

# BAYESIAN CLUSTERING FOR SYNCHRONIZED DIVING

by

Vivien Yat Wan Wong

Bachelor of Science, Simon Fraser University, 2007

Bachelor of Science, University of British Columbia, 2000

A PROJECT SUBMITTED IN PARTIAL FULFILLMENT  
OF THE REQUIREMENTS FOR THE DEGREE OF  
MASTER OF SCIENCE  
in the Department  
of  
Statistics and Actuarial Science

© Vivien Yat Wan Wong 2009  
SIMON FRASER UNIVERSITY  
Summer 2009

All rights reserved. This work may not be  
reproduced in whole or in part, by photocopy  
or other means, without the permission of the author.

## APPROVAL

**Name:** Vivien Yat Wan Wong  
**Degree:** Master of Science  
**Title of Project:** Bayesian Clustering for Synchronized Diving

**Examining Committee:** Dr. Derek Bingham  
Chair

---

Dr. Tim Swartz  
Senior Supervisor  
Simon Fraser University

---

Dr. Paramjit Gill  
Supervisor  
University of British Columbia Okanagan

---

Dr. Carl Schwarz  
External Examiner  
Simon Fraser University

**Date Approved:** \_\_\_\_\_

# Abstract

Synchronized diving is one of the most widely viewed Olympic events after its first appearance at the Sydney Olympic Games in 2000. It gives spectators the ability to compare performances of divers on their own without much understanding of the technical details of the sport. In this project, we develop methodology to investigate the complexity of judges' scores and the relative behaviour of judges from synchronized diving events. We explore a Bayesian clustering methodology as introduced in Gill, Swartz and Treschow (2007) to cluster judges. A model that captures the characteristics of the judges' scores is introduced and a dataset from the 12th FINA World Championships in Melbourne 2007 is fit using the proposed model. We demonstrate how the missing values raised from the judging system can be easily handled in a Bayesian analysis via implementation in WinBUGS. The analysis may reveal associations among judges.

**Key words:** Bayesian modelling; WinBUGS; Clustering; Markov chain Monte Carlo; Judging; Missing values; Synchronized diving

*In memory of my mother, Chi-Ching Law(1953-1992)  
and my grandmother, Kai-Fong Wong(1919-2001)*

# Acknowledgments

I am indebted to my senior supervisor, Dr. Tim Swartz, who made great effort to guide me throughout my studies in statistics. He was always ready and willing to help. He taught me how to be persistent in accomplishing the project. I really appreciate his dedication and patience. It is not possible to thank him enough for his support, guidance and encouragement throughout my studies at SFU. It is my great honour to be his student.

I would like to extend thanks to my co-supervisor, Dr. Paramjit Gill, who made valuable suggestions on my project. His expertise made programming in WinBUGS seems easy. Many thanks to Dr. Carl Schwarz for his comments and suggestions that led to great improvement in my project.

In the summer of 2008, I was very fortunate to work with Dr. Charmaine Dean, Dr. Jiguo Cao and Dr. Douglas Woolford for four months. This was a very valuable experience for me and I am grateful to Dr. Charmaine Dean for giving me the opportunity to work on her research project. Special thanks to Ian Bercovitz for affording me the trust to teach tutorials for the statistics course in the program in Liberal and Business Studies.

I would also like to thank the Department of Statistics and Actuarial Science for providing a supportive environment for me to pursue my studies in the Graduate Program. Sincere gratitude to the Department Chair, Dr. Richard Lockhart, who gave me the opportunity to pursue my studies at SFU, and Sadika, Kelly and Charlene for their kindness and assistance.

Thanks to all of the graduate students in the Department for making my studies enjoyable in the last couple of years. Special thanks to Saman Muthukumarana for many discussions on data modelling and statistical programming. I would also like to thank Suli Ma, Joslin Goh, Conghui Qu, Jingyu Chen, Zhong Wan, Donghong Wu, Monica Lu and Qifeng Jiang for their friendship.

Thank you my friend, Eugenia, for encouraging me to pursue my studies in the Graduate Program and standing by my side at all times. Last, but certainly not least, thank you Gary for your understanding and sacrifices, and my family for your endless support in so many ways.

# Contents

<b>Approval</b>	<b>ii</b>
<b>Abstract</b>	<b>iii</b>
<b>Acknowledgments</b>	<b>v</b>
<b>Contents</b>	<b>vii</b>
<b>List of Tables</b>	<b>ix</b>
<b>List of Figures</b>	<b>x</b>
<b>1 Introduction</b>	<b>1</b>
1.1 History of Synchronized Diving . . . . .	1
1.2 Scoring . . . . .	2
1.3 Motivation of the Project . . . . .	3
1.4 Organization of the Project . . . . .	3
<b>2 Bayesian Model Development</b>	<b>5</b>
2.1 Missing Scores . . . . .	6
2.2 Randomness in Parameters and Prior Information . . . . .	7
2.3 Components of the Bayesian Model . . . . .	8
<b>3 Bayesian Analysis via WinBUGS</b>	<b>10</b>
3.1 Simulated Data . . . . .	10
3.1.1 MCMC Analysis . . . . .	11
3.1.2 Interpretations of the Results . . . . .	13

3.2	FINA Data . . . . .	16
3.2.1	MCMC Analysis . . . . .	16
3.2.2	Interpretations of the Results . . . . .	20
<b>4</b>	<b>Bayesian Cluster Analysis</b>	<b>21</b>
4.1	Measure of Dissimilarity and Clustering Method . . . . .	21
4.2	Cluster Analysis . . . . .	22
<b>5</b>	<b>Conclusions and Remarks</b>	<b>26</b>
<b>A</b>	<b>The Judges' Nationalities</b>	<b>28</b>
<b>B</b>	<b>WinBUGS Code for the Simulated Data</b>	<b>29</b>
<b>C</b>	<b>WinBUGS Code for the FINA Data</b>	<b>32</b>
	<b>Bibliography</b>	<b>34</b>



# List of Tables

3.1	Posterior means of parameters for the simulated dataset. . . . .	13
3.2	Posterior mean estimates of $\sigma_{jj}$ and $\rho_{jk}$ for the FINA dataset. . . . .	19
4.1	Posterior probabilities that judges are singletons. . . . .	23
4.2	Posterior probabilities that pairs of judges are clustered. . . . .	24

# List of Figures

2.1	The judges' connectivity in the competition. . . . .	6
3.1	The BGR plot of $\rho_{23}$ . . . . .	12
3.2	The autocorrelation plot of the Markov chain output of $\rho_{23}$ (thin 5). . . . .	13
3.3	The density of $\rho_{23}$ . . . . .	14
3.4	The BGR plot of $\rho_{14}$ . . . . .	17
3.5	The autocorrelation plot of the Markov chain output of $\rho_{14}$ (thin 100). . . . .	18
3.6	The trace plot of $\rho_{14}$ . . . . .	18
4.1	The dendrogram for judges using <i>hclust</i> . . . . .	25

# Chapter 1

## Introduction

Synchronized diving is an aquatic event in the Summer Olympic Games and many other international aquatics competitions. It has become one of the most widely viewed events after its first appearance at the Sydney Olympic Games in 2000 as it gives spectators the ability to compare performances of two divers on their own.

This project began with a well-defined problem. We wanted to look at judges' scores from synchronized diving events and to develop Bayesian clustering methodology as introduced in Gill, Swartz and Treschow (2007) to cluster judges according to their scoring behaviours. The challenges involved data collection, modelling issues, treatment of missing values, the determination of a sensible dissimilarity measurement, a stopping criterion for clustering and implementation via WinBUGS. The choice of prior distributions was also a focus in this project.

### 1.1 History of Synchronized Diving

The first competitive diving event was a simple plunging competition in Great Britain in 1883. It was a standing dive made headfirst from a firm base. The goal of the diver was to travel the farthest distance while keeping the body motionless and facing downward. The sport we know today was developed from gymnastics rather than swimming in the 19th century. The Swedish and German gymnasts preferred to land in water rather than on hard floors during their practice. The aquatics-diving event appeared in the Games of the II Olympia, the former Olympic Games, in St. Louis, Missouri, the United States in 1904. There were two competitions: platform and plunge for distance where the Americans' and

Germans' divers were the competitors in these events. In the 1908 Summer Olympic Games in London, United Kingdom, a springboard diving event was introduced, while plunge for distance was removed. Women's platform and springboard events appeared in 1912 and 1920 respectively. Synchronized diving events were introduced in the 2000 Summer Olympics in Sydney, Australia.

