the sampling plans using Bayesian posterior estimation are sharper (Figure 6.3) than the plans from the sample information only. From this result, it can be concluded that the acceptance sampling plan derived using posterior distribution helps more for assurance of the consumer's risk.

Chapter VII

Modelling Failure to operation points for Canning process in a Fruit Juice Industry

7.1 Introduction

Canning is a process of preserving or packaging food or drink by putting it in sealed airtight container. It is mostly used to preserve extracted food materials in cans or jars for the use of extended time while the long-term use of the raw food or the product is not viable. The examination of the system failure in different units (points) of operation is the process of identification of critical control point (CCP) in HACCP. In this study, an attempt is made to model the failure rate in different operation points in a canning industry using Bayesian modelling approach. The effect of just in time is high in canning system, and so, data generated (collected) in the time basis that how many times it fails to operate properly is used for the modelling failure rate.

7.2 Model

7.2.1 Poisson-Gamma Hierarchical Model

Carlin and Louis (1996) proposed to use gamma distribution as the conjugate prior for the Poisson data. In hierarchical models, George et al (1993) suggested to use gamma conjugate prior for the first level hyper-parameters. For the data of the failure at different operation points, the number of failures *x(i)* is assumed to follow a Poisson distribution.

$$x(i) \sim Poisson\ (\lambda_i)$$

where, $\lambda_i = \theta_i \times t_i$ and $i = 1,2,.....,10$

The parameter of interest θ_i is the failure rate for operation points i and t_i is the length of the operation time of i^{th} operation point.

7.2.2 Specification of Hyper-parameter

The conjugate gamma prior is used for the failure rate θ_i.

$\theta_i \sim gamma\ (\alpha, \beta)$

α and β are the hyper-parameters and their specification are assumed as in George *et al*. (1993).

$\alpha \sim exponential\ (1.0)$

$\beta \sim gamma\ \beta\ (0.1, 1.0)$

which gives gamma posterior for β and a non standard posterior for α.

7.2.3 Graphical Model

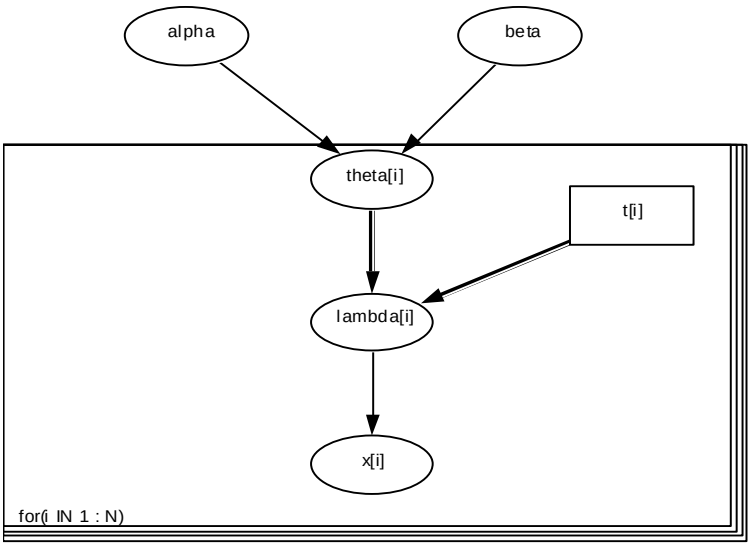


Figure 7.1 Graphic modelling of the parameters in failure to operation points

7.3 Sample and Data

An evaluation of failure in process of canning has been made in a fruit juice industry, Rijal Tasi Industry, Itahari, Sunsari. The sample evaluation has been designed for failure to operation in 10 different operation points of canning. The operation points / units considered are: (i) Raw material receiving station (ii)

Preparation of material (iii) Mixing ingredients (iv) Container cleaning (v) Closer parts (vi) Filling (vii) Sealing (viii) Retorting (ix) Cooling and (x) Storage.

The experimental data (given in Appendix A-3) have been adopted from the records of 720 hrs (45days×16hrs per day) of inspection. t_i is the length of the operation time, x_i is the number of failure. The parameter of interest is the failure rate x_i / t_i for operation points.

7.4 Analyzing Data

Initially, the failure rates per operation point *r(i)* have been estimated, which are presented in the Table 7.1. In classical analysis, the point estimates, r_i are found to be as the factual failure rate and *s.e.(i)* the standard errors of the estimates.

Table 7.1 Calculation of Failure rate per operation units /points

operation point (i)	1	2	3	4	5	6	7	8	9	10
r_i	0.1667	0.1250	0.0583	0.0417	0.0500	0.0611	0.0417	0.0521	0.0833	0.0926
s.e. (r_i)	0.1076	0.0675	0.0214	0.0167	0.0281	0.0179	0.0288	0.0227	0.0326	0.0279

7.5 Model Assessment

The WinBUGS software has been used for the graphical model, obtaining required posterior estimates of the proposed model and for the MCMC iteration. The WinBUGS code for the MCMC simulation is given in Appendix B-3. Starting from discarding first 1000 updates followed by a further 10000 updates gave the parameter estimates as given in Table 7.2. The density plots of posterior failure rates computed are given in the Figure 7.2.

Table 7.2 The posterior estimates of the parameters

node	mean	sd	MC error	2.50%	median	97.50%	start	sample
alpha	0.6299	0.2316	0.00401	0.2707	0.5994	1.1670	1001	10000
beta	3.2230	1.7310	0.02792	0.7026	2.9500	7.3490	1001	10000
theta[1]	0.1742	0.1089	0.001117	0.0293	0.1521	0.4458	1001	10000
theta[2]	0.1327	0.0705	6.44E-04	0.0329	0.1209	0.3028	1001	10000
theta[3]	0.0618	0.0226	2.45E-04	0.0259	0.0591	0.1127	1001	10000
theta[4]	0.0449	0.0174	1.91E-04	0.0176	0.0427	0.0850	1001	10000
theta[5]	0.0579	0.0305	2.88E-04	0.0144	0.0525	0.1313	1001	10000
theta[6]	0.0635	0.0187	1.71E-04	0.0324	0.0617	0.1049	1001	10000
theta[7]	0.0514	0.0319	3.03E-04	0.0086	0.0450	0.1296	1001	10000
theta[8]	0.0571	0.0239	2.66E-04	0.0204	0.0539	0.1130	1001	10000
theta[9]	0.0873	0.0336	3.35E-04	0.0352	0.0827	0.1642	1001	10000
theta[10]	0.2717	0.0839	8.28E-04	0.1316	0.2624	0.4600	1001	10000
lambda[1]	2.0900	1.3070	0.0134	0.3512	1.8260	5.3490	1001	10000
lambda[2]	3.1850	1.6920	0.0155	0.7899	2.9010	7.2670	1001	10000
lambda[3]	7.4180	2.7110	0.0294	3.1090	7.0890	13.5300	1001	10000
lambda[4]	6.4670	2.5080	0.0276	2.5290	6.1480	12.2400	1001	10000
lambda[5]	3.4760	1.8290	0.0173	0.8637	3.1510	7.8790	1001	10000
lambda[6]	11.4400	3.3640	0.0309	5.8280	11.1000	18.8900	1001	10000
lambda[7]	2.4650	1.5320	0.0146	0.4115	2.1610	6.2190	1001	10000
lambda[8]	5.4810	2.2930	0.0256	1.9550	5.1750	10.8500	1001	10000
lambda[9]	6.2840	2.4200	0.0242	2.5330	5.9530	11.8200	1001	10000
lambda[10]	9.7810	3.0220	0.0298	4.7390	9.4450	16.5600	1001	10000

7.6 Summary

From classical analysis, the raw estimates are found as the process average failure rate, just as a point estimate with the standard error. A complete posterior distribution of failure rate (θ_i) for each operation point has been obtained from the Gibbs sampler using WinBUGS. It has been observed that the posterior updated failure rates are slightly greater than the classical estimates, distributed with the slightly greater standard errors than the classical method. This result is might be due to the use of hyper priors with large variance. The 95% credible region for the estimated parameters and the density curves for each parameter have been obtained for a burn in of 1000 updates followed by a further 10000 updates. For the model verification, graphs of the posterior estimates are drawn and presented in the Figures 7.3 - 7.7.

Density plots:

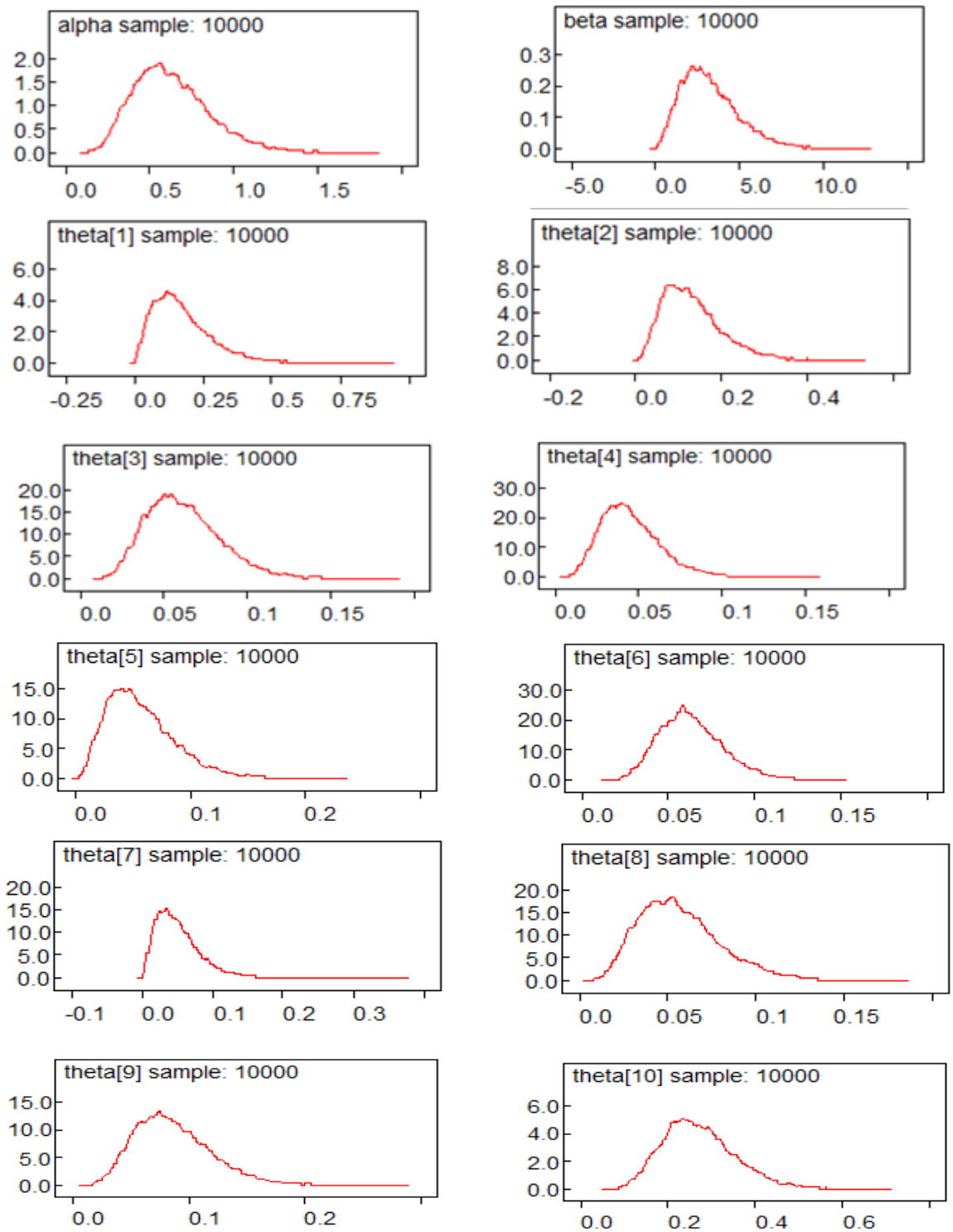


Figure 7.2 Kernel Density plots of the parameters in estimating failure to operation points

Box plots:

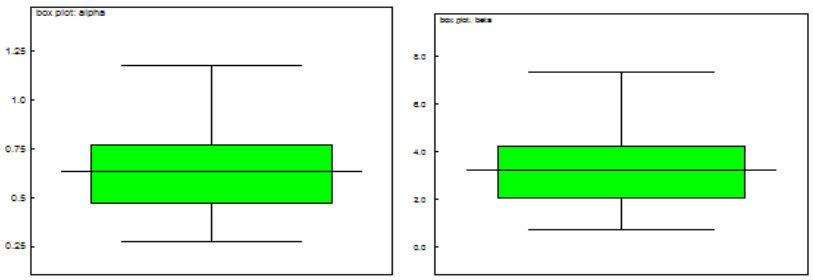


Figure 7.3 Box plots of predictive alpha and beta

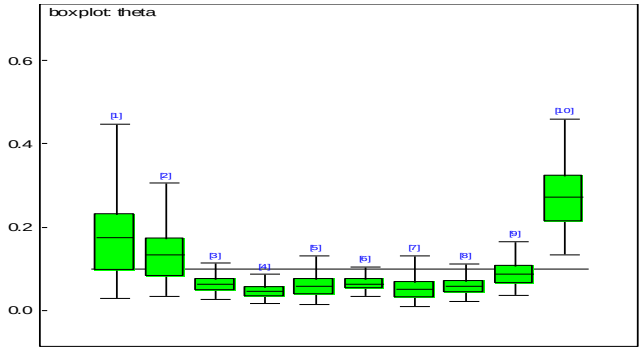


Figure 7.4 Box plots of posterior failure rate (theta|data)

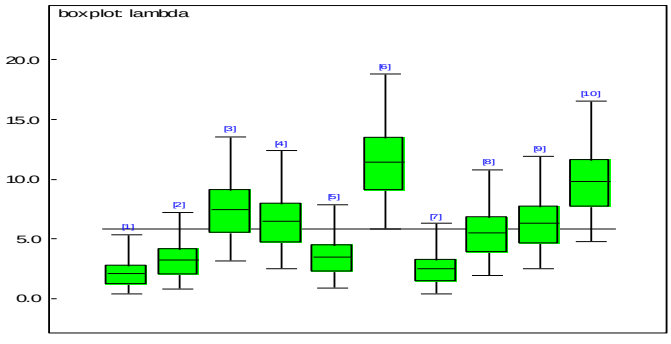


Figure 7.5 Box plots of posterior average failure rate (lambda|theta, t_i)

Scatter plot and Model fit

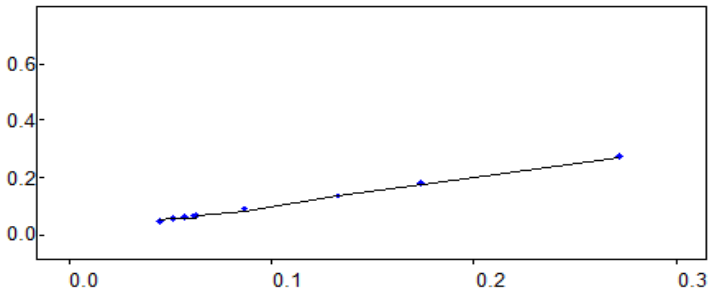


Figure 7.6 Scatter plots of posterior failure rate (theta|data)

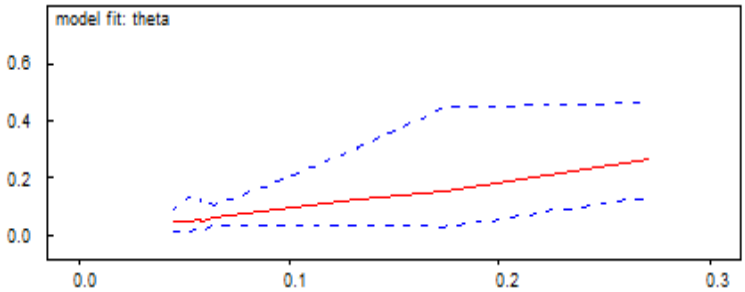


Figure 7.7 Model fit summaries for the predicted failure rate (theta|data)

Remarks:

To identify the priority points of the failure in operation points has been the main concern of this study, which should be corrected for the removal of the high frequency of the failure rate. From the posterior densities, it is observed that the failure rate $(\theta_i|data)$ is highest for the storage point, and the second and third points are raw material receiving station and material preparation point respectively (Figure 7.4). Highest average failure rate $(\lambda|data)$ is obtained for the points of closer parts and consequently storage point and ingredient mixing point respectively are obtained as second and third points (Figure 7.5). The modelling approach using the Bayesian method found to be advanced for the identification and determination of the failure rate with more variability. This method seemed to be outstanding for the prediction and to check out the effecting points, which reduce performance of the process. This method of modelling is expected to be helpful in identification of critical control points (CCP) while applying HACCP.

