Bayesian Modelling Approaches on some Issues of Agro-Food Production and Quality Control

A Thesis Submitted to

Central Department of Statistics

Institute of Science and Technology

Tribhuvan University, Kirtipur, Kathmandu, Nepal

For the Degree of

Doctor of Philosophy (Ph. D.) in Statistics

Ram Prasad Khatiwada

2010

Date: 28 / 07 / 2010

Recommendation

I have the pleasure of forwarding the thesis entitled "**Bayesian Modelling Approaches on some Issues of Agro-Food Production and Quality Control**" by Mr. Ram Prasad Khatiwada, for the award of the degree of Doctor of Philosophy (Ph. D.) in Statistics. The work presented in this thesis has been carried out under my supervision and has not been submitted for the award of a degree to any university.

Prof. Azaya Bikram Sthapit, Ph. D.

Supervisor,

Head,

Central Department of Statistics

Institute of Science and Technology

Tribhuvan University, Kathmandu, Nepal.

Declaration

I hereby declare that the work presented in this thesis has been done by myself and, has not been submitted for the award of any degree. All sources of information have been specially acknowledged by reference to the authors or institutions.

Ram Prasad Khatiwada

Central Campus of Technology

Institute of Science and Technology

Tribhuvan University

TRIBHUVAN UNIVERSITY
Institute of Science and Technology
DEAN'S OFFICE
Kirtipur, Kathmandu, Nepal

Institute of Science and Technology 2045

Reference No.:

Institute Of Science & Technology Dean's Office Kirtipur, 2045

Date:.....
July 10, 2011

TO WHOM IT MAY CONCERN

This is to certify that Mr. Ram Prasad Khatiwada has been successfully defended Ph.D. viva-voce examination held on July 7, 2011. The Research Committee of the Institute of Science and Technology has approved and recommended for the degree of Doctor of Philosophy (Ph.D.) in Statistics of Tribhuvan University, Nepal.

The topic of His thesis is **" Bayesian Modelling Approaches on some Issues of Agro-Food Production and Quality Control".**

Prof. Dr. Mukunda Prasad Gajurel
Dean

Telephone: 977-1-4330844 Fax: 977-1-4331755 Website: http://www.tribhuvan-university.edu.np
E-mail: deaniost@wlink.com.np

Acknowledgements

First of all, I would like to express my sincere gratitude to my respected supervisor Prof. Dr. Azaya Bikram Sthapit, Head, Central Department of Statistics, who devoted his valuable time in guiding, inspiring and encouraging me constantly.

I also wish to record my sincere gratitude to all the faculties as well as administrative staff of the Central Department. The moral support rendered to me by them during my work will be remembered forever.

I am also indebted to Institute of Science and Technology, Tribhuvan University and my duty station Central Campus of Technology, Dharan for providing me three years' study leave.

I would like to thank the teachers, the administrative staff and the students of Central Campus of Technology, Dharan, who have directly or indirectly helped me in different ways during my work.

I am also grateful to several people abroad, who have helped me providing suggestions and necessary materials to my work. I would especially like to thank Dr. B. M. Colosimo (Polytecnico di Milano, Italy), Prof. Jeffery Grynaviski (University of Chicago, US), Prof. Anthony O'Hagan (University of Sheffield, UK), and Prof. Richard Arnold (Victoria University, New Zealand).

I am grateful to Prof. S. K. Upadhyay, Banaras Hindu University, who has inspired me and provided me with valuable comments. I wish to express my thanks to the friends who encouraged me during the ATP on Bayesian Statistics at Banaras.

My special thanks go to my friends Rishi Raj Gautam, Ram Gautam, Dil K. Limbu, Kamal Maden, Rohit Dhungel and Dr. Tika Ram Aryal for their valuable support and suggestions.

I am grateful to my wife Khina (Shanta), daughters Bhumika and Srijana and son Hemanta, without whose sacrifices and constant support I could not have completed this work.

Finally, I would like to acknowledge Nepal Academy of Science and Technology (NAST) for providing me a partial financial support for the study with a research fellowship.