## 1.2 Scoring

Male divers have to perform six dives and female divers have to perform five dives in each of preliminary, semi-final and final rounds for individual events. For synchronized diving events, there are only preliminary and final rounds in most of the international competitions. Pairs of divers perform each dive simultaneously. Each pair of divers have to perform at least one dive consisting of forward facing take-off by both divers, at least one dive consisting of backward facing take-off by both divers and at least one dive with a combination of forward and backward facing take-off. Each dive is associated with a certain degree of difficulty depending on the type and the position of the dive and the number of half somersaults in the dive. In a synchronized diving competition, there is a panel of nine judges, two of whom judge execution of each diver and five judge synchronization. The execution adjudicators rate technique and execution of each dive, including the starting position, the run, the take-off, the flight, and the entry. The synchronization adjudicators consider the similarity of the divers on their approaches and take-offs, their heights during the dive, their coordinated timing of the movements during the flight, their angles of entry, their distances from the board at entry, and their coordinated timing of entry. The panel of judges score the dive immediately after each dive without communicating to each other. The scores range from 0 to 10 points with 0.5 points increments. The execution score (and synchronized score) is obtained by summing the judges scores with the highest and lowest scores removed, and then multiplying by 0.6 times the degree of difficulty of the dive. The degree of difficulty is a factor that typically ranges from 1.2 to 3.8 in high level competitions. A final score for a dive is the sum of the execution score and the synchronized score. The guidelines for the judges on scoring are as follows:

- Completely failed: 0 points
- Unsatisfactory: 0.5-2.0 points

- Deficient: 2.5-4.5 points
- Satisfactory: 5.0-6.0 points
- Good: 6.5-8.0 points
- Very Good: 8.5-10.0 points

### 1.3 Motivation of the Project

Synchronized diving is not the only sporting event which involves judging, but it is certainly one that is easy to understand by the public. The public can rate the synchronization of each pair of divers on their own without much understanding of the technical details of the sport. As we watch competitions, we also wonder about the “fairness” of the judging system and any associations among the judges and the divers. In the 2002 Olympic Winter Games in Salt Lake City, the pairs figure skating competition involved a scandal in scoring and judging ([www.olympic.org/uk/games/past/innovations\\_uk.asp?OLGT=2&OLGY=2002](http://www.olympic.org/uk/games/past/innovations_uk.asp?OLGT=2&OLGY=2002)) resulting in a suspension of a judge. Although we do not suggest that synchronized diving suffers from misconduct amongst judges, it would be interesting to develop methodology that investigates the results of judging. Little research has been done on the scoring of synchronized diving. Ding, Cheng, Lu and Zhou (2008) suggest computer synchronization analysis for synchronized diving videos.

In competitive diving, each dive involves a combination of skill, timing and coordination. In order to win in international synchronized diving events, divers have to perform consistently well through preliminary and final rounds that consist of 10 dives for women and 12 dives for men. A panel of judges are selected from a committee of renowned judges for each round of the competition. As it is not straightforward to observe any associations among judges by looking at the score sheet for each event, statistical modelling is needed to investigate the complexity of the scores and the relative behaviour of judges.

### 1.4 Organization of the Project

In Chapter 2, we consider a model that capture the characteristics of the judges’ scores in synchronized diving events. The model introduces the “missing” values raised from the judging system. In Chapter 3, we show how the observed and the missing values described

in Chapter 2 can be easily handled in a Bayesian analysis via implementation in WinBUGS. A simulated dataset and a dataset from the 12th FINA World Championships in Melbourne 2007 are fit using the proposed model. In Chapter 4, we propose a dissimilarity measurement and a stopping criterion on clustering judges for the diving event. The clustering algorithm is a Bayesian version of the unweighted paired group method of averaging (UPGMA). We provide some concluding remarks in Chapter 5.

## Chapter 2

# Bayesian Model Development

Having provided an introduction to synchronized diving in Chapter 1, we consider a model that captures the characteristics of the judges' scores in synchronized diving events. Consider the model:

$$\begin{pmatrix} y_{i1} \\ \vdots \\ y_{im} \end{pmatrix} \sim \text{MVN}_m \left[ \begin{pmatrix} \mu_i \\ \vdots \\ \mu_i \end{pmatrix}, \Sigma \right] \quad (2.1)$$

for  $i = 1, \dots, n$  where  $y_{ij}$  denotes the score of the  $j^{\text{th}}$  judge on the  $i^{\text{th}}$  dive. In (2.1), the parameter  $\mu_i$  represents the latent “truth” for the  $i^{\text{th}}$  dive, where we assume that judges' scores are stochastic but centred about the truth. In (2.1), the  $m \times m$  variance-covariance matrix  $\Sigma = (\sigma_{jk})$  describes the relationship between pairs of judges where judges  $j$  and  $k$  have a similar scoring behaviour if their correlation  $\rho_{jk} = \sigma_{jk}/\sigma_{jj}\sigma_{kk}$  is close to 1.0. A judge  $j$  who is less “consistent” would have a larger standard deviation  $\sigma_{jj}$  than other judges. In our problem, interest lies in the relative behaviour of judges as described by  $\Sigma$ , and we treat  $\mu_1, \dots, \mu_n$  as nuisance parameters. We assume independence amongst the dives  $i = 1, \dots, n$ .

In a Bayesian analysis, we require prior distributions on the parameters, and we let

$$\mu_1, \dots, \mu_n \sim \text{N}(\bar{y}_i, 1/\tau), \quad (2.2)$$

$$\Sigma^{-1} \sim \text{Wishart}(v_o, \Omega_o) \quad (2.3)$$

for  $i = 1, \dots, n$  where hyperparameters with a zero subscript are fully specified and  $\bar{y}_i$  is the sample mean score of the  $i^{\text{th}}$  dive. In (2.3), the density for  $\Sigma^{-1}$  is given by

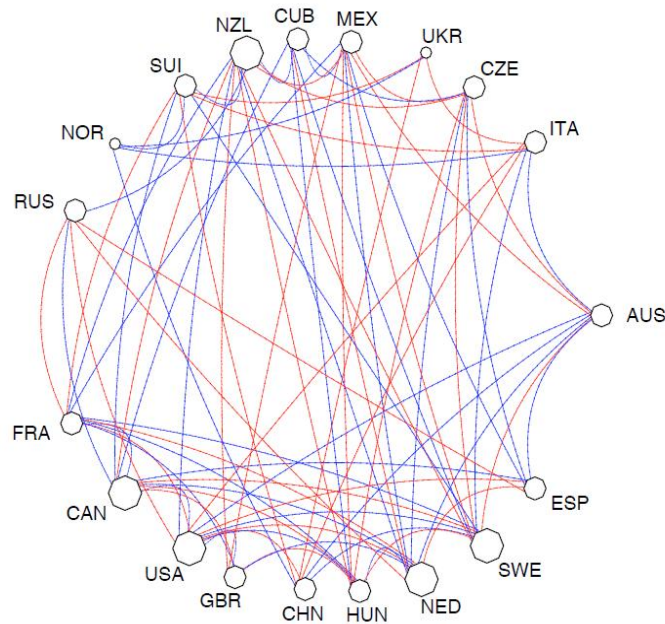
$$f(\Sigma^{-1}) \propto |\Omega_o|^{v_o/2} |\Sigma^{-1}|^{(v_o-m-1)/2} \exp^{-0.5tr(\Omega_o' \Sigma^{-1})}$$

as given in Congdon (2006). In section 2.2, we provide a detailed explanation on the choice of the prior distributions and the hyperparameters.

We believe that Bayesian methodology is the obvious approach to investigate the relative behaviour of judges as it allow us to incorporate our prior knowledge concerning the parameters.

## 2.1 Missing Scores

Figure 2.1: The judges' connectivity in the competition.



In Chapter 1, we mentioned that five judges are responsible for the synchronization scores for the dives. However, the five judges are selected from a committee of judges and the same five judges do not always adjudicate the same events (dives). In the FINA dataset, two judges are regarded as “connected” if they adjudicate on the synchronization of dives in the same event. Figure 2.1 shows the judges' connectivity in the competition where they



are labelled according to their nationalities. A list of the judges' nationalities is given in the Appendix A. The size of the node indicates the frequency that a judge adjudicates, and the line between two nodes shows that the two judges adjudicate in the same event at least once. Thus, the scores sheets do not contain all judges' scores for every dive. In order to simplify the model, we consider the missing scores of the judges who are not assigned to adjudicate the events.

The Bayesian approach treats every single missing value as a parameter that needs to be estimated. The missing value imputations can be easily embedded in the Bayesian computation algorithm. Posterior distributions for both the parameter and the missing data can be obtained as long as the required likelihood function and the prior distributions are provided. In this project, the missing scores and the actual scores from the dataset are the ingredients for specifying the likelihood function.

