BAYESIAN ANALYSIS OF DYADIC DATA ARISING IN BASKETBALL

by

Xin Lucy Liu

B.Sc., Shenyang University, 1996

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

SIMON FRASER UNIVERSITY
Spring 2007

APPROVAL

Name: Xin Lucy Liu

Degree: Master of Science

Title of project: Bayesian Analysis of Dyadic Data Arising in Basketball

Examining Committee: Dr. Richard Lockhart

Chair

Dr. Tim Swartz
Senior Supervisor
Simon Fraser University

Dr. Paramjit Gill
Co-Supervisor
Simon Fraser University

Dr. Tom Loughin
External Examiner

Date Approved:

Abstract

The goal of this project is to use statistical methods to identify players and combinations of players which affect a basketball team's performance. The traditional statistics which are recorded tell us only about the contribution of individual players (eg. points scored, rebounds, etc). However, there are subtle aspects of play such as defensive help, setting screens and verbal communication that are known to be important but are not routinely recorded. The model we propose is based on the Bayesian social relations model. The results help us identify aspects of player performance. Data from the NBA 2004 and NBA 2005 finals are used throughout the project to illustrate our approach.

Keywords: Bayesian social relations model, Dyadic data, WinBUGS

Dedication

In memory of my father, Zhong-Xuan Liu (1946-1994)

Acknowledgements

I would like to thank the people who have helped me go through this difficult, challenging yet exciting time of my life. Without their help, this project would not have been successful.

First of all, I can not express the gratitude I have for my supervisor Tim Swartz, who guided me through this project with so much patience. I learned more than statistics from him, I learned how to be kind and willing to help other people.

Secondly, it brings me great pleasure to thank my Co-Supervisor Paramjit Gill, who gave me excellent advice on this project, especially on how to use the WinBUGS software.

I also wish to thank my friend Claire Kirkpatrick with whom I work, who reviewed and edited the first draft of each chapter without a word of complaint. To my friends, Celes Ying, Chunfang Lin and Linda Lei, thank you for sharing your time with me, and standing by my side at all times.

Last but certainly not least, I am grateful to my mother, my husband and my son for all their support and care. I would have been lost without the strength that my mother gives to me. My life would never be fulfilled without my husband and my son.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Basketball history

As taken from Wardrop (1998), "Basketball was invented in 1891 by James Naismith, a physical education instructor at the YMCA Training School in Springfield, Massachusetts, USA. The game achieved almost immediate acceptance and popularity, and the first collegiate game, with five on each side, was played in 1896 in Iowa City, Iowa, USA." Although basketball is very much an "American game", James Naismith was a Canadian.

From its humble beginnings, basketball has gained nearly world-wide acceptance and is played professionally in many countries including the United States, Spain, Italy, Yugoslavia, Israel and Australia. However, the premier league for men's basketball is the National Basketball Association (NBA) where players have yearly salaries as much as $20 million US dollars. The interest in NBA basketball is therefore a big business enterprise.

1.2 Overview of basketball statistics

Considering the popularity of basketball, the amount of statistical research in basketball has been relatively small compared with other major sports such as baseball. Some research in refereed journals has focused on statistical models for shooting. Gilovich et al (1985) considered the modelling of shooting and this was followed up by Tversky and Gilovich

(1989). These papers examine shooting data from three sources: a controlled study of college basketball players, game data of NBA players, and free throw game data of NBA players. The research shows that on most occasions studied, the simple model of Bernoulli trials is adequate to describe the outcomes of a player's shots, but on some occasions the Bernoulli trials model is inadequate.

A key element in analyzing the success or failure of basketball teams is the study of its individual players. There are two aspects of player analysis: (1) the categorization of players by type; and (2) the rating of individual players using computations based on statistical methods.

Ghosh and Steckel (1993) analyzed data with the goal of categorizing players according to player type. The names and characteristics of their resultant clusters are provided below.

- Scorers: Scorers take the most shots, score the most points, and are the best free throw shooters.
- Dishers: Dishers lead in assists and steals, and are almost as good as scorers at shooting free throws. They are the worst at shot blocking and rebounding, and commit the fewest number of fouls.
- Bangers: Bangers lead in offensive and defensive rebounds, and have the second best field goal percentage. They are the worst free throw shooters.
- Inner Court: The inner court has the highest field goal percentage and attempts the most free throws. They are second in scoring, blocks and offensive rebounds, and third in defensive rebounds.
- Walls: The walls lead in blocked shots and fouls committed. They are second in defensive rebounds and third in offensive rebounds. They attempt the fewest field goals and free throws, and are the lowest in steals, assists, and points.
- Fillers: The fillers are a heterogeneous group. They are second highest in steals and fouls committed.

The team with the most points wins the game. But how can one measure an individual player's contribution to his team's success? That is a difficult question. Bellotti (1992), Heeren (1988) and Trupin and Couzens (1989) have contributed to this problem.

Bellotti (1992) and Heeren (1988) provide simple formula. For example, Bellotti (1992) simply adds and subtracts traditional statistics according to whether the statistics are good or bad. Bellotti (1992) uses the proposed formula to obtain the top NBA players of all time, and ranks them on a per minute basis. Trupin and Couzens (1989) assign different weights to the various good and bad statistics.

Perhaps the greatest non-refereed source for quantitative analyses in NBA basketball is available from the website *www.82games.com*. The articles investigate all sorts of questions related to basketball and generally use sound but simple statistical analyses. The site also provides resources and analysis tools for NBA front office executives.

Another interesting source of quantitative analyses in basketball is the book "Basketball on Paper" by Oliver (2004). It is a summary of one man's years of experience in reducing the game of basketball to numbers.

1.3 The goal of this project

The traditional statistics which are recorded for each game (eg. assists, rebounds, points scored, etc) only tell us about the contribution of individual players. However, there are subtle aspects of play (eg. defensive help, setting screens, verbal communication) that are known to affect the game but are not routinely recorded.

We intend to use statistical methods to identify players and combinations of players which affect the game. Fortunately, the response variables (1) points scored for and (2) points scored against are the obvious choices to study. We consider a Bayesian analysis of dyadic data using WinBUGS software where the data are taken from the 2004 and 2005 NBA finals. In Chapter 2, we introduce the software package WinBUGS. In Chapter 3, we review the traditional social relations model, and the Bayesian social relations model. In Chapter 4 we modify the BSRM to suit our NBA data sets and discuss the inclusion of covariates. We provide some concluding remarks in Chapter 5.

Chapter 2

Bayesian Analysis using WinBUGS

WinBUGS is a program for Bayesian model fitting that runs under Windows and can be freely downloaded from $www.mrc - bsu.cam.ac.uk/bugs$. The program is fully described on that website. This program's help facility also includes a large number of examples consisting of data sets and associated WinBUGS programs. As a stand alone program WinBUGS is usually run interactively through a series of menus and toolbars. However the current version, WinBUGS 1.4.1, includes a scripting language so that model fitting can be automated and controlled by a script file.

2.1 Bayesian model fitting using Markov chain methods

The scale of literature on Bayesian analysis is such that it is impossible to give a comprehensive review, but a brief account and references should be sufficient to enable anyone to follow their way through the WinBUGS program.