Chapter VIII

NIG Model for Predicting Porosity of a Rice Variety in terms of Moisture content and Bulk density

8.1 Introduction

Moisture content in grain is an indicator of its quality. It is usually expressed in proportion (or percent) of moisture present in the grain, which is measured either in wet basis or in dry basis. *Bulk density* is the density of a material when packed or stacked in bulk. Bulk density of particular solid is measured by allowing the sample to pour into a container of known dimensions. It is a volumetric analysis of quality measurement of the grain. *Porosity* is the property of some thing to which some thing is porous. Porosity is an important physical property characterizing the texture and the quality of dry and intermediate moisture foods. Porosity data is required in modelling and designing of various heat and mass transfer processes such as drying, frying, baking, heating, cooling, and extrusion. It is an important parameter in predicting diffusional properties of cellular foods (Sahin and Sumnu, 2006). Porosity is defined as the percentage of volume of intra-granular air space to the total volume of grain bulk. In this study, the porosity, bulk density, true density and moisture contents of a rice variety are obtained using the methods described in Sahin and Sumnu (2006) and an attempt is made to model the porosity of rice in terms of bulk density and moisture content using the Bayesian modelling.

8.2 Model

Normal regression model with hierarchical priors is used for modelling the relations of the characteristics of the given rice data in the Bayesian approach. Assuming that, Y_i denotes the response variable 'porosity' and X_{i1} and X_{i2} denote the explanatory variables 'moisture content' and 'bulk density' respectively, for $i = 1, \ldots\ldots, n$.

The likelihood of the data is modelled as:

$$Porosity \sim N(\mu_i, \sigma^2)$$

$$\mu_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} \qquad \text{for } i = 1, \ldots\ldots, n.$$

In normal regression using MCMC with WinBUGS, the precision parameter τ is used for $1/\sigma^2$ and β_j for $j=0,1,2$ are defined separately by single scalar node.

Prior Specification

As suggested by Spiegelhalter *et al.* (2003), in case of no information available, the choice of the prior mean is made zero, which corresponds to the no effects of *Xj* on *Y* and the prior variances of the effect *Bj* is set equal to large values to represent the high uncertainty or to denote prior ignorance. For τ, equal low prior parameter values are used for setting in this way that its prior mean equal to one and high prior variance. Simply, it is assumed as

$a = b=0.01$ for *gamma (a, b)*, which gives $E(\tau) =1$ and $V(\tau) = 100$.

$\sigma^2 \sim inv\_gamma(0.01,\ 0.01)$ as in Ntzoufras (2009)

Or, $\tau = 1/\sigma^2 \sim gamma(0.01,\ 0.01)$ and

$B_j \sim N(0,\ 0.001)$ for j= 0, 1,2.

An *inverse gamma prior* is specified for the parameters having normal data model, hence the modelling approach is called as NIG model. The WinBUGS code for the model specification is given in the Appendix B-4.

8.3 Sample and Data

The data set for modelling is adopted from the project work of a student of Food Technology. Random samples of Mansuli variety of rice have been drawn from four different local whole-seller rice mills of Dharan municipality. Seven replications are made for testing each characteristic. The moisture content, bulk-density, true density and porosity are taken as the test characteristics and calculated using the formula:

$$\text{Porosity}\,\% = \frac{\text{void density}}{\text{true density}} = \frac{\text{True density - Bulk density}}{\text{True density}} \times 100$$

The Table of data is given in the Appendix A-4.

8.4 Data Analysis

Experimental data of the moisture content, bulk density, true density and porosity are analysed to obtain summary and descriptive statistics**.** The summary statistics obtained so are given in the Appendix A-5. The average of the moisture content, bulk density and the porosity between groups found to be significantly different, since classical p-value is observed less than 0.01(Appendix A-6).

8.4.1 Scatter Plots, Correlation Coefficients and Regression Equations

The information obtained regarding correlation coefficients, regression coefficients (α- y intercept, β- slope) are summarized in the Table 8.1. The results regarding the coefficients of partial correlation and zero order correlation by classical method are presented in the Table 8.2 and that of multiple correlation and the multiple regression are summarized in the Tables 8.3 and 8.4 respectively.

The regression equation of bulk density (X_2) on the moisture content (X_1) is

$$X_2 = 0.569 + 0.021X_1,$$

The regression line of porosity (X_3) on bulk density X_2 is

$$X_3 = 35.171 - 3.254\, X_2$$

Table 8.1 Summary of the correlation and regression coefficients

between	*r*	*se(r)*	α	β	R^2
Bulk density[a] & Moisture Content	0.645	0.0264	0.569	0.021	0.416
Porosity[a] & Bulk density	-0.052	2.1811	35.171	-3.254	0.003

a Dependent variable

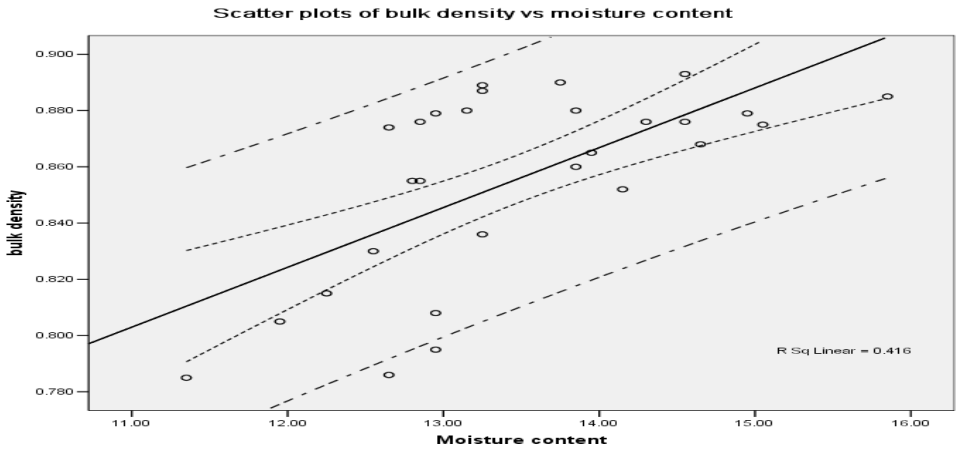


Figure 8.1 Scatter plots of moisture content vs bulk density

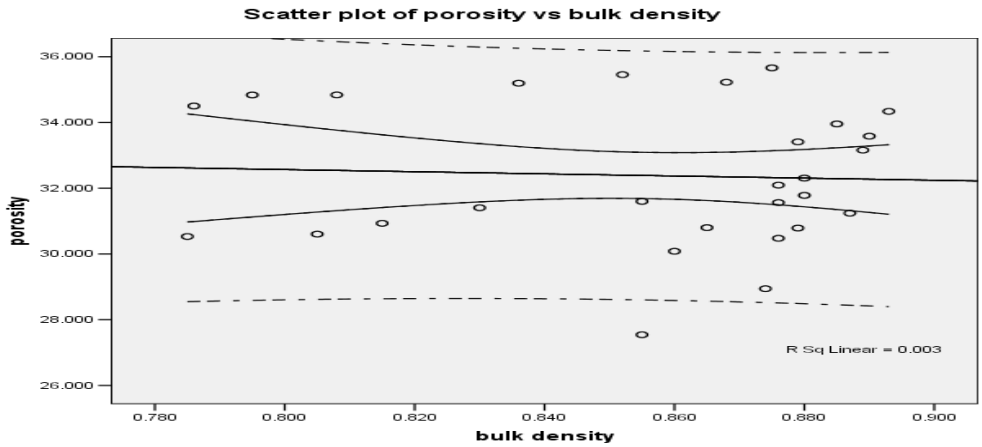


Figure 8.2 The scatter plots of bulk density vs porosity

8.4.2 Partial and Multiple Correlations and Generalized Regression Equation

A significant partial correlation (= -0.546) is observed between bulk density and porosity controlling the effect of moisture content, and a significant partial correlation (= 0.680) is observed between moisture content and porosity controlling the effect of bulk density. The zero order correlation coefficients between bulk density and porosity is found to be -0.052 (insignificant), the same between porosity and moisture content 0.485 (significant) and between moisture content and bulk density 0.645 (significant).

Table 8.2 Summary of the simple and partial correlation coefficients regarding Porosity, Moisture content and Bulk density

Correlations coefficients

Between		Zero-order	Partial	control variable
1	Moisture content & porosity	0.485	0.680	bulk density
2	Bulk density & porosity	-0.052	-0.546	moisture content
3	Moisture content& Bulk density	0.645	-	-

Table 8.3 Summary of the multiple correlation on porosity with moisture content and bulk density

Model Summary[a]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.681(a)	.463	.420	1.631919

a Predictors: (Constant), bulk density , Moisture content

Table 8.4 Summary of the linear regression coefficients on porosity with moisture content and bulk density

Coefficients(a)

Model		Un-standardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Contant)	41.208	8.031		5.131	.000
	Moisture content	1.850	.399	.888	4.631	.000
	Bulk density	-39.428	12.111	-.624	-3.255	.003

a Dependent Variable: porosity

The correlation coefficient of porosity and the joint effect of moisture content and bulk density $R_{p.bm}$ is obtained 0.681 with the standard error of estimate 1.6319. The coefficient of determination (adjusted), R^2, is 0.42. The linear regression constant (β_0)= 41.208 with s.e. = 8.031. The regression coefficient for moisture content (β_1) = 1.850 with s. e. = 0.399. The regression coefficient for bulk density (β_2) = -39.428 with s.e.= 12.111. Thus, the linear regression equation obtained in classical approach is

Porosity = 41.208 + 1.850 moist. cont. - 39.428 bulk dens.

8.5 Model Assessment

For the proposed model, the data have been analyzed using WinBUGS. The results of the posterior distribution so obtained are presented in the Table 8.5.

Table 8.5 WinBUGS results for the parameters in NIG model

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
beta0	35.87	7.7240	0.1387	19.770	36.11	50.650	1001	4000
beta1	1.71	0.4074	0.0067	0.901	1.71	2.492	1001	4000
beta2	-31.00	11.4300	0.1955	-53.100	-31.25	-7.585	1001	4000
s	1.70	0.2596	0.0053	1.302	1.66	2.282	1001	4000
s2	2.95	0.9462	0.0192	1.695	2.77	5.206	1001	4000
tau	0.37	0.1054	0.0021	0.193	0.36	0.590	1001	4000
mu[1]	30.95	0.7976	0.0150	29.31	30.95	32.46	1001	4000
mu[2]	31.35	0.6128	0.0116	30.12	31.36	32.50	1001	4000
mu[3]	31.56	0.5270	0.0100	30.48	31.56	32.55	1001	4000
mu[4]	31.60	0.4392	0.0081	30.73	31.60	32.46	1001	4000
mu[5]	31.34	0.4117	0.0066	30.54	31.34	32.17	1001	4000
mu[6]	31.08	0.5034	0.0073	30.11	31.07	32.10	1001	4000
mu[7]	32.28	0.3957	0.0055	31.52	32.27	33.08	1001	4000
mu[8]	31.26	0.4248	0.0068	30.43	31.25	32.10	1001	4000
mu[9]	32.90	0.3433	0.0054	32.22	32.89	33.58	1001	4000
mu[10]	32.91	0.3546	0.0054	32.21	32.91	33.62	1001	4000
mu[11]	33.17	0.4176	0.0062	32.34	33.17	34.02	1001	4000
mu[12]	33.60	0.4714	0.0072	32.66	33.59	34.55	1001	4000
mu[13]	34.19	0.5786	0.0091	33.04	34.18	35.32	1001	4000
mu[14]	35.54	0.8630	0.0140	33.83	35.54	37.22	1001	4000
mu[15]	30.41	0.5987	0.0091	29.24	30.40	31.61	1001	4000
mu[16]	30.69	0.5529	0.0083	29.60	30.68	31.79	1001	4000
mu[17]	30.77	0.5490	0.0081	29.69	30.76	31.87	1001	4000
mu[18]	31.03	0.5403	0.0079	29.99	31.03	32.13	1001	4000
mu[19]	30.97	0.5586	0.0082	29.89	30.96	32.10	1001	4000
mu[20]	31.80	0.4740	0.0067	30.87	31.79	32.75	1001	4000
mu[21]	33.07	0.5014	0.0074	32.11	33.07	34.09	1001	4000
mu[22]	33.14	0.7092	0.0137	31.68	33.16	34.49	1001	4000
mu[23]	33.37	0.6558	0.0126	31.99	33.39	34.63	1001	4000
mu[24]	32.97	0.5381	0.0105	31.84	32.97	34.00	1001	4000
mu[25]	32.61	0.3623	0.0068	31.87	32.61	33.32	1001	4000
mu[26]	33.66	0.4336	0.0073	32.78	33.66	34.52	1001	4000
mu[27]	34.02	0.5140	0.0082	32.99	34.02	35.03	1001	4000
mu[28]	34.48	0.6195	0.0099	33.27	34.48	35.69	1001	4000
SSE	76.12	8.6300	0.1523	67.23	73.71	98.53	1001	4000
MSE	2.72	0.3082	0.0054	2.40	2.63	3.52	1001	4000
R2	0.4613	0.1974	0.003338	0.1447	0.4388	0.9087	1001	4000
gainRE	0.3339	0.2136	0.004343	-0.1722	0.3754	0.6175	1001	4000

Plots:

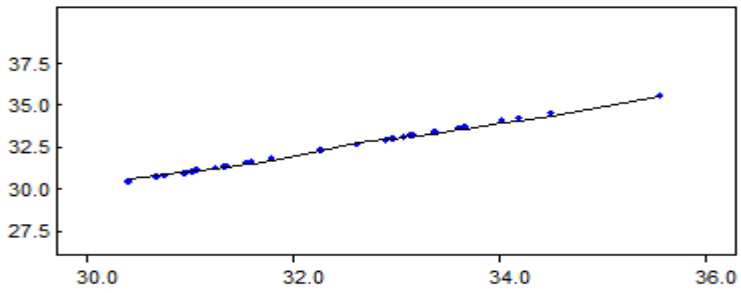


Figure 8.3 Scatter plot of the predicted values of porosity (mu)

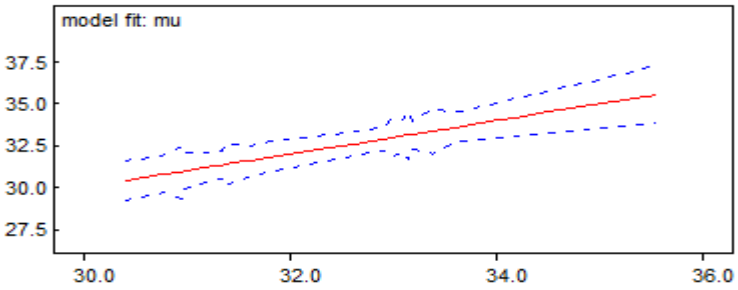


Figure 8.4 Plot of fitted model with credible region for predicted values of porosity