## 2.2 Randomness in Parameters and Prior Information

The basic components in a Bayesian analysis are the distribution of the data in the form of the likelihood and the prior distributions of the unknown parameters. The posterior distributions of the unknown parameters are then obtained and can be used to make inferences concerning the unknown parameters. One of the big differences between the Bayesian approach and the classical approach is that the unknown parameters are not treated as fixed values. The unknown parameters are random variables.

In this project, we believe that the  $\mu_i$ 's and  $\Sigma^{-1}$  are unknown parameters which arise from a Normal distribution and a Wishart distribution respectively. In (2.2), the parameter  $\mu_i$  represents the true score of the  $i^{th}$  dive, and the  $\bar{y}_i$  represents the sample mean score of the  $i^{th}$  dive given by the five judges where  $\mu_i$  has precision  $\tau$ . We believe that the true score of the  $i^{th}$  dive is centred about the sample mean. Hence, our approach is empirical Bayes. We also assume that  $\mu_i$  is normally distributed around  $\bar{y}_i$  with precision  $\tau$ , and  $\mu_i$  is bounded between 0.0 and 10.0. Hence, the Normal prior is really a truncated Normal.

In (2.1), the parameter  $\Sigma$  is the variance-covariance matrix and its conjugate prior is the inverse Wishart distribution which is the generalization of the inverse Gamma distribution. Thus, the conjugate prior of the precision matrix  $\Sigma^{-1}$  is the Wishart distribution. The degree of freedom parameter  $\nu_o$  and the scale matrix  $\Omega_o$  are two parameters used in specifying the Wishart distribution. The degree of freedom parameter  $\nu_o$  has to be equal to or larger than

$m$  for a proper prior; the scale matrix  $\Omega_o$  is symmetric and positive definite. The prior mean of the  $\Sigma^{-1}$  is  $v_o\Omega_o^{-1}$  (Congdon 2006). We set  $v_o = m$  and  $\Omega_o = I_m$  where  $I_m$  is the  $m \times m$  identity matrix. Summarizing, we have

$$\mu_i \sim N(\bar{y}_i, 1/\tau), \mu_i \in [0, 10]$$

$$\tau \sim \text{Gamma}(1.0, 0.5)$$

$$v_o = m$$

$$\Omega_o = \begin{pmatrix} 1 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & 1 & 0 \\ 0 & \cdots & \cdots & 0 & 1 \end{pmatrix}$$

The prior distribution of the  $\mu_i$  is a weakly informative prior. It represents our best knowledge about the parameter. The prior distribution of the precision matrix  $\Sigma^{-1}$  is a diffuse prior where the standard deviations of the judges' scores are roughly  $1/\sqrt{m}$  and the correlations between judges' scores are around 0.0. The parametrization of  $X \sim \text{Gamma}(a, b)$  is such that  $E(X) = ab$ .

In a Bayesian analysis, the justification of the prior distributions is always needed as the priors can influence the posterior estimates and the rate of convergence of sampling algorithms. The choice of prior distributions is discussed further in Chapter 3.

## 2.3 Components of the Bayesian Model

The framework of the Bayesian approach arises from Bayes' Rule accredited to Reverend Thomas Bayes (c. 1701-61). The posterior distribution is

$$\pi(\theta | y) = \frac{f(y | \theta)\pi(\theta)}{f(y)},$$

where  $\theta$  is the vector of parameters of interest including the missing values, and  $y$  is the observed data. The function  $f(y | \theta)$  is the likelihood of the data given  $\theta$  defined in (2.1).

The function  $\pi(\theta)$  is the prior density defined in (2.2) and (2.3). The function  $f(y)$  is the marginal density of the observed data that serves as the inverse of the norming constant of the posterior distribution and can be calculated as

$$f(y) = \int_{\theta} f(y | \theta) \pi(\theta) d\theta. \quad (2.4)$$

In general, we are interested in inferences about components of the parameter  $\theta$ . In order to do so, this requires the integration of the posterior density with respect to other parameters to obtain the marginal posterior density of a component of  $\theta$ . Typically, the posterior density is a complex function which we cannot integrate. Often, the posterior distribution cannot be fully specified as the norming constant (2.4) is usually not known and is difficult to obtain.

Evans and Swartz (1995) discuss the major techniques for the approximation of integrals in statistics when the integrations are difficult. Markov chain Monte Carlo (MCMC) methods are recommended for high-dimensional problems such as our research problem. MCMC methods are methods for the approximation of integrals using random variates which have the posterior as its equilibrium distribution. The Metropolis-Hastings and Gibbs sampling algorithms are two popular MCMC algorithms. In the next chapter, we discuss the details of model implementation and Bayesian inference using the software WinBUGS which utilizes the Metropolis-Hastings and Gibbs sampling algorithms.

## Chapter 3

# Bayesian Analysis via WinBUGS

The WinBUGS software was developed jointly by researchers at the MRC Biostatistics Unit and the Imperial College School of Medicine at St. Mary's Hospital, London for the BUGS (Bayesian inference Using Gibbs Sampling) project. The BUGS project is concerned with the development of flexible software for the Bayesian analysis of complex statistical models using MCMC methods. There are various software programs available to implement Bayesian analysis. WinBUGS is one of the programs which makes practical MCMC methods available for applied statistics, and it is free to download. The software can be downloaded from the BUGS Project web page at <http://www.mrc-bsu.cam.ac.uk/bugs/>.

In this chapter, we show how the high-dimensional problem discussed in Chapter 2 can be handled via WinBUGS. In order to test the model, an analysis on a simulated dataset is considered first; then a dataset from the 12th FINA World Championships in Melbourne 2007 is fit using the proposed model.

### 3.1 Simulated Data

It is helpful to know answers in advance and for that reason, we generate data from a specified distribution. In this case, we generate

$$\mu_1^*, \dots, \mu_n^* \sim N(7, 1)$$

where  $n = 25$ , and then generate

$$\begin{pmatrix} y_{i1}^* \\ \vdots \\ y_{im}^* \end{pmatrix} \sim \text{MVN}_m \left[ \begin{pmatrix} \mu_i^* \\ \vdots \\ \mu_i^* \end{pmatrix}, \Sigma^* \right]$$

for  $i = 1, \dots, n$  where  $m=3$  and

$$\Sigma^* = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (3.1)$$

Our simulated dataset is therefore based on  $m=3$  judges and  $n=25$  dives with no missing data.

### 3.1.1 MCMC Analysis

We fit the model described in (2.1) on this relatively simple dataset. The ratio (75/31) of data ( $y^*$ ) to parameters ( $\mu_i, \Sigma$ ) from the simulated dataset suggests no problems. The WinBUGS code for the analysis of the simulated data is provided in Appendix B. To perform the MCMC simulation, we only need to specify the likelihood and the prior distributions in the model specification of WinBUGS and provide the dataset in a matrix form. The WinBUGS software allows us to provide the initial values for the parameters to carry out the MCMC simulation or it can assign initial values as described in the WinBUGS manual.

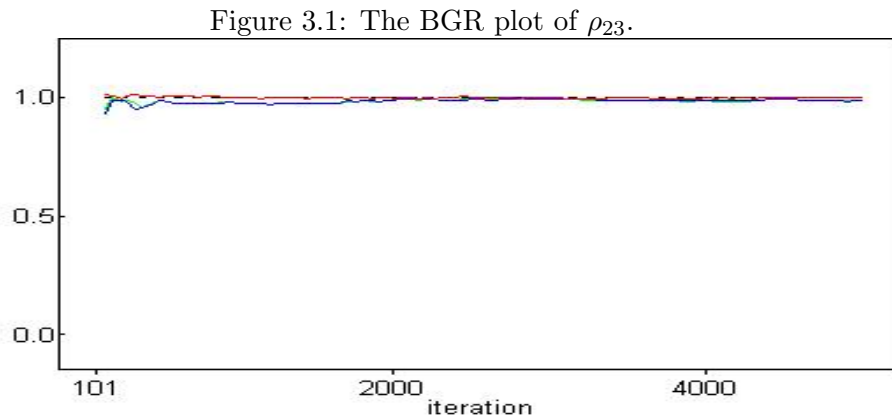
The form of the likelihood function is the same for the simulated data and the FINA data. The likelihood is given by

$$f(y \mid \underline{\mu}, \Sigma) = \prod_{i=1}^n \frac{1}{(2\pi)^{m/2} |\Sigma|^{1/2}} \exp \left\{ -\frac{1}{2} \sum_{i=1}^n \left[ (\underline{y}_i - \underline{\mu}_i \mathbf{1})' \Sigma^{-1} (\underline{y}_i - \underline{\mu}_i \mathbf{1}) \right] \right\}. \quad (3.2)$$

For the  $\mu_i$ 's, a Normal prior is used with mean  $\bar{y}_i$ 's and precision  $\tau$  and a Wishart prior is assigned to the precision matrix as described in section 2.2. We use the initial values provided by WinBUGS in the MCMC simulation.

