MMathStat Project

Bayesian inference for ranking models applied to Formula One data

Author:
Liam Kirrane

Supervisor:
Dr. Daniel Henderson

May 1, 2014

Abstract

In all professional sport we are constantly trying to find out which player or team is the best. Some sports use points systems where others apply knock-out tournaments but flaws can be found in both systems. Formula One motor racing adopts a points system which is changing for the 2014 season and many people don't feel the new points system will help decide on a deserved winner of the sport. This report will look at using full Bayesian analysis of some real life Formula One data by applying MCMC methods to the Plackett-Luce model and then extend further to other ranking models. We will use these models to analyse the 2013 Formula One season and to see if we believe that the points system used in Formula One accurately reflects the driver abilities and gives a deserved winner.

Contents

Chapter 1

Introduction

1.1 Formula One motor racing

Formula One, also known as Formula 1 or F1 and referred to officially as the FIA Formula One World Championship is fast becoming one of the most popular sports on the planet. The 2013 season was the most viewed season ever as millions watched as races in 19 different countries all across the world took place. In each race in the 2013 season, there were 11 teams competing, with each team having 2 drivers. There were 23 drivers in total who competed across the entire season. The list of drivers together with the results of all 19 races is given in Appendix A.

The aim over the course of the season is to collect points in each race and the person who collects the most points wins the Championship. The Formula One points system in 2013 was as follows

Position	1^{st}	2^{nd}	3^{rd}	4^{th}	5^{th}	6^{th}	7^{th}	8^{th}	9^{th}	10^{th}
Points	25	18	15	12	10	8	6	4	2	1

A question that is commonly asked each season is 'Who is going to win the championship?'. In 2013 that person was Sebastian Vettel, the bookies favourite, and we shall be analysing whether or not he deserved his title as world champion.

We will be analysing data from the 2013 Formula One season to see if the points system reflects the abilities of the drivers, if Sebastian Vettel deserved his first place finish and whether there are any other interesting results throughout the season. When analysing sport data, we often are looking for who is the best player or team. Here we will use rankings models to compare the ability of each driver. We focus on the ranks (finishing positions) rather than times or some other measure of performance as the races took place on different tracks over different distances. The first model that we will look at is the Plackett-Luce model (Plackett, 1975; Luce 1959) which is a commonly used model to estimate competitor strengths based on ranks.

We will be adopting a Bayesian approach throughout this report and using Markov chain Monte Carlo (MCMC) methods to sample from the posterior distribution. Bayesian methods and MCMC are reviewed briefly in Section 1.2 before we describe the Plackett-Luce model in Section 1.3.

1.2 Bayesian inference

Bayesian inference is the process of fitting a probability model to a set of data and summarising the result by a probability distribution on the parameters of the model and on unobserved quantities such as predictions for new observations (Gelman et al, 2004). Bayesian statistical conclusions about a parameter Θ are made in terms of probability statements conditional on the observed value of the data y, $p(\Theta \mid y)$. This conditioning on observed data is where Bayesian inference differs from the usual statistical inference approach. However, for simple analyses, it can be seen that similar conclusions can be drawn between the two approaches. However for more complex problems, Bayesian analyses can be easily extended whereas the traditional statistical inference approach can have problems. It is this reason why we will be using a Bayesian approach in this report.

For our model we will be using MCMC to sample our posterior values. Markov Chain Monte Carlo (MCMC, Brooks et al, 2011) is a general method based on approximating a target posterior distribution, $p(\Theta \mid y)$ by sequentially drawing values of Θ, therefore the draws form a Markov chain, from the approximate distributions and correcting those draws. This method is useful as the approximate distributions are improved at each step in the simulation, in the sense of converging to the target distribution. The Gibbs sampler, defined in terms of Θ, is an example of a Markov chain algorithm. If we suppose parameter vector Θ can be divided into d components, $\Theta = (\Theta_1, ..., \Theta_d)$, each iteration of the Gibbs sampler cycles through the components of Θ drawing each subset conditional on the value of all the others. Each component of Θ is ordered for each iteration t, and each Θ_j^t sampled from the conditional distribution given all other components of Θ. So each component is updated conditional on the latest values of the other components of Θ, which are the already updated components for the iteration t values and the iteration $t-1$ values for the others.

1.3 Plackett-Luce Model

We assume that $p_i \leq K$ drivers compete in race i, where $i = 1, 2, ..., n$. We write the result of race i as $\rho_i = (\rho_{i1}, \rho_{i2}, ..., \rho_{ip_i})$ where ρ_{i1} is the driver who finished first, ρ_{i2} is the driver who finished second, etc.
The Plackett-Luce model assumes that in a race with k drivers, labelled $1, 2, \ldots, k$, the probability that the j^{th} driver wins is

$$\frac{\lambda_j}{\lambda_1 + \lambda_2 + ... + \lambda_k} = \frac{\lambda_j}{\sum_{m=1}^{k} \lambda_m}$$

where the j^{th} driver has 'ability' $\lambda_j > 0$. So now we know who is first, we think of the person who came second as winning out of the remaining drivers, and so on until last place.

Let's say for example that there were only 4 races, $n = 4$, and 4 drivers, $k = 4$, in a season, and that the outcome of those races were as follows.
Race 1: $\to \rho_1 = \{3, 1, 2, 4\}$
Race 2: $\to \rho_2 = \{1, 2, 4\}$
Race 3: $\to \rho_3 = \{3, 4, 1\}$
Race 4: $\to \rho_4 = \{4, 1, 2, 3\}$.

The results above tell us that in race 1, the winner was driver number 3, then in second place was driver 1, followed by driver 2 and in last place was driver 4. We also see in races 2 and 3 that not every driver competed. The probability that these races finished in this ordering based on the Plackett-Luce model are;

$\Pr(\rho_1 \mid \lambda) = \frac{\lambda_3}{\lambda_3+\lambda_1+\lambda_2+\lambda_4} \times \frac{\lambda_1}{\lambda_1+\lambda_2+\lambda_4} \times \frac{\lambda_2}{\lambda_2+\lambda_4}$

$\Pr(\rho_2 \mid \lambda) = \frac{\lambda_1}{\lambda_1+\lambda_2+\lambda_4} \times \frac{\lambda_2}{\lambda_2+\lambda_4}$

$\Pr(\rho_3 \mid \lambda) = \frac{\lambda_3}{\lambda_3+\lambda_4+\lambda_1} \times \frac{\lambda_4}{\lambda_4+\lambda_1}$

$\Pr(\rho_4 \mid \lambda) = \frac{\lambda_4}{\lambda_4+\lambda_1+\lambda_2+\lambda_3} \times \frac{\lambda_1}{\lambda_1+\lambda_2+\lambda_3} \times \frac{\lambda_2}{\lambda_2+\lambda_3}$.

In general, the Plackett-Luce model gives the probability of the observed ranking as

$$\Pr(\rho_i \mid \lambda) = \prod_{j=1}^{p_i} \frac{\lambda_{\rho_{ij}}}{\sum_{k=j}^{\rho_i} \lambda_{\rho_{ik}}} = \prod_{j=1}^{p_i-1} \frac{\lambda_{\rho_{ij}}}{\sum_{m=j}^{\rho_i} \lambda_{\rho_{im}}},$$

where $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_K)$, and so it can be shown that over the course of a season the probability of observing all the results that we did is

$$p(D|\lambda) = \prod_{i=1}^{n} \Pr(\rho_i|\lambda) = \prod_{k=1}^{K} \lambda_k^{w_k} \left(\prod_{i=1}^{n} \prod_{j=1}^{p_i-1} \sum_{m=j}^{p_i} \lambda_{\rho_{im}} \right)^{-1},$$

where w_k is the number of races where the k^{th} driver is not in last place. This is what we will use to calculate our likelihood values.

So for our example from earlier,

$$\begin{aligned}
\Pr(D|\lambda) &= \prod_{i=1}^{n} \Pr(\rho_i|\lambda) = \Pr(\rho_1 \mid \lambda) \times \Pr(\rho_2 \mid \lambda) \times \Pr(\rho_3 \mid \lambda) \times \Pr(\rho_4 \mid \lambda) \\
&= \frac{\lambda_3}{\lambda_3 + \lambda_1 + \lambda_2 + \lambda_4} \times \frac{\lambda_1}{\lambda_1 + \lambda_2 + \lambda_4} \times \frac{\lambda_2}{\lambda_2 + \lambda_4} \times \frac{\lambda_1}{\lambda_1 + \lambda_2 + \lambda_4} \\
&\quad \times \frac{\lambda_2}{\lambda_2 + \lambda_4} \times \frac{\lambda_3}{\lambda_3 + \lambda_4 + \lambda_1} \times \frac{\lambda_4}{\lambda_4 + \lambda_1} \times \frac{\lambda_4}{\lambda_4 + \lambda_1 + \lambda_2 + \lambda_3} \\
&\quad \times \frac{\lambda_1}{\lambda_1 + \lambda_2 + \lambda_3} \times \frac{\lambda_2}{\lambda_2 + \lambda_3} \\
&= \lambda_1^3 \lambda_2^3 \lambda_3^2 \lambda_4^2 \times \{(\lambda_3 + \lambda_2 + \lambda_2 + \lambda_4) \times \cdots \times (\lambda_2 + \lambda_3)\}^{-1}.
\end{aligned}$$

and so our w_k values would be w_1=3, w_2=3, w_3=2 and w_4=2.

