STATISTICAL APPROACH ON GRADING: MIXTURE MODELING

ZAIRUL NOR DEANA BINTI MD DESA

A dissertation submitted in partial fulfillment of the
requirements for the award of the degree of
Master of Science (Mathematics)

Faculty of Science
Universiti Teknologi Malaysia

APRIL 2006

Especially for my beloved parents

Mak, Tok and Tokwan

who taught me to trust myself and love all things great and small

My siblings ~ Abg Am, Dila, Fatin and Aida

Zaha

&

All teachers, lecturers and friends

ACKNOWLEDGEMENT

First of all, I would like to thank Allah S.W.T, the Lord Almighty, for the health and perseverance needed to complete this thesis.

I would like to express my appreciation to the supervisor Dr. Ismail B. Mohamad for his meticulous and painstaking review and his tireless effort in reading on an earlier draft of this thesis. In addition, I thank the chairperson Dr. Arifah Bahar and both internal examiners Dr. Zarina Mohd Khalid and Tn. Hj. Hanafiah Mat Zin for their helpful detailed comments that improved the final report. My sincere appreciation also extends to the lecture and my fellow postgraduate colleagues of the Department of Mathematics at Universiti Teknologi Malaysia who provided me with valuable input.

Finally, I gratefully acknowledge the support of my family and anonymous referees for their patience and understanding.

ABSTRACT

The purpose of this study is to compare results obtained from three methods of assigning letter grades to students' achievement. The conventional and the most popular method to assign grades is the Straight Scale method. Statistical approaches which use the Standard Deviation and conditional Bayesian methods are considered to assign the grades. In the conditional Bayesian model, we assume the data to follow the Normal Mixture distribution where the grades are distinctively separated by the parameters: means and proportions of the Normal Mixture distribution. The problem lies in estimating the posterior density of the parameters which is analytically intractable. A solution to this problem is using the Markov Chain Monte Carlo method namely Gibbs sampler algorithm. The Gibbs sampler algorithm is applied using the WinBUGS programming package. The Straight Scale, Standard Deviation and Conditional Bayesian methods are applied to the examination raw scores of 560 students. The performance of these methods are compared using the Neutral Class Loss, Lenient Class Loss and Coefficient of Determination. The results showed that Conditional Bayesian performed out the Conventional Method of assigning grades.

ABSTRAK

Tujuan kajian ini adalah untuk membandingkan keputusan yang didapati daripada tiga kaedah memberi gred kepada pencapaian pelajar. Kaedah konvesional yang popular adalah kaedah Skala Tegak. Pendekatan statistik yang menggunakan kaedah Sisihan Piawai dan kaedah *Bayesian* bersyarat dipertimbangkan untuk memberi gred. Dalam model Bayesian, dianggapkan bahawa data adalah mengikut taburan Normal Tergabung dimana setiap gred adalah dipisahkan secara berasingan oleh parameter-parameter; min-min dan kadar bandingan dari taburan Normal Tergabung. Masalah yang timbul adalah sukar untuk menganggarkan ketumpatan *posterior* bagi parameter-parameter tersebut secara analitik. Satu penyelesaian masalah ini adalah dengan menggunakan kaedah Markov Chain Monte Carlo iaitu melalui algorithm persampelan Gibbs. Algorithm persampelan Gibbs diapplikasikan dengan menggunakan pekej perisian pengaturcaraan WinBUGS. Kaedah Skala Tegak, kaedah Sisihan Piawai dan kaedah Bayesian bersyarat dijalankan terhadap markah mentah peperiksaan daripada 560 orang pelajar. Pencapaian ketiga-tiga kaedah dibandingkan melalui nilai Kehilangan Kelas Neutral, Kehilangan Kelas Tidak Tegas dan Pekali Penentuan. Didapati keputusan yang diperolehi menunjukkan bahawa kaedah Bayesian Bersyarat menunjukkan pencapaian yang lebih baik dibandingkan dengan kaedah Skala Tegak dan kaedah Sisihan Piawai.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDIXES