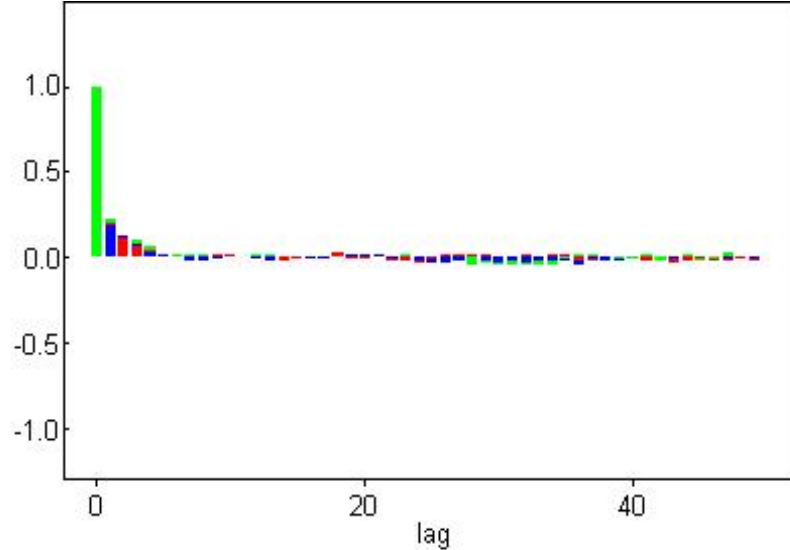
Three independent chains are simulated to investigate the convergence of the Markov chain. We use the functions provided in the ‘‘Sample Monitor Tool’’ in WinBUGS to obtain summaries for the parameters. As described in the WinBUGS manual, WinBUGS provides

the Brooks-Gelman-Rubin (BGR) plot to assess convergence. It calculates the Gelman-Rubin convergence statistic as modified by Brooks and Gelman (1998). The three lines on the plot represent the between chain variation, the within chain variation, and their ratio. As the ratio converges to 1 and both the between and within chain variation are stable (not necessary close to 1), the chains converge. The BGR plot for  $\rho_{23}$  is given in Figure 3.1. There is no evidence of lack of convergence since all three lines are stable and close to 1.0. The BGR plots of the other parameters show similar patterns.



As a Markov chain is a stochastic process with the property that the present state  $i$  depends only on the previous state  $i - 1$ , the dependency between successive variates of the Markov chain has to be assessed. “Thinning” is a standard strategy used to obtain “independent” variates simulated from a Markov chain as variates with high autocorrelations lead to wrong estimation of the standard deviation of the estimates. We use the “auto cor” tool provided by WinBUGS and the autocorrelation plot of  $\rho_{23}$  (thin 5) as given in Figure 3.2. It appears that the correlation between every 5<sup>th</sup> value is not strong. It is therefore appropriate to treat every 5<sup>th</sup> value from the Markov chain as independent. The other parameters have similar autocorrelation plots.

Posterior estimates of the parameters are obtained based on 10000 iterations after a 2000 iterations for burn-in and thinning. The burn-in period is applied such that variates are considered after the Markov chain has reached an equilibrium stage, and the dependency of the variates to the early stage of MCMC simulation has no effect on the estimates. The posterior estimates of the variances  $\sigma_{jj}$  and the correlations  $\rho_{jk}$  are summarized in Table

Figure 3.2: The autocorrelation plot of the Markov chain output of  $\rho_{23}$  (thin 5).

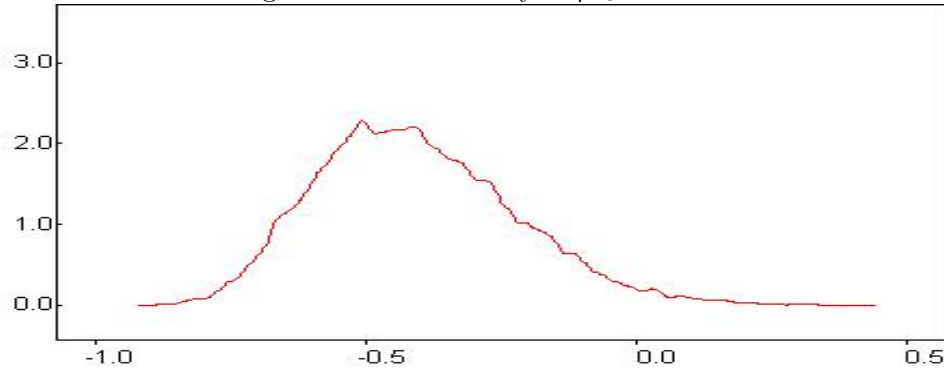
3.1. Figure 3.3 shows the density of  $\rho_{23}$ . The densities of  $\rho_{12}$  and  $\rho_{13}$  have similar shapes which are unimodal and symmetric around the mean value. The summary statistics and the densities of the posterior estimates can be easily obtained using the “stats” and “density” tools in WinBUGS. These results provide clear pictures of the estimates and the error of the estimated values. By looking at the density of the simulated values, we are also able to assess the model and the assumptions on the parameters.

Table 3.1: Posterior means of parameters for the simulated dataset.

Parameter	Mean	SD	Parameter	Mean	SD
$\sigma_{11}$	0.67	0.22	$\rho_{12}$	-0.46	0.17
$\sigma_{22}$	0.69	0.22	$\rho_{13}$	-0.37	0.19
$\sigma_{33}$	0.60	0.20	$\rho_{23}$	-0.41	0.18

### 3.1.2 Interpretations of the Results

As the purpose of working on a simulated dataset is to assess whether the analysis is trustworthy, we consider the implications of Figure 3.1, Figure 3.2, Figure 3.3 and Table 3.1.

Figure 3.3: The density of  $\rho_{23}$ .

The sample size and the number of parameters of interest are small in the simulated dataset, so the number of iterations needed to reach convergence is also small as expected. Moreover, the autocorrelations between variates from MCMC are not high and we are able to use the standard thinning method to solve the dependency issue between successive values. We would expect to see higher autocorrelations from the MCMC results for the FINA dataset as a consequence of the missing value imputations. Thus, a larger number of iterations may be needed in order to offset the loss of sample values from thinning.

From Table 3.1, the posterior estimates of the parameters are within reasonable ranges with similar standard deviations. The  $\hat{\sigma}_{jj}$ 's are around 0.67 and the  $\hat{\rho}_{jk}$ 's are around  $-0.4$  with standard deviations roughly 0.22. We expect as the number of dives  $n$  increases, the standard deviation for the posterior estimates should decrease. Since the simulated dataset only considered  $m = 3$  judges, we believe that it is reasonable to have all negative values for the  $\rho$ 's.

To gain an understanding of why all of the covariances are negative, it is instructive to look at the full conditional distributions. Full conditional distributions are used in the Gibbs sampling algorithm as the distributions from which variates are generated. WinBUGS uses the Gibbs sampling method for well-behaved models with log concave densities and the



Metropolis-Hastings method for non-standard models. The full conditional for  $\mu_i$  is

$$\begin{aligned}
[\mu_i|\cdot] &\propto f(y|\mu, \Sigma) = \prod_{i=1}^n f(y_i|\mu_i, \Sigma) \propto f(y_i|\mu_i, \Sigma) \propto \exp\left\{-\frac{1}{2}(y_i - \mu_i\mathbf{1})'\Sigma^{-1}(y_i - \mu_i\mathbf{1})\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left(\mu_i^2\mathbf{1}'\Sigma^{-1}\mathbf{1} - 2\mu_i\mathbf{1}'\Sigma^{-1}y_i\right)\right\} = \exp\left\{-\frac{1}{2/(\mathbf{1}'\Sigma^{-1}\mathbf{1})}\left(\mu_i^2 - 2\mu_i\frac{\mathbf{1}'\Sigma^{-1}y_i}{\mathbf{1}'\Sigma^{-1}\mathbf{1}}\right)\right\} \\
&\propto \exp\left\{-\frac{1}{2/(\mathbf{1}'\Sigma^{-1}\mathbf{1})}\left(\mu_i - \frac{\mathbf{1}'\Sigma^{-1}y_i}{\mathbf{1}'\Sigma^{-1}\mathbf{1}}\right)^2\right\} \sim N\left(\frac{\mathbf{1}'\Sigma^{-1}y_i}{\mathbf{1}'\Sigma^{-1}\mathbf{1}}, \frac{1}{\mathbf{1}'\Sigma^{-1}\mathbf{1}}\right) \quad (3.3)
\end{aligned}$$