Our aim is to infer the ability parameters λ based on the observed data. As stated in Section 1.2 we use a Bayesian approach which involves first specifying a prior distribution for the unknown parameters. We use independent Gamma prior distributions to express beliefs about driver abilities. We take $\lambda_k \sim Gamma\,(a, b)$ independently for $k = 1, 2, \dots, K$ and so the joint prior density is

$$p(\lambda) = \prod_{k=1}^{K} p(\lambda_k) = \prod_{k=1}^{K} \frac{b^a \lambda_k^{a-1} e^{-b\lambda_k}}{\Gamma(a)}.$$

The driver ability parameters are positive and the gamma distribution is a natural prior distribution for a positive quantity such as this. Inferences are based on the posterior distribution

$$p(\lambda|D) \propto p(\lambda)p(D|\lambda).$$

However, this distribution is not available analytically and so we resort to MCMC (see Section 1.2). Specifically, Caron and Doucet (2012) show that introducing a particular set of latent variables produces a complete-data likelihood that is easier to work with. This allows us to construct a Gibbs sampler for sampling from $p(\lambda|D)$. In particular, for $i = 1, ..., n$ and $j = 1, ..., p_i - 1$, we introduce the following latent variables $Z = \{Z_{ij}\}$ sampled from the exponential distribution:

$$Z_{ij} \sim Exp\left(\sum_{m=j}^{\rho_i} \lambda_{\rho_{im}}\right)$$

so that the full conditional distribution of the latent variables is

$$p(Z \mid D, \lambda) = \prod_{i=1}^{n} \prod_{j=1}^{p_i - 1} \left(\sum_{m=j}^{p_i} \lambda\rho_{im}\right) \exp\left(-\left(\sum_{m=j}^{p_i} \lambda\rho_{im}\right) Z_{ij}\right).$$

The complete-data is therefore

$$p(D \mid \lambda)\, p(Z \mid D, \lambda) = p(Z, D \mid \lambda).$$

We use our gamma prior distribution and the complete-data likelihood to calculate the full joint density of the parameters and latent variables:

$$p(\lambda, Z|D) \propto p(\lambda)\, p(Z, D \mid \lambda).$$

To sample from $p(\lambda, Z|D)$ we use the following Gibbs sampler as specified in Caron and Doucet (2012). At iteration t: for $i = 1, ..., n$ and $j = 1, ..., p_i - 1$, sample

$$Z_{ij}^{(t)} \mid D, \lambda^{(t-1)} \sim Exp\left(\sum_{m=j}^{p_i} \lambda_{\rho_{im}}^{(t-1)}\right).$$

For $k = 1, ..., K$, sample

$$\lambda_k^{(t)} \mid D, Z^{(t)} \sim Gamma\left(a + w_k, b + \sum_{i=1}^{n} \sum_{j=1}^{p_i - 1} \delta_{ijk} Z_{ij}^{(t)}\right)$$

where δ_{ijk} is the indicator of the event that driver k finishes no better than j^{th} in the i^{th} race.

$$\delta_{ijk} = \begin{cases} 1, & \text{if } k \in \{\rho_{ij}, ..., \rho_{ip_i}\}, \\ 0, & \text{otherwise.} \end{cases}$$

To improve the mixing of the MCMC algorithms, an additional sampling step can be added where we normalise the current parameter estimate at iteration t, λ^t, and then rescale them randomly using a prior draw from $\Lambda = \sum_{i=1}^{K} \lambda_i$.

For i=1,...,K, set

$$\lambda_i^{*(t)} = \frac{\lambda_i^t}{\sum_{j=1}^{K} \lambda_j^t} \Lambda^{(t)}$$

where $\Lambda^{(t)} \sim Gamma(Ka, b)$. This is possible because the data provide no information about the parameter Λ; multiplying all the λ values by a constant will produce the same likelihood. This step can dramatically improve the mixing of the Markov chain.

We are now in a position to write a Gibbs sampler in R (R Development Core Team, 2012) for the Plackett-Luce model and in the next chapter we use this code to analyse data from the 2013 Formula One season. The R code is given in Appendix B.

Chapter 2

Plackett-Luce Model

2.1 Introduction

In this chapter we fit the Plackett-Luce model to data from the 2013 Formula One season. We have set our prior distribution for each driver to be a $Gamma(a, b)$ distribution with $a = b = 1$, i.e., an exponential distribution. We could have chosen other priors but chose this one as it is a fairly uninformative choice and represents weak prior beliefs. Our prior for all drivers is equal so we are making no prior assumptions that any driver is better than any other. The prior distribution gives flexibility to include prior beliefs if we have them by changing the a parameter. We ran the Gibbs sampler described in Chapter 1 for 10000 iterations so that means that all 23 drivers have 10000 sampled posterior λ values on which to base our inferences. From Figure 2.1 it seems that all the drivers traces have converged to their stationary distribution.

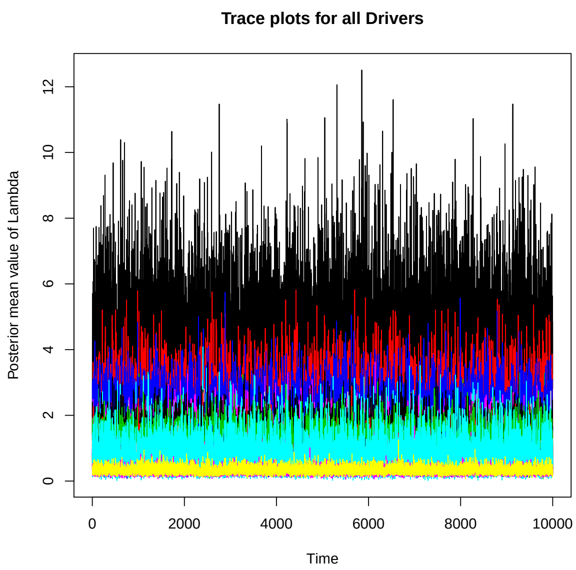


Figure 2.1: Trace plot for all drivers in the Formula One season 2013

2.2 Model results

We begin by comparing each driver's overall finishing position against their posterior mean value of λ for the 2013 Season; see Figure 2.2. There is obvious correlation between

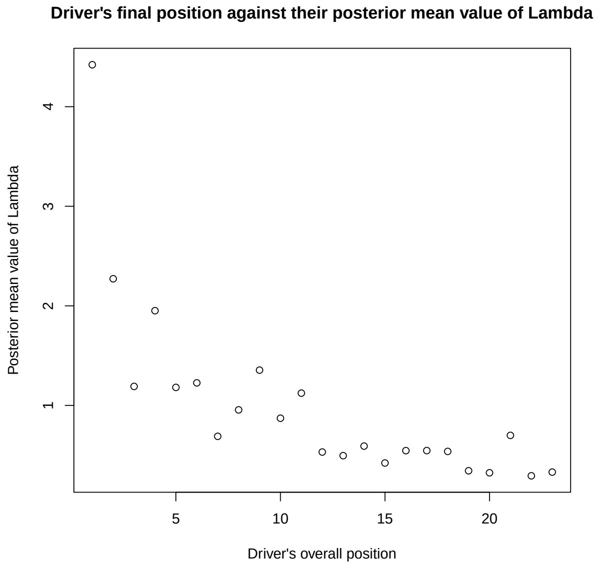


Figure 2.2: Posterior mean against finishing position

driver's overall position and their λ coefficient however there are some results which are unexpected. The posterior mean values of λ for Jenson Button (driver 9) and Heikki Kovaleinen (driver 21) are larger than their overall position would suggest, and drivers Mark Webber (driver 3) and Romain Grosjean (driver 6) have lower λ values than we would expect, based on their points finish. We shall look into the reason for this later on.

Table 2.1 shows us the order of the drivers overall finishing positions next to the driver's posterior mean value of λ and the standard deviation of λ. We see that generally the higher the standard deviation, the higher the mean of λ. We know that for a $Gamma(a, b)$ distribution that the mean is a/b and the variance is a/b^2 so clearly if the mean is increased, that means either a has increased or b has decreased, so that the standard deviation would increase too. The posterior distributions are mixtures of gamma distributions and so we use the above logic to explain these findings.

We see that Kovalainen has a larger standard deviation (4th highest) than some drivers whose mean (11th largest) value of λ is larger than his. This is because Kovalainen only participated in the last two races with most other drivers participating in all 19 races and so his posterior will be very similar to his prior distribution as it is based on less data and so will have a larger standard deviation.

Position	Driver	Mean λ	Standard deviation λ
1	Vettel	4.42	1.34
2	Alonso	2.27	0.75
3	Webber	1.19	0.38
4	Hamilton	1.95	0.63
5	Raikkonen	1.18	0.41
6	Rosberg	1.23	0.38
7	Grosjean	0.69	0.24
8	Massa	0.96	0.32
9	Button	1.35	0.42
10	Hulkenberg	0.87	0.29
11	Perez	1.12	0.34
12	di Resta	0.53	0.18
13	Sutil	0.50	0.17
14	Ricciardo	0.59	0.20
15	Vergne	0.42	0.14
16	Gutierrez	0.55	0.17
17	Bottas	0.55	0.17
18	Maldonado	0.54	0.18
19	Bianchi	0.34	0.12
20	Pic	0.32	0.11
21	Kovalainen	0.70	0.43
22	Garde	0.29	0.10
23	Chilton	0.33	0.10

Table 2.1: Posterior summaries for all 23 drivers

2.3 Vettel

We can see from the posterior values of λ in Table 2.1 that this model predicts that Vettel deservedly was crowned champion of the 2013 season. And it is hard to argue with that point, as anyone who follows Formula One knows of Vettel's complete dominance in this season. We can compare his posterior distribution to that of his closest competitor, Fernando Alonso, however we expect this will just show more just how much Vettel deserved this title. From Figure 2.3 of the posterior density of each driver's λ values, we see just how dominant Vettel was. If a posterior mean value of λ of 4.42 for Vettel compared to Alonso's 2.27 doesn't show he deserved his title, then the fact that Vettel's minimum sampled value of λ is only marginally smaller than Alonso's mean value shows that this was no fluke.

So now that we have that Vettel deserved top spot, we shall look into other interesting results, such as the four drivers mentioned earlier whose λ values weren't in accordance to their finishing position.

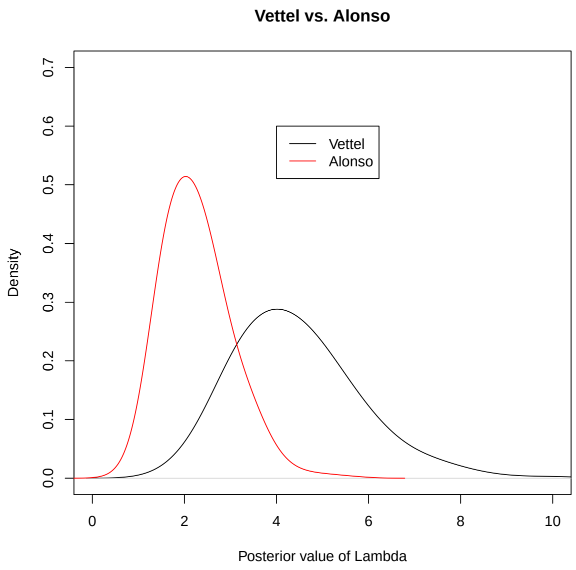


Figure 2.3: Posterior density plots for Vettel and Alonso

2.4 Other drivers of Interest

Four drivers whose λ coefficients are not as expected given their overall finishing position are Webber, Grosjean, Button and Kovalainen. Graphs below look into the reason behind these four driver's unexpected results.