NOMENCLATURE

GC	-	Grading on Curves
GB	-	Conditional Bayesian Grading
MCG	-	Multi-Curve Grading
MCMC	-	Markov Chain Monte Carlo
G	-	Grade Sample Space
N	-	Number of Students in a class
n_g	-	Number of Students for Grade g
B	-	Burn-In Period
T	-	Number of Iterations
$h\{\theta \mid x\}$	-	Conditional Probability Density of Prior
$L\{x \mid \theta\}$	-	Conditional Likelihood Function of Raw Score
$p(\cdot \mid \mathbf{x})$	-	Conditional Distribution of Conjugate Prior or Posterior Density
$\pi(\theta)$	-	Prior Distribution
$m(x)$	-	Marginal Density of Raw Score
$p(x_i)$	-	The Probability Distribution of Raw Score
π_g	-	Component Probability of Component g
θ	-	Parameter of Interest (Conjugate Prior)
$\boldsymbol{\Theta}$	-	Vector of Parameter of Interest
$N(\cdot,\cdot)$	-	Normal Distribution

$IG(\cdot,\cdot)$	-	Inverse Gamma Distribution
$Di(\cdot)$	-	Dirichlet Distribution
$C(\cdot)$	-	Categorical Distribution
R	-	Ratio in Gelman-Rubin Statistics
R^2	-	Coefficient of Determination
$C(y_i, \hat{y}_i)$	-	Loss Function
CC	-	Class Loss
LF	-	Leniency Factor

CHAPTER 1

RESEARCH FRAMEWORK

1.1 Introduction

At the end of a course, educators intend to convey the level of achievement of each student in their classes by assigning grades. Students, university administrators and prospective employers use these grades to make a multitude of different decisions. Grades cause a lot of stress for student; this exhibits the fact of education life. Grades reflect personal philosophy and human psychology, as well as effort, to measure intellectual progress with standardized objective criteria.

There are many ways to assign student's grades which all seem to have their advantages and disadvantages. The educators or graders are the most proficient persons to form a personal grading plan because it incorporates the personal values, beliefs, and attitudes of a particular educator. For that reason, a philosophy of grading in establishing a grading plan must be shaped and influenced by current research evidence, prevailing lore, reasoned judgement and matters of practicality. However, a more professional approach should be developed with the ability to be applied at any grade level and in any subject matter area where letter grades are assigned to students at the end of reporting period.

1.2 Statement of the Problem

Most approaches in grading plan require additional effort and varying degrees of mathematical expertise. The educator has to assign a score, which meaningfully assign a letter grade, such as A, B- or C, to each student. There is no standard answer to questions like: What should an "A" student grade mean? What percent of students in my class should received a "C"?. University or faculty regulations encourage a uniform grading policy so that grade of A, B, C, D and E will have the same meaning, independent of the faculty or university awarding the grade. Other campus units usually know the grading standard of a faculty or university.

For example, a "B" in a required course given by Faculty X might indicate that the student have an ability in developing most of the skills and referred to as prerequisites for later learning. A "B" in required course given by Faculty Y might indicate that the student is not a qualified candidate for graduate school in the related fields. Nevertheless, the faculty and educator may be using different grading standards. Course structure may seem to require a grading plan which differs from faculty guidelines or the educator and faculty may hold different ideas about a function of grading. Therefore a satisfactory grading plan must be worked out in order to meet the objective measurement and evaluation in education.

Since both philosophies and instructional approaches change as curriculum changes, educators need to adjust their grading plans accordingly. In this study, we are not comparing faculty regulations on their grading methods but we attempt to differentiate each letter grade based on the overall raw score of the student from the beginning of a semester to the end of the semester period. Statistically based method is used in this research which takes into account the grading philosophy with respect to conditions of measurement and evaluation of students' achievement. The students' final grades intend to have a norm-referenced meaning. By definition, a norm-reference grade does not tell what a student can do; there is no content basis other than the name of the subject area associated with the grade. Furthermore, the distinctions and relationship of several grading methods; conventional and futuristic are discussed carefully.

1.3 Research Objectives

The objectives of this study are to understand the grading philosophy, grading policies, grading methods and exploring the appropriate grading methods. The philosophy and policy are viewed as educational principles and the grading methods were driven by statistical procedures. The primary objectives are to develop mathematical models on grading system for both conventional and future approaches and finally we carry out the programming method in statistical analysis on Bayesian Grading method of assigning grades. The data on the past years record is used in this study.

1.4 Scope of the Study

Assigning a grade to student can be done in various ways. At present, most instructors assign grade conventionally through the raw score from the test given in class and final examination raw scores at the end of a semester period. The grades may be assigned based on the instructors' "feel" throughout the instructors experience with their students. To avoid the "unfair" judgment on student performance, the new approach in grading method which is statistically based is adjusted to the conventional grading plan.