As  $\mu_i$  is generated near  $\bar{y}_i$ , the centre of the data  $y_{i1}, \dots, y_{im}$  for each  $i = 1, \dots, n$ , this suggests that the sample correlation between  $y_{ji}$  and  $y_{ki}$ ;  $i = 1, \dots, n$  is likely to be negative since large (small)  $y_{ji}$ 's tend to occur with small (large)  $y_{ki}$ 's. To see this more clearly, consider the case of  $m = 2$  judges and again assume that  $\mu_i \approx \bar{y}_i$  for  $i = 1, \dots, n$ . Then the correlation between  $y_{1i}$  and  $y_{2i}$  is given by

$$\begin{aligned}
&\frac{\sum_{i=1}^n (y_{i1} - \mu_i)(y_{i2} - \mu_i)}{\sqrt{\sum_{i=1}^n (y_{i1} - \mu_i)^2 \sum_{i=1}^n (y_{i2} - \mu_i)^2}} \approx \frac{\sum_{i=1}^n (y_{i1} - \bar{y}_i)(y_{i2} - \bar{y}_i)}{\sqrt{\sum_{i=1}^n (y_{i1} - \bar{y}_i)^2 \sum_{i=1}^n (y_{i2} - \bar{y}_i)^2}} \\
&= \frac{\sum_{i=1}^n \left(\frac{y_{i1} - y_{i2}}{2}\right) \left(\frac{y_{i2} - y_{i1}}{2}\right)}{\sqrt{\sum_{i=1}^n \left(\frac{y_{i1} - y_{i2}}{2}\right)^2 \sum_{i=1}^n \left(\frac{y_{i2} - y_{i1}}{2}\right)^2}} = \frac{\sum_{i=1}^n (y_{i1} - y_{i2})(y_{i2} - y_{i1})}{\sqrt{\sum_{i=1}^n (y_{i1} - y_{i2})^2 \sum_{i=1}^n (y_{i2} - y_{i1})^2}} \\
&= \frac{-\sum_{i=1}^n (y_{i1} - y_{i2})^2}{\sum_{i=1}^n (y_{i1} - y_{i2})^2} = -1
\end{aligned}$$

Therefore, without deriving the full conditional for  $\Sigma$ , we see the pattern that emerges. The  $\mu_i$ 's are generated near their data means, and then  $\Sigma$  is generated with strong negative correlations. And from (3.3), the stronger the correlations in  $\Sigma$ , the larger  $\mathbf{1}'\Sigma^{-1}\mathbf{1}$ , and a repeated cycle of generating  $\mu_i$  near  $\bar{y}_i$  is obtained.

We have used an empirical Bayes approach for the simulated dataset by assigning a prior that depends on data. The  $\mu_i$ 's represent the truth of the performances of dives in an event and the empirical Bayes method allows us to capture the best knowledge that we have about the parameters based on the dataset. We apply a vague prior on the  $\Sigma^{-1}$ . It is known that a small value of  $v_o$  provides a vague prior and a large value provides an informative prior (Carlin and Louis 2000, Congdon 2006). Thus,  $v_o$  is usually chosen to be the smallest possible value that gives a proper vague prior; in our case, it is  $v_o = m$ .

## 3.2 FINA Data

The Fédération Internationale de Natation (FINA) is the International Federation recognized by the International Olympic Committee for administering international competitions in aquatic sports. The World Aquatic Championships is one of the international aquatic competitions run by FINA. It has been held every two years since 2001. The dataset from the 12th FINA World Championships in Melbourne 2007 is collected from the Diving Plongeon Canada website ([www.diving.ca](http://www.diving.ca)). The synchronized diving data are extracted from the whole dataset and is fit using the proposed model in (2.1). Data from three completely failed dives are removed as these low scores do not provide any information about the relative behaviour of judges. Appropriate prior distributions were used in the Bayesian analysis as described in section 2.2.

The objective of fitting the FINA dataset is to explore the relative behaviour of the judges based on the synchronization scores given in the synchronized diving events. As we described in section 2.1, the actual scores from the dataset are not “complete” in the sense that only five judges scores are available per dive. We consider the missing scores as parameters but use the missing scores as data so that the likelihood is easily specified.

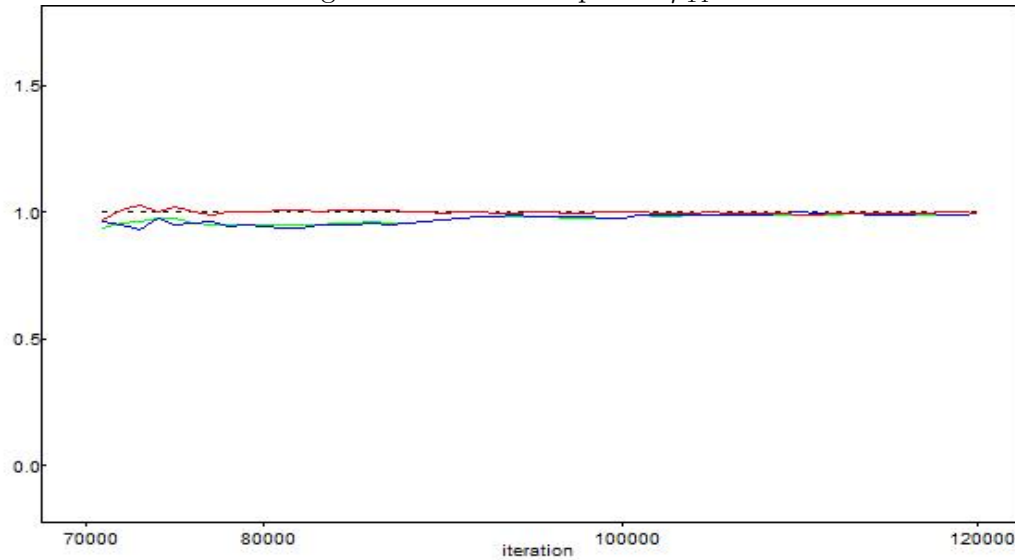
The dataset is composed of the synchronization scores of 607 dives given by 19 judges from the committee of judges. There were 23 judges in the committee and 4 judges did not adjudicate the synchronization of dives. The synchronization scores of 607 dives were obtained from diving events categorized by 3m springboard and 10m platform, preliminary and final rounds, for females and males in the competition. The effect of the categories of the events is not discussed in this project. We assume that the scoring behaviour of judges is the same throughout the whole competition regardless of the types of events.

### 3.2.1 MCMC Analysis

We fit the model described in (2.1) on the dataset. It seems that the ratio (3035/9295) of data ( $y^*$ ) to parameters and missing scores ( $\mu, \Sigma, y_{miss}$ ) from the dataset might suggest problems. The missing scores are treated as parameters in the process of missing values imputation. The parameters of interest in our model are the  $\mu$  and  $\Sigma$ , and there are  $607 + 190 = 797$  of them. Appendix C provides the WinBUGS code for the analysis of the FINA data.

Three independent chains are simulated to investigate the convergence of the estimated

parameters. We use similar tools as described in section 3.1.1 to obtain the summary statistics of the posterior estimates. The BGR plot for  $\rho_{14}$  is given in Figure 3.4. It appears that the between chain variation plot and the within chain variation plot are stable and their ratio is close to one after 90000 iterations. As the BGR plots for the other parameters show similar patterns, we believe convergence is obtained after 90000 iterations.

Figure 3.4: The BGR plot of  $\rho_{14}$ .

The dependency between successive variates of the Markov chain is examined. The autocorrelation plot of  $\hat{\rho}_{14}$  (thin 100) is given in Figure 3.5. It appears that the correlations between every 100<sup>th</sup> variate from the Markov chain are not strong. It is therefore appropriate to treat every 100<sup>th</sup> variate as independent. The other parameters of interest have similar autocorrelation plots. Figure 3.6 provides the trace plot of  $\rho_{14}$  obtained from the MCMC algorithm. It appears that the variates are getting more stable as the MCMC iterations move on.

Posterior estimates of the parameters are obtained based on 125000 iterations after a burn-in of 90000 iterations (thin 100) from three independent chains. Table 3.2 summarizes the posterior estimates of the variances  $\sigma_{jj}$  and the correlations  $\rho_{jk}$  obtained from MCMC.