Figure 2.4 shows the four driver's posterior density plots. Notice that for Kovalainen's

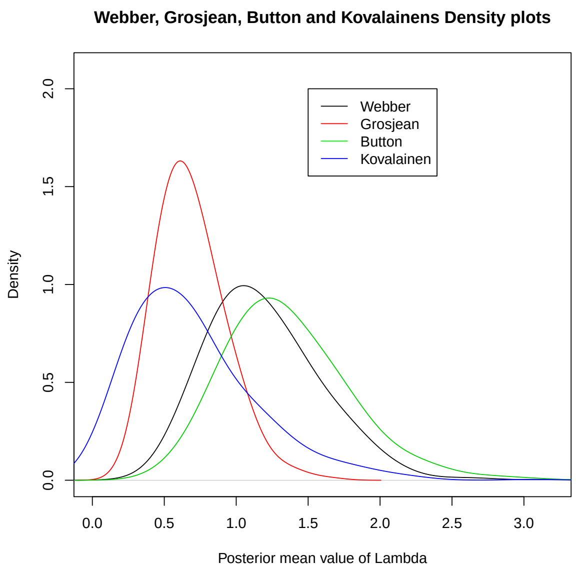


Figure 2.4: Posterior density plot for drivers of interest

plot that the density is not equal to zero at λ equal to zero, this assumes that we can have a positive density for negative λ which isn't true. This is simply caused by the kernel density estimation when smoothing our plots. Unfortunately these plots do not tell us too much and they definitely do not explain why their posterior mean values for λ are not as we would expect based on the points finishing position. We shall look at different types of plots to try to find an explanation.

Unfortunately these plots have not explained our strange results, so we will try another method. As we saw in Chapter 1, the Formula One points system only awards points to the first 10 finishing positions in each race, and then positions 11 to 22 receive 0 points. So that would mean a driver finishing in 11th place will get the same score as someone who finished in 22nd place.

Now we will look at all four drivers' finishing positions.

Webber:	6	2	20	7	5	3	4	2	7	4	5	3	15	21	2	20	2	3	2
Grosjean:	10	6	9	3	22	17	13	19	3	6	8	8	21	3	3	3	4	2	22
Button:	9	17	5	10	8	6	12	13	6	7	6	10	7	8	9	14	12	10	4
Kovalainen:																		14	14

Button, who has a larger posterior mean value for λ than his overall position would suggest, seems to be quite a consistent driver. In 15 of his 19 races he achieved a finishing position between 6th and 12th so why would this leave him with the 3rd highest mean λ score?

Grosjean, who has a smaller posterior mean value for λ than his overall position would expect, had a very strong ending to the Formula One 2013 season, finishing 4th or better on 5 of the last 6 races. However prior to this fantastic finish, he was very inconsistent finishing 3rd one race and 22nd the next, and it seemed that if he did finish outside the Formula One points places (top 10), then he would finish very low.

Webber seems quite a consistent driver, but again when he didn't finish in a points position, it seemed that he would finish terribly. The 4 times he finished outside the points places his finishing positions were 15th, 20th, 20th and 21st. But still he was a very consistent driver, so how come the Plackett-Luce model gave him such a low posterior mean value of λ?

Kovalainen is a hard driver to judge, seeing as he only participated in the last two races. His posterior mean values of λ would have been draws from the prior distribution, which has a mean value of 1, until he had participated. He then finished 14th in the last two races, so you may expect the Plackett-Luce model to give him close to the 14th highest mean λ value, which it did. Not only does Kovalainen have a high posterior mean value of λ, he also has one of the highest standard deviations, as discussed earlier. However as the Formula One points system awards 0 points to 14th place, he is ranked 21st position using the Formula One system.

This is where the Formula One points system differs from the Plackett-Luce model. The Plackett-Luce model will rank all 22 drivers in each race and there can be no ties. So whereas finishing 11th or 22nd means nothing in the Formula One points system, it makes a large difference to the λ value you receive for each and every position.

Therefore, the reason Button has such a large posterior mean value of λ is because when he does not finish in the top 10, he still had a high finishing position. Whereas drivers Webber and Grosjean seemed to ease off if they knew they were not finishing inside the top

10. Now this is where you could say the Plackett-Luce model is unsuitable for Formula One data as it obviously can't take into account each drivers individual targets. For example, we saw that when Webber did not finish inside the top 10, he would finish very low, but we do not know if Webber was still trying his hardest for these races, or if he had realised that he was not going to be finishing in a points position so he decided to ease off to make sure he does not damage his car or crash and harm himself. These cars cost millions of pounds to produce so drivers and owners do not want them damaged if it will have no effect on their overall points total. So Mark Webber has the luxury to be able to drive more safely and take a poor finishing position, as it makes no difference to him whether he finishes 11th or 22nd. Also Webber has the no real incentive to beat his team mate (Vettel) as Vettel was clear favourite from early on in the season. However some other drivers, who may not be competing to finish 1st place or aiming for a podium finish may have different incentives. They might be a new young driver on the Formula One scene who is trying to impress to get a better team next season, or just to make sure they still have a job next season! So to some drivers, there is a big difference to finishing 11th or 22nd, however to the better drivers, who know that they will still have a job the following season the difference is less clear and they can ease off for the cars and their own safety.

2.5 Summary

So now we have seen why the results using the Plackett-Luce model are different from that of the actual Formula One 2013 results, but who's to say which results are better? If there were points offered for all 22 finishing positions, would it surely not make a more interesting race, as each and every finishing position would matter and drivers could not slow down for the safety of their cars and themselves? Or can we alter the Plackett-Luce model so that the ranking system works similar to that of the Formula One points system?

Chapter 3

Truncated Plackett-Luce model

3.1 Introduction

We discussed at the end of the last chapter about altering the Plackett-Luce model we had used so that it is more like the Formula One points system. But how could we do this? And will this affect the accuracy or realism of the Plackett-Luce model?

In this chapter we consider a truncated version of the Plackett-Luce model. By truncating the model we mean that we would only include ranks up to a certain number into our model, so if we truncated up to the 5th value, that would mean we would only record the drivers who finished 1st, 2nd, 3rd, 4th and 5th, and we would consider every other result as not in the top 5. So the most important factor in this is what finishing position do we truncate our model up to? As the Formula One points system awards points to the first 10 positions then we will produce a truncated Plackett-Luce model for the first 10 drivers. However we will discuss later why truncating for the top 10 finishing positions may not necessarily be the fairest way to rank the drivers.

3.2 Truncated Plackett-Luce model

The truncated Plackett-Luce model, truncated to the first r values, assumes that the probability of the observed finishing order, for $r < p_i$, is

$$\Pr\left(\rho_i \mid \lambda\right) = \prod_{j=1}^{r} \frac{\lambda_{\rho_{ij}}}{\sum_{m=j}^{\rho_i} \lambda_{\rho_{im}}},$$

and so it can be shown that the likelihood based on the results of the whole season

$$p(D|\lambda) = \prod_{i=1}^{n} \Pr(\rho_i|\lambda) = \prod_{k=1}^{K} \lambda_k^{w_k} \left(\prod_{i=1}^{n} \prod_{j=1}^{r} \sum_{m=j}^{p_i} \lambda_{\rho_{im}} \right)^{-1},$$

where w_k is the number of races where the k^{th} driver finishes is in the top r positions. This is what we will use to calculate our likelihood values.

As in the original Plackett-Luce model, it is easier to work with the complete-data likelihood, so we need to calculate

$$p\left(D \mid \lambda\right) p\left(Z \mid D, \lambda\right) = p\left(Z, D \mid \lambda\right).$$

For $i = 1, ..., n$ and $j = 1, ..., r$, we introduce the following latent variables $Z = \{Z_{ij}\}$ sampled from the exponential distribution:

$$Z_{ij} \sim Exp\left(\sum_{m=j}^{\rho_i} \lambda_{\rho_{im}}\right)$$

giving

$$p\left(Z \mid D, \lambda\right) = \prod_{i=1}^{n} \prod_{j=1}^{r} \left(\sum_{m=j}^{p_i} \lambda \rho_{im}\right) \exp\left(-\left(\sum_{m=j}^{p_i} \lambda \rho_{im}\right) Z_{ij}\right).$$

We use a $Gamma(1,1)$ prior distribution to calculate the joint posterior distribution, the same as we did for the Plackett-Luce model. We use this prior distribution and the complete-data likelihood to calculate:

$$p\left(\lambda, Z | D\right) \propto p\left(\lambda\right) p\left(Z, D \mid \lambda\right).$$

To sample from $p\left(\lambda, Z | D\right)$ we use the following Gibbs sampler. At iteration t: For $i = 1, ..., n$ and $j = 1, ..., r$, sample

$$Z_{ij}^{(t)} \mid D, \lambda^{(t-1)} \sim Exp\left(\sum_{m=j}^{p_i} \lambda_{\rho_{im}}^{(t-1)}\right).$$

For $k = 1, ..., K$, sample

$$\lambda_k^{(t)} \mid D, Z^{(t)} \sim Gamma\left(a + w_k, b + \sum_{i=1}^{n} \sum_{j=1}^{r} \delta_{ijk} Z_{ij}^{(t)}\right)$$

where δ_{ijk} is the indicator of the event that driver k finishes no better than j^{th} in the i^{th} race. So this has not changed from our original model. The R code for the truncated model is listed in Appendix C.

3.3 Results

The Gibbs sampler for the truncated Plackett-Luce model was run for 10000 iterations and the usual checks on convergence were performed. From Figure 3.1 we can see the trace plots for all 23 drivers who competed in the 2013 season for the truncated model. We see that all appear to converge to their stationary distribution.