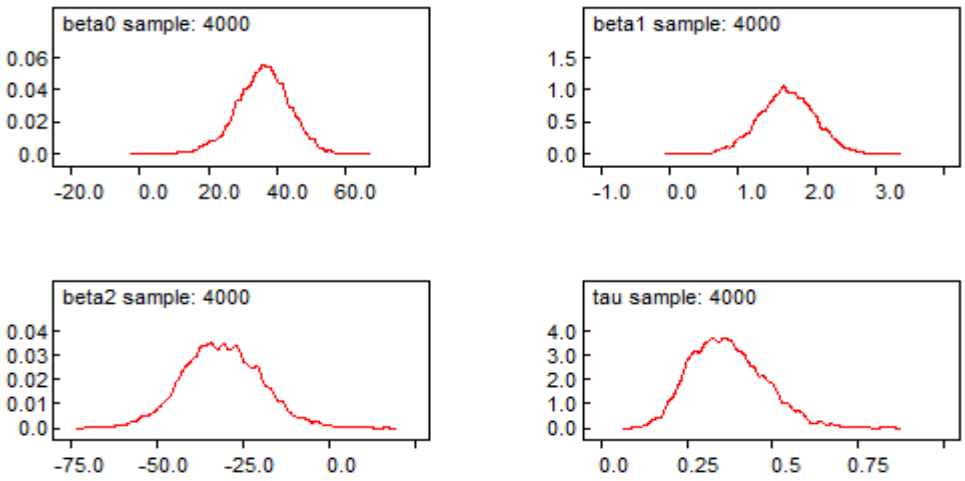


Figure 8.5 Density plots of regression coefficients and precision

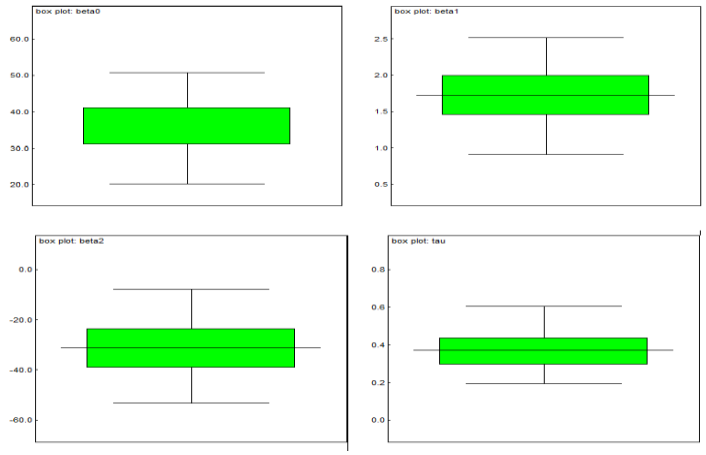


Figure 8.6 Box plots of the regression coefficients and precision

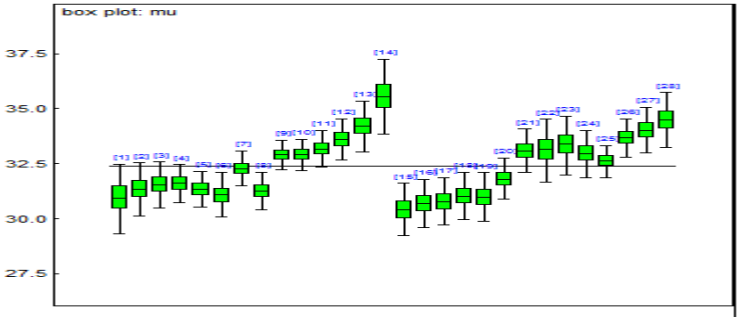


Figure 8.7 Box plots of the predicted values of porosity

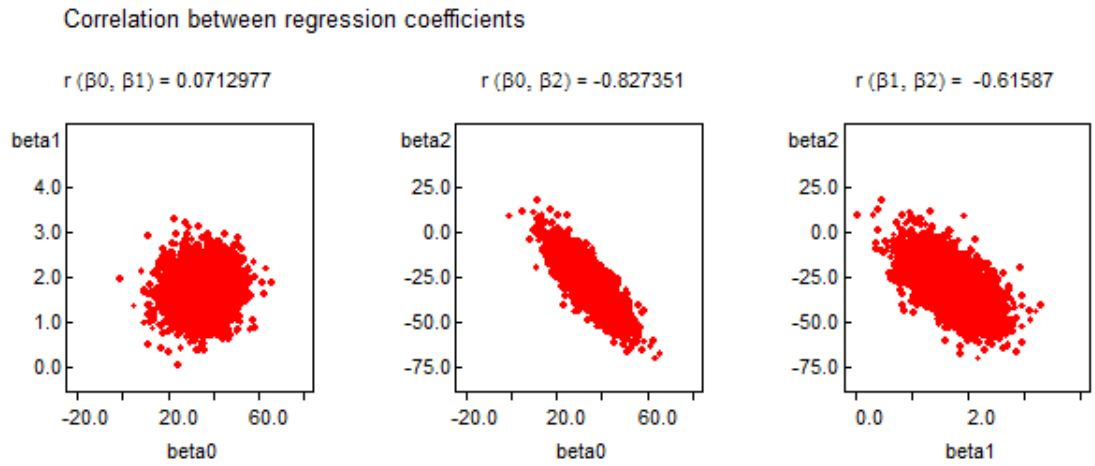


Figure 8.8 Correlation plots of regression coefficients

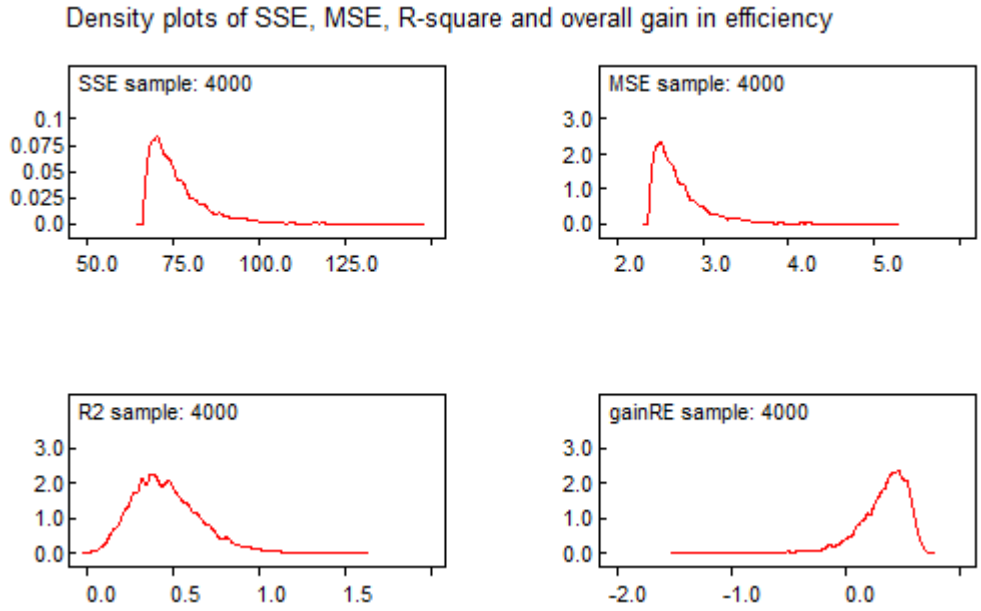


Figure 8.9 Density plots of SSE, MSE, R-square and overall gain in relative efficiency of the fitted model

8.5.4 Summary

The posterior summaries and densities are presented in Table 8.5, after running MCMC algorithm for 5000 iterations and discarding initial 1000 ones. The analysis of the posterior distribution indicates a considerable improvement of the precision in the prediction of porosity when bulk density and the moisture content are taken as independents for the rice variety. The Bayesian analysis gives the linear regression model of porosity in terms of moisture content and bulk density as

$$Porosity_b = 35.87 + 1.71\ moist.\ Cont. - 31.0\ bulk\ dens\ .$$

This model has the regression coefficients smaller in comparison to classical model. The standard deviations of β_0 and β_2 in Bayesian estimation are smaller than that of the β_0 and β_2 in classical estimation. The standard deviation of β_1 is equivalent to that of the β_1 for classical estimate. The gain in relative efficiencies for β_0 and β_2 are 7.5% and 10.92% respectively. No gain in relative efficiency is observed on the parameter β_1.

The Bayesian analysis gives complete information regarding the distributions of the estimates. The complete information concerning all uncertain mean values of the regression model is obtained in this approach. The overall gain in relative efficiency of the model is calculated and found to be 0.3339 with a sd of 0.2136. The 95% credible region for the gain in relative frequency is -0.1722 to 0.6175. The Bayesian version of Mean squared error (MSE) and R square (R^2) are other common diagnostics for the regression model, the value of MSE is computed as 2.71 with sd 0.3082 and that of R-square is 0.46 with sd 0.1974. The graphs for the model fitting are given in the Figures 8.3 to 8.7. The correlation plots of the regression coefficients of the fitted model are given in the Figure 8.8. The density plots of SSE, MSE, R-square and overall relative efficiency of the fitted model are given in the Figure 8.9.

Remarks

The main concern of this study has been to modelling porosity of the Mansuli rice in terms of moisture content and bulk density to improve the design of heat and mass transfer process for increasing the drying, frying and baking quality. The modelling by normal regression using Bayesian approach seemed to be more precise. Thus, it is concluded that the results obtained from such modelling can help better in designing of heat and mass transfer process to increase the baking quality of rice.

Chapter IX

Normal Regression Models with two different priors for Estimation of Gluten content relating to Protein content in a Variety of Wheat

9.1 Introduction

The protein content in flour is the most important characteristics of wheat because of its relation to baking quality. Proteins are surface active compounds and comparable with low molecular weight emulsifiers (surfactants). They result in lowering of interfacial tension of fluid interfaces. Proteins emulsify an oil phase in water and stabilize the emulsion (Sahin and Sumnu, 2006). Gluten is the main structure forming protein in wheat flour. It helps in increasing flavour and self- life of the product. It helps to make the product soft (ibid).

In this study, two normal regression models are set up for modelling the percentage of gluten content in a variety of wheat in terms of gluten content. The proposed models are updated using the Bayesian method and the models are compared to show which model fits well.

9.2 Models

The gluten content in the wheat flour is expressed in terms of proportion (percentage), so the normal model is adopted for the samples drown from the lot of large size in same variety. For normal regression model, two different priors are selected and the posterior densities of the parameters are summarized. Based on Bayesian version of MSE and R-square the models are compared using WinBUGS.

Model 1: Normal Regression Model with Uniform (a flat) Prior

Likelihood: The proportion (percentage) of gluten content in a wheat variety,

$Yi \sim N(\mu_i, \sigma^2)$, where, $\mu i = \alpha + \beta x_i$ (x_i is the proportion (or percentage) of protein content)

Priors: $\alpha \sim Uniform(0, 1)$, $\beta \sim Uniform(0, 1)$ and $1/\sigma^2 = \tau \sim Gamma(0.1, 0.1)$

Model 2: Normal Regression Model with Normal (noninformative /vague) Prior

Likelihood: The proportion (percentage) of gluten content in a wheat variety,

$Yi \sim N(\mu_i, \sigma^2)$, where, $\mu i = \alpha + \beta x_i$ (x_i is the proportion (or percentage) of protein content)

Priors: $\alpha \sim Normal(0, 1000)$, $\beta \sim Normal(0, 1000)$ and $\tau \sim gamma(0.1, 0.1)$

The WinBUGS code for the computation for normal regression model with uniform prior is given in the Appendix B-5 and that for the normal regression model with non-informative normal prior is given in the Appendix B-6.

9.3 Sample and Data

Independent samples of size 20 were selected for a variety of wheat to study the relationship between protein and gluten presented in the Appendix A-7. The samples were drawn from the Mahalaxmi flour mill, Sonapur.

9.4 Summarizing Data

The summary statistics of protein and gluten content is computed and given in the Table 9.1. The summary of the correlation and regression coefficients between proportion (percentage) of protein content (x) and gluten content (y) with the standard errors of the estimates are calculated and presented in the Table 9.2. The regression equation of the gluten content in terms of protein content is obtained as

$Y=0.29+0.38X$.

Table 9.1 Summary statistics of the percentage protein and gluten content

Content	Mean	S. D.	S. E. of Mean	Variance	Min	Max	Range
protein	13.3205	2.27762	.50929	5.188	9.11	16.67	7.56
gluten	5.3600	.92653	.20718	.858	3.44	7.21	3.77

Table 9.2 Summary of the correlation and regression coefficients between percentage of protein and gluten content

Between	r	se (r)	R^2	Regression coefficients y_1 on x_1	se(α)	se (β)
(x_1 , y_1)	0 .935*	0.337	0.875	α=.29, β= 0.38	0.458	0.034

*significant correlation

9.5 Model Assessment

9.5.1 Assessment of Model 1

The posterior densities of the parameters of the first model are obtained using WinBUGS, as given in Table 9.3. The density plots and trace plots the distribution of alpha, beta and tau for the model 1 are given in the Figures 9.1-9.3. The box plots of alpha and beta are given in the Figure 9.4. Based on the posterior distribution of the values of the parameters, the fitted model1 is

$$\hat{Y} = 0.4896 + 0.4955X$$

Table 9.3 Summary of the posterior density of the parameters in model 1

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
alpha	0.4896	0.2905	0.004221	0.02324	0.4834	0.9729	1	5000
beta	0.4955	0.2888	0.004459	0.02575	0.4931	0.9743	1	5000
tau	0.9764	3.1250	0.041170	2.86E-16	0.0047	10.19	1	5000
mu[1]	5.004	2.649	0.04046	0.64	4.982	9.43	1	5000
mu[2]	6.713	3.641	0.05578	0.76	6.676	12.75	1	5000
mu[3]	5.995	3.224	0.04934	0.70	5.962	11.35	1	5000
mu[4]	7.620	4.168	0.06392	0.82	7.575	14.53	1	5000
mu[5]	8.368	4.603	0.07064	0.85	8.319	16.00	1	5000
mu[6]	6.228	3.359	0.05143	0.72	6.197	11.80	1	5000
mu[7]	8.259	4.540	0.06966	0.85	8.215	15.79	1	5000
mu[8]	8.165	4.485	0.06882	0.85	8.122	15.61	1	5000
mu[9]	7.605	4.160	0.06379	0.82	7.560	14.50	1	5000
mu[10]	5.514	2.945	0.04503	0.67	5.484	10.41	1	5000
mu[11]	8.750	4.825	0.07407	0.87	8.695	16.75	1	5000
mu[12]	6.124	3.299	0.05049	0.71	6.093	11.60	1	5000
mu[13]	6.406	3.463	0.05303	0.73	6.371	12.15	1	5000
mu[14]	7.774	4.258	0.06530	0.83	7.731	14.84	1	5000
mu[15]	6.366	3.440	0.05267	0.72	6.332	12.07	1	5000
mu[16]	6.332	3.420	0.05236	0.72	6.299	12.01	1	5000
mu[17]	8.343	4.589	0.07042	0.85	8.296	15.95	1	5000
mu[18]	6.114	3.293	0.05040	0.71	6.082	11.58	1	5000
mu[19]	8.532	4.699	0.07211	0.86	8.480	16.32	1	5000
mu[20]	7.590	4.151	0.06365	0.82	7.545	14.47	1	5000

Plots:

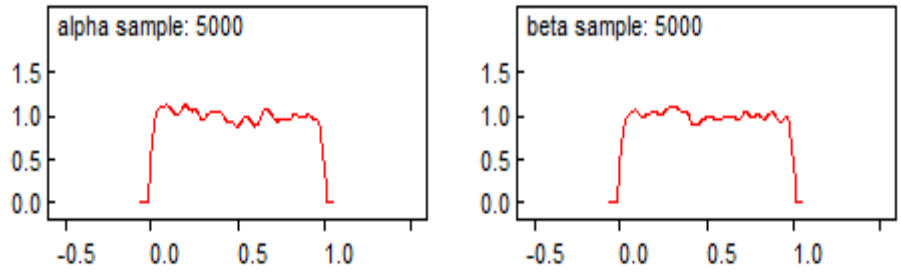


Figure 9.1 Posterior density plots of alpha and beta in model 1

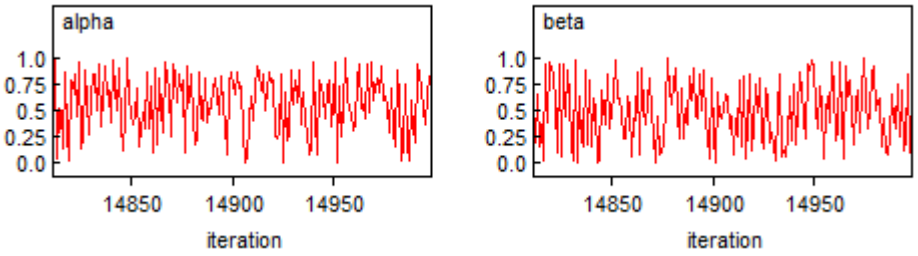


Figure 9.2 Trace plots of alpha and beta for the 200 iterations in model 1

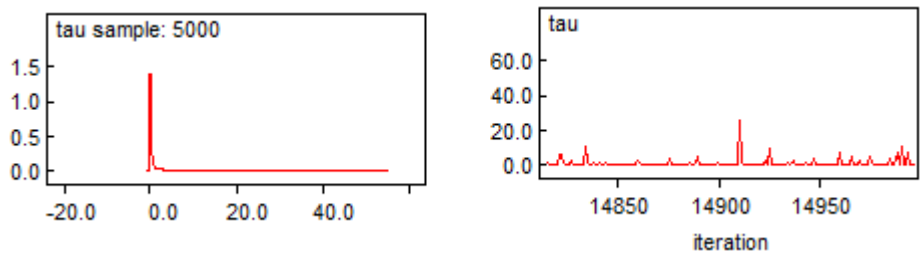


Figure 9.3 The density plot and trace of the tau in model 1

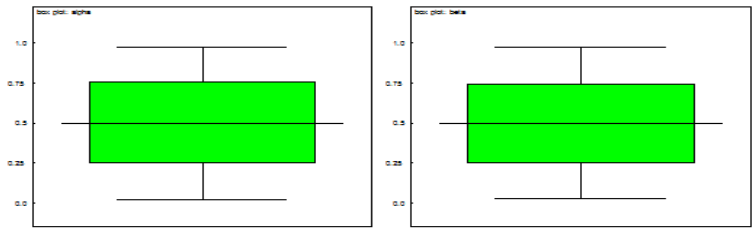


Figure 9.4 Box plots of alpha and beta for model 1

9.5.2 Assessment of Model 2

The posterior densities for the second model obtained using WinBUGS are given in the Table 9.4. The density plots and trace plots of alpha, beta and tau are given in the Figures 9.5 – 9.7. The box plots of alpha and beta are given in the Figures 9.8. Based on the posterior distribution of the values of the parameters, the fitted model 2 is