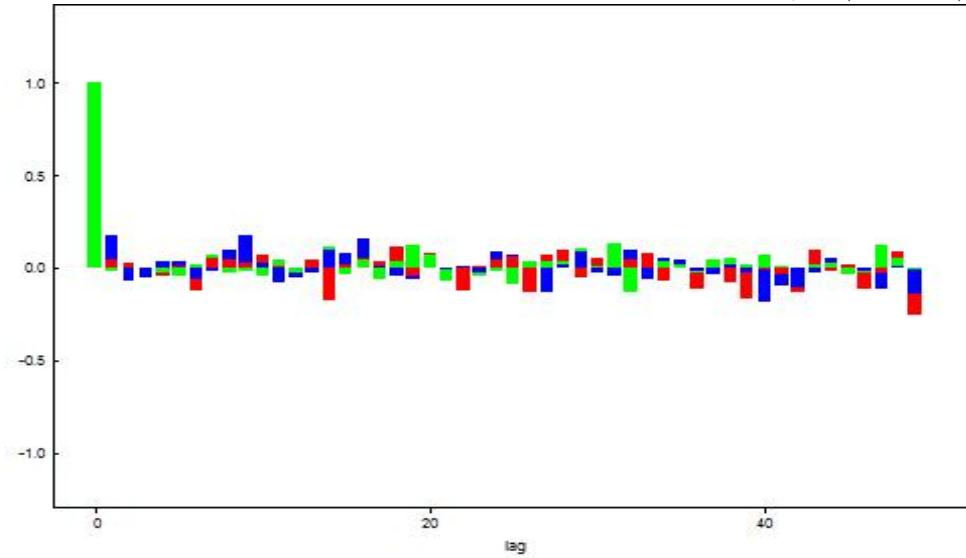
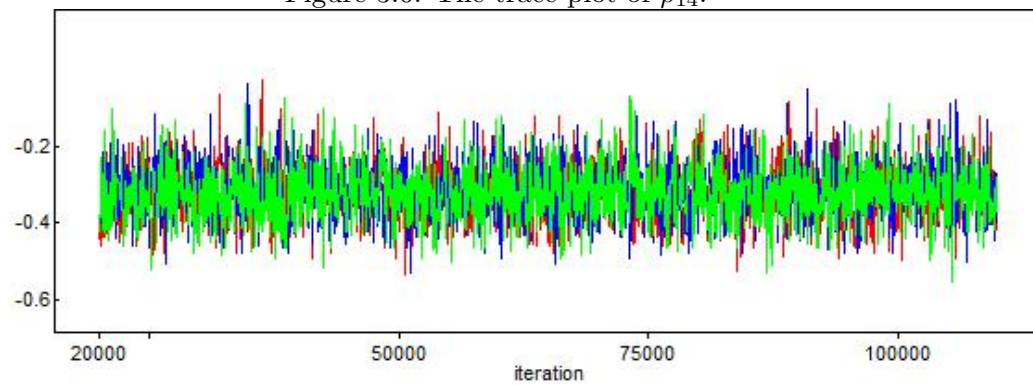
Figure 3.5: The autocorrelation plot of the Markov chain output of  $\rho_{14}$  (thin 100).Figure 3.6: The trace plot of  $\rho_{14}$ .

Table 3.2: Posterior mean estimates of  $\sigma_{jj}$  and  $\rho_{jk}$  for the FINA dataset.

	FRA	CAN	USA	GBR	CHN	HUN	NED	SWE	ESP	AUS	ITA	CZE	UKR	MEX	CUB	NZL	SUI	NOR	RUS
FRA	0.36																		
CAN	-0.12	0.40																	
USA	-0.32	-0.37	0.47																
GBR	-0.33	-0.23	0.11	0.38															
CHN	0.33	0.36	-0.37	-0.30	0.35														
HUN	-0.09	-0.30	-0.10	0.18	-0.17	0.34													
NED	-0.19	-0.20	0.24	-0.19	-0.33	-0.06	0.33												
SWE	-0.16	-0.08	0.00	-0.03	-0.32	-0.10	0.61	0.36											
ESP	0.33	-0.22	-0.09	-0.25	0.19	-0.08	-0.05	0.06	0.41										
AUS	-0.06	-0.33	-0.20	0.27	-0.26	0.49	-0.04	-0.02	-0.20	0.39									
ITA	0.15	0.38	-0.30	-0.03	0.14	-0.15	-0.35	-0.31	-0.01	-0.32	0.37								
CZE	0.08	0.36	0.02	0.15	0.23	-0.29	-0.33	-0.41	-0.32	-0.36	0.41	0.41							
UKR	0.12	0.05	-0.08	-0.28	0.04	-0.13	0.17	0.23	0.15	-0.03	-0.19	-0.19	0.39						
MEX	-0.13	0.60	-0.12	-0.15	0.25	-0.14	-0.39	-0.27	-0.31	-0.24	0.31	0.23	-0.06	0.38					
CUB	0.30	-0.32	0.10	0.18	-0.01	-0.14	-0.20	-0.18	0.16	0.08	-0.01	-0.06	-0.08	-0.12	0.27				
NZL	-0.07	-0.28	0.10	-0.26	0.09	-0.03	-0.11	-0.20	0.41	0.06	-0.13	-0.30	0.10	-0.19	-0.02	0.33			
SUI	-0.40	-0.37	0.41	0.54	-0.45	0.32	0.02	-0.13	-0.34	0.25	-0.12	0.05	-0.38	-0.10	0.11	-0.16	0.36		
NOR	-0.17	-0.34	0.25	0.08	-0.36	0.12	0.34	0.34	0.00	0.26	-0.52	-0.36	-0.20	-0.31	0.01	0.08	0.12	0.44	
RUS	-0.11	-0.14	-0.31	0.49	-0.19	0.43	-0.11	0.02	-0.36	0.50	0.00	-0.03	-0.19	-0.06	0.04	-0.32	0.35	0.05	0.42

( $\sigma_{jj}$  on the diagonal and  $\rho_{jk}$  on the off-diagonal)

### 3.2.2 Interpretations of the Results

For the FINA dataset, the sample size is small as compared to the number of parameters, so the number of iterations needed to reach convergence is high as expected. The autocorrelations between variates from MCMC are high and we therefore use every 100<sup>th</sup> value from the simulation. Thus, a larger number of iterations is needed for inference.

The posterior estimates of the parameters appear reasonable. The posterior  $\hat{\sigma}_{jj}$ 's are around 0.40 and the posterior  $\hat{\rho}_{jk}$ 's range from  $-0.52$  to  $0.61$  with standard deviations roughly 0.20. It is expected that more than half of the correlations should be negative for the reasons given in section 3.1.2. Different priors on  $\Sigma^{-1}$  have been applied to the model and the posterior estimates of  $\sigma_{jj}$ 's and  $\rho_{jk}$ 's are similar.

The posterior estimate of each  $\rho_{jk}$  provides a measure of the linear relationship between two judges' scores. However, we want to examine the relative behaviour of judges collectively. A Bayesian cluster analysis is used to carry out the analysis and the relative behaviour of judges is revealed by the groupings resulting from the cluster analysis.

## Chapter 4

# Bayesian Cluster Analysis

Cluster analysis refers to methods that groups data so as to capture the natural structure of the data. The idea of clustering is basic in the sense that it makes no assumptions concerning the number or the structure of groups resulting from the analysis. A difficulty with clustering methodologies is that the total number of ways to partition objects into groups increases enormously as the number of objects increases (Johnson and Wichern 2007). For the FINA dataset, there are over five trillion ways to cluster the judges into groups of varying sizes. Thus, instead of enumerating and assessing the complete list of possible clusters, clustering methods have been developed to find good partitions without looking at all possible clusters.

In general, the basic ingredients for clustering are a measure of dissimilarity (similarity) and the clustering method. However, the results obtained by different clustering methods can differ and no statistical significance test has been developed to test the fit of the clusters. We introduce a Bayesian approach to provide a way to assess the “credibility” of the resulting clusters. This approach is then applied on the results obtained from Chapter 3 and the relative behaviour of judges is revealed by the Bayesian clustering analysis.

### 4.1 Measure of Dissimilarity and Clustering Method

The correlation between judges’ scores is an obvious choice for the measure of dissimilarity in the clustering method as it represents the linear relationship between two judges’ scoring behaviour. We use

$$d_{jk} = 1 - \rho_{jk} \tag{4.1}$$

as the dissimilarity measure in our algorithm such that a large positive correlation becomes a small difference between two judges  $j$  and  $k$ , and a small or negative correlation becomes a large difference between the two judges.

We obtain a  $19 \times 19$  symmetric matrix of similarities  $\mathbf{D}=(d_{jk})$  from (4.1) using a single iteration of MCMC output. We apply the agglomerative hierarchical clustering algorithm of Lance and Williams (1996) known as UPGMA (unweighted paired group method of averaging) to determine the merging of objects in successive steps. Two objects (an individual or a cluster) with the smallest average  $d_{jk}$  merges to form a new cluster in each step.

In most clustering software packages, the resulting clusters can be represented by a dendrogram or a tree graph that shows the level at which the clusters merge. In most clustering algorithms, clustering continues leaving a single entity at the last step of the process. We adopt the idea from Gill, Swartz and Treschow (2007) to impose a stopping criterion for the clustering algorithm and obtain probabilities of the clusters obtained.