The Bayesian approach to statistics is fundamentally different from the classical approach. In the classical approach, the parameter, θ, is thought to be an unknown but fixed quantity. Data X are drawn from a population indexed by θ, and based on the observed value X = x, knowledge about the value of θ is obtained. In the Bayesian approach, θ is considered to be a random quantity whose variation can be described by a probability distribution (called the *prior distribution*). This is a subjective distribution, based on the experimenter's belief, and is formulated before the data are seen. Data are then taken from a population indexed by θ and the prior distribution is updated with this data information. The updated prior distribution is called the *posterior distribution*.

If we denote the prior distribution by $\pi(\theta)$ and the density by $f(x|\theta)$, then the posterior distribution, the conditional distribution of θ given x, is

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

where f(x) is the marginal distribution of X, that is

$$f(x) = \int f(x|\theta)\pi(\theta)d\theta.$$

The marginal distribution of X often presents a problem in that it may be very difficult to calculate; typically this requires either a very large summation or a multidimensional integral. For many years the difficulty of evaluation of the marginal distribution effectively restricted Bayesian analyses to simple problems in which the integration was tractable. Then, in the later 1980's, Bayesian statisticians started to implement Monte Carlo integration, that is integration by simulation. Suppose then that you are able to generate a sample θ_1, θ_2, ..., θ_n from the posterior distribution. Then the integral

$$I(m) = \int m(\theta)\pi(\theta|x)d\theta$$

can be approximated by

$$\hat{I}(m) = \frac{\sum_{i=1}^{n} m(\theta_i)f(x|\theta_i)\pi(\theta_i)}{\sum_{i=1}^{n} f(x|\theta_i)\pi(\theta_i)}$$

where the function $m = m(\theta)$ provides a posterior quantity of interest. For example, $m = m(\theta) = \theta$ gives $I(m)$ equal to the posterior mean.

Now it is rarely a simple matter to generate a sample θ_1, θ_2, ..., θ_n directly from the posterior. Instead, it may be easier to construct a Markov chain θ_1, θ_2, ..., θ_n whose equilibrium distribution is the posterior. This procedure is known as Markov chain Monte Carlo (MCMC) and WinBUGS is a software program that constructs Markov chains.

2.2 Specifying models in WinBUGS

WinBUGS has its own language for specifying models and this is described in detail in the WinBUGS manual (*http://mathstat.helsinki.fi/openbugs/*), although many people find it easier to learn the language by following the accompanying WinBUGS examples. We will consider a simple model that illustrates how WinBUGS is run. Suppose that we have a random variable y that is the number of successes in n independent trials with success probability θ, the candidate model is,

$$y \sim Binomial(n, \theta)$$

For a Bayesian analysis we must state our prior belief for the parameter θ. For example, we may consider prior distribution

$$\theta \sim Uniform(0, 1)$$

We could describe this model in WinBUGS using the more or less self-explanatory code

```
model {
   y~ dbin(theta, n)
theta ~ dunif(0, 1)
   }
```

The WinBUGS language is closely modelled on S-Plus and R in which the combined symbol $<-$ is used to denote assignment. The symbol $\sim$ denotes that the variable to the left has a probability distribution given by the distribution on the right. Notice that the order of the binomial parameters in WinBUGS is reversed.

WinBUGS requires two other pieces of information before it can fit the model: the data and some initial values for the MCMC chain. Once again WinBUGS adopts the R style for data structures and so data are usually given as lists. For our example the data might be supplied as

```
list( y = 5, n = 12)
```

and the initial values could be

```
list( theta = 0).
```

2.3 Running WinBUGS

WinBUGS uses compound files with the .odc extension. A compound file contains various types of information (formatted text, tables, formulae, plots, graphs, etc) displayed in a single window and stored in a single file. The tools needed to create and manipulate these information types are available, so there is no need to continuously move between different programs. WinBUGS has been designed so that it produces output directly to a compound file and can get its input directly from a compound file. Let us assume that our model, data and initial values are stored in text files called **Mybugwithouti.txt**, **data2004.txt**, and **inits2004.txt**, and that these files are stored in the folder **c:/lucyproject**. To run the model all we need is another text file containing the script.

A basic script for our problem might be stored as **script.txt** and consist of

```
display(‘log’)
check(‘c:/lucyproject/Mybugwithouti.txt’)
data(‘c:/lucyproject/data2004.txt’)
compile(1)
inits(1, ‘c:/lucyproject/inits2004.txt’)
update(10000)
set(‘alpha’)
set(‘beta’)
update(100000)
coda(*.‘:/lucyproject/out’)
quit()
```

Line 1 opens a WinBUGS log window in which we will be able to follow the progress of our script and possibly see any error messages. Line 2 asks WinBUGS to check that our model description is syntactically correct; it is at this stage that we pick up the typing errors, missed brackets, etc. Line 3 reads the data and then line 4 compiles our model into a program for analyzing our problem. The ’1’ in the compile statement tells WinBUGS that we only intend running one chain. However running several parallel chains is a good way of discovering whether the chains have converged. Line 5 reads the initial values for chain 1 and then on line 6 a MCMC chain of length 10000 is created. This initial chain is called the *burn-in* and will be discarded by WinBUGS because we have not told it to store any results. The *burn-in* is intended to allow the chain to stabilize and so remove the effects of the initial values. Deciding on the appropriate length of the *burn-in* is an art in itself. The two **set** commands in line 7 and line 8 tell WinBUGS to store α and β in subsequent simulations. Then in line 9 a further chain of length 100000 is run. The stored values of α and β are written to an output file in **coda** format. **Coda** is a program written in S-plus that was designed for examining MCMC output. In particular, **Coda** can be used to help assess whether the chain has converged. Finally the **quit** command closes WinBUGS. If the **quit** command is omitted then WinBUGS stays active after the script is executed. This can be helpful when you need to debug a program.

To run **script.txt**, we could open **script.txt** in a running WinBUGS, click Model $->$ Script on the main menu.

Chapter 3

Modelling

3.1 The social relations model

The social relations model (SRM) was developed by Larry La Voie and David A Kenny and named after the interdisciplinary social science department at Harvard University that no longer exists. The model describes dyadic relationships when variables are measured on a continuous scale. Data from two-person interactions and rating or sociometric studies can be used. The level of measurement should be interval (eg. seven-point scales) and not categorical (e.g. Yes/No). Generally the data are collected from people but the dyadic units can be animals, groups, organizations, cities, countries, etc.

The SRM is a special case of generalization theory (Cronbach *et al.*, 1972) with the basic model being a two-way random-effects analysis of variance with three major types of effects: actor, partner, and relationship effects. The actor effect represents a person's average level of a given behaviour in the presence of a variety of partners. The partner effect represents the average level of a response which a person elicits from a variety of partners. The relationship effect represents a person's behaviour toward another individual in particular, above and beyond their actor and partner effects. For example, in the study of interpersonal attraction, the actor effect is how much a person likes others. The partner effect is how much the person is liked in return. Relationship effects are directional or asymmetric. To differentiate relationship effects from error variance, multiple indicators of the construct, either across time or with different measures, are necessary.

The focus in the SRM is not on estimating the effects for specific persons and relationships but in estimating the variance due to effects. So, in a study of how intelligent people see each other, interest is in whether there are actor, partner, and relationship variances. Actor variance would assess if people saw others as similar in terms of intelligence, partner variance would assess whether people agree with each other in their ratings of intelligence, and relationship variance would assess the degree to which perceptions of intelligence are unique. The mathematics of the SRM can be found in the book "Interpersonal Perception: A Social Relations Analysis" (*http://davidakenny.net/ip/interbok.htm*) by David A. Kenny (1994).