Figure 3.2 is a plot of the drivers' overall finishing positions against their posterior mean values of λ calculated from the Plackett-Luce model truncated at 10th position. We see from this plot just how dominant Sebastian Vettel was in this season, with a posterior

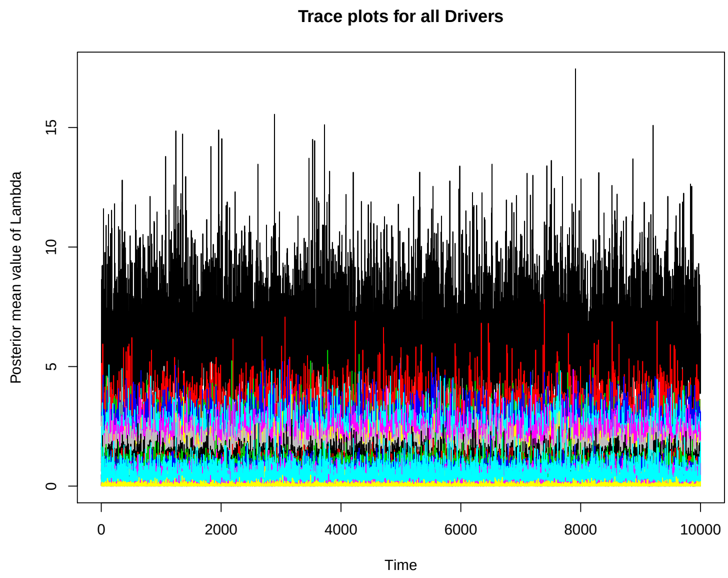


Figure 3.1: Trace plot for all drivers

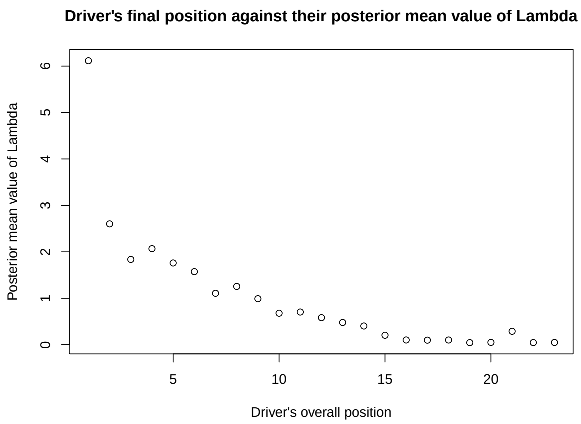


Figure 3.2: Posterior mean against finishing position

mean higher than double his closest competitors. There is definite correlation here and there are fewer unexpected values than there were in the original Plackett-Luce model so it appears from this plot that truncating our model matches up better with the Formula One points system. The only unexpected result that stands out on this plot is Heikki Kovalainen (driver 21) has a larger λ value again so we shall look into why this is later on.

Figure 3.3 shows the density plots for Vettel and Alonso for the Plackett-Luce model and the truncated Plackett-Luce model. Vettel's dominance was clear from the Plackett-Luce model, and using the truncated model only shows more of how deserving he was to be crowned champion as there is less overlap between these densities. Vettel's mean posterior value of λ is only slightly smaller than Alonso's maximum sampled posterior λ value.

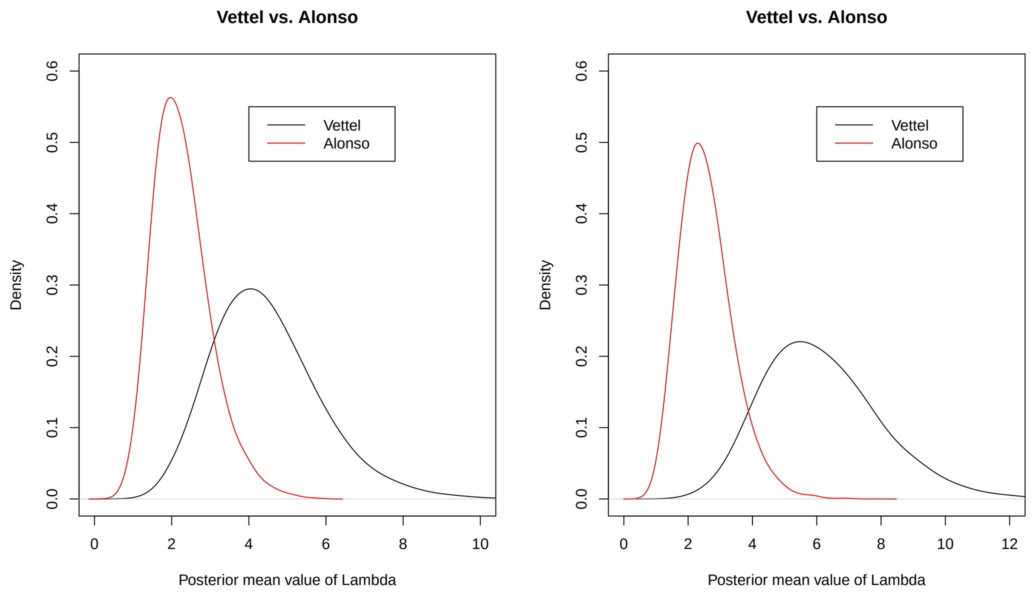


Figure 3.3: Posterior density plots for Vettel and Alonso. Left: Plackett-Luce model. Right: Truncated Plackett-Luce model

3.4 Truncated model vs Original model

In this chapter so far we have looked at the results of the truncated Plackett-Luce model, but we haven't looked into whether or not it is an improvement on our last model. To compare these models we will look at the posterior mean values of λ for each model.

Table 3.1 is a table of the drivers ordered by their overall finishing position and their posterior mean value of λ calculated using the original Plackett-Luce model and a posterior mean value of λ calculated using the truncated Plackett-Luce model. Figure 3.4 shows plots of finishing position against posterior mean value of λ for each graph. We see from

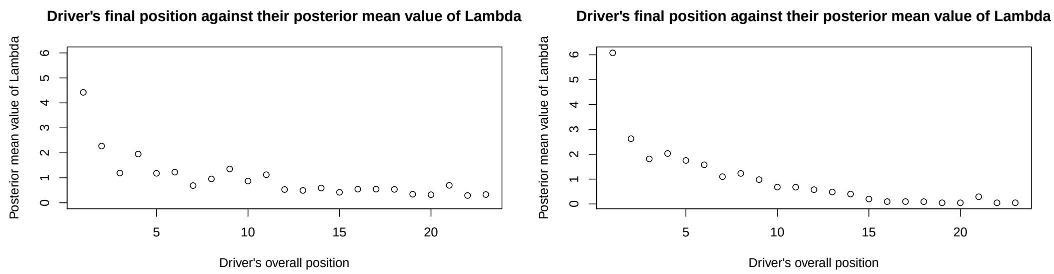


Figure 3.4: Plots of overall finishing position against posterior mean value of λ. Left: Plackett-Luce model. Right: Truncated Plackett-Luce model

Table 3.1 that it appears that the λ values of the better drivers, who finished in the higher positions, increase whereas the drivers in 9th place or worse, apart from one exception, decrease using the truncated model. So now we need to find a logical reason why the truncated model is giving these better drivers much larger λ values than our original model. The difference in the models is that the truncated model takes into account the

Overall Position	Driver	Overall Points	Mean λ Original model	Mean λ Truncated model
1	Vettel	397	4.42	6.12
2	Alonso	242	2.27	2.60
3	Webber	199	1.19	1.83
4	Hamilton	189	1.95	2.07
5	Raikkonen	183	1.18	1.76
6	Rosberg	171	1.23	1.57
7	Grosjean	132	0.69	1.11
8	Massa	112	0.96	1.26
9	Button	73	1.35	0.99
10	Hulkenberg	51	0.87	0.68
11	Perez	49	1.12	0.70
12	di Resta	48	0.53	0.58
13	Sutil	29	0.50	0.48
14	Ricciardo	20	0.59	0.40
15	Vergne	13	0.42	0.20
16	Gutierrez	6	0.55	0.10
17	Bottas	4	0.55	0.10
18	Maldonado	1	0.54	0.10
19	Bianchi	0	0.34	0.05
20	Pic	0	0.32	0.05
21	Kovalainen	0	0.70	0.29
22	Garde	0	0.29	0.05
23	Chilton	0	0.33	0.05

Table 3.1: Comparison of posterior means under the two models

top 10 finishes and then counts all other drivers as 'not in the top 10' equally, but that would not mean an increase for the better drivers alone necessarily. This could be a similar problem as we saw in Chapter 2, where the better drivers are giving up and driving safer if they know they are not finishing in a points position. Let's take Vettel as an example. His mean λ value changes from 4.42 (normal model) to 6.12 (truncated model) and he finishes in the top 10 for every race except 1. In the British Grand Prix (race 8), he had to retire and finished in 21st position. So our original Plackett-Luce model would have used that result when calculating his mean λ value, however the truncated model would just consider it as not in the top 10. So it appears that when the better drivers finished outside the top 10, they would have had a very poor finish, as they know they will not be picking up large amounts of points so would rather protect themselves and their cars. As discussed in previous chapter, not every driver has this luxury and most of the drivers know that they can not drive safer and that they have to still try their hardest to finish with as many points as possible, even if it is only picking up one point at a time. Another good example to look at is Jenson Button. In the previous chapter we discussed why he had a much larger posterior mean value for λ than his point total would suggest and explained it was due to his consistency of finishing between 6th and 12th. However with the truncated model, it does not make any difference if you finished 11th consistently or 22nd consistently, as they will be considered the same. So Button has a lot of scores finishing between 6th and 10th, which is still impressive, but it lowers his mean λ value considerably to a value that we would expect based on his points total.

Another interesting result from the truncated model is that there are 5 drivers who never finished inside the top 10, so we would expect all 5 to have the same mean λ value. Four of the drivers have a mean λ value of 0.05, where Kovalainen has a λ value of 0.29, which is higher than 8 drivers in total. The reason for this is because Kovalainen only participated in the last 2 races so his posterior will be similar to his prior. As the prior mean is 1, which is larger than some of posterior values for the worse drivers, the truncated model has given him a larger posterior mean value of λ than the Formula One points system suggests he should have.

The top 8 drivers have increased λ values for the truncated model but the remaining drivers have decreased values, except for Paul di Resta. Di Resta is an interesting case as he finished in the top 10 on 9 occasions so we would have expected a high mean value for λ from the original Plackett-Luce model, however on 5 separate occasions he finished 20th position or worse, so this is why he has such a small posterior mean λ value for our original model. So it seems that di Resta thought he had the luxury to take it safe if he was not finishing in a points position too, however this clearly was not the case as Paul di Resta lost his place on the team and is not currently participating in the 2014 Season.