Abbreviations/Acronyms

AEL	Absolute Error Loss
AIC	Akaike Information Criteria
BE	Bayes Estimator
BF	Bayes Factor
BIC	Bayesian Information Criteria
BUGS	Bayesian Analysis using Gibbs Sampling
CBS	Central Bureau of Statistics
CCP	Critical Control Point
cdf	cumulative density function
DAG	Directed Acyclic Graphs
FAO	Food and Agricultural Organization
FDA	Food and Drug Administration
HACCP	Hazard Analysis and Critical Control Point
HPD	Highest Probability Density
IG	Inverse Gamma
iid	independently and identically distributed
ISBA	International Society for Bayesian Analysis
EM	Expectation and Maximization
LEL	Lowest Expected Loss
LQL	Limiting Quality Level
MA	Metropolis Algorithm

MC	Monte Carlo
MCMC	Markov Chain Monte Carlo
MHA	Metropolis Hastings Algorithm
ML	Maximum Likelihood
MLE	Maximum Likelihood Estimator
MSE	Mean Squared Error
NIG	Normal Inverse-Gamma
pdf	probability density function
pmf	probability mass function
RE	Relative Efficiency
sd	standard deviation
SEL	Squared Error Loss
SE (or se)	standard error
SSE	Sum of Squared Error
TQM	Total Quality Management
WinBUGS	Windows version of Bayesian Analysis using Gibbs Sampling
WFP	World Food Program
WHO	World Health Organization

List of Tables

Table 3.1	Some conjugate priors for common likelihood functions	28
Table 5.1	Summary of the estimated parameters and confidence intervals using classical method	63
Table 5.2	Summary of the prior density, likelihood and the posterior density with precisions	64
Table 5.3	Summary of the posterior distribution using normal informative prior	64
Table 5.4	Summary of the predictive distribution using normal informative prior	64
Table 5.5	Summary of the posterior density using normal informative prior	66
Table 5.6	Summary of the posterior density using non-informative prior	67
Table 6.1	Sampling plan for $\beta = 0.1$ using binomial cdf	72
Table 6.2	Summary of the Beta (2, 38) distribution	73
Table 6.3	Sampling plan for and $\beta = 0.1$ using posterior distribution	74
Table 6.4	Summary of posterior distribution $Be(a+c, n+b-c)$	74
Table 7.1	Calculation of failure rate per operation units /points	78
Table 7.2	The posterior estimates of the parameters of failures to operation points	79
Table 8.1	Summary of the correlation and regression coefficients	85
Table 8.2	Summary of the Simple and partial correlation coefficient regarding porosity, moisture content and bulk density	87
Table 8.3	Summary of the multiple correlation of porosity with moisture content and bulk density	87
Table 8.4	Summary of the linear regression coefficients on porosity with moisture content and bulk density	87
Table 8.5	WinBUGS results for the parameters in NIG model	88
Table 9.1	Summary statistics of the percentage protein and gluten content of a variety of wheat flour	94
Table 9.2	Summary of the correlation and regression coefficients between percentage of protein and gluten content	94
Table 9.3	Summary of the posterior density of the parameters in model 1	95
Table 9.4	Summary of the posterior density of the parameters in model 2	97

List of Figures

Figure 3.1	An Outline of Bayesian method	17
Figure 5.1	Plot of the posterior density of average weight *(θ)*	65
Figure 5.2	Triplot of the prior density, posterior density and likelihood	65
Figure 5.3	Plot of the predictive density of the weight (*X*)	65
Figure 5.4	Kernel density plots of the posterior distribution through MCMC using WinBUGS	66
Figure 5.5	Trace of the posterior distribution through MCMC using WinBUGS	66
Figure 5.6	The density plots of the posterior distribution for 5000 and 30000 iteration	67
Figure 5.7	The trace of the last 400 iteration of the posterior distribution	67
Figure 6.1	Distribution of sample nonconforming proportion	72
Figure 6.2	Plot of the prior distribution of average proportion nonconforming items	73
Figure 6.3	Plots of posterior distribution	75
Figure 7.1	Graphic modelling of the parameters in failure to operation points	77
Figure 7.2	Kernel Density plots of the parameters in estimating failure to operation points	80
Figure 7.3	Box plots of predictive alpha and beta	81
Figure 7.4	Box plots of posterior failure rate (theta|data)	81
Figure 7.5	Box plots of posterior average failure rate (lambda|theta, t_i)	81
Figure 7.6	Scatter plots of posterior failure rate (theta|data)	82
Figure 7.7	Model fit summaries for the predictive failure rate (theta|data)	82
Figure 8.1	Scatter plots of moisture content vs bulk density	86
Figure 8.2	The scatter plots of bulk density vs porosity	86
Figure 8.3	Scatter plot of the predicted values of porosity (mu)	89
Figure 8.4	Plot of fitted model with credible region for predicted values of porosity	89