There has been extensive research using the SRM. Consult the "Interpersonal Perception: A Social Relations Analysis" page to see the type of questions that can be answered using the model. Any response that is dyadic can be studied using the model.

The most common social relations design is the round-robin research design. In this design, each person interacts with or rates every other person in the group, and data are collected from both members of each dyad. The scores of the two people are usually different. Usually there are multiple round robins.

We restrict our attention to the situation where the response is measured on a continuous scale, at least approximately. For continuous data, the model expresses the paired responses in an additive fashion

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk}$$

$$y_{jik} = \mu + \alpha_j + \beta_i + \gamma_{ji} + \epsilon_{jik} \qquad (3.1.1)$$

where μ is the overall mean, α_i is the effect of subject i as an actor, β_j is the effect of subject j as a partner, γ_{ij} is an interaction effect representing the special adjustment which subject i makes for subject j as a relationship effect and ϵ_{ijk} represents the error term which picks up measurement error and/or variability in behaviour on different occasions. The expected responses $E(y_{ijk})$ and $E(y_{jik})$ differ as the actor and partner have different parameters. We refer to μ, the α's, the β's and the γ's as first-order parameters With i =

1, ..., m, j = 1, ..., m and k = 1, ..., n_{ij}, there are m^2+m+1 such parameters and $\sum_{i \neq j} n_{ij}$ observations. Thus, even in the simple structure (3.1.1) where relatively few observations are available to identify parameters, a Bayesian approach suggests itself.

The SRM goes on to assume that the overall mean μ is fixed but that the other terms in (3.1.1) are random. Specifically, it is assumed that

$$E(\alpha_i) = E(\beta_j) = E(\gamma_{ij}) = E(\epsilon_{ijk}) = 0$$

$$var(\alpha_i) = \sigma^2_\alpha,\ var(\beta_j) = \sigma^2_\beta,\ var(\gamma_{ij}) = \sigma^2_\gamma,\ var(\epsilon_{ijk}) = \sigma^2_\epsilon$$

$$corr(\alpha_i,\ \beta_i) = \rho_{\alpha\beta},\ corr(\gamma_{ij},\ \gamma_{ji}) = \rho_{\gamma\gamma},\ corr(\epsilon_{ijk}, \epsilon_{jik}) = \rho_{\epsilon\epsilon} \tag{3.1.2}$$

and all other covariances are zero. The parameters $\{\sigma^2_\alpha,\ \sigma^2_\beta,\ \sigma^2_\gamma,\ \sigma^2_\epsilon,\ \rho_{\alpha\beta},\ \rho_{\gamma\gamma},\ \rho_{\epsilon\epsilon}\}$ are called the variance-covariance parameters (or components). As the subjects are a sample from a population, the variance-covariance population parameters are of primary interest. These parameters model the variability and co-variability of social/psychological phenomena in a population of human subjects.

The interpretation of the variance-covariance parameters is naturally problem specific. However, for the sake of illustration, suppose that the response y_{ijk} is the measurement of how much subject i likes subject j based on their kth meeting. In this case, $\rho_{\alpha\beta}$ represents the correlation between α_i and β_i, i = 1,..., m, and we would typically expect a positive value. That is, an individual's positive (negative) attitude towards others is usually reciprocated. The interpretation of $\rho_{\gamma\gamma}$ is typically more subtle. In this example, a positive value of $\rho_{\gamma\gamma}$ may be interpreted as the existence of a special kind of "sympatico" when two individuals hit it off and vice-versa. In the social psychology literature, this is referred to as dyadic reciprocity (Kenny 1994).

3.2 The Bayesian social relations model

When developing the Baysian analogue BSRM of the SRM, Gill and Swartz (2001) maintain the first and second moment assumptions as in (3.1.2), but go further by assigning distributional forms to the parameters. Specifically, let $\mu_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij}$ and assume conditionally

$$\begin{pmatrix} \alpha_i \\ \beta_i \end{pmatrix} \sim Normal_2 \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \sum\nolimits_{\alpha\beta} \right]$$

$$\begin{pmatrix} \gamma_{ij} \\ \gamma_{ji} \end{pmatrix} \sim Normal_2 \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \sum\nolimits_{\gamma} \right]$$

$$\begin{pmatrix} y_{ijk} \\ y_{jik} \end{pmatrix} \sim Normal_2 \left[\begin{pmatrix} \mu_{ij} \\ \mu_{ij} \end{pmatrix}, \quad \sum\nolimits_{\epsilon} \right]$$

where k = 1, ..., n_{ij}, 1 $\leq$ i $\neq$ j $\leq$ m and

$$\sum\nolimits_{\alpha\beta} = \begin{pmatrix} \sigma_\alpha^2 & \rho_{\alpha\beta}\sigma_\alpha\sigma_\beta \\ \rho_{\alpha\beta}\sigma_\alpha\sigma_\beta & \sigma_\beta^2 \end{pmatrix}$$

$$\sum\nolimits_{\gamma} = \sigma_\gamma^2 \begin{pmatrix} 1 & \rho_{\gamma\gamma} \\ \rho_{\gamma\gamma} & 1 \end{pmatrix}$$

$$\sum\nolimits_{\epsilon} = \sigma_\epsilon^2 \begin{pmatrix} 1 & \rho_{\epsilon\epsilon} \\ \rho_{\epsilon\epsilon} & 1 \end{pmatrix} .$$

Up to this point, the Bayesian model has not introduced any new parameters. However, it is sensible to express our uncertainty in the variance-covariance parameters and to also regard μ as random. Therefore, following conventional Bayesian protocol for linear models (Gelfand, Hills, Racine-Poon and Smith 1990).

$$\mu \sim Normal \left[\theta_\mu, \quad \sigma_\mu^2\right], \qquad \theta_\mu \sim Normal \left[\theta_0, \quad \sigma_{\theta_0}^2\right],$$

$$\sigma_\mu^{-2} \sim Gamma \left[a_0, \quad b_0\right], \qquad \Sigma_{\alpha\beta}^{-1} \sim Wishart_2 \left[(v_0 R)^{-1}, \quad v_0\right] \tag{3.2.1}$$

where X $\sim$ $Gamma \left[a, \quad b\right]$ implies E(X) = a/b.

The parameters subscripted with a 0 in (3.2.1) are referred to as hyper-parameters and are often set to give diffuse prior distributions for the parameters θ_μ, σ_μ^2 and $\sum_{\alpha\beta}$. Diffuse distributions are useful when a user does not have strong prior opinions regarding parameters. The choices are robust in the sense that inferences do not change dramatically when the hyperparameters are perturbed. For more information on the setting of the hyperparameters, see Gill and Swartz (2004)

3.3 The Bayesian social relations model in basketball

Although there is great interest in basketball performance, there are no analyses based on statistical models which investigate the performance of players and combinations of players. In the previous sections of this chapter, we discussed the SRM and the BSRM. Based on the BSRM developed by Gill and Swartz (2004), we propose the following model for NBA finals data which will help us look closer at the performance of players and combinations of players. By extensive viewing of videotape, we collected scoring data and lineup data during different time intervals. During each time interval, there is one stable lineup and scoring results are recorded. Note that scoring records were not recorded during the last minute of matches where excessive fouling often occurs.