It appears from Figure 3.4 that the truncated model seems to suit the Formula One points system much more than the original Plackett-Luce model; however there are still some issues that we must consider. Truncating the model for the first 10 positions seems logical as it is the points places, and as we explained some of the better drivers seemed to slow down if they knew they were not finishing in a points position, but at what exact position do these drivers decide to slow down? If a driver is coming in 11th position surely he still believes he can finish in the points places, or even 12th or slightly higher will believe they have a chance. So should we not truncate the model until at least 12th position, as these drivers are still trying their best to finish as high as possible? The problem is that there is no correct answer to that question, each and every driver will have different aims and goals for each race, so there is no level we could truncate the model to which will fairly take into account every driver's targets. For example if Vettel was in 11th place towards one of the latter races, he may slow down because he knows collecting 1 or 2 points is meaningless to him, whereas many other drivers would still try their hardest to reach a points position.

3.5 Prediction

We can compare the two models in terms of their predictive ability by computing the probability of the result in the final race of the season, given only the results prior to that race. So in the final race (race 19), the ordering from first to last was Vettel (Driver 1), Webber (Driver 3), Alonso (Driver 2), ..., Pic (Driver 21), Bottas (Driver 17) and Grosjean

(Driver 7) last. So our predictive probability is obtained by calculating the values of

$$\frac{\lambda_1}{\lambda_1+\lambda_3+\lambda_2...+\lambda_{21}+\lambda_{17}+\lambda_7} \times \frac{\lambda_3}{\lambda_3+\lambda_2...+\lambda_{21}+\lambda_{17}+\lambda_7}$$
$$\times \frac{\lambda_2}{\lambda_2...+\lambda_{21}+\lambda_{17}+\lambda_7} \times ... \times \frac{\lambda_{21}}{\lambda_{21}+\lambda_{17}+\lambda_7} \times \frac{\lambda_{17}}{\lambda_{17}+\lambda_7}$$

for each of the 10000 sampled posterior values of λ and then taking the mean of these values. We use our likelihood functions for the original Plackett-Luce model and the truncated Plackett-Luce model to calculate the probability of the final race. So the probability that the final race finished with all 22 drivers in the positions they actually finished in was 9.00 $\times 10^{-20}$ for the Plackett-Luce model and 1.92 $\times 10^{-21}$ for the truncated model, so we see that the original model has a larger, even though still very small, probability of correctly predicting the final race. The probabilities are small as we are multiplying 22 probabilities together and there are 22! possible finishing permutations of the 22 drivers in the final race. Looking at the ratio of the probabilities under the two models gives a ratio of 46.8:1 in favour of the original model. This might be expected as the Plackett-Luce model uses all the results whereas the truncated model throws away some information on the worse drivers. However, if you were trying to model the data for financial gain then the betting markets only look at predicting the winner, the podium finishers and the points finishers, not the overall finishing order of all 22 drivers, and so it is these aspects of the models that we turn to now.

We shall now look at calculating the probability of correctly predicting the winner, the podium finishers and the drivers finishing in a points position for the final race using each model. Vettel was the winner of this race and the Plackett-Luce model gave Vettel a 19.3% probability of winning the final race, whereas the truncated model gave Vettel a probability of 27.7% (a ratio of 1:1.44 in favour of the truncated model). For correctly predicting the podium finish (Vettel 1st, Webber 2nd, Alonso 3rd) the Plackett-Luce model has a probability of 0.170% whereas the truncated model has a probability of 0.575% (a ratio of 1:3.38 in favour of te truncated model). The Plackett-Luce has a probability of 2.16 $\times 10^{-10}$ for correctly predicting the outcome of the points positions (Vettel, Webber, Alonso, Button, Rosberg, Perez, Massa, Hulkenburg, Hamilton, Ricciardo) and the truncated model has a probability of 2.99 $\times 10^{-09}$ (a ratio of 1:13.84 in favour of the truncated model).

We may have expected the Plackett-Luce model to be more accurate at predicting the final outcome of all 22 positions, as some drivers never finished inside the top ten positions so the truncated model would have no information for these drivers. It appears here that the truncated model gives a much better prediction for the outcome of winner, podium finish and points finish of the final race, which we assumed from the scatter plots of overall finishing position against mean λ values for both models. However this does not necessarily make the truncated model a better model as it is only based on the prediction for a single race.

3.6 Summary

So the major advantages of using the truncated version of the Plackett-Luce model are that it helps analyse the data for the better drivers. We can see from the predictions made above that the truncated model has a much larger chance of matching the real life results. By ignoring data outside the points positions, where many of these better drives slow down for safety reasons, it seems to give a mean posterior λ value closer to what the Formula One points system would suggest for these drivers.

There are some disadvantages too of course. The truncated model does not tell us much at all about drivers who don't regularly finish in the points positions. Some drivers did not finish inside the top 10 once out of all 19 races so our truncated model gives us absolutely no information about which of these drivers is a better driver. There were also some drivers, who only finished inside the top 10 once or twice and again the truncated model could barely differentiate these drivers. This is why our truncated model gave a very small probability of being able to calculate the final race in the correct order.

In conclusion, it would appear that if you wanted to predict who is going to win, finish in a podium position, or finish in a top 10 position, for future races then using the truncated version of the Plackett-Luce model we have used here would be appropriate and give more reliable results than the Plackett-Luce model (based on only the results of the last race). However if you wanted to rank all 23 drivers then we would use the original Plackett-Luce model as the truncated model can't differentiate between the drivers who finish below 10th regularly.

Chapter 4

Sequential Modelling

4.1 Introduction

In this chapter we will analyse races sequentially to predict an outcome for the following race. For example, we will analyse the data from the first race to predict the outcome of the second race. Then we will analyse the data for the first two races to predict the outcome of the third race, and we will continue to do this to until we analyse the first 18 races to predict the outcome of the final race. In particular we will compare the sequential predictions under the two models we have looked at so far. We start by looking at the original Plackett-Luce model. After each race we run the Gibbs sampler to obtain a new posterior distribution based on the data so far. The usual checks on convergence did not highlight any problems.

4.2 Plackett-Luce model

From Figure 4.1 we can see how each driver's posterior mean for λ changes with each resulting race. It is hard to get much information from this graph as it is so tight, so we shall separate certain drivers to find more information.

4.2.1 Top four drivers

We see from Figure 4.2 which plots the sequential posterior means for the top four drivers, that all four drivers start with a similar mean λ value as all four drivers performed well in the first race (coming 3rd, 2nd, 5th and 6th position). Then after analysing the first two races we see the first major change. Vettel, Webber and Hamilton had very strong races, coming 1st, 2nd and 3rd respectively, whereas Alonso had his worst performance of the season finishing in last position, so we see that his mean value of λ drops massively and much below the other three drivers. Alonso then performs well consistently for the next 13 races, where his worst position in that time is 8th. We see his mean λ value growing slowly each time and he eventually overtakes both Webber and Hamilton.

Webber's plot takes a similar trend to Alonso's for most of the season, as one poor result

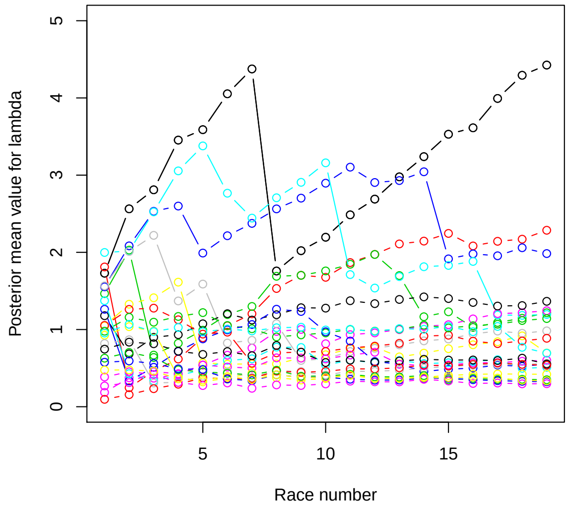


Figure 4.1: Plot of all 23 drivers changing posterior mean values of λ over the 19 races in the 2013 season

(race 3 in China where he finished 20th) drags his mean value of λ down significantly. He, again similar to Alonso, performs well consistently for a large number of races which slowly increases his λ scores race by race, but, unfortunately for himself, he seems to have a poor 4 race spell (finishing 15th, 21st, 2nd and 20th) which almost halves his posterior mean value of λ so sees him finish with a lower score than we would expect.

Hamilton's plot is increasing until one small dip (upon finishing 12th in Spain, race 5), but then continues increasing. However after finishing 20th (in Japan, race 15) he takes a large drop which takes him below Alonso, where he stays for the remaining 4 races. It is an interesting fact that if Hamilton had finished slightly higher in the Japanese race, however still not in a points position, then he may well have finished with a higher posterior mean value for λ than Alonso, even though in the Formula One points system, Alonso finished with many more points that Hamilton did.

Vettel's plot increases rapidly as he started the season in fantastic form (finishing in the top 4 for the opening 7 races, including 3 1st places) however we see his posterior mean λ value drop massively from 4.38 to 1.76 in Vettel's only poor race of the season (finishing 21st in Great Britain, race 8). This takes his λ value below that of Hamilton's, even though he has almost double the amount of Formula One points for the season at this point. But Vettel shows this was just a one-off mistake and not a drop in form, as he finishes 1st place in 10 of the last 11 races. Interestingly though, he only overtakes Hamilton at race 13, where Vettel has already accumulated over 100 more points than Hamilton has at this point. Clearly these facts show that these results for the Plackett-Luce model are not perfect, as Vettel should surely already have a much larger posterior mean value for λ at an earlier stage as he was so dominant.