$$\hat{Y} = -0.9193 + 0.4358X \ .$$

Table 9.4 Summary of the posterior density of the parameters in model 2

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
alpha	-0.9193	31.77	0.4556	-63.07	-0.9823	60.66	501	5000
beta	0.4358	31.73	0.5518	-60.56	0.1735	63.80	501	5000
tau	0.9889	3.097	0.0418	2.39E-15	0.0049	9.57	501	5000
mu[1]	3.051	291.40	5.0710	-558.90	1.171	587.60	501	5000
mu[2]	4.554	400.40	6.9690	-772.80	0.966	808.70	501	5000
mu[3]	3.922	354.60	6.1710	-681.10	0.316	717.00	501	5000
mu[4]	5.352	458.30	7.9770	-883.50	1.401	925.20	501	5000
mu[5]	6.010	506.10	8.8090	-973.40	1.282	1021.00	501	5000
mu[6]	4.127	369.40	6.4300	-711.70	0.334	746.80	501	5000
mu[7]	5.914	499.20	8.6880	-959.80	1.580	1008.00	501	5000
mu[8]	5.831	493.20	8.5830	-948.10	1.516	995.50	501	5000
mu[9]	5.339	457.40	7.9610	-881.80	1.476	923.40	501	5000
mu[10]	3.500	324.00	5.6370	-620.70	0.930	654.80	501	5000
mu[11]	6.345	530.50	9.2340	-1019.00	1.071	1069.00	501	5000
mu[12]	4.036	362.80	6.3140	-698.00	0.125	733.50	501	5000
mu[13]	4.284	380.80	6.6280	-734.20	0.818	769.50	501	5000
mu[14]	5.487	468.20	8.1480	-900.90	1.344	944.30	501	5000
mu[15]	4.249	378.30	6.5840	-729.30	0.711	764.50	501	5000
mu[16]	4.219	376.10	6.5450	-724.90	0.576	760.00	501	5000
mu[17]	5.988	504.60	8.7820	-970.30	1.401	1018.00	501	5000
mu[18]	4.027	362.20	6.3030	-696.70	0.194	732.20	501	5000
mu[19]	6.154	516.60	8.9910	-993.60	2.019	1042.00	501	5000
mu[20]	5.326	456.40	7.9440	-880.10	1.550	921.50	501	5000

Density Plots

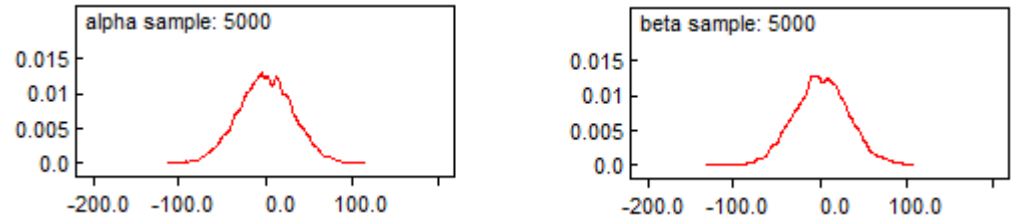


Figure 9.5 Posterior density plots of alpha and beta in model 2

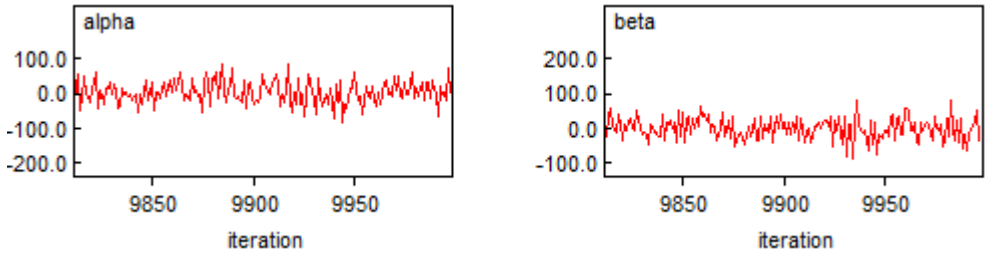


Figure 9.6 Trace plots of alpha and beta for the 200 iterations in model 2

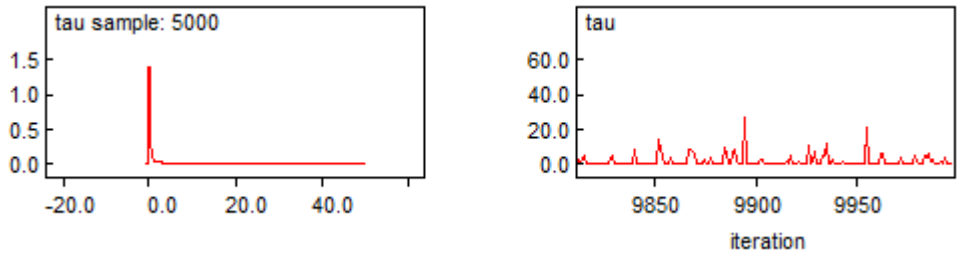


Figure 9.7 The density plot and trace of the tau in model 2

Box plots

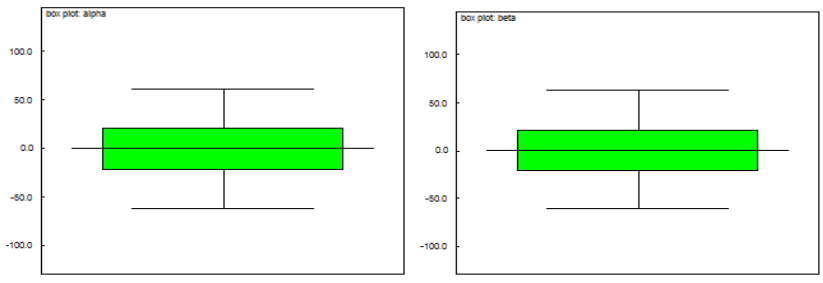


Figure 9.8 Box plots of alpha and beta for model 2

9.6.4 Model comparison and Summary

For the model comparison, the Bayesian version of MSE and R-square are calculated, which are found to be 3.16 and 0.318 respectively for the first model and 0.3419 and 0.903 for the second model respectively.

From results, it can be concluded that the second model seems to be fitted better than the first model, since it has large R^2 value and less mean square errors of the estimate than the first model. The second model gives the predicted values very close to that obtained in classical regression due to the use of a non-informative normal prior with large variance. However, model 2 has heavy tail distribution because of the large values of standard deviations of the estimates. The first model can be considered fitted well in this grounds that it has the posterior density of the parameter sharper than the second model. Whatever, the values of MSE and R^2 are obtained, it can predict the parameter with great precision than the second because of more consistency in the distribution. For the further analysis, the Bayesian information criteria (BIC) and Akaike Information criteria (AIC) are calculated using WinBUGS for both the models. The BIC and AIC for model1 and model2 are obtained as

Model	BIC	AIC
model 1	236.3	232.3
model 2	231.4	227.4

According to Kontoleon and Yabe (2006), the model 2 is selected as best fitted model which has minimum BIC value than that of the model 1. The box plots of the predicted y_i (i. e., mu) are given in the Figures 9.9 -9.10. The scatter plots and the fitted line for mu are given in the Figures 9.11-9.12 and 913- 9.14 respectively.

Plots for model comparison:

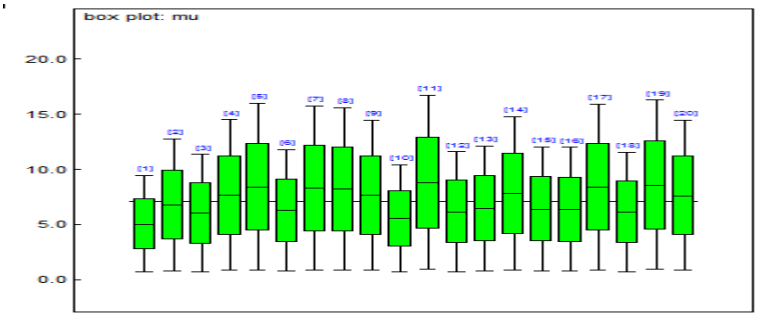


Figure 9.9 Box plots of posterior mu in model 1

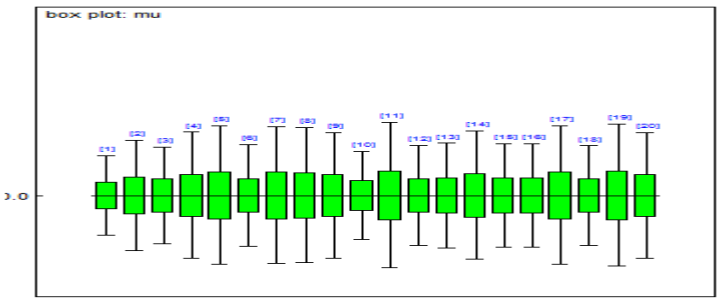


Figure 9.10 Box plots of posterior mu in model 2

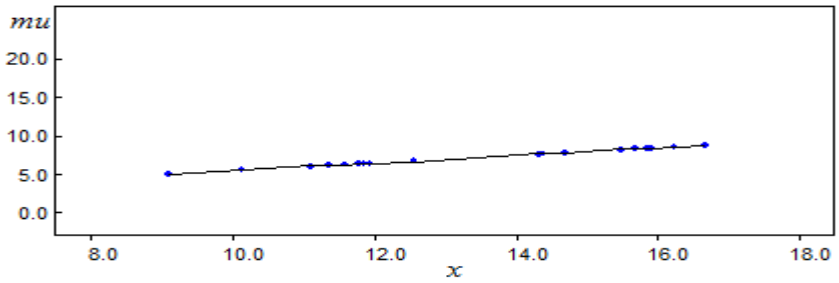


Figure 9.11 Scatter plot of mu in model 1

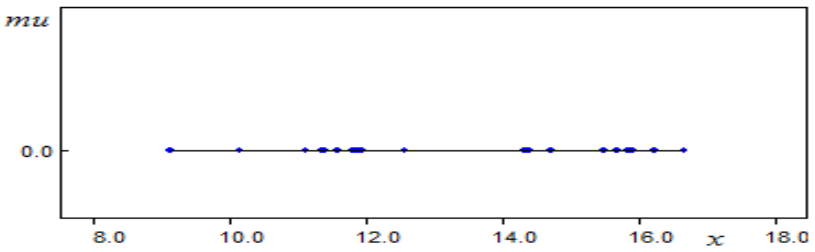


Figure 9.12 Scatter plot of mu in model 2

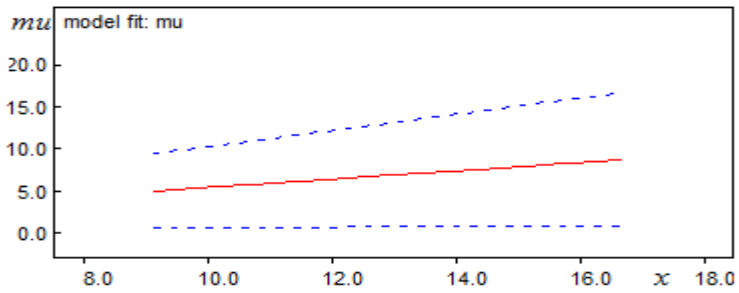


Figure 9.13 Plots of fitted model with 95% credible region for model 1

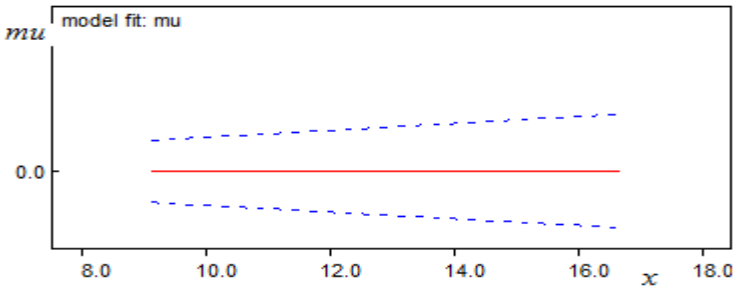


Figure 9.14 Plots of fitted model with 95% credible region for model 2

Remarks:

The major question of this study has been to obtain the better model for determination of gluten content in terms of protein content in a variety of wheat flour to increase the self life and the flavour of the product. From the results of the study, it can be concluded that the modelling the gluten content in terms of protein content is better in normal regression model with normal non-informative prior.

Chapter X

Conclusions and Recommendations

10.1 Preliminary

Bayesian statistical methods are quite unknown methods in Nepal; one may rarely meet these methods even in the scientific literature. Moreover, Bayesian method is being as an uninterested approach due to ignorance, lack of the computational skills, not including it in academic curriculum and misinformation about it. Applications of this method for quality control problems and industrial decisions are nonexistence.

Consciously, knowing the challenges of using Bayesian methods, this study has been made deeply affected by emotionally inspired for its use. The most demeaned point in application of Bayesian approach was the word 'prior'. More about priors and their uses have been discussed in chapter 3. In most of the contributions, non-informative priors have been used to mitigate the influence of priors on their results. However, other priors such as subjective/conjugate priors (as discussed in chapter 3.2.6) are equally useful, in case the sources of the priors are based on the prior experimental results. The other main anxiety in application of Bayesian analysis is treating parameter as a random. Treating theta (parameter) as a random variable does not necessarily mean that theta is random; to a certain extent, it expresses ones uncertainty about theta. It is the quantity of interest.

10.2 Conclusions:

In this study, different techniques of predicting the parameters with a complete probabilistic distribution have been introduced rather than building a complicated model to match every portion of the observed data or taking in unnecessarily collection of expensive test data.

Graphic exploration and interactive discovery have been exposed, which can help in identifying patterns in the data that may be hidden by descriptive statistics alone. The techniques of modeling using Bayesian paradigm have been investigated and models have been built to help increase the accuracies of predictants. Further, it has been concluded and suggested that these procedures will be helpful for modelling product quality as consumer's concern.

It is commonly known to us that the classical statistical tools provide less precise support for decision making in presence of high uncertainty regarding parameter space and a small sized sample. In the Bayesian approaches, several techniques can be found set up that enable to handle these conditions in straightforward way.

It has been exposed that the Bayesian decision model with appropriate adaptation of Gibbs sampler with the development of MCMC methods provides adequate tools for solving the quality concerned problems. Also, it has been made known that the Bayesian prediction model offers support for decision related to the designing and quality issues of the product.

In the study, it is also exposed that the implementation of the hierarchical model can make the application of conjoint analysis possible, which is difficult in traditional approaches even when hierarchical set up exists in the concerned problems. Conclusively, it has been recommended that the application of the Bayesian procedures using MCMC technique via WinBUGS shows us the way to better inference, modelling and precise decision.

10.3 Recommendations for Further Areas of Research

During this research it has been known that there is a good neighbouring area, related to this research, having high options for the applications of the Bayesian methods in prediction, decision making and modelling uncertainties are applicable. Some of them are:

- Application of the suggested methods in plant based studies, usage of these methods in companies' decision-making
- The study of management aspects of quality (TQM) in a Bayesian approach
- The study of the consumer behaviour, when questions about the characteristics of the consumers are available, by the use of the random coefficients model
- The study of the investigation of conjoint models in quality control aspects
- Application of Bayesian methods using MCMC technique to offline process control
- The study of quality heterogeneity in food and food products using GLM based extensions
- Adoption of Mixture models for the analysis of quality control problems and consumer preferences
- Application of Bayesian methods using MCMC technique to a real process modelling
- Application of Bayesian methods using MCMC technique to sampling inspection of the products
- Bayesian approach to time series data in quality control and industrial sectors

Broader application of the Bayesian methods is open for all in all branches of sciences and knowledge, where statistical procedures are applicable.