Figure 8.5 Density plots of regression coefficients and precision 89
Figure 8.6 Box plots of the regression coefficients and precision 90
Figure 8.7 Box plots of the predicted values of porosity 90
Figure 8.8 Correlation plots of regression coefficients 90
Figure 8.9 Density plots of SSE, MSE, R-square and overall gain in relative efficiency of the fitted model 91
Figure 9.1 Posterior density plots of alpha and beta in model 1 96
Figure 9.2 Trace plots of alpha and beta for the 200 iterations in model 1 96
Figure 9.3 The density plot and trace of the tau in model 1 96
Figure 9.4 Box plots of alpha and beta for model 1 96
Figure 9.5 Posterior density plots of alpha and beta in model 2 98
Figure 9.6 Trace plots of alpha and beta for the 200 iterations in model 2 98
Figure 9.7 The density plot and trace of the tau in model 2 98
Figure 9.8 Box plots of beta and alpha for model 2 98
Figure 9.9 Box plots of posterior mu in model 1 100
Figure 9.10 Box plots of posterior mu in model 2 100
Figure 9.11 Scatter plot of mu in model 1 100
Figure 9.12 Scatter plot of mu in model 2 100
Figure 9.13 Plots of fitted line with 95% credible region for model 1 101
Figure 9.14 Plots of fitted line with 95% credible region for model 2 101

Contents

Part One

Chapters	Topics	Page(s)
Chapter I		
	Introduction	1 - 8
	1.1 Background	1
	1.2 Statement of the Research Problem	2
	1.3 Rationale of the Research	3
	1.4 Research Question	5
	1.5 Review Practicalities	6
	1.6 Structure of the Thesis	8
Chapter II:		
	Objectives	9 -11
	2.1 Objectives of the Study	9
	2.2 Major Issues and Variables	9
	2.3 Sources of Sample and Data	10
	2.4 Delimitations of the Study	11
Chapter III:		
	Literature Review	12-42
	3.1 Quality	12-15
	3.1.1 Introduction	12
	3.1.2 Principles of Quality Control / Management	13
	3.1.3 Review of Bayesian Approaches in Food Quality Control	14
	3.2 Bayesian Statistics	15-34
	3.2.1 An Introduction to Bayesian statistics	15
	3.2.2 Underlying Principle of the Bayesian Approach	17
	3.2.3 Steps in Bayesian Method	17

3.2.4 Mathematical Notations and Foundation 17
3.2.5 Likelihood Function and Bayes' Theorem 21
3.2.6 Priors: An Overview 22
3.2.7 Some useful Terminologies in Bayesian Analysis 30

3.3 Computational Methods in Bayesian Paradigm 35-42

3.3.1 Introduction 35
3.3.2 Markov Chain Monte Carlo 35
3.3.3 Monte Carlo Method 37
3.3.4 Importance Sampling 38
3.3.5 Gibbs Sampling 39
3.3.6 The Metropolis Algorithm (MA) 41
3.3.7 The Metropolis-Hasting Algorithm (MHA) 41
3.3.8 Computer Programs used for Analysis 42

Chapter IV:

Methodology 43-60

4.1 Bayesian Inference 43-50
4.1.1 Preliminaries 43
4.1.2 Bayesian Estimation 44
4.1.3 Hypothesis Testing and Bayes Factor 49

4.2 Bayesian Modelling 51-60
4.2.1 Introduction to Statistical Models 51
4.2.2 Modelling in Bayesian Paradigm 52
4.2.3. Graphical Models 54
4.2.4 Normal Regression Model 55
4.2.5 Random Effect Model 56
4.2.6 Generalized Linear Model 58
4.2.7 Bayesian Models with Hierarchical Priors 59
4.2.8 Model Checking in Bayesian Paradigm 60

Part Two

Chapter V:

An Application of Bayesian Method in Estimating the Weights of Packaged Food 61-69

Chapter VI:

Analysis of Beta Prior Distribution for an Acceptance Sampling Plan of Food Product 70-75

Chapter VII:

Modelling Failure to operation points for Canning process in a Fruit Juice Industry 76- 82

Chapter VIII:

NIG Model for Predicting Porosity of a Rice Variety in terms of Moisture content and Bulk density 83-92

Chapter IX:

Normal Regression Models with two different priors for Estimation of Gluten content relating to Protein content in a Variety of Wheat 93-101

Chapter X:

Conclusion and Recommendations 102-104

10.1 Preliminary 102

10.2 Conclusions 103

10.4 Recommendations for Further Areas of Research 104

Summary 105- 109

Appendix A: Data Tables 110-114

Appendix B: WinBUGS' Codes 115-119

Appendix C: Theoretical Methods of Computations 120-129

References: 130-141