For a given line in our data set corresponding to t seconds of play, let y_1 be the number of points scored by team 1 and let y_2 be the number of points scored by team 2. We assume

$$y_1 \sim \text{Poisson}(\text{t}\theta_1)$$

where

$$log\theta_1 = \alpha_1^{(1)} + ... + \alpha_5^{(1)} + \beta_1^{(2)} + ... + \beta_5^{(2)} + \tau^{(1)} \tag{3.3.1}$$

and

$$y_2 \sim \text{Poisson}(\text{t}\theta_2)$$

where

$$log\theta_2 = \alpha_1^{(2)} + ... + \alpha_5^{(2)} + \beta_1^{(1)} + ... + \beta_5^{(1)} + \tau^{(2)}. \tag{3.3.2}$$

Our model assumes that the number of points scored by team 1 and team 2 are Poisson with the expected number of points proportional to the length of the time interval during which the lineups for both teams are constant. Although in basketball, points can be scored in increments of 1, 2 or 3 points, the Poisson assumption may be reasonable as the time intervals are never too short. We might think of a time interval as consisting of a large number of possessions where each possession has the potential of having points scored. The total number of points over all of the possessions in a time interval may therefore be approximately normal due to the Central Limit Theorem. We prefer the Poisson distribution due to its discreteness and skewness. In a more comprehensive analysis, we would like to use statistical methods to check the propriety of the Poisson assumption.

In (3.3.1) and (3.3.2), we express the scoring rates θ_1 and θ_2 in a log-linear fashion. The log term is introduced to ensure that scoring rates are non-negative. With respect to θ_1, (3.3.1) stipulates that scoring for team 1 depends on the five players on team 1 ($\alpha_1^{(1)}$, ..., $\alpha_5^{(1)}$) and the five players on team 2 ($\beta_1^{(2)}$, ..., $\beta_5^{(2)}$) who are on the court during the particular time interval. The α's therefore represent offensive ability with large values corresponding to good offensive players. Similarly the β's represent defensive ability with large values corresponding to bad defensive players. Although only five α's and five β's show up in the model for each line of our data set, we actually have twelve α's and twelve β's corresponding to a team's roster. For WinBUGS programming, we use subscripts one to twelve for team 1 players and thirteen to twenty-four for team 2 players.

There is another widely acknowledged element that affects scoring. It is the home court advantage. We add the parameters $\tau^{(1)}$ and $\tau^{(2)}$ to our model structure. The parameter $\tau^{(1)}$ is the home court advantage for team 1 where $\tau^{(1)} = \tau = log(4.0/48)$ when the game is played in team 1's city, otherwise it is set to 0. The parameter $\tau^{(2)}$ is the home court advantage for team 2 where $\tau^{(2)} = \tau = log(4.0/48)$ when the game is played in team 2's city, otherwise it is set equal to 0. The parameter τ is set as a constant since it is well known that the NBA home court advantage is worth roughly 4.0 points per game.

We consider the parameters to have the same prior distributions as the parameters in the BSRM of Gill and Swartz (2007)

$$\begin{pmatrix} \alpha_i^k \\ \beta_i^k \end{pmatrix} \sim Normal_2 \left[\begin{pmatrix} 0.05 \\ 0.05 \end{pmatrix}, \quad \textstyle\sum_{\alpha\beta} \right]$$

where i = 1, ..., 24, and k = 1,2 with

$$\textstyle\sum_{\alpha\beta} = \begin{pmatrix} \sigma_\alpha^2 & \rho_{\alpha\beta}\sigma_\alpha\sigma_\beta \\ \rho_{\alpha\beta}\sigma_\alpha\sigma_\beta & \sigma_\beta^2 \end{pmatrix} \qquad \textstyle\sum_{\alpha\beta}^{-1} \sim Wishart_2 \left[(v_0)^{-1}I, \quad v_0 \right].$$

We also looked at some more complex models that involved interaction terms between players. First, we considered adding all the interaction terms in the model, but this lead to convergence problems. This was because we had too many interaction terms and not enough data. As a result, our model became unidentifiable. Let us explain it another way. In regression analysis, there is a full rank assumption for the design matrix. In order for $(X'X)$ to be invertible ($(X'X)^{-1}$ exists), the X matrix needs to have full (column) rank. This means that all of the columns of X must be linearly independent. We have a similar problem with our model. Our log-linear relationship (either (3.3.1) or (3.3.2)) can be expressed using a design matrix X. Each column in the matrix represents a parameter in our model, each row in the matrix represents a stable lineup in our data set. When we include all of the interaction terms in our model we add $\binom{24}{2} = 276$ more parameters, but we have just 129 rows for the NBA 2004 finals and 204 rows for the NBA 2005 finals. In these cases, the column rank exceeds the number of rows leading to an unidentifiable nodel with convergence problems.

We therefore tried to fit the model with some interaction terms that we found interesting but still maintaining an identifiable model. For the NBA 2004 finals, we chose Shaquille ONeal and Kobe Bryant as our interesting players. Because Shaquille ONeal and Kobe Bryant have extreme styles and each demand the ball, they are controversial. Kobe Bryant tends to shoot too much and not pass, Shaquille ONeal is dominant because of his size. We chose 19 interaction terms which were combinations of Shaquille ONeal and Kobe Bryant with other players. For the NBA 2005 finals, we chose Bruce Bowen and Tayshaun Prince as our interesting players since they are known as defensive specialists. We chose 17 interaction terms which were combinations of Bruce Bowen and Tayshaun Prince with other players. In both cases, we did not have convergence problems, but all the interaction term effects were small compared to the main effects. We therefore decided to stick to our original model.

Chapter 4

Analysis of Basketball Data

We collected basketball data from the NBA 2004 finals and the NBA 2005 finals. We used WinBUGS as the software package for implementing MCMC. With WinBUGS' built-in graphical and analytical capabilities to check convergence, it is easy for us to find a satisfactory burn-in period. We simulated 100,000 samples for parameter estimation after the burn-in period. We focus on the estimation of the α_i, the β_i, the α_i - β_i and τ in this section. Figure 4.1 is the convergence trace plots of α_1 to α_6 for the NBA 2004 data set. All of the other parameters had similar convergence trace plots as Figure 4.1. Figure 4.1 suggests that a burn-in period of 10,000 iterations is more than adequate.

4.1 NBA 2004 finals

In the NBA 2004 finals, the two teams were the Los Angeles Lakers and the Detroit Pistons. The season ended with the Detroit Pistons upsetting the heavily favoured Los Angeles Lakers four games to one.