***

Summary

The purpose of this research work is to devise appropriate methods that would lead us to monitor the quality of food statistically. The main concerns of the study are to formulate methods of inferences, develop prediction method and model the relations between the characteristics regarding some aspects of food production and quality, using Bayesian paradigm.

Bayesian methods are used in this research, with the anticipation that it can motivate us to recognize new methods for inference in food production and quality concerns and to make the predictions in straightforward way. The two major challenges of this work were (i) to set up the prior and combine it to the likelihood in such a way that the posterior would be a proper probability distribution and (ii) to use MCMC via WinBUGS and interpret posterior density.

This study focuses mainly on the functional aspects, which means it is looked at what the functions of Bayesian methods are, and how these functions can be used to prediction and modelling some aspects of food quality concerns.

The definitions of functions of Bayesian methods have been reviewed from different literatures (Chapter III). The methodologies of drawing inference, prediction and modelling along with technical aspects of computation, simulations and modelling have been discussed in Chapter IV. The major contributions of the research work have been presented in chapters V - IX.

In the main contributions (part II) of the thesis, the Bayesian methods are employed for analyzing data and predicting the relations of different variables related to food quality. Particular features have been emphasized for drawing inferences, formulating prediction equations and modelling the food quality characteristics using Bayesian paradigm. Further, the necessary development of model checking for the proposed models has also been worked out.

In the following sections, the main features of each methodology implemented to the study are discussed briefly.

Normal Model with Normal Prior

The normal-normal model is used for the purpose of estimating the weights of the packaged food (ghee). It is put forward to look into the consumers' concern whether the weights of the pouches are underweight.

Initially, an attempt has been made for drawing inference with full posterior normal density regarding the weights of the packaged food using an informative (conjugate) normal prior. The Gibbs sampler with MCMC method (using WinBUGS) has supported the belief that the success of this model is largely being able to predict accurately when prior information is available based on expert's opinion.

From the posterior distribution, using conjugate normal prior with known variance it is observed that the process fails to meet the specification. Thus, a remark has been made that the process should be adjusted to the direction that could increase the average weight of the pouches.

Further, the work is done for the same using a non-informative prior. Clearly, an indication is observed that the non-informative prior has the negligible impact on the model. It is also noticed that the Bayesian method works only as the process of strengthening the updated inference. The software used facilitates mainly in generating data from the posterior distribution. As the iteration increases, the posterior density of the parameter of the interest observed to be more precise.

Beta Prior Distribution to Acceptance Sampling Plan

This study is designed for the purpose of obtaining an acceptance sampling plan for the attributes of non-conformance of the packaging of a food product (instant noodle). Firstly, the acceptance sampling plan has been developed solely based upon data taken from online sample observations. Then, an analysis of the beta prior distribution has been done from the past factory records and combined together with experimental data to develop sampling plan using the Bayesian method.

From the simulation, it is observed that when the Bayesian approach is used to defining parameters and to develop acceptance plan, there is substantial reduction in the number of samples needed for inspection. From the results, it is exposed that the sampling plans using posterior estimation are sharper than the plans from the sample

information alone. From this study, it has been remarked that the acceptance sampling plan derived using posterior distribution helps more for the assurance of the consumers' risk.

Use of Poisson-Gamma Hierarchical Model

The Poisson-Gamma Hierarchical model is used for modelling failure to operation points for a canning process in a fruit juice industry. At first, estimates of the true failure rate with the standard errors per operation point have been estimated in classical point of view.

In the Bayesian approach, a hierarchical model is developed and the data fitted with Poisson distribution. A directed acyclic graph (DAG) is developed and posterior estimates of the parameters of failures to operation points are obtained. From the Gibbs sampler using WinBUGS, a complete posterior distribution is obtained for the failure rate (θ_i) for each operation point.

The main concern of this study is to identify the priority points of the failure in operation points (CCPs) that should be corrected for the removal of high failure rate. From the result of the posterior densities, it is observed that the failure rate is highest for the storage and respectively raw material receiving station and material preparation point as second and third highest points. Average failure rate (λ) is obtained highest at the points of closer parts and respectively second and third points are storage and ingredient mixing points.

The application of the Bayesian method has been found to be superior for the identification and determination of the failure rate. It has been appeared advanced for the prediction and to check out with priority the effecting points that reduce performance of the process.

Normal Inverse-Gamma Model for the Prediction

The NIG model is used for predicting porosity of a rice variety in terms of moisture content and bulk density. The main concern of this study is to model the relations, which can be used for improving the design of heat and mass transfer process for the increase of the drying, frying and baking quality of rice.

At first, a regression model is obtained in classical approach. Secondly, the attempt is made to applying Bayesian NIG model. The posterior summaries and densities are obtained after running MCMC algorithm via WinBUGS.

A considerable improvement of the precision in the prediction of porosity has been indicated by the analysis of the posterior distribution when bulk density and the moisture content are taken into account for a rice variety. The regression coefficients for this model have been observed smaller in comparison to classical model with less standard deviation to each estimate.

In this approach, complete information has been obtained concerning the parameters of the regression model. The proposed model has been validated using graphical tools and found to be good fitted. The proposed NIG model has been put forward to use for the better designing of the concerned quality.

Normal Regression Model with Two Different Priors

This study is intended to obtain the better model for the determination of gluten content in terms of protein content in a variety of wheat flour, which the gluten, is important to increase the self life and the flavour of the product. Normal regression models with two different priors have been used for predicting proportion (percentage) of gluten content in terms of protein content.

The two priors (i) uniform (a flat prior) and (ii) non-informative normal (a vague prior) are selected as prior distributions for the regression parameter. The posterior densities of the parameters are summarized for each model and the models are compared based on Bayesian versions of MSE and R-square.

It is observed that the second model is fitted better, because of having large R^2 value and less MSE of the estimate, than the first model. It is also observed that the second model gives the predicted values very close to that the classical regression gives. It might be observed due to the use of a non-informative prior with large variance. However, this model has heavy tail distribution because of the large values of standard deviations of the estimates. For the robust decision, BIC and AIC are computed for both the models and model 2 is selected as better fitted model which has minimum BIC value than that of the model 1.

From this study, it has been concluded that the modelling the gluten content in terms of protein content seems to be better in normal regression model with normal non-informative prior in case of prior ignorance.

Towards the end

In this thesis, different techniques of predicting the parameters with a complete probabilistic distribution have been introduced rather than building a complicated model to match every portion of the observed data, or taking in unnecessarily collection of expensive test data. Graphic exploration and interactive discovery have been exposed, which can help in identifying patterns in the data that may be hidden by descriptive statistics alone.

Obviously, in case of a small sized sample and in presence of high uncertainty regarding parameter space the existing statistical method based upon data alone provide less precise support for decision making. In the proposed Bayesian approaches, several techniques can be found that enable to handle these conditions in straightforward way.

In this study, the techniques of modeling using Bayesian paradigm have been investigated and models have been built to help increase the accuracies of predictants. It is exposed that the implementation of the hierarchical model can make the application of conjoint analysis possible, which is difficult in classical approaches even when hierarchical set up exists in the concerned problems.

From the study, it is also observed that the Bayesian estimation using MCMC technique via WinBUGS leads us to better inference and precise decision. Thus, it has been concluded and suggested that these procedures will be helpful for modelling product quality as consumer's concern.

At last, some recommendations have been made for the use of Bayesian methods to the area interrelated to this research, where the prediction, decision making and modelling uncertainties are applicable.

***

Appendix

Appendix A: Data Tables

Appendix A-1:

Table showing the observed weight of pouched ghee

sample	obs.1	obs. 2	obs. 3	obs. 4	obs.5		Sum	Average	Range
1	900	902	900	898	900		4500	900	4
2	905	901	906	909	904		4525	905	8
3	915	911	918	916	910		4570	914	8
4	917	912	912	912	912		4565	913	5
5	928	926	925	926	930		4635	927	5
6	912	916	916	911	920		4575	915	9
7	902	903	895	897	903		4500	900	8
8	916	912	920	922	920		4590	918	10
9	909	911	907	906	907		4540	908	5
10	914	913	919	916	918		4580	916	6
11	915	920	922	911	922		4590	918	11
12	924	920	924	922	930		4620	924	10
13	925	926	928	922	924		4625	925	6
14	936	935	931	932	936		4670	934	5
15	929	930	930	931	925		4645	929	6
16	919	921	920	922	918		4600	920	4
17	921	923	925	929	927		4625	925	8
18	930	927	933	931	929		4650	930	6
19	925	929	930	934	932		4650	930	9
20	928	932	930	930	930		4650	930	4
21	924	924	926	926	920		4620	924	6
22	928	932	930	935	925		4650	930	10
23	932	933	938	935	932		4670	934	6
24	922	919	925	921	923		4610	922	6
25	935	934	931	935	935		4670	934	4
					average		921	921	6.76
					std dev		10.3277 (for all data)	10.1694 (sample means)	2.1656

Source: Experimental (Laboratory) Records, maintained by DDC, Balaju, March, 2008

Appendix A-2

Proportion of nonconforming items

Sample size	16	20	20	24	24	28	28	28	32	32	32	36	36	36	40	40	40	40	total
Number of item nonconforming	1	1	2	1	2	1	2	3	1	2	3	1	2	3	2	3	4	5	
frequency	3	8	2	6	1	3	4	2	2	1	2	1	6	1	3	3	1	1	50
proportion	.0625	.0500	.1000	.0417	.0833	.0357	.0714	.1071	.0313	.0625	.0938	.0278	.0556	.0833	.0500	.0750	.1000	.1250	

Source: Experimental data from Himalayan snacks, July 2009

Appendix A-3

Failures in different operation points in a time base

Operation point	Length of operation time (t_i) in hrs	Number of failures (x_i)
1	12	2
2	24	3
3	120	7
4	144	6
5	60	3
6	180	11
7	48	2
8	96	5
9	72	6
10	108	10

Source: Records obtained from seminar paper on 'In Plant study 2008' submitted to CCT, Dharan.

Appendix A-4

The sample values of different characteristics of Mansuli rice

Mill code	Sub sample	Moisture content (%)	Bulk density	True density	Porosity	Average moisture content	Average Bulk density	Average porosity
A	As1	11.35	0.785	1.13	30.531	12.564	0.836	31.31
	As2	11.95	0.805	1.16	30.603			
	As3	12.25	0.815	1.18	30.932			
	As4	12.55	0.830	1.21	31.405			
	As5	12.85	0.855	1.25	31.600			
	As6	13.15	0.880	1.29	31.783			
	As7	13.85	0.880	1.30	32.308			
B	Bs1	12.80	0.855	1.18	27.542	14.321	0.871	31.35
	Bs2	13.85	0.860	1.23	30.081			
	Bs3	13.95	0.865	1.25	30.800			
	Bs4	14.30	0.876	1.28	31.563			
	Bs5	14.55	0.876	1.29	32.093			
	Bs6	14.95	0.879	1.32	33.409			
	Bs7	15.85	0.885	1.34	33.955			
C	Cs1	12.65	0.874	1.23	28.943	13.321	0.884	31.79
	Cs2	12.85	0.876	1.26	30.476			
	Cs3	12.95	0.879	1.27	30.787			
	Cs4	13.25	0.887	1.29	31.240			
	Cs5	13.25	0.889	1.33	33.158			
	Cs6	13.75	0.890	1.34	33.582			
	Cs7	14.55	0.893	1.36	34.338			
D	Ds1	12.65	0.786	1.20	34.500	13.664	0.831	35.1
	Ds2	12.95	0.795	1.22	34.836			
	Ds3	12.95	0.808	1.24	34.839			
	Ds4	13.25	0.836	1.29	35.194			
	Ds5	14.15	0.852	1.32	35.455			
	Ds6	14.65	0.868	1.34	35.224			
	Ds7	15.05	0.875	1.36	35.662			

Source: Data adopted from the paper presented by R. R Gautam and R. Baral in seminar organized by Food Technology Subject committee, CCT Dharan, January 2010.

Appendix A-5

Summary statistics of the different characteristics of rice data

Source of rice sample	Mean	S. D	Variance	Minimum	Maximum	Range	S E of
1. Moisture content							
Mill A	12.5643	.81941	.671	11.35	13.85	2.50	.30971
Mill B	14.3214	.95388	.910	12.80	15.85	3.05	.36053
Mill C	13.3214	.64734	.419	12.65	14.55	1.90	.24467
Mill D	13.6643	.94415	.891	12.65	15.05	2.40	.35686
Total	13.4679	1.02895	1.059	11.35	15.85	4.50	.19445
2. Bulk density							
Mill A	.83571	.037129	.001	.785	.880	.095	.014033
Mill B	.87086	.010976	.000	.855	.885	.030	.004149
Mill C	.88400	.007528	.000	.874	.893	.019	.002845
Mill D	.83143	.035636	.001	.786	.875	.089	.013469
Total	.85550	.033932	.001	.785	.893	.108	.006413
3. True density							
Mill A	1.21714	.065247	.004	1.130	1.300	.170	.024661
Mill B	1.27000	.054772	.003	1.180	1.340	.160	.020702
Mill C	1.29714	.047509	.002	1.230	1.360	.130	.017957
Mill D	1.28143	.062297	.004	1.200	1.360	.160	.023546
Total	1.26643	.062552	.004	1.130	1.360	.230	.011821
4. Porosity							
Mill A	31.30889	.653325	.427	30.531	32.308	1.777	.246934
Mill B	31.34907	2.161574	4.672	27.542	33.955	6.413	.816998
Mill C	31.78932	1.945704	3.786	28.943	34.338	5.395	.735407
Mill D	35.10125	.400697	.161	34.500	35.662	1.162	.151449
Total	32.38713	2.143218	4.593	27.542	35.662	8.119	.405030

Appendix A-6

Classical ANOVA for the different characteristics of rice sample

source	Sum of Squares	df	Mean Square	F	Sig.
1. Moisture content					
Between groups (Combined)	11.235	3	3.745	5.180	.007*
Within Groups	17.351	24	.723		
Total	28.586	27			
2. Bulk density					
Between groups (Combined)	.014	3	.00471	6.669	.002*
Within Groups	.017	24	.00071		
Total	.031	27			
3. True density					
Between groups (Combined)	.025	3	.008	2.515	.082
Within Groups	.080	24	.003		
Total	.106	27			
4. Porosity					
Between groups (Combined)	69.748	3	23.249	10.281	.000*
Within Groups	54.273	24	2.261		
Total	124.021	27			

*significance at 1% level of significance

Appendix A-7

Percentage of protein content and gluten content in variety of wheat

sample no.	Protein	Gluten		sample no.	Protein	Gluten
1	9.11	3.44		11	16.67	7.21
2	12.56	5.48		12	11.37	4.22
3	11.11	4.49		13	11.94	5.10
4	14.39	5.74		14	14.70	5.75
5	15.90	6.45		15	11.86	4.47
6	11.58	4.35		16	11.79	5.15
7	15.68	6.35		17	15.85	6.41
8	15.49	5.76		18	11.35	4.68
9	14.36	5.78		19	16.23	6.03
10	10.14	4.77		20	14.33	5.57

Source: Mahalaxmi Flour Mill, Sonapur

Appendix B: WinBUGS Code

Appendix B-1

WinBUGS code for Bayesian Inference with Informative Normal prior

```
    # Model
        {
        for(i in 1:N)
        {
        wt [i]~dnorm(theta, phai)
        }
        theta~dnorm(930, tau)
        phai<-1/pow(sigma,2)
        tau<-k/pow(sigma,2)
        sigma<-10
        k<-5
        }
    # Initial values
        list(theta=930)
    # Data
        list(wt=c(900,905,914,913,927,915,900,918,908,916,922,924,925,934,929,920,925,
        930,930,930,924,930,934,922,934),N=25)
```

Appendix B-2

WinBUGS code for Bayesian inference using Non-informative Prior

```
# Model
        {
        for(i in 1:N)
        {
        wt[i]~dnorm(theta, phai)
        }
        theta~dnorm(0, 0.000001)
        phai<-1/pow(sigma, 2)
        sigma<-10
        }
# Initial value
        list(theta=0)
# Data
        list(wt=c(900,905,914,913,927,915,900,918,908,916,922,924,925,934,929,920,925,
        930,930,930,924,930,934,922,934),N=25)
```