This method has a scientific evidence in assigning grade as compared to using only instructors' personal feels.

A model called *Bayesian Grading (GB)* method is developed to assign the grades. A *Bayesian Inference* based on decision making is an important tool to classify the letter grade into its particular class or component. The *Gibbs Sampler* is used in estimating the optimal class to the grade when the students' raw scores are assumed to be normal and form a bell-shaped distribution. Adjustments to the raw score which take into account the instructor's leniency factor is to allow the educator to vary the leniency of his or her evaluation. Based on the information, we calculate the probability that each student's raw score corresponds to each possible letter grade. The grader's (or instructor's) degree of leniency is used to specify educators' loss function which is used to assign the most optimal letter grades.

These categories of grading are built upon earlier understanding of student raw scores, and it combine the raw scores with current data measure in a way that update the degree of belief (*subjective probability*) of the educator. With this principle, the student raw scores are assumed to be independent of the other students.

1.5 Significance of the Study

In this research, the Bayesian Grading (GB) method of assigning letter grades to student based on their whole semester raw score is described. The GB categorize the marks into several different classes and each class is assigned a different letter grade. The methods take into account the educator degree of leniency in categorizing the raw scores into several classes.

This instructional statistical designed is to help prospective, intermediate and beginning educators to sort out the issues involved in formulating their grading plan and to help experienced educators to reexamine the fairness and defensibility of their current grading practices. It also can be applied at any level of school, college or university.

1.6 Research Layout

Chapter I is intended to introduce basic terminology and a framework of the study. Chapter II, include literatures on some basic grading policies and grading plan for conventional grading methods and a futuristic grading method that will be used throughout the dissertation.

Chapter III presents a more specific grading method which is the grading based on curve and a introducing the basic Bayesian grading method which include a discussion on finding the probability distribution of letter grades. A Bayesian inference in setting the prior and estimating the posterior of probability figured theoretically by including the proofs for readers understanding.

In Chapter IV, we carefully discuss the model parameters estimation that were drawn from the mixture models using Gibbs Sampler. In addition, an estimation of the letter grades which take into account the instructors' loss function is shown to find the optimal letter grades. The simulation is developing uses the WinBUGS (the recently developed software package: the MS **Win**dows operating system version of **B**ayesian Analysis **U**sing **G**ibbs **S**ampling). This is a flexible software for the Bayesian analysis in analyzed complex statistical models uses Markov chain Monte Carlo (MCMC) methods. The URL address to download the free version of the software is www.mrc-bsu.cam.ac.uk/bugs/.

The significance of the results in real life will be judged by several selected instructors which uses real raw scores data. Furthermore, the result will be compared between the conventional grading methods and the Bayesian Grading method.

Finally, Chapter V includes the conclusion and suggestion for further research on grading methods.

REFERENCES

Alex, S. (2003). *A Method for Assigning Letter Grades: Multi-Curve Grading*. Dept. of Economics, University of California - Irvine, 3151 SSPA, Irvine.

Ash, R.B., (1972), *Real Analysis and Probability*. New York: Academic Press Inc.

Berger, J.O. (1985), *Statistical Decision Theory and Bayesian Analysis*. 2nd Edition. New York; Springer-Verlag New York, Inc.

Birnbaum, D.J., (2001). *Grading System for Russian Fairy Tales*. www.clover.slavic.pitt.edu/~tales/02-1/grading.html

Box, G.E.P. and Tiao, G.C. (1973). *Bayesian Inference in Statistical Analysis*. Massachusetts: Addison-Wesley Publishing Company, Inc.

Casella, G., and George, E. I., (1992). *Explaining Gibbs Sampler*. Vol. 46. No. 3. pg. 167-174.The American Statistical Assosiation.

Congdon, P (2003). *Applied Bayesian Modelling*. West Sussex, England: John Wiley & Son Ltd.

Cornebise, J., Maumy, M., and Philippe, G. *A*.(2005) *Practical Implementation of the Gibbs Sampler for Mixture of Distribution: Application to the Determination of Specifications in Food Industry*. www.stat.ucl.ac.be/~lambert/BiostatWorkshop2005/slidesMaumy.pdf.

Cross, L.H., (1995). Grading Students. ED398239. ERIC Clearinghouse on Assessment and Evaluation, Washington DC. www.ericfacility.net/ericdigest.