We first inspect the data using a traditional method. In Table 4.1 we list the 24 players in the NBA 2004 finals together with their minutes and their plus/minus statistics. The number "PlusMinus" in Table 4.1 is the total points that the team gained when the corresponding player was on the court. The number "PM/Game" is the "PlusMinus" statistic extended to a full game (ie. 48 minutes). Higher values of PM/Game suggests that a player

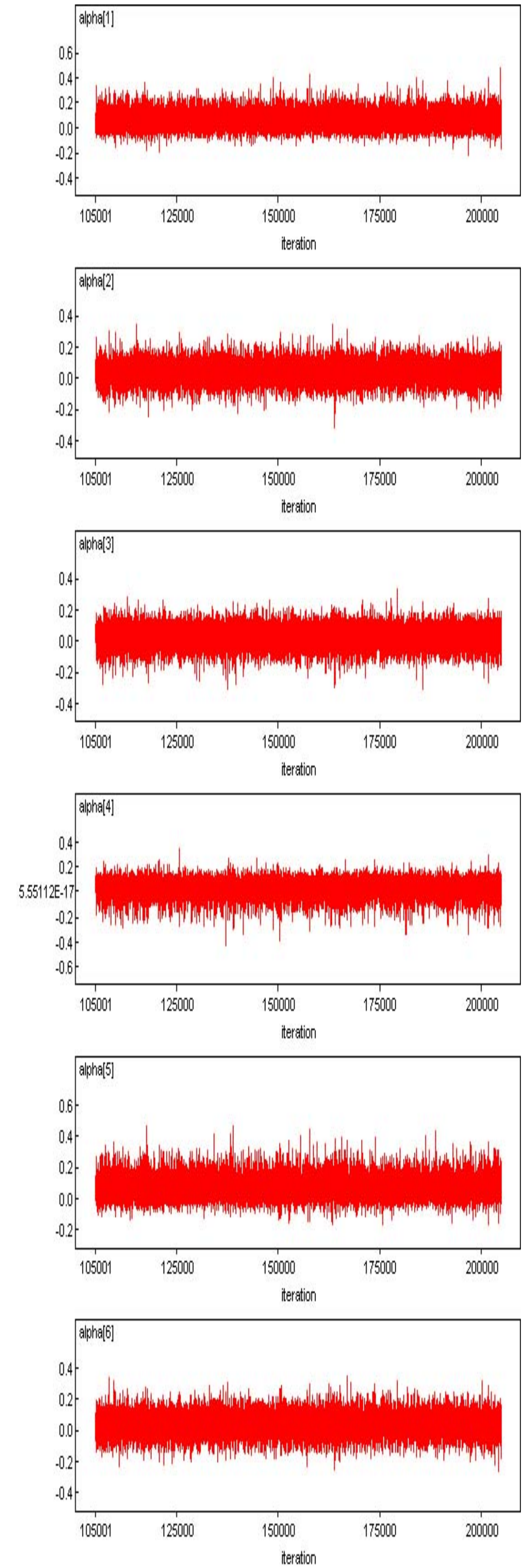

Figure 4.1: The convergence trace plots of α_1 to α_6 for the NBA 2004 finals data.

makes a greater contribution to winning than does his teammates. From Table 4.1, we see that Luke Walton, Derek Fisher, Shaquille ONeal and Karl Malone's performances were the best for LA and Ben Wallace, Tayshaun Prince, Chauncey Billups and Elden Campbell's performances were the best for Detroit. When fitting our model or looking at the plus/minus statistics, it might be reasonable to not put too much faith on values for players who had very little court time. We see that overall the numbers for Detroit were higher than LA, this is because Detroit won the series, and Detroit scored more than LA.

Table 4.1: NBA 2004 finals plus/minus player summaries

Player	Minutes	PlusMinus	PM/Game
Shaquille ONeal	213.2	-31.0	-7.0
Karl Malone	124.7	-20.0	-7.7
Devean George	106.4	-22.0	-9.9
Kobe Brayant	231.1	-52.0	-10.8
Gary Payton	166.8	-48.0	-13.8
Slava Medvendenko	70.9	-21.0	-14.2
Derek Fisher	100.4	-4.0	-1.9
Luke Walton	78.8	6.0	3.7
Kareen Rush	75.9	-11.0	-7.0
Rick Fox	28.8	-11.0	-18.3
Brian Cook	21.2	-1.0	-2.3
Bryon Russell	6.8	-15.0	-105.9
Ben Wallace	205.1	64.0	15.0
Rasheed Wallace	152.5	30.0	9.4
Tayshaun Prince	198.3	49.0	11.9
Richard Hamilton	223.8	41.0	8.8
Chauncey Billups	193.7	45.0	11.2
Elden Cambell	68.4	14.0	9.8
Lindsey Hunter	64.3	3.0	2.2
Corliss Williamson	52.8	-4.0	-3.6
Mike James	17.4	2.0	5.5
Darvin Ham	7.2	-2.0	-13.3
Mehmut Okur	37.5	-6.0	-7.7
Darko Milicec	3.9	-6.0	-73.2

Let us now look at the estimates from our model.The estimation of the parameter α, β and α - β and their standard deviations (SD) is given in Table 4.2. In Table 4.2, α is the offensive rating and β is the defensive rating for players, The quantity α - β is the overall rating for a player. For α - β, the higher the value is, the better the player performance. From Table 4.2, we can see that Shaquille ONeal and Gary Payton were the best offensively

Table 4.2: Posterior means and standard deviations for the NBA 2004 finals data.

	α		β		α - β	
Player	Mean	SD	Mean	SD	Mean	SD
Shaquille ONeal	0.067	0.048	0.038	0.055	0.029	0.077
Karl Malone	0.047	0.043	0.046	0.049	0.001	0.066
Devean George	0.038	0.044	0.038	0.052	0.001	0.067
Kobe Bryant	0.029	0.051	0.059	0.053	-0.030	0.070
Gary Payton	0.070	0.047	0.115	0.066	-0.045	0.080
Slava Medvendenko	0.048	0.045	0.060	0.048	-0.013	0.065
Derek Fisher	0.056	0.046	0.062	0.052	-0.006	0.066
Luke Walton	0.063	0.044	0.047	0.050	0.015	0.067
Kareem Rush	0.040	0.045	0.048	0.049	-0.008	0.066
Rick Fox	0.042	0.047	0.066	0.052	-0.023	0.071
Brian Cook	0.047	0.049	0.071	0.054	-0.024	0.075
Bryon Russell	0.044	0.050	0.088	0.063	-0.044	0.082
Ben Wallace	0.086	0.048	0.024	0.055	0.063	0.076
Rasheed Wallace	0.051	0.044	0.045	0.049	0.007	0.065
Tayshaun Prince	0.072	0.045	0.061	0.050	0.011	0.066
Richard Hamilton	0.078	0.047	0.078	0.058	-0.001	0.073
Chauncey Billups	0.051	0.050	0.042	0.052	0.009	0.067
Elden Campbell	0.053	0.043	0.020	0.055	0.033	0.070
Lindsey Hunter	0.038	0.048	0.027	0.055	0.011	0.071
Corliss Williamson	0.070	0.046	0.058	0.053	0.012	0.071
Mike James	0.072	0.047	0.057	0.055	0.015	0.072
Darvin Ham	0.060	0.048	0.051	0.057	0.009	0.075
Mehmut Okur	0.047	0.047	0.078	0.056	-0.031	0.074
Darko Milicec	0.050	0.053	0.051	0.059	-0.001	0.081

for LA, and Ben Wallace, Tayshaun Prince and Richard Hamilton were the best offensively for Detriot. Thes results are noteworthy because both Gary Payton and Ben Wallace are not high scorers, yet their contribution to team offence is considerable. This highlights a benefit of the methodology since traditional statistics do not readily show the contribution that Payton and Wallace make offensively. Shaquille ONeal and Devean George were the best defensively for LA, and Ben Wallace was the best defensively for Detroit. From the α - β values, we can see that Shaquille ONeal was distinguished in LA and Ben Wallace was by far the best overall contributor for Detroit. There are a few players that we like to address a little more. Gary Payton was very good at offence - he was the best offensive player in LA, but he was the worst defensive player in the NBA 2004 finals. That is why his overall rating in the NBA 2004 finals was not good. This is interesting as Payton had earned the nickname "the Glove" for his supposed defensive prowess in his early years in the NBA. Tayshaun