## 4.2 Cluster Analysis

The “coda” tool from WinBUGS is used to obtain 1000 correlation matrices from the MCMC output. We adopt a clustering program written in the Fortran programming language by Dr. Tim Swartz and carry out the clustering algorithm as described in section 4.1. We propose a stopping criterion

$$d_{max} = 0.70$$

such that merging of any two objects  $j$  and  $k$  (an individual or a cluster) stops when the minimum dissimilarity amongst the current clusters

$$d_{jk} > d_{max}. \tag{4.2}$$

In other words, the clustering algorithm stops when the smallest average correlation between objects is smaller than 0.30. Cohen (1988) defines that a positive correlation between 0.30 to 0.50 is regarded as a medium correlation. Thus, we use this guideline to propose the stopping criterion. Another way to set the stopping criterion is to see any natural gaps among all the correlations. However, it is an ad hoc method to set the criterion after viewing the data and the values obtained from MCMC.

Table 4.1 and Table 4.2 show the clustering results. Table 4.1 shows the posterior probabilities that judges are singletons. A judge  $j$  is a singleton when the judge is not clustered



with another object (another judge or a cluster of judges) at the end of the clustering algorithm. Since 1000 variates of MCMC output are used, the posterior probability that judge  $j$  is a singleton can be approximated by the proportion of time that judge  $j$  is a singleton. A high probability suggests that judge  $j$ 's scoring behaviour is different from the other judges' in the competition. Table 4.2 shows the posterior probability that a pair of judges is clustered. In other words, judges who are clustered as a group have similar scoring behaviour. For example, judges from Canada and Mexico have a similar scoring behaviour 93% of the time as the posterior probability that they are clustered is 0.93, while judges from Mexico and the United States have a similar scoring behaviour 1% of the time as the posterior probability that they are clustered is 0.10.

Table 4.1: Posterior probabilities that judges are singletons.

Judge	Probability	Judge	Probability	Judge	Probability	Judge	Probability
CAN	0.00	SWE	0.00	SUI	0.00	AUS	0.01
NED	0.01	RUS	0.01	GBR	0.02	MEX	0.04
NOR	0.05	ITA	0.05	ESP	0.06	HUN	0.10
FRA	0.11	CHN	0.12	CZE	0.16	UKR	0.23
USA	0.27	NZL	0.30	CUB	0.34		

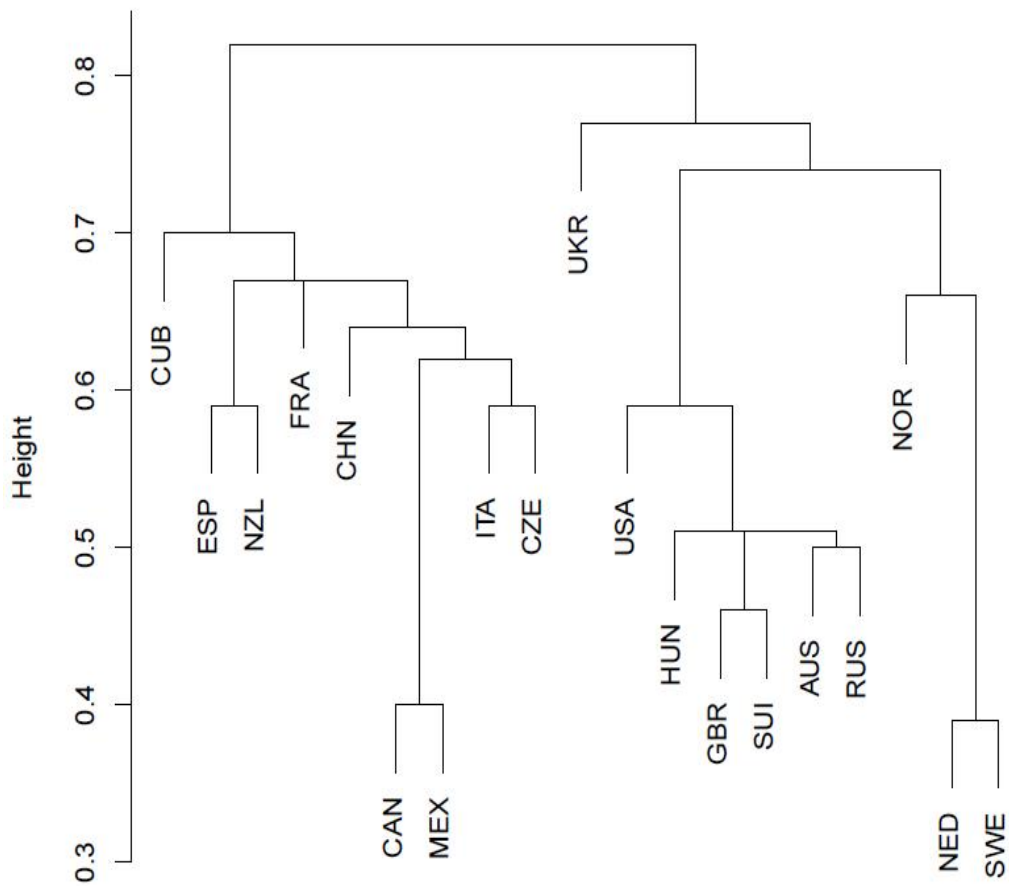
The results from both Table 4.1 and Table 4.2 show that most of the judges agree with other judges. Based on the 1000 correlation matrices obtained from MCMC output, the judges from Canada, Sweden and Switzerland agree with other judges more than 99% of the time as the posterior probabilities that they are singletons are 0.00 ( $< 0.01$ ). Conversely, the posterior probability that the judge from New Zealand is a singleton is 0.30 and the posterior probability that the judge from Cuba is a singleton is 0.34. These two judges are not clustered with the other judges more than 30% of the time; thus, they do not agree with other judges more than 30% of the time. Table 4.2 shows the posterior probabilities that some pairs of judges are clustered are small. For example, the posterior probabilities that the judge from Canada clusters with other judges are small, except with the judges from Mexico, Italy, Czech Republic and China. Similarly, the posterior probabilities that the judge from the Netherlands clusters with other judges are small, except with the judges from Sweden, Norway, Ukraine and the United States. Thus, there exists distinct clusters of judges that characterized judges' scoring behaviours based on the posterior probabilities.

Table 4.2: Posterior probabilities that pairs of judges are clustered.

	FRA	CAN	USA	GBR	CHN	HUN	NED	SWE	ESP	AUS	ITA	CZE	UKR	MEX	CUB	NZL	SUI	NOR
CAN	0.00																	
USA	0.00	0.00																
GBR	0.00	0.00	0.16															
CHN	0.31	0.40	0.00	0.00														
HUN	0.00	0.00	0.00	0.46	0.00													
NED	0.00	0.00	0.04	0.00	0.00	0.00												
SWE	0.00	0.00	0.04	0.00	0.00	0.00	0.98											
ESP	0.30	0.00	0.00	0.00	0.20	0.00	0.00	0.01										
AUS	0.02	0.00	0.00	0.49	0.00	0.82	0.00	0.00	0.00									
ITA	0.20	0.48	0.00	0.01	0.11	0.00	0.00	0.00	0.01	0.00								
CZE	0.08	0.41	0.10	0.04	0.20	0.00	0.00	0.00	0.00	0.00	0.59							
UKR	0.11	0.06	0.03	0.01	0.03	0.03	0.30	0.32	0.16	0.05	0.00	0.02						
MEX	0.00	0.93	0.01	0.00	0.36	0.00	0.00	0.00	0.00	0.00	0.46	0.35	0.06					
CUB	0.40	0.00	0.07	0.09	0.06	0.01	0.00	0.00	0.16	0.07	0.10	0.00	0.04	0.00				
NZL	0.00	0.00	0.01	0.00	0.11	0.00	0.00	0.00	0.55	0.01	0.02	0.00	0.15	0.00	0.00			
SUI	0.00	0.00	0.42	0.71	0.00	0.35	0.00	0.00	0.00	0.34	0.00	0.05	0.00	0.01	0.06	0.00		
NOR	0.01	0.00	0.22	0.12	0.00	0.21	0.37	0.37	0.06	0.26	0.00	0.00	0.00	0.00	0.07	0.10	0.08	
RUS	0.00	0.00	0.00	0.67	0.00	0.72	0.00	0.01	0.00	0.78	0.01	0.01	0.04	0.00	0.06	0.00	0.44	0.19

Figure 4.1 shows a dendrogram based on the posterior means of the dissimilarity matrix obtained from a clustering function in R (*hclust*) using the single linkage method. The results obtained from the Bayesian cluster analysis and the dendrogram are fairly similar. The dendrogram also shows that the judges are divided into two large clusters. The scoring behaviour of the judge from Ukraine appears slightly different from the others. However, no probability statement can be obtained using *hclust*.