Appendix B-3

WinBUGS code for Poisson-gamma hierarchical Model

```
#Model
        {
        for(i in 1:N)
        {
        theta[i] ~ dgamma(alpha, beta)
        lambda[i] <- theta[i] * t[i]
        x[i] ~ dpois(lambda[i])
        }
        alpha ~ dexp(1)
        beta ~ dgamma(0.1, 1.0)
        }
        #Data
        list(t = c(12, 24, 120, 144, 60, 180, 48, 96, 72, 36),
            x = c( 2,     3,     7,   6,     3,    11,     2,      5,    6,    10), N = 10)

#initials
        list(alpha = 1, beta = 1)
```

Appendix B-4

WinBUGS code for the NIG model specification in normal regression of the characteristics of rice data

```
#likelihood
model{
        for(i in 1:n){
        porosity[i]~dnorm(mu[i], tau)
        mu[i]<-beta0+beta1*mois[i]+beta2*bulkd[i]
        numerator[i]<-(mu[i]-mean(porosity[]))*(mu[i]-mean(porosity[]))
        denominator[i]<-(porosity[i]-mean(porosity[]))*(porosity[i]-mean(porosity[]))
        se[i]<-(mu[i]-porosity[i])*(mu[i]-porosity[i])
}
#prior
        tau~dgamma(0.01, 0.01)
        beta0~dnorm(0.0, 1.0E-3)
        beta1~dnorm(0.0, 1.0E-3)
        beta2~dnorm(0.0, 1.0E-3)
        s2<-1/tau
        s<-sqrt(s2)
        R2<-sum(numerator[])/sum(denominator[])
        SSE<-sum(se[])
        MSE<-mean(se[])
        gainRE<-1-((s2)/(sum(denominator[])/n))
}
#initials
        list(tau=1, beta0=1, beta1=0, beta2=0)
#Data
        list(n=28,
        porosity=c(30.531, 30.603, 30.932, 31.405, 31.600, 31.783, 32.308, 27.542, 30.081,
        30.800, 31.563, 32.093, 33.409, 33.955, 28.943, 30.476, 30.787, 31.240, 33.158,
        33.582, 34.338, 34.500, 34.836, 34.839, 35.194, 35.455, 35.224, 35.662),
        mois=c(11.35, 11.95, 12.25, 12.55, 12.85, 13.15, 13.85, 12.80, 13.85, 13.95, 14.30,
        14.55, 14.95, 15.85, 12.65, 12.85, 12.95, 13.25, 13.25, 13.75, 14.55, 12.65, 12.95,
        12.95, 13.25, 14.15, 14.65, 15.05),
        bulkd=c(.785, .805, .815, .830, .855, .880, .880, .855, .860, .865, .876, .876, .879,
        .885, .874, .876, .879, .887, .889, .890, .893, .786, .795, .808, .836, .852, .868,
        .875))
```

Appendix B-5

Normal regression model with uniform (a flat) prior

```
# Model{
for(i in 1:n){
glut[i]~dnorm(mu[i], tau)
mu[i]<-αλπηα+βετα*prot[i]
log.like[i] <- -0.5*log(2*Pi)-0.5*log(s2)-0.5*(glut[i]-mu[i])*(glut[i]-mu[i])/s2
like[i] <- exp( log.like[i] )
}
Pi<-3.141593
dm <- 2 + g[1] + g[2]
                    Deviance <- -2*sum(log.like[1:n])
                    AIC <- Deviance + dm*2
                    BIC <- Deviance + dm*log(n)
                    L <- prod( like[1:n] )
# Prior
tau~dgamma(0.1, 0.1)
αλπηα~dunif(0,1)
βετα~dunif(0, 1)
s<-1/sqrt(tau)
s2<-1/tau
pi<-3.141593
}
#Data
list(n=20, g=c(1,1),prot=c(9.11, 12.56, 11.11, 14.39, 15.90, 11.58, 15.68, 15.49, 14.36,
10.14, 16.67, 11.37, 11.94, 14.70, 11.86, 11.79, 15.85, 11.35, 16.23, 14.33)
glut=c(3.44, 5.48, 4.49, 5.74, 6.45, 4.35, 6.35, 5.76, 5.78, 4.77, 7.21, 4.22, 5.10, 5.75, 4.47,
5.15, 6.41, 4.68, 6.03, 5.57))
# Initials
list( tau=1, αλπηα=1, βετα=0)
# use g to define the fitted model
# m1 g=c(0,0)
# m2 g=c(1,0)
# m3 g=c(0,1)
# m4 g=c(1,1)
```

Appendix B-6

Normal regression model with non-informative prior

```
# Model{
for(i in 1:n){
glut[i]~dnorm(mu[i], tau)
mu[i]<-αλπηα+βετα*prot[i]
log.like[i] <- -0.5*log(2*Pi)-0.5*log(s2)-0.5*(glut[i]-mu[i])*(glut[i]-mu[i])/s2
like[i] <- exp( log.like[i] )
}
Pi<-3.141593
dm <- 2 + g[1] + g[2]
                    Deviance <- -2*sum(log.like[1:n])
                    AIC <- Deviance + dm*2
                    BIC <- Deviance + dm*log(n)
                    L <- prod( like[1:n] )
# Prior
tau~dgamma(0.1, 0.1)
αλπηα~dnorm(0.0,1.0E-3)
βετα~dnorm(0.0, 1.0E-3)
s<-1/sqrt(tau)
s2<-1/tau
pi<-3.141593
}
#Data
list(n=20, g=c(1,1),prot=c(9.11, 12.56, 11.11, 14.39, 15.90, 11.58, 15.68, 15.49, 14.36,
10.14, 16.67, 11.37, 11.94, 14.70, 11.86, 11.79, 15.85, 11.35, 16.23, 14.33)
glut=c(3.44, 5.48, 4.49, 5.74, 6.45, 4.35, 6.35, 5.76, 5.78, 4.77, 7.21, 4.22, 5.10, 5.75, 4.47,
5.15, 6.41, 4.68, 6.03, 5.57))
# Initials
list( tau=1, αλπηα=1, βετα=0)
# use g to define the fitted model
# m1 g=c(0,0)
# m2 g=c(1,0)
# m3 g=c(0,1)
# m4 g=c(1,1)
```

Appendix C: Theoretical Methods for Computation

Appendix C-1

The overview of the method of Monte Carlo simulation

Let, f(x) be a function and we have to compute a complex integral

$$I = \int_a^b f(x)\ dx$$

If we can decompose *f(x)* into production of a function *g(x)* and a probability density function *p(x)* defined over the interval *(a, b),* then we have

$$I = \int_a^b f(x)\ dx = \int_a^b g(x)\ p(x)\, dx \quad = \quad E_{p(x)}[g(x)]$$

$E_{p(x)}[g(x)]$ is the expectation of g(x) over the density p(x).

If we draw $x_1, x_2, \ldots\ldots\ldots\ldots, x_n$ random variables, large n, from the density p(x) then

$$I = \int_a^b f(x)\ dx = \int_a^b g\ (x)\ p(x)\, dx \quad = \quad E_{p(x)}[g(x)],$$

It is known as Monte Carlo integration.

In case of Bayesian analysis, Monte Carlo integration can be used to approximate marginal posterior distribution.
Let,

$$I\ (y) = \int g(y|x)\ \ p(x)\ \ dx,$$

we approximate it as

$$\hat{I}(y) = \frac{1}{n}\sum_i^n g(y|x_i),$$

x_i*'s* are drawn from the density p(x).
The standard error for the estimation in MC method is

$$SE\{\hat{I}(y)\} = \sqrt{\frac{1}{n}\left[\frac{1}{n-1}\sum_{i=1}^{n}\left(g\left(y\,\middle|\,x_i\right) - \hat{I}(y)\right)^2\right]}$$

The up to date over view of this method is well available in Smith and Robert (1993).

*

Appendix C-2

Method of Importance Sampling:

Let, the density *g(x)* is the density of interest which is approximated roughly by *p(x)* then,

$$\int g(x)\ q(x)\ dx = \int g\ (x) \left(\frac{q(x)}{p(x)}\right) p(x)\ dx$$

$$= E_{p(x)} \left[g\ (x)\ \frac{q\ (x)}{p\ (x)} \right]$$

By using Monte Carlo integration, we can write

$$\int g(x)\ q(x)\ dx \simeq \frac{1}{n} \sum_{i=1}^{n} g\ (x_i) \left(\frac{q\ (x_i)}{p\ (x_i)}\right),$$

where x_i's are drawn from the distribution given by *p(x)* [i.e., $x_i \sim p(x)$]. It works well when sampling from proposal is easy and target is hard. It is the basis of Importance Sampling.

The marginal density as a function of y is given by

$$J(y) \simeq \frac{1}{n} \sum_{i=1}^{n} g\left(y \middle| x_i\right) \left(\frac{q\ (x_i)}{p\ (x_i)}\right)$$

An alternative form of the importance sampling (Carlin and Louis, 1996) is defined with weight function *w,*

$$\int g(x)\ q\ (x)\ dx \simeq \hat{I} = \frac{\sum_{i=1}^{n} w_i\ g(x_i)}{\sum_{i=i}^{n} w_i},$$

where, the weight $w_i = \frac{q(x_i)}{p\ (x_i)}$

It works well when *q* (*x*) is known up to a multiplicative constant.

It has an associated Monte Carlo variance of

$$Var\ (\hat{I}) = \frac{\sum_{i=1}^{n} w_i \left(g\ (xi) - \hat{I}\ \right)^2}{\sum_{i=1}^{n} w_i}$$

*

Appendix C-3

Method of Gibbs sampling:

Gibbs sampler can be thought of a stochastic analogue to be EM (Expectation and Maximization) approach used to obtain likelihood function when missing data are present. In the Gibbs sampler random sampling replaces the expectations and maximization step, (Walsh, 2004).

Gibbs sampling is Markov Chain Monte Carlo (MCMC) method like Metropolis Hastings algorithm. It is a special case of Metropolis Hastings algorithm in which the random value $\alpha = 1$, is always expected to be accepted. Gibbs sampler can be used to generate random sample (observation) from a joint distribution or marginal distribution, given all the associated complete conditional distribution.

Let, $\pi(\theta) = \pi\ (\theta_1, \theta_2, ...\theta_k), \quad \theta \in R^n$ denote a joint density and

$\pi(\theta_i \mid \theta_{j \neq i}),\ i = 1,2,.....k,\ \ 1 < \mathrm{k} \leq \mathrm{n}$

is the complete conditional density for each of components θ_i .

Besag (1974) confirmed under mild condition, these conditional distributions uniquely determine the full joint distribution $\pi(\theta_1, \theta_2, ...\theta_k)$ and hence marginal distribution $\pi\ (\theta_i)\ \ \forall\, i = 1,2,....n$.

Let,

$\theta^{(0)} = \left(\theta_1^{(0)}, \theta_2^{(0)}, ..., \theta_k^{(0)}\right)$ are arbitrary starting values.

The successive random draws are made from full conditional distribution $\pi(\theta_i \mid \theta_j, j \neq i)$; $\ i = 1,2,.....k$, as follows.

We draw, $\theta_1^{(1)}$ from $\pi\left(\ \theta_1 \mid \theta_2^{(0)}, \theta_3^{(0)} ....., \theta_k^{(0)}\ \right)$;

$\theta_2^{(1)}$ from $\pi\ \left(\theta_2 \mid \theta_1^{(1)}, \theta_3^{(0)} ....., \theta_k^{(0)}\ \right)$;

$\theta_3^{(1)}$ from $\pi\left(\theta_3 \mid \theta_1^{(1)}, \theta_2^{(1)}, \theta_4^{(0)} ....., \theta_k^{(0)}\ \right)$;

. . ….. …..

$\theta_k^{(1)}$ from $\pi\left(\theta_k \mid \theta_1^{(1)}, \theta_2^{(1)}, ..., \theta_{k-1}^{(1)}\ \right)$;

This completes one iteration or a transition from.

$\theta^{(0)} = \left(\theta_1^{(0)}, \theta_2^{(0)}, ...\theta_k^{(0)}\ \right)$ to

$\theta^{(1)} = \left(\theta_1^{(1)}, \theta_2^{(1)}, ..., \theta_k^{(1)}\ \right)$.

Iteration of such cycle of random variable generation from each of the full conditional distributions in turn produces a sequence $\theta^0, \theta^1, \theta^2, ..., \theta^t$; which is realization of Markov chain with transition probability from θ^t to θ^{t+1}, given by

$$K_\theta\left(\theta^t, \theta^{t+1}\right) = \prod_{j=1}^{k} \pi\left(\theta_i^{t+1} \mid \theta_j^{t} \text{ for } j > i, \ \theta_j^{t+1} \text{ for } j < i\right).$$

The key features of Gibbs sampling algorithm is that the sample can be drawn only from full conditional distribution. Geman and Geman (1984) suggested that under mild condition for a large number of iteration (i.e., t $\to \infty$) k-tuple converges to a random observation from joint density $\pi(\theta_1, \theta_2, ...\theta_k)$. So for large t we simply write $(\theta_1, \theta_2, ...\theta_k)$ instead of $(\theta_1^{(1)}, \theta_2^{(1)}, ...\theta_k^{(1)})$.

Drapper (2000) noted that the Gibbs sampler usually produces chains with smaller autocorrelation than the other MCMC samplers do. The auto correlation of order k is a measure of how correlated the k values apart are for series $\{X_1, X_2 .....X_k\}$. It is

$$\rho_k = \frac{\text{cov}(X_t, X_{t+k})}{\text{var}(X)} \text{ and sample correlation function is}$$

$$r_k = \frac{\sum_{t=1}^{n-k} (X_t - \overline{X})(X_{t+k} - \overline{X})}{\sum_{t=1}^{n} \left(X_t - \overline{X}\right)^2}$$

The autocorrelation time is the time required the terms of the number of values apart, for the correlation to go to zero, (Goodman and Sokal, 1989). It is defined as

$$\tau = 1 + 2\sum_{t=1}^{\infty} \rho_t$$

The estimation of τ is given by $\tau \geq 1 + 2\sum_{t=1}^{3\tau} r_t$

The effective number of sample, which we want to make as larger as possible is given by

$$n_{eff} = \frac{\text{totalnumberof sample}}{\text{autocorrelationtime}}$$

The standard error is calculated by

$$SE = \sqrt{\frac{Var(X)}{n_{eff}}}, \text{ and}$$

$$\text{Var}(X) = \frac{1}{n-1} \sum_{j=1}^{n} \left(X_j - \overline{X}\right)^2$$

Different illustrations of the application of the Gibbs sampler and the procedural information are found in Sokal (1989), Besag et al. (1995) and Lee (1997).

*

Appendix C-4

The Expressions of the Metropolis algorithm

The algorithm continues as follows

- Draw the starting Value $\theta^0 \sim p_0(\theta)$

- For t = 1, 2,…….

- Draw $\theta^* \sim J_t(\theta^* \mid \theta^{t-1})$. The jumping distribution $J_t(\theta^* \mid \theta^{t-1})$ must be symmetric, i.e., $J_t(\theta^k \mid \theta^l) = J_t(\theta^l \mid \theta^k)$

- Calculate the importance ratio $r = \dfrac{p(\theta^* \mid x)}{p(\theta^{t-1} \mid x)}$

- Set $\theta^t = \theta^*$ with probability r

- The algorithm requires the ability to calculate r and to draw θ^* from the jumping distribution $J_t(\theta^* \mid \theta)$

Appendix C-5

The expression for the Metropolis-Hastings Algorithm

The M-H algorithm goes on as:

- Draw the starting Value $\theta^0 \sim p_0(\theta)$
- For t = 1, 2,…….
  - Draw the jumping distribution $\theta^* \sim J_t(\theta^* \mid \theta^{t-1})$; jumping distribution need not to be symmetric $J_t(\theta^k \mid \theta^l) \neq J_t(\theta^l \mid \theta^k)$
  - To correct the asymmetry the importance ratio changes to $r = \frac{p(\theta^* \mid x) / J_t(\theta^* \mid \theta^{t-1})}{p(\theta^{t-1} \mid x) / J_t(\theta^{t-1} \mid \theta^*)}$
  - If $J(\theta^* \mid \theta) = p(\theta^* \mid x)$ for all θ then *r=1* and θ^t are a sequence of independent draws from *p (θ | x)*

We start with initial value θ_0, such that $g(\theta_0) > 0$; using current θ values we sample a candidate point θ^*. Let, $q(\theta_1, \theta_2)$ denote transition probability function or the jumping distribution, which is the probability of returning a value θ_2 given a process value θ_1. It is also referred to as the proposal distribution or candidate generating distribution (Chib & Greenberg, 1995). The jump density in the Metropolis algorithm is symmetric i.e., $q(\theta_1, \theta_2) = q(\theta_2, \theta_1)$.