Ebel, R. L. and Frisbie, D. A. (1991). *Essentials of Educational*. 3th ed. Englewood Cliffs, NJ.: Prentice-Hall, Inc.

Figlio, D.N., and Lucas, M.E. (2003). *Do High Grading Standards Effect Student Performance?*. Journal of Public Economics 88 (2004) 1815-1834.

Frisbie, D.A and Waltman, K.K.(1992). *Developing a personal grading plan*. *Educational Measurement: Issues and Practice*. Iowa: National Council on Measurement in Education.

Gelman, A., Carlin, J.B., Stern, H.S., and Rubin, D.B. (1995). *Bayesian Data Analysis*. London: Chapman & Hall.

Glass, G.V., and Stanley, J.C.(1970). *Statistical Methods in Education and Psychology*. Englewood Cliffs, New Jersey: Prentice-Hall, Inc.

Hogg, R.V. and Craig, A.T. (1978). *Introduction to Mathematical Statistics*. 4th ed. New York: MacMillan Publishing Co.,Inc.

Hogg, R.V., McKean, J.W. and Craig, A.T. (2005). *Introduction to Mathematical Statistics*. 6th ed. New Jersey: Pearson Prentice Hall.

Jasra, A., Holmes, C.C., and Stephens, D.A. (2005). *Markov Chain Monte Carlo Methods and the Label Switching Problem in Bayesian Mixture Modeling*. Vol. 20. No.1, 50-57. Statistical Sciece; Institute of Mathematical Statistics.

Johnson, B. and Christensen, L.(2000). *Educational Research: Chapter 5 -Quantitative and Qualitative Approaches*. 2nd ed. Alabama ,Allyn and Bacon Inc.

Jones, P.N., (1991). *Mixture Distributions Fitted to Instrument Counts*. Rev. Sci. Instrum. 62(5); Australia: American Institute of Physics.

Lawrence D.A (2005). A Guide to Teaching & Learning Practices*: Chapter 13-Grading*. Tallahassee , Florida State University. Unpublished.

Lawrence, H.C.(1995). *Grading Students*. ED398239. ERIC Clearinghouse on Assessment and Evaluation, Washington, DC.

Lee, P.M., (1989). *Bayesian Statistics*. New York. Oxford University Press.

Martuza, V.R. (1977). *Applying Norm-Referenced and Criterion-Referenced: Measurement and Evaluation*. Boston Massachusetts: Allyn and Bacon Inc.

Merle, W.T. (1968). *Statistics in Education and psychology: A First Course*. New York: The Macmillan Company.

Newman, K. (2005). *Bayesian Inference*. Lecture Notes. http://www.creem.st-and.ac.uk/ken/mt4531.html.

Peers, I.S.(1996). *Statistical Analysis For Education and Psychology Researchers*. London: Falmer Press.

Pradeep, D. and John, G. (2005). *Grading Exams: 100, 99, ... , 1 or A, B, C? Incentives in Games of Status*. 3-6.

Press, S.J. (2003). *Subjective and Objective Bayesian Statistic: Principle, Models and Application.* New Jersey: John Wiley & Son, Inc.

Raftery, A.E (1996), *Hypothesis Testing and Model Selection,* in S.R.W.R. Gilks & D.Spiegelhalter, eds, “Markov Chain Monte Carlo in Pratice”, Chapman and Hall, pg. 163-188.

Robert, L.W. (1972). *An Introduction to Bayesian Inference and Decision*. New York: Holt, Rinehart and Winston, Inc.

Robert, M.H. (1998). *Assessment and Evaluation of Developmental Learning: Qualitative Individual Assessment and Evaluation Models*. Westport: Greenwood Publishing Group Inc.

Spencer, C. (1983). *Grading on the Curve*. Antic, 1(6): 64

Stanley, J.C. and Hopkins, K.D. (1972). *Educational Psychological Measurement and Evaluation*. Englewood Cliffs, NJ.: Prentice Hall, Inc.

Stephens, M. (2000). *Bayesian Analysis of Mixture Models with An Unknown Number of Component-An Alternative To Reversible Jump Methods*. University of Oxford. The Annals Statistics. Vol. 28, No.1, 40-74.

Walsh, B. (2004). *Markov Chain Monte Carlo*. Lecture Notes for EEB 581.

Walvoord, B.E. and Anderson, V.J. (1998). *Effective Grading: A Tool for Learning and Assessment*. San Francisco: Jossey-Bass Publisher.