Prince is famous for his defensive skill, but in the NBA 2004 finals, his defensive skills were not so obvious. For Shaquille ONeal, when we inspected his "PlusMinus" number, he was not the best player in LA, but in our model his α - β indicated that he was the best player in LA. This is probably because he frequently played with some bad players or bad combination of players. This was confirmed by looking at the data more closely. It shows that LA had great trust in ONeal. It again demonstrates that our methodology can pick out patterns that are not immediate from using traditional statistics. We mention that we ignored the estimates for Kareem Rush, Rick Fox, Brian Cook, Bryon Russell, Corliss Williamson, Mike James, Darvin Ham, Mehmut Okur and Darko Milicec since they played so much less compared to the other players. Parameter estimation for these players may not be reliable.

4.2 NBA 2005 finals

In the NBA 2005 finals, the two teams were the San Antonio Spurs and the Detroit Pistons. The season ended with the San Antonio Spurs defeating the defending champion Detroit Pistons four games to three. Table 4.3 lists the 24 players in the NBA 2005 finals together with their minutes and their plus/minus statistics. We see that Manu Ginobili had the highest plus/minus statistics for San Antonio and Rasheed Wallace had the highest plus/minus statistics for Detroit.

Table 4.4 shows the estimates of the parameters α, β and α - β and their standard deviations (SD) for the NBA 2005 finals. There were not too many surprises when we compared our model's results and the "PlusMinus" results. The highest overall ratings for San Antonio belonged decisively to Manu Ginobili. The highest overall ratings for Detroit belonged decisively to Rasheed Wallace. Manu Ginobili and Rasheed Wallace have similar profiles. They were the best defensive players on each team (ie lowest value of β). Robert Horry was the best offensive player on San Antonio and Richard Hamiltion was the best offensive player on Detroit. Compared to the NBA 2004 finals, Ben Wallace's defense was much worse.

Table 4.3: NBA 2005 finals plus/minus player summaries

Player	Minutes	PlusMinus	PM/Game
Tim Duncan	285.2	-8.0	-1.3
Tony Parker	264.8	-3.0	-0.5
Bruce Bowen	270.1	-14.0	-2.5
Manu Ginobill	251.4	30.0	5.7
Nazr Mohammed	159.2	-12.0	-3.6
Robert Horry	202.3	12.0	2.8
Brent Barry	143.4	-24.0	-8.0
Devin Brown	34.8	-14.0	-19.3
Tony Massenburg	8.6	-8.0	-44.7
Rasho Nesterovic	25.2	-7.0	-13.3
Beno Udrih	46.2	-20.0	-20.8
Glenn Robinson	13.9	3.0	10.4
Ben Wallace	281.4	1.0	0.2
Rasheed Wallace	225.6	41.0	8.7
Tayshaun Prince	275.5	9.0	1.6
Richard Hamilton	294.5	16.0	2.6
Chauncey Billups	279.8	11.0	1.9
Elden Campbell	138.5	-6.0	-2.1
Lindsey Hunter	155.9	-20.0	-6.2
Antonio McDyess	8.7	4.0	22.2
Ronald Dupree	29.6	5.0	8.1
Darvin Ham	1.1	-1.0	-42.4
Carlos Arroyo	7.7	0.0	0.0
Darko Milices	6.8	5.0	35.6

Table 4.4: Posterior means and standard deviation for the NBA 2005 finals data

	α		β		α - β	
Player	Mean	SD	Mean	SD	Mean	SD
Tim Duncan	0.052	0.043	0.060	0.045	-0.008	0.063
Tony Parker	0.057	0.046	0.068	0.045	-0.011	0.065
Bruce Bowen	0.069	0.043	0.083	0.045	-0.015	0.062
Manu Ginobili	0.069	0.041	0.015	0.051	0.054	0.067
Nazr Mohammed	0.036	0.044	0.040	0.045	-0.004	0.064
Robert Horry	0.089	0.044	0.076	0.045	0.012	0.062
Brent Barry	0.041	0.043	0.069	0.043	-0.028	0.061
Devin Brown	0.046	0.047	0.062	0.050	-0.016	0.069
Tony Massenburg	0.045	0.053	0.054	0.055	-0.009	0.077
Rasho Nesterovic	0.052	0.048	0.054	0.053	-0.003	0.071
Beno Udrih	0.037	0.052	0.066	0.050	-0.029	0.072
Glenn Robinson	0.050	0.051	0.033	0.059	0.017	0.079
Ben Wallace	0.068	0.042	0.068	0.046	-0.001	0.063
Rasheed Wallace	0.059	0.041	0.010	0.055	0.049	0.069
Tayshaun Prince	0.044	0.044	0.037	0.047	0.006	0.064
Richard Hamilton	0.077	0.047	0.063	0.047	0.015	0.068
Chauncey Billups	0.047	0.043	0.064	0.046	-0.017	0.062
Elden Campbell	0.061	0.049	0.072	0.057	-0.011	0.074
Lindsey Hunter	0.055	0.041	0.055	0.045	-0.001	0.061
Antonio McDyess	0.046	0.042	0.066	0.044	-0.021	0.061
Ronald Dupree	0.057	0.050	0.044	0.059	0.013	0.077
Darvin Ham	0.058	0.050	0.046	0.058	0.012	0.076
Carlos Arroyo	0.069	0.047	0.043	0.056	0.025	0.073
Darko Milicec	0.056	0.051	0.047	0.058	0.009	0.076

Chapter 5

Concluding Remarks

This project provides a preliminary study of the use of a variation of the Bayesian social relations model to investigate player performance in basketball.

One of the practical difficulties in the approach is that data are not readily available. The author spent close to 240 hours transcribing 12 games of videotape into the required data format. In practice, we would want even more data to make reliable player evaluations.

In future work, we would also like to investigate model selection to determine which are the best covariates, whether a Poisson distribution is best, etc.

However, we believe that this project is the first attempt at complex statistical modelling in basketball. The response variables are clearly sensible and the model attempts to recognize player contributions in a team setting. Contributions beyond the traditional statistics may be realized and a player's worth can be assessed from both an offensive and defensive perspective.

Appendix A

WinBUGS code for the NBA 2004 finals

This appendix provides WinBUGS code for the NBA 2004 finals. WinBUGS code for the NBA 2005 finals is the same except for a different data file.