For a candidate point θ^* , we define

$$\alpha = \frac{p(\theta^*)}{p(\theta_{t-1})} = \frac{g(\theta^*)}{g(\theta_{t-1})} ,$$

here $p(\cdot)$ is such that the normalizing constant k cancels out.

If the jump increases the density $(\alpha > 1)$, accept the candidate point (set $\theta^t = \theta^*$) and repeat the process. If $\alpha < 1$, accept the candidate point with the probability α, else it repeat again.

We can summarize the metropolis sampling as first computing

$$\alpha = \min\left(\frac{g(\theta^*)}{g(\theta_{t-1})}, 1 \right)$$

and then accepting a candidate point with the probability of move (α).

This generates a Markov chain $(\theta_0, \theta_1, \theta_2, ..., \theta_k, ...)$ as the transition probabilities from θ_t to θ_{t+1} depends only on θ_t and not on $(\theta_0, \theta_1, \theta_2, ..., \theta_{t-1})$. The chain approaches its stationary distribution in sufficient burn in period in the steps.

Hasting (1970) generalized this algorithm using the equilibrium distribution of interest $p(\theta)$. The distribution $p(\theta)$ only enters $p(\theta, \theta^1)$ through the ratio $\frac{p(\theta^1)}{p(\theta)}$. When $p(\theta)$ is posterior distribution, the knowledge of proportionality of the distribution is sufficient for the implementation of the algorithm.

Hasting (1970) generalized Metropolis algorithm using arbitrary probability, such $q(\theta_1, \theta_2) = \Pr(\theta_1 \to \theta_2)$ and setting the acceptance probability for a candidate point as

$$\alpha = \min\left(\frac{f(\theta^*)\ q(\theta^*, \theta_{t-1})}{f(\theta_{t-1})\ q(\theta_{t-1}, \theta^*)} \ , \quad 1 \right)$$

It is the Metropolis Hasting Algorithm. The originality of the Metropolis Algorithm is well again if the proposal distribution is symmetric (i.e., $(x, y) = q(y, x)$) (Bennett *et al.,* 1995).

*

Appendix C-6

Intrinsic Estimation:

The conventional loss function focus on the distance between the estimate $\hat{\theta}$ and the true value θ rather than the distance between the probability model they level. The concept of Intrinsic Estimation is found in Bernardo (2003). Intrinsic loss is the function which focuses on the probability model and the intrinsic difference between the results they produce. Intrinsic losses try to focus on how the probability model $p(X \mid \theta, \lambda)$ is as of its closest approximation within the family $\{p(X \mid \hat{\theta}, \lambda_i), \lambda_i \in \Lambda\}$ and typically produce invariant solutions. The intrinsic loss $\delta(\hat{\theta}, \theta)$ is defined as $\delta(\hat{\theta}, \theta) = \min\{K(\hat{\theta} \mid \theta), K(\theta \mid \hat{\theta})\}$

where $K(\theta_i \mid \theta_J) = \int_{\tau} p\left(t \mid \theta_j\right) \log \frac{p\left(t \mid \theta_j\right)}{P\left(t \mid \theta_i\right)}\, dt$ and $t = t(X) \in \tau$ is any sufficient statistics.

With some nuisance parameter 'λ'

$\delta\left\{\hat{\theta}, (\theta, \lambda)\right\} = \min_{\lambda_i = \Lambda} \delta\left\{(\hat{\theta}, \lambda_i), (\theta, \lambda)\right\}$ and the posterior (expected) intrinsic discrepancy is $d(\hat{\theta} \mid X) = \int_{\Lambda}\int_{\Theta} \delta\left\{\hat{\theta}, (\theta, \lambda)\right\}\, p(\theta, \lambda \mid X)\, d\theta\, d\lambda$.

The intrinsic estimator $\theta^* = \theta^*(X)$ is the corresponding Bayes estimator which minimizes the posterior expected loss $\theta^*(X) = \arg\min_{\hat{\theta} \in \Theta} d(\hat{\theta} \mid X)$

Intrinsic Credible Region:

For any loss function $l(\hat{\theta}, \theta)$ a q credible lowest expected loss (LEL) region may be defined as q credible R^*_q such that

$$l(\theta_i \mid X) < l(\theta_j \mid X) \quad \forall \theta_i \in R^*_q, \forall \theta_j \notin R^*_q$$

where $l(\hat{\theta} \mid X) = \int_{\Theta} l(\hat{\theta}, \theta)\ \ p(\theta \mid X)\ \ d\theta$ is the posterior expected loss from using $\hat{\theta}$ instead of using θ. The loss function used is variant under parameterization (Bernardo, 2003). The LEL credible region will also be invariant.

*

Appendix C-7

#Theorem1: If the hypothesis is simple, the Bayes factor in favour H_0 verses H_1 is just the ratio of likelihood under H_0 to that of the H_1. ie,. $B_0 = \frac{L(\theta_0)}{L(\theta_1)} = \frac{p(x|\theta_0)}{p(x|\theta_1)}$

Its proof is given as:

Define a posterior density $p_i = p(\theta_i)p(x|\theta_i) = \pi_i p(x|\theta_i)$ for *i =1,2*

$$\frac{p_0}{p_1} = \frac{\pi_0 p(x|\theta_0)}{\pi_1 p\ (x|\theta_1)}$$

So, $$B_0 = \frac{p_0}{p_1} \cdot \frac{\pi_1}{\pi_0} = \frac{\pi_0\ \ p\ (x|\theta_0) \times \pi_1}{\pi_0 \times \pi_1\ \ p(x|\theta_1)} = \frac{p(x|\theta_0)}{p(x|\theta_1)} = \frac{L(\theta_0)}{L(\theta_1)}$$

Theorem2: If the hypothesis is composite, prior distribution of θ conditioned on H_0 verses H_1 is $p_i(\theta) = \frac{p(\theta)}{\pi_i}$ or $p(\theta) = \pi_i\ p(\theta_i)$ for $i = 1,2$

i.e., $p_0(\theta) = \frac{p(\theta)}{\pi_0}$ *for* $\theta \in \Theta_0$ and $p_1(\theta) = \frac{p\ (\theta)}{\pi_1}$ *for* $\theta \in \Theta_1$

where $p(\theta)$ is the prior density of θ and $p_i(\theta)$ is the restriction of $p(\theta)$ to Θ_i normalized to give a probability density over Θ_0 and similarity for $p_1(\theta)$.

Its proof is given as:

$$p_i = \Pr(\theta \in \Theta_i \mid x) = \int_{\theta \in \Theta_i} p\ (\theta \mid x)\ dx$$

$$\propto \int_{\theta \in \Theta_i} p_i(\theta)\ \ p(x|\theta_i)\ d\theta$$

$$= \pi_i \times \int_{\theta \in \Theta_i} p(x|\theta)\ \ p_i(\theta) \quad d\theta$$

Thus, the Bayes factor in favour the null Hypothesis H_0 is

$$B_0 = \frac{p_0}{p_1} \cdot \frac{\pi_1}{\pi_0} = \frac{\int_{\theta \in \Theta_0} p(x|\theta)\ \ p_0(\theta) \quad d\theta}{\int_{\theta \in \Theta_1} p(x|\theta)\ \ p_1(\theta) \quad d\theta},$$

which is a ratio of weighted likelihood of Θ_0 and Θ_1.

*

Appendix C-8

Poison- Gamma Hierarchical Modelling:

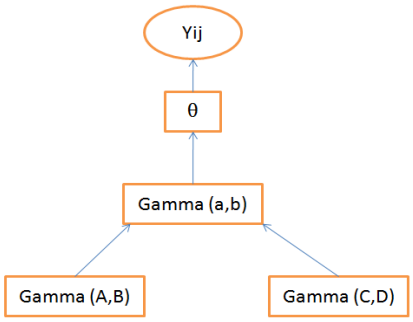
For a Poisson –gamma hierarchical modelling we assume,

$$X_i \mid \theta_i \sim Poisson\ (\theta_i)$$

$$\theta_i \mid a,b \sim gamma(a,b)$$

$$a \sim gamma(A,B)$$

$$b \sim gamma(C,D)$$



The hierarchical modelling of Regression model is given by

$$Y_{ii} \sim N\left(\alpha_i + \beta_i X_{ij},\ \sigma_\varepsilon^{\,2}\right)$$

$$\alpha_i \sim N\left(\mu_\alpha,\ \sigma_\alpha^{\,2}\right)$$

$$\beta_i \sim N\left(\mu_{\beta,}\ \sigma_\beta^{\,2}\right)$$

$$\frac{1}{\sigma_\varepsilon^{\,2}} \sim gamma\ (A,B)$$

$$\mu_\alpha \sim N\left(0,\ \varsigma_\alpha^{\,2}\right)$$

$$\mu_\beta \sim N\left(0,\ \varsigma_\beta^{\,2}\right)$$

$$\frac{1}{\sigma_\alpha^{\,2}} \sim gamma(C_1,\ D_1)$$

$$\frac{1}{\sigma_\beta^{\,2}} \sim gamma(C_2,\ D_2)$$

*

References

Adhikari, P. (2002). A review of integrated pest management program for sustainable agricultural in Nepal. In *Agriculture and Environment***,** HMG, Ministry of Agriculture and Cooperative, Singh Darbar, Kathmandu, Nepal.

Bajracharya, D. N. (2001). *Science and Technology in Nepal*, RONAST, Nepal.

Barendsz, A.W. (1998). Food safety and total quality management. *Food Control,* 9(2-3): 163-170.

Barker, G. C. (2003). Application of Bayesian belief network models to food safety science. In *Bayesian Statistics and Quality Modelling in the Agro-food Production Chain***,** ed. Van Boekel, M. A. J. S., Stein, A. et al, pp.117-128. Kluwer academic Publication, The Netherlands.

Beal, M. J. (2003). *Variational Algorithm for Approximate Bayesian Inference*. Ph. D. thesis Neuroscience unit, university of London. www.cse.buffalo.edu

Bennett, J. E., Racine-Poon, A. and Wakefield, J. C. (1995). Markov Chain Monte Carlo for nonlinear hierarchical models. In M*arkov Chain Monte Carlo in Practice,* ed. W. R. Gilks, S. Richardson, and D. J. Spiegehalter, pp. 140-162. Chapman Hall, NewYork.

Berg, B. A. (2004). *Markov Chain Monte Carlo Simulations and Their Statistical Analysis*. World Scientific, Singapore.

Berger, J. O., (1985). *Statistical Decision Theory and Bayesian Analysis*, 2nd ed., Springer series in statistics, Springer, New York.

Bernardo, J. M. and Smith, A. F. M. (1994). *Bayesian Theory*, John Wiley and Sons Ltd., Chichester.

Bernardo, J. M. (2003). Bayesian Statistics. *Probability and Statistics* (ed. R. Viertl,) of the *Encyclopaedia of life support system*. Oxford, UK: UNESCO.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of Royal Statistical Society, B,* 36:192–236.

Besag, J. and Clifford, P. (1989). Generalized Monte Carlo significance tests. *Biometrika,* 76 (4): 633-642.

Besag, J. and Clifford, P. (1991). Sequential Monte Carlo p-values. *Biometrika,* 78 (2): 301-304.

Besag, J. and Green, P. J. (1993). Spatial statistics and Bayesian computation (with discussion). *Journal of the Royal Statistical Society, B,* 55 (1): 25-27.

Besag, J., York, J. and Mollie, A. (1991). Bayesian restoration with two applications in spatial statistics (with discussion). *Annals of the Institute of Statistical Mathematics,* 43: 1-59.

Besag, J., Green, P. J., Higdon D., and Mengersen, K. L. M. (1995). Bayesian computation and stochastic systems (with discussion). *Statistical Science,* 10: 3–66.

Birkes, D., Dodge, Y. (1993). *Alternative Methods of Regression*, John Wiley and Sons, New York.

Bourlakis, M., Weightman, P. (2004). *Food Supply System Management*, Blackwell Publication, UK.

Campbell, G. (2000). Bayesian statistics in a regulatory environment, In *ICSA Bulletin,* July 2000 issue, pp. 38-52.

Carlin, B. P. (1996). Hierarchical longitudinal modelling. In *Markov Chain Monte Carlo in Practice*, ed. W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, pp. 303–319, Chapman & Hall, London.

Carlin B. P. and Louis T. A. (1996). *Bayes and Empirical Bayes Methods for Data Analysis,* Chapman and Hall, New York.

Carlin, B. P. and Gelfand, A. E. (1991). An iterative Monte Carlo method for nonconjugate Bayesian analysis. *Statistics and Computing,* 1: 119-128.

Carlin, B. P., Gelfand, A. E. and Smith, A. F. M. (1991). Hierarchical Bayesian analysis of change point problems. *Applied Statistics,* 41: 389-405.

Carlin, B. P., Polson N. G. and Stoffer, D. S. (1992). A Monte Carlo approach to nonnormal and nonlinear state space modelling. *Journal of the American Statistical Association,* 87: 493-500.

Casella, G. and George, E. I. (1992). Explaining the Gibbs sampler. *The American Statistician*, 46 (3): 167-174.

CBS (1998). *Census of Manufacturing Establishments, National Level 1996-97,* Central Bureau of Statistics, HMG, Kathmandu, Nepal.

------ (2002). *Population Census 2001: National report*, Central Bureau of Statistics, HMG, Kathmandu, Nepal.

Chen, M. and Shao, Q. (1998). Monte Carlo methods for Bayesian analysis of constrained parameter problems. *Biometrika*, 85 (1): 73-87.

Chen, M. and Shao, Q. (1999). Monte Carlo estimation of Bayesian credible and HPD intervals. *Journal of Computational and Graphical Statistics*, 8 (1): 69-92.

Chib, S. and Greenberg, E. (1995). Understanding the Metropolis-Hastings algorithm. *The American Statistician*, 49 (4): 327-335.

Condra, L. W. (1993). *Reliability Improvement with Design of Experiments*, Marcel Dekker, Inc. New York.

Congdon, P. (2003). *Applied Bayesian modelling*, New York: Wiley.

Corney, D. P. A. (2000). Designing food with Bayesian Belief Networks. In *Adaptive computing in design and manufacture*, ed. I. Parmee, pp. 83-94, Springer-Verlag,.
[http://www.cs.ucl.ac.uk/staff/D.Corney/FoodDesign.html]

Cowles, M. K. and Carlin, B. P. (1996). Markov chain Monte Carlo convergence diagnosis: a comparative review. *Journal of the American statistical Association*, 91: 883-904.

Cowell, R. G., Dawid, A. P. and Lauritzen, S. (1999). *Probabilistic Networks and Expert Systems*, Statistics for engineering and information science, Springer, New York.

Crosby, P. (1979). *Quality Is Free*, Penguin Group, New York.

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

Draper, D. (2000). *Bayesian Hierarchical Modelling*. Draft version on the web: [http://www.bath.ac.uk/»masdd/], 17/03/2008

Duncan, A. J. (1970). *Quality Control and Industrial Statistics*, Richard D. Irwin Inc, Homewood, Illinois.

Edwards, W., Lindman, H., and Savage, L. J, (1963). Bayesian statistical inference for psychological research. *Psychological review*, 70: 193-242

English, L. P. (1999). *Improving Data Warehouse and Business Information Quality: Methods for Reducing Costs and Increasing Profits*. Wiley, New York.

FAO-WHO, (2000). Safety Aspects of genetically modified foods of plant origin. Report of a *Joint FAO/WHO Expert Consultation on Foods Derived from Biotechnology*. Geneva, Switzerland.

FAO -WHO (2002). From the Farm to the Fork, In: *Food quality and safety*, paper in the site of the pan-European conference on food safety and quality (organised by the FAO and WHO and held in Budapest in February, 2002), accessed on 24/06/ 2006 from http://www.foodsafetyforum.org/paneuropean/index_en.htm,

FDA (2006). Guidance for the use of Bayesian Statistics in Medical Device Clinical Trials, *Draft guidance*, accessed on May 23, 2006 from http://www.fda.gov/cdrh/osb/guidance/1601.pdf

Fearn, T. (2003). Bayesian statistics and the agro-food production chain, In *Bayesian Statistics and Quality Modeling in the Agro- food Production Chain***,** **e**d. Van Boekel, M. A. J. S., Stein, A. et al, pp. 11-16, Kluwer academic Publication, The Netherlands.

Ferguson, T. S. (1973). A Bayesian analysis of some non-parametric problems. *Annals of Statistics,* l :209-30.

Fishman, G. S. (1995). *Monte Carlo: Concepts, Algorithms, and Application,* Springer Verlag New York.