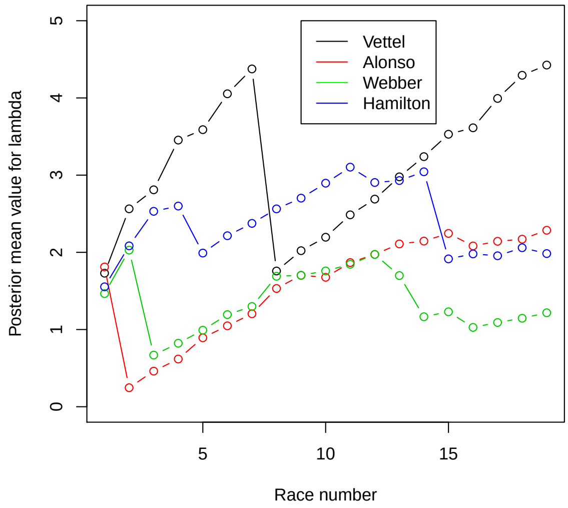


Figure 4.2: Plot of top 4 drivers changing posterior mean values of λ over the 19 races in the 2013 season

4.2.2 Other drivers

We shall now look at some results for some other drivers. In Figure 4.3 we have picked out drivers Grosjean and Kovalainen who finished in positions 7th and 21st in the 2013 season of Formula One. These drivers have been chosen as they both have interesting

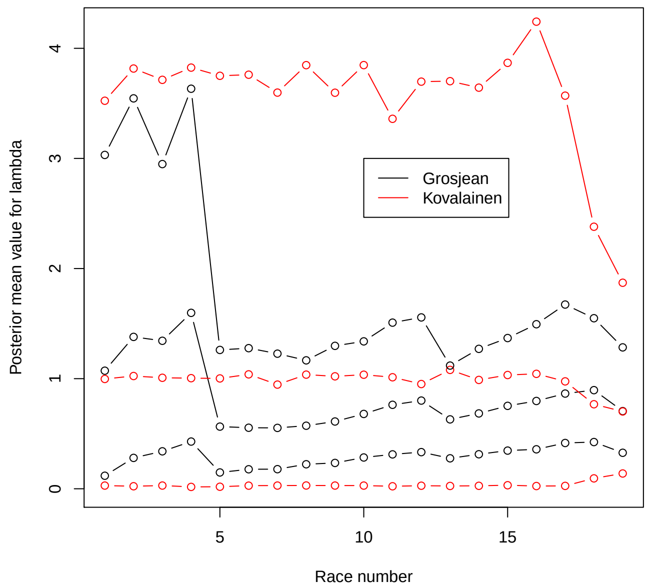


Figure 4.3: Plot of Grosjean and Kovalainen's changing posterior mean values of λ over the 19 races in the 2013 season, with 95 % credible intervals

aspects which change their mean λ values. But what makes these drivers so interesting? Well Grosjean finished in 7th position overall, yet his overall posterior mean value for λ is the same as the driver who finished 21st. Grosjean's mean λ is rising until race 5 (Spain where he finished in last position) where it then drops from 1.61 to 0.57. Grosjean never comes close to reaching this high again and slowly increases to 0.77 before another poor race (21st in Singapore, race 13) drops his posterior mean down to 0.65 and then again increases before finishing last in the last race leaving him on an overall posterior mean value of λ of 0.70. Now compare Grosjean to Kovalainen, who does not participate in the first 17 races so his plot stays around 1, which is the mean of the prior distribution, before coming 14th in both of the last two races to drop his mean λ value to 0.69. The upper bound for Kovalainen's 95% interval is very interesting as it is approximately 3.8 for the first 17 races and then drops massively once he participates in the last two races. Also his lower bound is almost zero throughout the entire season so clearly Kovalainen's results have a large amount of variability.

4.3 Truncated Plackett-Luce

From Figure 4.4 we can see how each driver's mean λ changes with each resulting race. We will now look at the same results we looked at for the original Plackett-Luce model

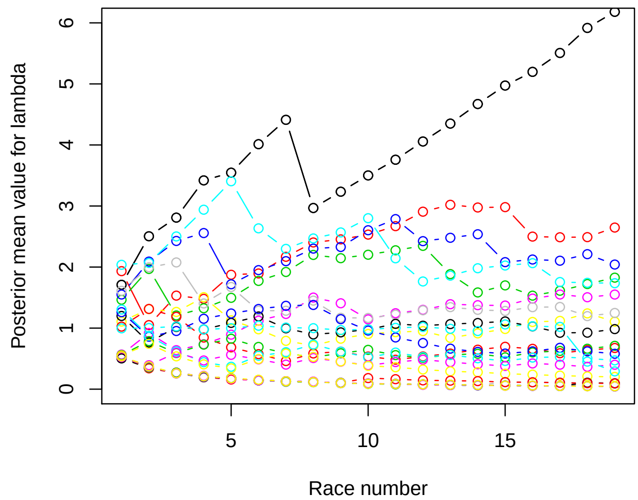
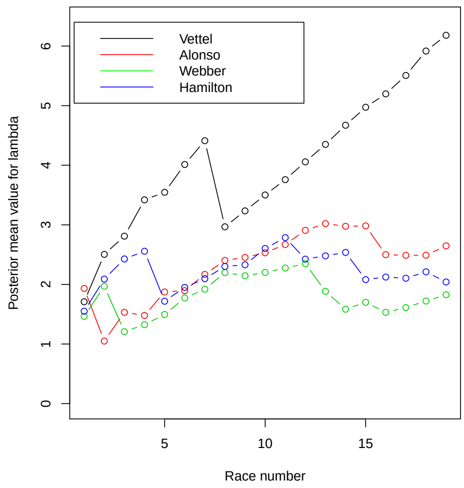


Figure 4.4: Plot of all 23 drivers' changing posterior mean values of lambda over the 19 races in the 2013 season for the truncated Plackett-Luce model

to see the differences between the two models.

We can compare the top 4 drivers using the two different models, see Figure 4.5. On the truncated model, we can see that the race Vettel finished 21st does not affect his posterior mean value for λ as heavily as it did using our original model, so Vettel maintains the largest λ values from race 2 throughout the rest of the season.



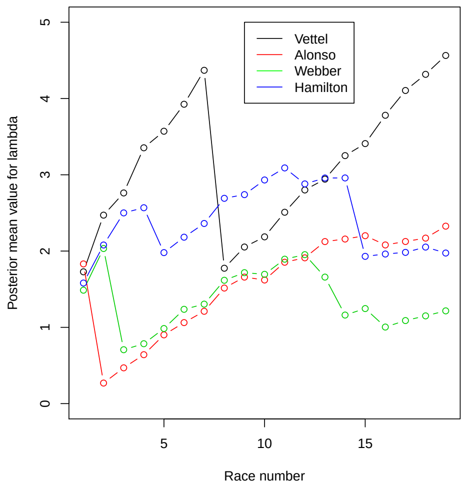


Figure 4.5: Plot of top 4 drivers changing posterior mean values of lambda over the 19 races in the 2013 season. Left: Truncated model. Right: Original Plackett-Luce model

On the truncated model, we see that when a driver has an excellent or a terrible race, it doesn't seem to affect the overall mean as much as it does for the original model. Alonso, Webber and Hamilton all finish with slightly higher mean values of λ but in the same order in the truncated model. This further shows that the truncated model would be a good model for analysing the better drivers in Formula One data as having a one-off poor race doesn't affect his overall posterior mean as dramatically so can still give an accurate score for the driver.

Grosjean's results are quite interesting because for our original Plackett-Luce model his mean λ value dropped massively from race 4 and then steadily increases. However the truncated model does not take as much of a fall for his poor performance in race 5, but it slowly decreases until he has a score lower than the prior mean. However it appears the later races where Grosjean was consistently in the top 10 increase his mean lambda value more in the truncated model than it does the original model. Grosjean's 95% interval seems to be wider for the truncated model than it does for the original Plackett-Luce model.

Kovalainen's results are almost identical for the first 17 races as he did not participate so they are all draws from his prior distribution. However then we can see the difference not finishing inside the top 10 makes for these two models. Kovalainen finishes 14th in the final two races which slightly decreases his posterior mean value of λ in the original model, however in the truncated model this is just recorded as finishing outside the top 10 in both races, and so his λ value drops massively. This is where the truncated model can give misleading results as Kovalainen's mean value would have dropped massively whether he had finished 11th in both races or 22nd. Kovalainen's upper 95% limit has a peak at over 4.0 but drops way below Grosjean's upper bound for the truncated model, however it remains larger for the original model.

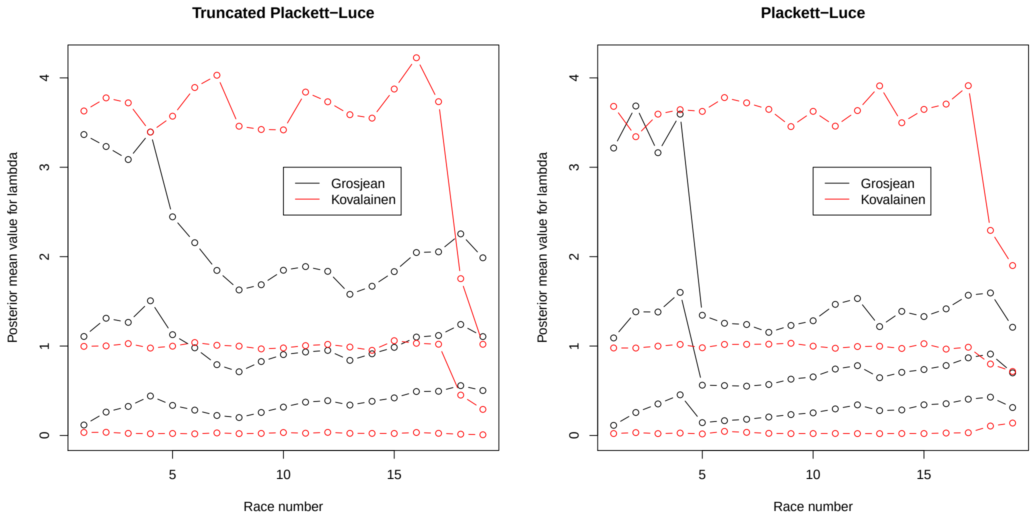


Figure 4.6: Plot of Grosjean and Kovalainen's changing posterior mean values of λ over the 19 races in the 2013 season, with 95 % credible intervals. Left: Truncated model. Right: Original Plackett-Luce model

4.4 Truncated model vs Original model

Looking at data sequentially, we shall look at the probability of the actual results occuring for the winner, finishing in a podium position and finishing in a points position for both our Plackett-Luce model and our truncated Plackett-Luce model.