♯ Some priors used here are different than in Gill and Swartz (2001) but can be easily changed

♯ m=number of subjects

♯ Data are read as subject1 subject2 subject3 ... subject10 score1 score2 from a rectagular data file

```
model {
for (i in 1:nobs) {     y[i, 1] <- score1[i]
    y[i, 2] <- score2[i]
    y[i,1] ~ dpois(Ey[i,1])
    y[i,2] ~ dpois(Ey[i,2])
    log(Ey[i,1]) <- alpha[s1[i]] + alpha[s2[i]]+ alpha[s3[i]]+ alpha[s4[i]]+ alpha[s5[i]] +
beta[s6[i]] + beta[s7[i]]+ beta[s8[i]]+ beta[s9[i]]+ beta[s10[i]]+ log(time[i])    + t1[i]log(4.0/48)
+ t2[i]log(4.0/48)
    log(Ey[i,2]) <- alpha[s6[i]] + alpha[s7[i]]+ alpha[s8[i]]+ alpha[s9[i]]+ alpha[s10[i]] +
beta[s1[i]] + beta[s2[i]]+ beta[s3[i]]+ beta[s4[i]]+ beta[s5[i]]+ log(time[i])    + t1[i]log(4.0/48)
+ t2[i]log(4.0/48)
}
♯ Prior for alphas and betas
```

```
for (i in 1:m){ alpha[i] < − aa[i,1]; beta[i] < − aa[i,2]
    aa[i,1:2] ∼ dmnorm(z1[1:2],S2[,])}
for (i in m+1:2*m){ alpha[i] < − bb[i,1]; beta[i] < − bb[i,2]
    bb[i,1:2] ∼ dmnorm(z2[1:2], S2[,])}
for (i in 1:2*m) { q[i] < − alpha[i] -beta[i]}
z1[1] < − 0.05; z1[2] < − 0.05
z2[1] < − 0.06; z2[2] < − 0.06
♯ S2 is precision matrix of (alpha,beta) Sigma[1:2 , 1:2] < − inverse(S2[1:2 , 1:2 ])
siga < − Sigma[1,1] ; sigb < − Sigma[2,2] corrab < − Sigma[1,2]/(sqrt(Sigma[1,1]*Sigma[2,2]))
S2[1:2,1:2] ∼ dwish(Omega[1:2,1:2], 2)
}
♯Initial values
list( S2=structure(.Data=c(10,0,0,10), .Dim=c(2,2))) list(S2=structure(.Data=c(1,0,0,1),
.Dim=c(2,2)))
♯Data file
list( nobs=129, m=12, Omega = structure(.Data = c(0.003, 0, 0, 0.003), .Dim = c(2,
2)), )
s1[] s2[] s3[] s4[] s5[] s6[] s7[] s8[] s9[] s10[] score1[] score2[] time[] t1[] t2[]
1 2 3 4 5 13 14 15 16 17 8 12 34 9 1 0
1 2 3 4 5 18 14 15 16 17 6 2 114 1 0
1 2 3 4 5 18 13 15 16 17 1 4 70 1 0
1 2 3 4 7 18 13 15 16 17 0 0 72 1 0
1 2 9 4 7 18 13 15 16 19 2 3 84 1 0
2 6 9 4 7 23 13 20 16 19 4 0 55 1 0
2 6 9 4 7 23 13 20 17 19 0 1 33 1 0
2 1 9 10 7 23 13 20 17 19 2 2 47 1 0
2 1 4 10 7 23 13 20 17 19 0 2 74 1 0
2 1 4 10 5 23 13 20 17 16 3 1 99 1 0
2 1 4 3 5 15 13 18 17 16 5 5 209 1 0
2 1 4 3 5 20 13 18 17 16 6 7 152 1 0
2 1 4 3 5 17 13 14 15 16 11 18 511 1 0
2 1 4 3 7 17 13 18 15 19 4 2 118 1 0
9 1 4 6 7 16 13 18 15 19 2 4 91 1 0
9 2 4 6 7 16 13 14 15 19 0 4 119 1 0
```

```
9 2 4 6 7 16 20 14 22 19 0 3 14 1 0
9 2 4 1 7 16 20 14 22 19 6 2 154 1 0
9 2 4 1 7 16 15 14 18 17 2 1 122 1 0
9 2 4 1 7 16 15 13 18 17 0 0 10 1 0
9 2 4 1 7 16 15 13 14 17 2 0 24 1 0
9 2 4 1 5 16 15 13 14 17 2 5 106 1 0
7 2 4 1 5 16 15 13 14 17 2 4 90 1 0
7 9 4 1 5 16 15 13 14 17 3 4 81 1 0
3 2 5 4 1 16 15 13 14 17 11 10 510 1 0
8 2 5 4 1 16 15 13 14 17 1 3 23 1 0
8 2 5 4 1 16 15 13 18 17 1 3 113 1 0
8 2 7 4 1 16 20 13 18 19 2 0 45 1 0
2 9 7 4 6 23 20 13 17 19 0 2 61 1 0
1 9 7 4 6 23 20 13 17 19 3 2 92 1 0
1 9 7 8 6 15 18 13 17 19 0 2 29 1 0
1 9 7 8 6 15 18 23 17 19 5 0 115 1 0
1 9 7 8 6 15 18 23 17 16 1 2 21 1 0
1 4 7 8 6 15 18 23 17 16 0 2 25 1 0
1 4 2 8 5 15 18 23 17 16 6 5 118 1 0
1 4 2 8 5 15 13 23 17 16 11 5 244 1 0
1 4 2 3 5 15 14 13 17 16 15 22 525 1 0
1 4 2 3 5 15 14 18 17 16 1 1 49 1 0
1 4 2 3 5 15 14 23 17 16 8 5 111 1 0
1 4 9 6 7 15 13 23 16 19 0 2 105 1 0
1 11 9 8 7 20 13 23 16 19 2 3 86 1 0
1 4 9 8 7 20 13 14 16 19 5 0 62 1 0
1 4 9 8 7 15 13 14 16 17 0 2 90 1 0
2 4 9 8 7 15 13 14 16 17 2 8 192 1 0
1 4 9 8 7 15 13 14 16 17 6 6 116 1 0
1 4 2 8 7 15 13 14 16 17 3 2 33 1 0
1 4 2 8 7 15 13 14 16 17 10 2 218 1 0
3 4 2 8 7 15 13 14 16 17 0 0 64 1 0
4 1 2 3 5 17 13 15 14 16 13 17 521 0 1
```

```
4 1 2 8 5 17 13 15 14 16 1 2 94 0 1
4 1 2 8 5 20 13 18 19 16 0 2 36 0 1
4 1 2 8 7 20 13 18 19 16 2 0 43 0 1
4 2 9 8 7 20 13 21 19 18 0 1 49 0 1
4 1 9 8 7 20 13 16 19 18 0 2 74 0 1
4 1 9 6 7 20 13 16 19 18 0 0 8 0 1
4 1 9 6 7 20 23 16 19 18 1 2 115 0 1
4 1 9 6 7 20 23 16 17 18 2 2 129 0 1
5 1 9 6 3 13 23 16 17 15 5 2 128 0 1
5 1 4 6 3 13 18 16 17 15 0 0 18 0 1
5 1 4 6 3 13 18 19 17 15 2 0 82 0 1
5 1 4 6 7 13 20 16 17 15 6 6 115 0 1
5 1 4 2 3 17 13 16 15 14 10 17 359 0 1
5 1 4 6 3 17 13 16 15 14 2 4 132 0 1
11 8 4 6 7 19 13 16 15 14 3 0 115 0 1
11 8 4 6 7 19 13 16 15 20 4 3 114 0 1
11 8 4 6 7 17 18 16 19 20 0 2 63 0 1
1 8 4 6 7 17 18 16 19 20 0 5 87 0 1
1 8 5 6 9 17 18 16 19 20 1 0 7 0 1
1 4 5 6 9 17 18 16 19 20 0 2 62 0 1
1 4 5 6 9 17 15 16 13 14 5 2 54 0 1
1 4 5 8 9 17 15 16 13 14 5 6 229 0 1
6 4 5 8 9 17 15 16 13 14 2 4 48 0 1
11 4 5 8 9 17 15 16 13 14 3 0 62 0 1
11 4 5 8 9 19 21 23 22 24 1 0 45 0 1
11 12 5 8 9 19 21 23 22 24 0 4 63 0 1
5 4 1 2 3 17 16 14 15 13 0 0 90 0 1
5 4 1 2 10 17 16 14 15 13 16 18 408 0 1
5 4 1 2 7 17 16 14 15 13 2 0 63 0 1
5 4 1 2 7 19 16 14 15 18 4 3 125 0 1
6 4 1 9 7 19 17 14 20 18 3 2 133 0 1
6 4 1 9 7 19 17 14 20 13 0 0 50 0 1
8 4 1 5 7 19 18 16 20 13 3 6 58 0 1
```