Fayyad, U. M., Piatesky-Shapiro, G. and Smyth P. (1996). From Data Mining to knowledge Discovery: An Overview. In *Advances in Knowledge Discovery and Data Mining*, ed. U. M. Fayyad, G. Piatesky-Shapiro, P. Smyth and R. Uthurusamy, pp. 1-37, AAAI-Press. Menlo Park, Canada**.**

Gamerman, D. (1997). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference,* Chapman & Hall, London.

Gelfand, A. E., Hills, S. E., Racine-Poon, A. and Smith, A. F. M. (1990). Illustration of Bayesian inference in normal data models using Gibbs sampling. *Journal of the American Statistical Association,* 85: 972-985.

Gelfand, A. E. and Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association,* 85: 398-409.

Gelfand, A. E., Smith, A. F. M., and Lee, T. M. (1992). Bayesian analysis of constrained parameter and truncated data problems using Gibbs sampling. *Journal of the American Statistical Association,* 87: 523-532.

Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science,* 7 (4): 457-472.

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images. *IEEE Transactions, Pattern Analysis and Machine Intelligence,* 6: 721-741.

George, E., McCulloch, R., and Tsay, R. (1996). Two approaches of Bayesian model selection with Applications. In *Bayesian Statistics in Statistics and Econometrics,* ed. D. Berry, K. chalnoner, and J. Geweke, pp. 339-348. Essays in Honour of Arnold Zellner, Willey, New York.

Geyer, C. J. (1992). Practical Markov Chain Monte Carlo. *Statistical Science,* 7 (4): 473-483.

Ghosh, S. K. (2006). Prior distribution of the variable selection problem. *BSWG seminar, Raleigh NC*. North Carolina State University, from http//:www.stat.ncsu.edu.bswg/

Gilks, W. R. (1992). Derivative-free adaptive rejection sampling for Gibbs sampling. In *Bayesian Statistics,* ed. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, 4: pp. 641-649, Oxford University Press, Oxford.

Gilks, W. R., Oldfield, L. and Rutherford, A. (1989). Statistical analysis, In *Leucocyte Typing IV: White Cell Differentiation Antigens,* ed. W. Knapp, B. Dorken, W. R. Gilks, E. P. Rieber, R. E. Schmidt, H. Stein and A. E. G. Kr. von dem Borne, pp. 6-12. Oxford University Press, Oxford.

Gilks, W. R., Richardson, S. and Spiegelhalter, D. J. (1996). *Markov Chain Monte Carlo in Practice,* Chapman & Hall, London.

Gilks, W. R. and Wild, P. (1992). Adaptive rejection sampling for Gibbs sampling. *Applied Statistics*, 41 (2): 337-348.

Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 82: 711-732.

Griffiths, T. L., Tenenbaum, J. B. & Kemp, C. (2008). Theory-based Bayesian models of inductive learning and reasoning. *Trends in Cognitive Science, 10,* 309-318, a tutorial accessed from [http://www.cocosci berkley.edu/tom/bays/ pdf. tutorial] on 05/03/2008.

Goodman, J. and Sokal, A. (1989). Multigrid Monte Carlo method: Conceptual foundations. *Physical Review D*, 40 (6): 2035-2071.

Hammersley, J. M. and Handscomb, D. C. (1964). *Monte Carlo Methods*, Methuen London.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57: 97-109.

Hawthorn, J., Schweigert, B., Stewart, G., Gacula, M. Jr., Singh, J. (1984). *Statistical Methods in Food and Consumer Research***,** Academic press, USA. Accessed on July 8, 2006, from http://books.elsevier.com/bookscat/links/details.asp?isbn=0122720504,

Heckerman, D. (1996). Bayesian networks for knowledge discovery. In *Advances in Knowledge Discovery and Data Mining*, ed. U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth and R. Uthurusamy, pp. 273-306. AAAI/MIT Press.

Hills, B. P. (2001). The power and pitfalls of deductive modelling, *Food Process Modelling*, ed. Tijskens, L. L. M., A. T. O, BV., Hertog, T. M., et al. Woodhead Publishing House, England.

Horchner, P. M., D. Brett, B. Gormley, I. Jenson and A. M. Pointon (2006). HACCP-based approach to the derivation of an on-farm food safety program for the Australian red meat industry. *Food Control* 17(7): 497-510

Hubbard, M. R. (2003). *Statistical Quality Control for the Food Industry*, 3rd ed. Kluwer Academic, New York

Imai, M. (1997). *Gemba Kaizen: A Commonsense, Low Cost Approach to Management*, McGraw-Hill, New York.

Imai, M. (1989). *Kaizen: The Key to Japan's Competitive Success*. Random House, New York.

ISBA (1992). What is Bayesian Analysis?, wave page of *International Society of Bayesian Analysis.* [http://www. bayesian.org.openingpage.html], July15,2006

Ishikawa, K. (1986). *Guide to Quality Control* (Industrial Engineering & Technology), Asian Productivity Organization.

Ishikawa, K. (1988). *What is Total Quality Control? The Japanese way,* trans. David Lu, Prentice-Hall, Englewood Cliffs New Jersey.

Jeffreys, H. (1961). *Theory of probability*, 3rd ed. Clarendon press, Oxford.

Juran, J. M. (1988). *Juran on Planning For Quality*. Macmillan Free Press, New York,

Jordan, M. I., and Weiss, Y. (2002). Graphical models: Probabilistic inference. In *The Handbook of Brain Theory and Neural Networks,* ed. M. Arbib, 2nd ed. MIT Press Cambridge, MA.

Kass, R. and Raftery, A. (1995). Bayes Factors. *Journal of the American Statistical Association,* 90: 773-795.

Kass, R. and Washerman, L. 1995. A reference Bayesian test for nested hypotheses and its relationship to the Shewarz criterion. *Journal of the American Statistical Association,* 90: 928-934

K.C., J. B. (2002). Status of the food industries in Nepal, *Journal of Institute of Science and Technology,* T. U., Nepal, 12: 21-31.

Kennedy, M. C., Clough H. E. and Joanne T. (2009). Case studies in Bayesian microbial risk assessments, In *Environmental Health BioMed Central*, open access available from: http://www.ehjournal.net/content/8/S1/S19

Khatiwada, R. P. and Sthapit, A. (2008). An application of Bayesian method in packaged food quality control. *Nepal Journal of Science and Technology*, 9: 41-48.

Khatiwada, R. P. (2009). An alternative simplified attribute sampling plan for food product control. *Journal of Food Science and Technology Nepal,* 9: 103-109.

Khatiwada, R. P. (2010). Analysis of Beta prior distribution for an acceptance sampling plan in food quality: A Bayesian approach. *Journal of Institute of Science and Technology*, T. U., Nepal, 16: 64-74.

Kontoleon, A. and Yabe, M. (2006). Market segmentation analysis of preference for GM derived animal foods in the UK. *Journal of Agriculture and Food Industrial Organization*,4: article 8.

Kullback, S. and Leibler, R. A. (1968). On information and sufficiency. *Annals of Mathematical Statistics*, 22: 79-86.

Lee, P. M., (1997). *Bayesian Statistics: An Introduction*, Arnold, London. Accessed on July 2006 from http://wwwusers.york.ac.uk/~pml1/bayes/book.htm.

Lindley, D. V. (1965). *Introduction to Probability and Statistics from a Bayesian Viewpoint* (part 1), Cambridge University press, Cambridge.

Lindley, D.V. (1970). *Introduction to Probability and Statistics from a Bayesian View point* (part 2 Inference). Cambridge University press, Great Britain.

Lindley, D. V. (1971). *Making Decisions*, John Wiley, London.

Liechty, J. C. and Roberts, G. O. (2001). Markov chain Monte Carlo methods for switching diffusion models. *Biometrika*, 88 (2): 299-315.

Lunning, P. A. (2002). Food quality management: A techno-managerial approach. In *Food Quality Management Handbook from C.H.I.P.S.*, Hospitality Industry Publications, Texas US.

Mack, W., Langholz, B. and Thomas, D. C. (1990). Survival models for familial aggregation of cancer. *Environmental Health Perspective*, 87: 27-35.

Martens, H. (1983). Understanding food research data. In *Food Research and Data Analysis*, ed. Martens, H. and Russwurm, H. Jr., pp. 5-38. Norwegian Food Research Institute As, Applied Science Publisher, New York.

McCullagh, P. and Nelder J. (1989). *Generalized Linear Models: Monographs on Statistics and Applied Probability*, vol. 37, 2nd ed, Chapman and Hall, Cambridge, UK.

Malakoff, D. (1999). Bayes offers a new way to make sense of numbers. *Science*, 286: 1460-1464.

Meeker, W. Q. and Escobar, L. A. (1998). *Statistical Methods for Reliability Data*, Wiley, New York.

Mendenhall, W. and Sincich, T. (1995). *Statistics for Engineering and the Sciences*, 4th ed., Prentice Hall, Inc. Upper Saddle River, NJ.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. and Teller, E. (1953). Equation of state calculations by fact computing machines. *Journal of chemical Physics*, 21: 1087-92.

Metropolis N. and Ulam, S. (1949). The Monte Carlo method. *Journal of the American Statistical Association*, 44 (247): 335-341.

Mitasova H. and Mitas L. (1998). Process modelling and simulations, *NCGIA Core Curriculum in GI Science*, accessed on July 10, 2006 from: http://www.ncgia.ucsb.edu/giscc/units/u130/u130.html

Mitra, A. (2001). *Fundamentals of Quality Control and Improvement*, 2nd ed. Indian reprint, Pearson Education, India.

Notermans, S. and Mead, G. C. (1996). Incorporation of elements of quantitative risk analysis in the HACCP system. *International Journal of Food Microbiology* 30:157-173.

Ntzoufras, I. (2009). *Bayesian Modelling using WinBUGS,* John Wiley and Sons, New Jersey, Canada.

O'Hagan, A. (1995). Fractional Bayes factors for model Comparison. *Journal of the Royal Statistical Society*, B 57: 57-69

O'Hagan, A. (2003). Bayesian Statistics: Principles and Benefits. In *Bayesian Statistics and Quality Modelling in the Agro- Food Production Chain*, **e**d. Van Boekel, M. A. J. S., Stein, A. et al, pp.31-45. Kluwer academic Publication, The Netherlands.

Owen A. and Zhou, Y. (2000). Safe and effective importance sampling. *Journals of the American Statistical Association*, 95(449): 135-143.

Pennello, G. (2006). Planning a Bayesian trial: Prior distributions and hierarchical models. On behalf of FDA *Public Meeting on Draft Bayesian Guidance.* Accessed on July 27, 2006 from www.fda.gov/default.htm.

Rasmussen, C. E. and Ghahramani, Z. (2003). *Bayesian Monte Carlo*, Gatsby Computational Neuroscience Unit, University College London. Accessed on 03/08/2009 from http://www.gatsby.ucl.ac.uk

Richardson, S. and Best, N. (2003). Bayesian hierarchical models in ecological studies of health–environment effects. *Environmetrics* 14: 129–147.

Robert, C. (2004). Bayesian computational methods. In *Handbook of Computational Statistics* (Vol. I), ed. Gentle, J., Hardle, W. and Mori, Y. Chap. 3, Springer-Verlag, Heidelberg.

Robert, C. P. and Casella, G. (2004). *Monte Carlo Statistical Methods*, 2nd ed. Springer-Verlag, New York.

Roberts, G. O., and Polson, N. G. (1990). A note on the geometric convergence of the Gibbs sampler, *Technical Report*, Dept. of Mathematics, University of Nottingham.

Sahin S. and Sumnu, S. G.(2006). *Physical Properties of Foods*, Springer, US.

Schervish, M. J., and Carlin, B. P. (1992). On the convergence of successive substitution sampling. *Journal of Computational and Graphical Statistics*, 1(2): 111-127.

Shao, J. (1989). Monte Carlo approximations in Bayesian decision theory. *Journal of the American Statistical Association, 84* (407): 727- 732.

Shewhart, W. (1986). *Statistical Methods from the Viewpoint of Quality Control*. Dover Publications, New York, NY.

Smith, A. F. M. and Roberts, G. O. (1993). Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society, B,* 55 (1): 3-23.

Sokal, A. D. (1989). *Monte Carlo methods in statistical mechanics: Foundations and new algorithms*, Lecture notes from Course de Troisième Cycle de la Physique en Suisse Roman de. Accessed on 15/ 04/2006 from [http//:www.physics. nyu.edu/faculty/sokal_html]

Spiegelhalter D. J., Best N. G., Carlin B. P. and van der Linde, A. (2002). Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society, B,* 64**:** 583-640.

Spiegelhalter, D. J., Thomas, A., Best, N., and Lunn, D. (2003). *WinBUGS Version 1.4 User Manual.* MRC Biostatistics Unit, Cambridge, UK. Freely available from http://www.mrc-bsu.cam.ac.uk/bugs/.

Stein, A. (2003). Bayesian networks and food security – an introduction. In *Bayesian Statistics and Quality Modelling in the Agro- Food Production Chain,* **e**d. Van Boekel, M. A. J. S., Stein, A. et al. pp. 107-116, Kluwer Academic Publication, The Netherlands.

Taguchi, G. (1986). *Introduction to Quality Engineering, Designing Quality into Products and Processes*, Unipub., White Pains, NY.

Taguchi, G. (1987). *Experimental Design*, Maruzen, Tokyo.

Tanner, M. A. (1991). Tools for Statistical Inference: Observed data and data augmentation methods. *Lecture Notes in Statistics*, vol. 67, Springer-Verlag, New York.

Tanner, M. A. (1996). *Tools for Statistical Inference*, 3rd ed. Springer-Verlag, New York.

Tierney, L. (1994). Markov chains for exploring posterior distributions (with discussions). *The Annals of Statistics*, 22 (4): 1701-1728.

Thomas, A., Spiegelhalter, D. J., and Gilks, W. R. (1992). BUGS: A program to perform Bayesian inference using Gibbs sampling, In *Bayesian Statistics*, ed. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, 4: 837-842, Oxford University Press, Oxford.

Van Beek, P. (2003). Quantitative modelling in design and operation of food supply systems. In *Bayesian Statistics and Quality Modelling in the Agro- Food Production Chain,* **e**d. Van Boekel, M. A. J. S., Stein, A. et al. pp. 141-144, Kluwer Academic Publication, The Netherlands.

Van Boekel, M. A. J. S. (2003). Bayesian solutions for Food-Science problems. In *Bayesian Statistics and Quality Modelling in the Agro- Food Production Chain,* **e**d. Van Boekel, M. A. J. S., Stein, A. et al. pp. 17-27, Kluwer Academic Publication, The Netherlands.

Van der Voet, H., and Paulo, M. J. (2003). Some explorations into Bayesian modelling of risks due to pesticide intake from food. In *Bayesian Statistics and Quality Modelling in the Agro- Food Production Chain, e*d. Van Boekel, M. A. J. S.; Stein, A. et al. pp. 145-162, Kluwer Academic Publication, The Netherlands.

Van der Vorst, J. G. A. J. (2005). Performance measurement in agri-food supply chain networks: an overview. In *Quantifying Supply Chain Performance*. Kluwer, Dordrecht.

Vose, D. (1996). *Quantitative Risk Analysis: A guide to Monte Carlo Simulation,* John Wiley & Sons, ISBN 0-471-95803-4.

Vose, D. (2000). *Risk Analysis: A Quantitative Guide*, 2nd ed. Wiley, Chichester

Walsh, B. (2004). *Markov Chain Monte Carlo and Gibbs Sampling*. Lecture Notes for EEB 581, version : 26 April 2004 accessed on 4/12/2006 from [http://www.nitro.biosci.arizona.edu./personal/st/ jbwalsh _lecturenote],

York, J. (1992). Use of the Gibbs sampler in expert systems, *Artificial Intelligence*, 56: 115-130.

Young, T. M. and Guess, F. M. (2002). Mining Information in Automated Relational Databases for Improving Reliability in Forest Products Manufacturing. *International Journal of Reliability and Applications*, **3** (4), 155-164.

Zeger, S. L. and Karim, M. R. (1991). Generalized linear models with random effects; a Gibbs sampling approach. *Journal of the American Statistical Association,* 86:79-86

Zellner, A. (1971). *An Introduction to Bayesian Inference in Econometrics*. NewYork: Wiley.

Zellner, A. (1986). On assessing prior distribution and Bayesian regression Analysis using g-prior distribution, In *Bayesian inference and Decision Techniques: Essays in Honour of Bruno de Finetti*, ed. P Goel and A Zellner, pp. 233-243 North Holand, Amsterdam .