Figure 4.7 shows the probability that the real life winner of each race, won the race based on analysing all the races that happened prior to that race, for both the Plackett-Luce model and for the truncated Plackett-Luce model. For example, the first point is the probability of actual winner (Vettel) winning the second race, based on the data from the first race. We see that the probabilities are close for the first 9 points for the two models, but then the last 9 points (in which Vettel won each race) the truncated model has a higher probability of correctly choosing the winner (Vettel). For most of the season, the truncated model gives a better predictive probability for the winner than the original Plackett-Luce model.

Figure 4.8 looks at the probabilities of correctly predicting the podium finishes for each race by sequential analysis, for the two models. The probability under the truncated model seems to be higher for the vast majority of races, especially in the later races where it becomes much greater. The truncated model seems much better for predicting the podium positions when the better drivers finish in these positions, for example the final race where the top 3 drivers (Vettel, Alonso, Webber) finish in the podium positions and the truncated model predicts this with a 0.58% chance whereas the original model predicts this outcome with a probability of just 0.17%.

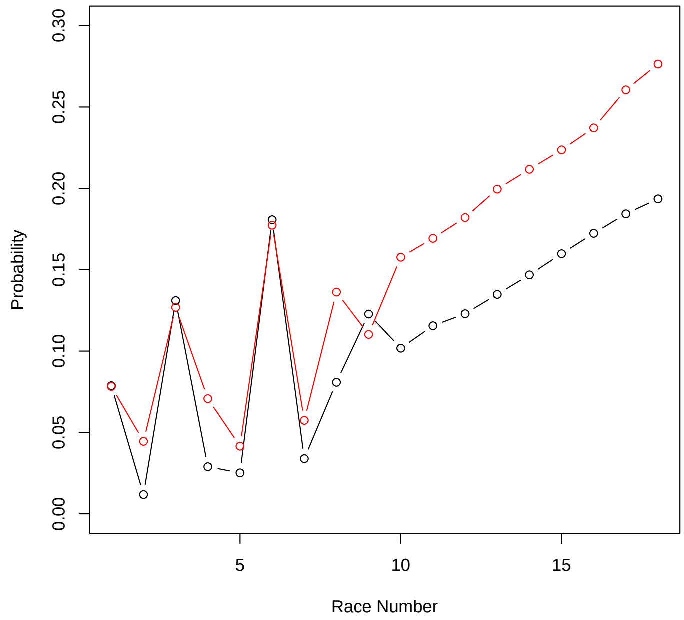


Figure 4.7: Plot of the predictive probability of the real life winner of each race occuring by analysing data sequentially. Plackett-Luce in black. Truncated Plackett-Luce in red

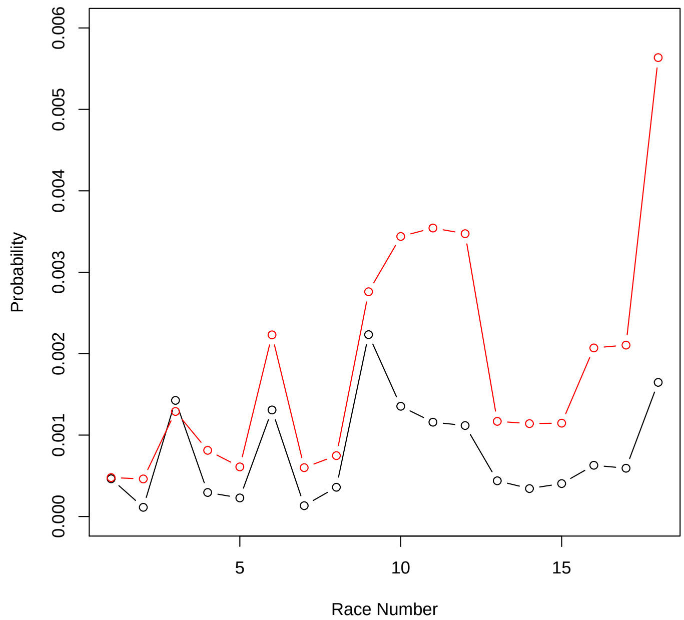


Figure 4.8: Plot of the predictive probability of the real life podium positions of each race occuring by analysing data sequentially. Plackett-Luce in black. Truncated Plackett-Luce in red

Figure 4.9 looks at the probabilities of correctly predicting the points finishes for each race by sequential analysis, for the two models. For most races these probabilities are

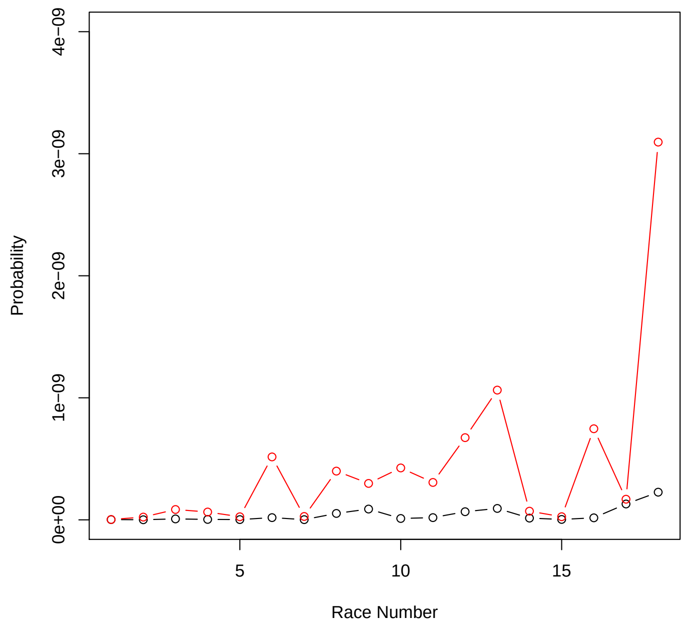


Figure 4.9: Plot of the predictive probability of the real life points positions of each race occuring by analysing data sequentially. Plackett-Luce in black. Truncated Plackett-Luce in red

very small and don't tell us much information but again it is the truncated model which gives a larger probability of correctly predicting these outcomes.

We also looked at predicting the entire outcome of each race by analysing the data sequentially for each model (results not shown). Not surprisingly the original model generally provided a better prediction than the truncated model.

4.5 Summary

These results have further cemented our theory that the truncated model is better used to analyse data for the better drivers as a poor race doesn't affect the results as much. However the truncated model tells us very little about the drivers who regularly finish outside the top 10 so if we wanted information on those drivers then we would have to use our original Plackett-Luce model.

Chapter 5

Gamma Ranking Models

5.1 Introduction

In this chapter we explore some other models for data in the form of ranks and apply them to the Formula One data. Recall that in the Plackett-Luce model, the probability of the observed ranking (finishing order) is

$$p_{PL}(\rho_i|\lambda) \equiv \Pr(\rho_i|\lambda) = \prod_{j=1}^{p_i-1} \frac{\lambda_{\rho_{ij}}}{\sum_{m=j}^{p_i} \lambda_{\rho_{im}}}.$$

This probability can also be obtained by considering the order statistics of exponential latent variables. Let $Z_j \sim Exp(\lambda_j)$ for all j drivers in the ith race, then it can be shown that

$$p_{PL}(\rho_i|\lambda) \equiv \Pr(Z_{\rho_{i1}} < Z_{\rho_{i2}} < \cdots < Z_{\rho_{ip_i}}).$$

We can interpret the latent variables Z_j as the finishing times of the drivers in the race; the driver with the quickest time comes first, the driver with the next fastest time finishes second, and so on. These latent variables are different from the latent variables described in Chapter 1, but the resulting model for the ranks is equivalent.

5.2 Gamma latent variables

We could consider other distributions for the latent variables. For example, Stern (1990) proposed gamma latent variables for analysing ranking data. Specifically, let $Z_j \sim Gamma(r, \lambda_j)$ for all j drivers in the ith race, and define the probability of the observed ranking as

$$p_G(\rho_i|\lambda, r) \equiv \Pr(Z_{\rho_{i1}} < Z_{\rho_{i2}} < \cdots < Z_{\rho_{ip_i}}).$$

Clearly when $r = 1$, $Z_j \sim Gamma(1, \lambda_j) \equiv Exp(\lambda_j)$ and so the gamma latent variable model reduces to the Plackett-Luce model. For integer $r > 1$, Stern (1990) derived an analytic expression for the likelihood $p_G(\rho_i|\lambda, r)$ but the formula is very complicated and is computationally very demanding for r greater than about 2 or 3. However, we can use the latent variable representation to construct a Gibbs sampler to sample from the posterior distribution of λ for fixed r.

The model is

$$\begin{aligned}
\lambda_k|a,b &\sim \text{Gamma}(a,b), \quad k=1,2,\ldots,K,\\
Z_{ij}|\lambda_j &\sim \text{Gamma}(r,\lambda_j), \qquad i=1,2,\ldots,n, \quad j=1,\ldots,p_i,\\
\rho_i|Z_i &= \text{order}(Z_i),
\end{aligned}$$

where $Z_i = (Z_{i1},\ldots,Z_{ip_i})$ and order() is shorthand for a function which returns a permutation which rearranges its argument into ascending order, e.g. order((3,1,4,2))=(2,4,1,3).

This model can be fitted using rjags (Plummer, 2014). Rjags is an R library which links to the program JAGS (Just Another Gibbs Sampler) which samples automatically from the posterior distribution of a specified model by using MCMC (typically Gibbs sampling). The rjags code for the model is given in Appendix D. We were able to sample 100,000 values of λ, for each driver from their posterior distribution. We have a burn in period; to determine how many iterations are needed before the Gibbs sampler has reached its stationary distribution, of 10000. Also we have thinned by only taking every 10th iterate to obtain a posterior sample whose autocorrelations are roughly zero.

In JAGS, the first half of the update iterations are used to optimally tune the MCMC scheme and the second half is then run with these optimized tuning parameters. After using R to calculate the effective sample size for each parameter which we want to be close to the length of the converged chain as that means they're very low autocorrelations.

Convergence can be assessed further by running multiple chains from different starting points. We have used three chains from different starting points which are all drawn randomly from the Gamma prior distribution for λ. The trace plots (not included) were in the same region of parameter space with similar variation along the trace plot.

5.3 Models with different shape parameter

We were able to make it so that we can change the gamma shape parameter, r, with ease so we can see if the different models are better at explaining the Formula One results, for this we shall look at the predictive probabilities of a set of different values of r. We know that when r=1, our gamma ranking model is the same as our Plackett-Luce model, so we will look at r values ranging from 2 up to 20.