8 3 1 5 7 19 18 16 20 13 0 0 15 0 1

8 3 1 5 7 17 18 16 20 13 4 2 73 0 1

8 3 1 5 4 17 15 16 14 13 0 0 19 0 1

8 3 1 5 4 19 15 16 14 13 0 2 31 0 1

2 3 1 5 4 19 15 16 14 13 4 0 130 0 1

2 9 1 5 4 21 15 16 14 13 0 4 101 0 1

2 10 1 5 4 21 15 16 14 13 3 4 110 0 1

2 3 1 5 4 17 15 16 14 13 7 8 270 0 1

6 3 1 5 4 17 15 16 14 13 4 0 115 0 1

6 10 1 5 4 17 15 16 14 13 0 5 93 0 1

6 10 1 5 4 17 15 16 14 18 2 0 54 0 1

8 10 1 5 4 17 15 16 14 18 4 2 152 0 1

8 10 7 1 4 17 15 16 13 18 0 4 82 0 1

8 5 7 1 4 17 15 16 13 18 2 0 36 0 1

8 5 7 1 4 17 15 16 13 14 4 5 133 0 1

8 5 7 1 3 17 15 16 13 14 3 5 106 0 1

6 5 4 1 3 17 15 16 13 14 0 0 18 0 1

6 5 4 1 7 17 15 16 13 14 6 11 170 0 1

12 5 4 1 7 17 15 16 13 14 1 0 10 0 1

6 5 4 1 7 17 15 16 13 14 2 2 37 0 1

6 5 4 1 3 17 15 16 13 14 2 1 39 0 1

6 5 4 1 7 17 15 16 13 14 2 2 29 0 1

3 5 4 1 7 17 15 16 13 14 0 2 51 0 1

3 6 1 5 4 15 14 13 17 16 14 7 360 0 1

3 6 11 5 4 15 14 13 17 16 0 8 105 0 1

3 6 1 5 4 15 20 13 17 16 4 8 130 0 1

3 8 1 5 4 15 20 18 17 16 4 2 63 0 1

7 8 1 9 4 14 20 18 19 16 0 3 63 0 1

7 8 1 9 4 13 20 18 19 16 7 7 177 0 1

7 8 1 5 4 13 20 18 19 16 0 2 37 0 1

10 8 1 5 4 13 15 18 17 16 0 0 18 0 1

10 6 1 5 4 13 15 18 17 16 2 3 80 0 1

10 6 1 5 4 13 15 23 21 16 3 6 129 0 1

```
10 6 11 5 4 13 15 23 21 16 2 0 49 0 1
9 6 11 5 4 13 15 23 21 16 7 9 146 0 1
1 6 3 5 4 13 15 14 17 16 8 7 294 0 1
1 6 3 5 4 13 15 18 17 16 0 2 91 0 1
1 12 3 5 4 13 15 18 17 16 6 10 179 0 1
1 12 10 5 4 13 15 18 17 16 0 2 41 0 1
1 12 10 5 4 13 15 18 17 19 0 2 42 0 1
1 12 10 7 4 13 15 18 17 19 0 2 25 0 1
8 4 11 7 9 16 13 17 20 19 2 6 88 0 1
8 4 1 7 9 16 13 17 20 19 4 0 51 0 1
8 4 1 7 9 16 13 17 15 14 5 6 243 0 1
8 4 11 7 9 16 13 17 15 14 4 4 162 0 1
8 4 11 7 9 16 23 21 15 19 5 2 57 0 1
8 10 11 7 9 16 24 21 20 19 3 0 52 0 1
8 10 11 7 9 16 24 21 20 19 3 0 52 0 1
NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA
END
```

Bibliography

[1] Bellotti, B. *The Points Created Basketball Book, 1991 - 92*, New Brunswick, New Jersey: Night Work Publishing (1992).

[2] Cronbach, L.J., Gleser, G.C., Nanda, H. and Rajaratnam, N. *The dependabiliy of behavioral measurements: Theory of generalizability of scores and profiles.* New York: Wiley (1972).

[3] Gelfand, A. E., Hills, S. E., Racine-Poon, A. and Smith, A. F. M. Illustration of Bayesian inference in normal data models using Gibbs sampling. *Journal of the American Statistical Association*, **85**, 927 - 985 (1990).

[4] Ghosh, A. and Steckel, J. Roles in the NBA: there's always room for a big man, but his role has changed. *Interfaces*, **23(4)**, 43 - 55 (1993).

[5] Gill, P.S. and Swartz, T.B. Bayesian analysis of dyadic data. *American Journal of Mathematical and Management Sciences*, **26(3-4)**, (2007).

[6] Gill, P.S. and Swartz, T.B. Bayesian analysis of directed graphs data with applications to social networks. *Journal of the Royal Statistical Society, Series C*, **53(2)**, 249 - 260 (2004).

[7] Gill, P.S. and Swartz, T.B. Statistical analyses for round robin interaction data. *The Canadian Journal of Sttistics*, **29**, 321 - 331 (2001).

[8] Gilks, W.R., Richardson, S. and Spiegelhalter, D.J. (editors). *Markov Chain Monte Carlo in Practice.*, London: Chapman and Hall.

[9] Gilovich, T., Vallone, R. and Tversky, A. The hot hand in basketball: on the misperception of random sequences. *Cognitive Psychology*, **17**, 295 - 314 (1985).

[10] Heeren, D. *The Basketball Abstract.* Englewood Cliffs, New Jersey Prentice Hall (1988).

[11] Kenny, D.A. *Interpersonal Perception: A Social Relations Analysis.* New York: Guilford Press (1994).

[12] Oliver, D. *Basketball on Paper.* Brasseys INC. Washington, D. C. (2004).

[13] Tversky, A. and Gilovich, T. *The "hot hand": stististical reality or cognitive illusion?* Chance, New Directions ofr Statistics ans Computing. **2 (1)**, 16 - 21 (1989).

[14] Trupin, J. and Couzens, G. S. *HoopStats: The Basketball Abstract.* New York: Bantam Books (1989).

[15] Wardrop, R.L. *Basketball in Statistics in Sport* (editor Jay Bennett). Arnold Publishers, 65 -82 (1998).