Figure 4.1: The dendrogram for judges using *hclust*.



## Chapter 5

# Conclusions and Remarks

In this project, we develop a Bayesian model to capture the complexity of judges' scores and explore the relative behaviour of judges from synchronized diving events. We carry out a Bayesian analysis via implementation in WinBUGS. The Bayesian model becomes the keystone of the analysis as it has the ability to impute missing values. The empirical Bayes method allows us to incorporate our knowledge of the parameters based on the dataset. The posterior mean estimates of the correlations obtained from MCMC provide us with the average relative scoring behaviour of judges from synchronized diving events. As expected, more than half of the estimated correlations are negative because an observation falling above the truth is most likely accompanied by an observation falling below the truth. To further investigate the relative behaviour of judges, we carry out a Bayesian cluster analysis on the MCMC output. The results from the Bayesian cluster analysis suggest that two distinct groups of judges exist. The judge from Ukraine has a slightly different scoring behaviour compared to the other judges. However, the results do not prove that the judge from Ukraine is unqualified. This approach only provides an alternative scientific way to assess the judges' scoring. The posterior estimate of the standard deviation of a judge indicates the consistency of judging across the entire event.

In our model described in Chapter 2, we assume that the scores given by judges have a multivariate Normal distribution. This model assumption implies that the scores are continuous variables rather than discrete variables. We believe the reason that the scores are assigned in discrete values is to let the judges choose some convenient values for the scores as the judges have to score each dive immediately after the dive.

For the analysis on the FINA dataset, the data to parameters ratio is about 1 : 3. This

may suggest that the model has insufficient data to provide good estimates for the parameters. However, since we apply a Bayesian approach in the analysis, the prior information can be viewed as a supplement to the insufficient dataset. In addition, the primary parameters of interest in our model is  $\Sigma$  as described in Chapter 3 and replicate information is available for  $\Sigma$ .

Our project is based on the synchronization scores of synchronized diving events in the 12th FINA World Championships in Melbourne 2007. We believe analyses for other international competitions can provide more information on the judges' scoring behaviours. It would be interesting to compare the results we obtained in this project with other results involving the same judge from Ukraine. However, we find that data is not easy to obtain. The judges' results are not usually available directly from the score sheets. Furthermore, it would be interesting to investigate the associations among judges and divers in the competition.

## Appendix A

# The Judges' Nationalities

Key	Nationality
FRA	France
CAN	Canada
USA	The United States
GBR	Great Britain
CHN	China
HUN	Hungary
NED	The Netherlands
SWE	Sweden
ESP	Spain
AUS	Australia
ITA	Italy
CZE	Czech Republic
UKR	Ukraine
MEX	Mexico
CUB	Cuba
NZL	New Zealand
SUI	Switzerland
NOR	Norway
RUS	Russia

## Appendix B

# WinBUGS Code for the Simulated Data

```
#the simulated dataset
model
{
  for(i in 1:N)
  {
    Y[i,1:m]~dmnorm(mu[i,],G[,])
  }

  lower <- 0
  upper <- 10

  for(i in 1:N)
  {
    for(j in 1:m)
    {
      mu[i,j]<-mumu[i]
    }
    mumu[i]~dnorm(gmumu[i],gv)I(lower,upper)
  }

  gv ~dgamma(1,0.5)
```

```

G[1:m,1:m] ~ dwish(R0[,],m)
  for(j in 1:m)
  {
    R0[j,j] <- 1
  }
  for(i in 1:m-1)
  {
    for(j in i+1:m)
    {
      R0[i,j]<- 0.0
      R0[j,i]<- 0.0
    }
  }
}

Sigma[1:m,1:m] <- inverse(G[,])
Corr[m,m] <- Sigma[m,m]
for (i in 1:(m-1))
{
  Corr[i,i] <- Sigma[i,i]
  for (j in (i+1):(m))
  {
    Corr[i,j] <- Sigma[i,j]/(sqrt(Sigma[i,i])*sqrt(Sigma[j,j]))
  }
}
}#end of model

list(N=25, m=3,
gmumu=c(
9.5 , 9.0 , 7.1 , 7.0 , 8.0 , 4.6 , 5.9 , 6.7 , 7.6 , 5.0 ,
6.0 , 5.5 , 7.3 , 6.7 , 4.5 , 8.2 , 6.4 , 9.0 , 7.3 , 5.5 ,
5.8 , 6.2 , 5.9 , 6.6 , 6.2),
Y = structure(.Data=c(
10.0,9.5 ,9.0 ,
9.4 ,9.6 ,8.1 ,
8.0 ,7.2 ,6.1 ,
7.1 ,8.0 ,6.1 ,
8.6 ,6.6 ,8.7 ,

```



```
4.2 ,4.9 ,4.7 ,  
5.5 ,5.4 ,7.0 ,  
7.0 ,7.1 ,6.0 ,  
8.1 ,6.5 ,8.1 ,  
4.7 ,4.7 ,5.7 ,  
6.3 ,5.6 ,6.0 ,  
5.1 ,5.8 ,5.5 ,  
8.0 ,6.7 ,7.2 ,  
6.5 ,7.2 ,6.4 ,  
3.5 ,5.0 ,4.8 ,  
6.5 ,10.0,8.0 ,  
5.8 ,5.4 ,8.1 ,  
9.0 ,8.8 ,9.0 ,  
6.8 ,7.8 ,7.2 ,  
4.8 ,5.9 ,5.8 ,  
4.3 ,6.5 ,6.7 ,  
6.2 ,5.8 ,6.6 ,  
7.9 ,4.8 ,5.1 ,  
7.0 ,6.1 ,6.7 ,  
6.4 ,7.6 ,4.7  
) ,  
.Dim=c(25,3)  
)
```

## Appendix C

# WinBUGS Code for the FINA Data

```
model
{
  for(i in 1:N)
  {
    Y[i,1:m] ~ dnorm(mu[i,], G[,])

    lower <- 0
    upper <- 10

    for(i in 1:N)
    {
      for (j in 1:m)
      {
        mu[i,j] <- mumu[i]
      }
      mumu[i] ~ dnorm(gmumu[i],gv)I(lower, upper)
    }
    gv ~ dgamma(1,0.5)

    G[1:m,1:m] ~ dwish(R0[,],m)
  }
}
```

```
for(j in 1:m)
{
  R0[j,j] <- 1
}

for(i in 1:m-1)
{
  for(j in i+1:m)
  {
    R0[i,j]<- 0.0
    R0[j,i]<- 0.0
  }
}
Sigma[1:m,1:m] <- inverse(G[,])

Corr[m,m] <- sqrt(Sigma[m,m])

for (i in 1:(m-1))
{
  Corr[i,i] <- sqrt(Sigma[i,i])
  for (j in (i+1):(m))
  {
    Corr[i,j] <- Sigma[i,j]/(sqrt(Sigma[i,i])*sqrt(Sigma[j,j]))
  }#end of j-for-loop
}#end of i-for-loop

} #end of model
```

# Bibliography

- [1] Brooks, S.P., and Gelman, A. Alternative methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, 7:434–455, 1998.
- [2] Carlin, B.P. and Louis, T.A. *Bayes and Empirical Bayes Methods for Data Analysis (2nd edition.)*. Chapman and Hall, London, 2000.
- [3] Cohen, J. *Statistical Power Analysis for the Behavioral Sciences (2nd edition)*. Erlbaum, Hillsdale, 1988.
- [4] Congdon, P. *Bayesian Statistical Modelling (2nd edition.)*. John Wiley & Sons, Chichester, 2006.
- [5] Ding, H., Cheng, J., Lu, H. and Zhou, Z. Synchronization analysis for synchronized diving videos. *Proceedings on 2008 IEEE International Conference on Multimedia and Expo.*, pages 897–900, 2008.
- [6] Evans, M. and Swartz, T.B. Methods for approximating integrals in Statistics with special emphasis on Bayesian integration problems. *Statistical Science*, 10:254–272, 1995.
- [7] Gill, P.S., Swartz, T.B, and Treschow, M. A stylometric analysis of King Alfred’s literary works. *Journal of Applied Statistics*, 34:1251–125, 2007.
- [8] Johnson, R.A. and Wichern, D.W. *Applied Multivariate Statistical Analysis (6th edition.)*. Prentice Hall, New Jersey, 2007.
- [9] Lance, G.N. and Williams, W.T. A general theory of classificatory sorting strategies: 1. Hierarchical systems. *The Computer Journal*, 9:373–380, 1966.
- [10] Spiegelhalter, D., Thomas, A. and Best, N. WinBUGS User Manual. *Cambridge: MRC Biostatistics Unit*, 2003.