We will first look at plots of driver's finishing positions against their posterior mean value of λ for various r values. The first thing that stands out from Figure 5.1 is that Vettel (Driver 1) has the highest posterior mean of λ for each value of r, which is what we would expect. However it appears that as r increases, his posterior mean decreases, as does for most of the better drivers. Conversely the drivers who finished in the lower down positions posterior mean values of λ seem to slowly increase as r is increased. The range of posterior means for r=2 is 4.09 whereas for r=20 is 0.96. So even though the relative structure of higher finishing position receiving a higher posterior mean value of λ throughout all values of r holds, the values get much more condensed as you increase r.

Next we will look at the density plots for the top 4 finishing drivers for various values of r. We see quite clearly from Figure 5.2 that both Vettel's mean and variance drops every time that r is increased. The same can be said about Alonso and Hamilton but to

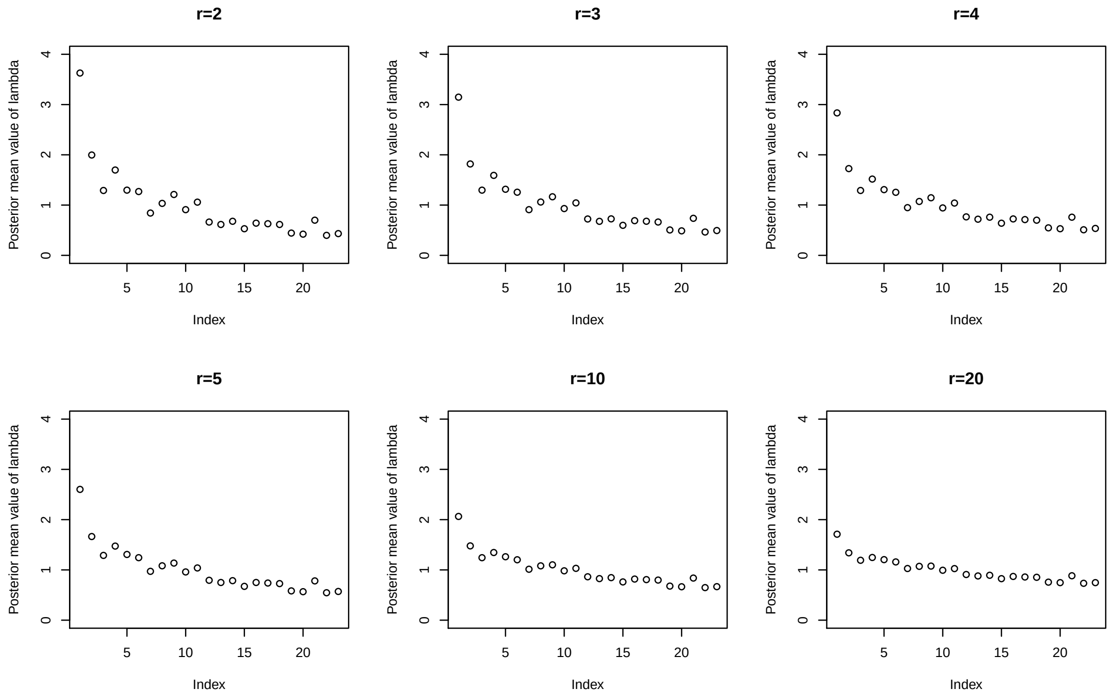


Figure 5.1: Posterior mean against finishing position for various values of r

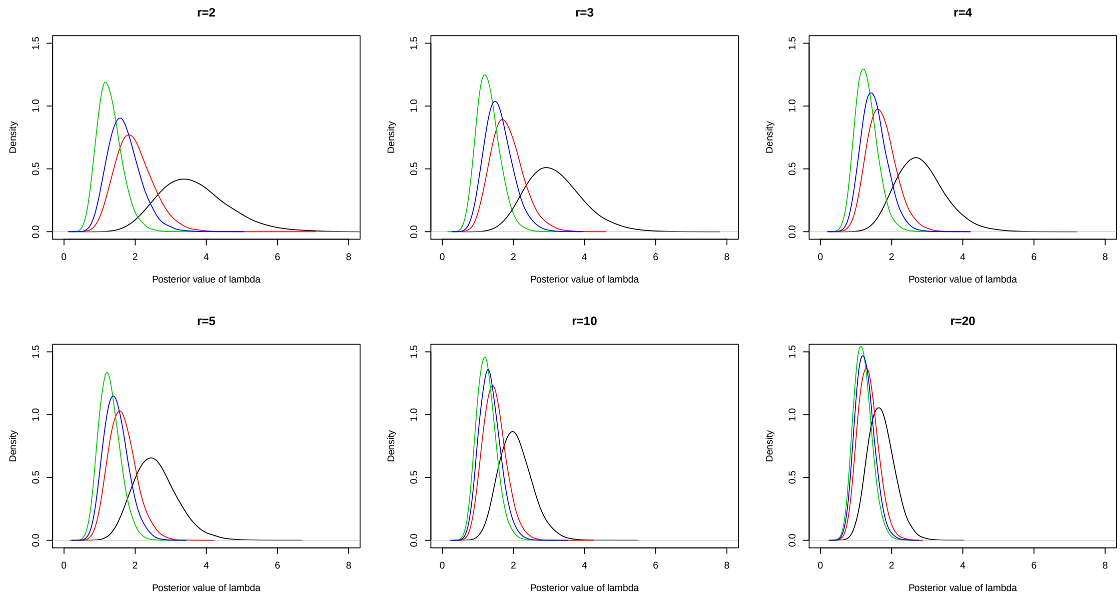


Figure 5.2: Posterior density plots for top 4 drivers. Vettel: Black, Alonso: Red, Webber: Green, Hamilton: Blue

a smaller degree. However Webber's posterior mean value of λ doesn't increase, it stays at a similar figure for each value of r; however the variance still seems to decrease.

Next we will look at plots of the probability of each driver winning, finishing in a podium position (top 3) or finishing in a points position (top 10) in the final (19th) race by simulating 10,000 hypothetical races, each race using a value of λ sampled from its posterior distribution. We use simulation here as there is no explicit formula for all of these probabilities (as there was in the Plackett-Luce model). This involves, at each iteration t, sampling the gamma latent variables for each driver from $Z_j^t \sim Gamma(r, \lambda_j^t)$ and then ordering the values from the smallest to the largest to determine the finishing positions, using the information from the first 18 races.

From Figure 5.3 we see the probability of each driver winning the final race of the 2013 Formula One season. The first thing that stands out is that Raikkonen has a probability

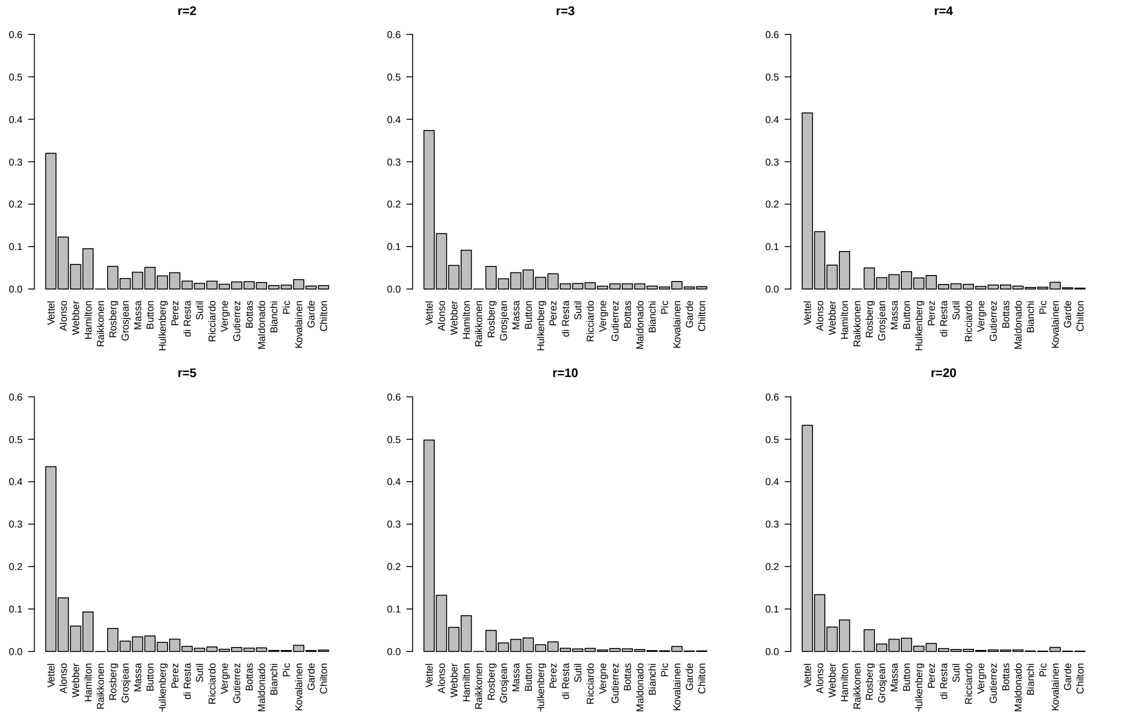


Figure 5.3: Probability of each driver winning the final race in the 2013 Formula One season, based on data on the previous 18 races, and Gamma ranking models with different shape parameters (r)

of zero, which he actually does for all the following plots, but this is because he did not participate in the final race, therefore has no chance of winning it. We see that as the shape parameters increases, the probability of Vettel, and some of the other better drivers, winning the final race increases. Also when r=2, every driver has a chance of winning, however when r increases, certain drivers start having a 0% chance of winning and by r=20, there are approximately a third of the drivers with 0% chance of winning the final race. So it appears that increasing the shape parameter increases the better driver's probabilities of winning the final race and decreases the probabilities of the worse drivers. There are other factors to consider though that these models can not take into consideration, such as the fact that Vettel had won the previous 8 Grand Prix and was hot favourite with the bookmakers to carry on that fantastic form. However this model just looks throughout the entire season and doesn't take form into consideration.

We shall now look to see if there is a similar effect to the probabilities of drivers finishing in a podium position in the final race of the 2013 Formula One season. We can see clearly from Figure 5.4 that the probabilities follow a similar pattern to that of Figure 5.3 in the

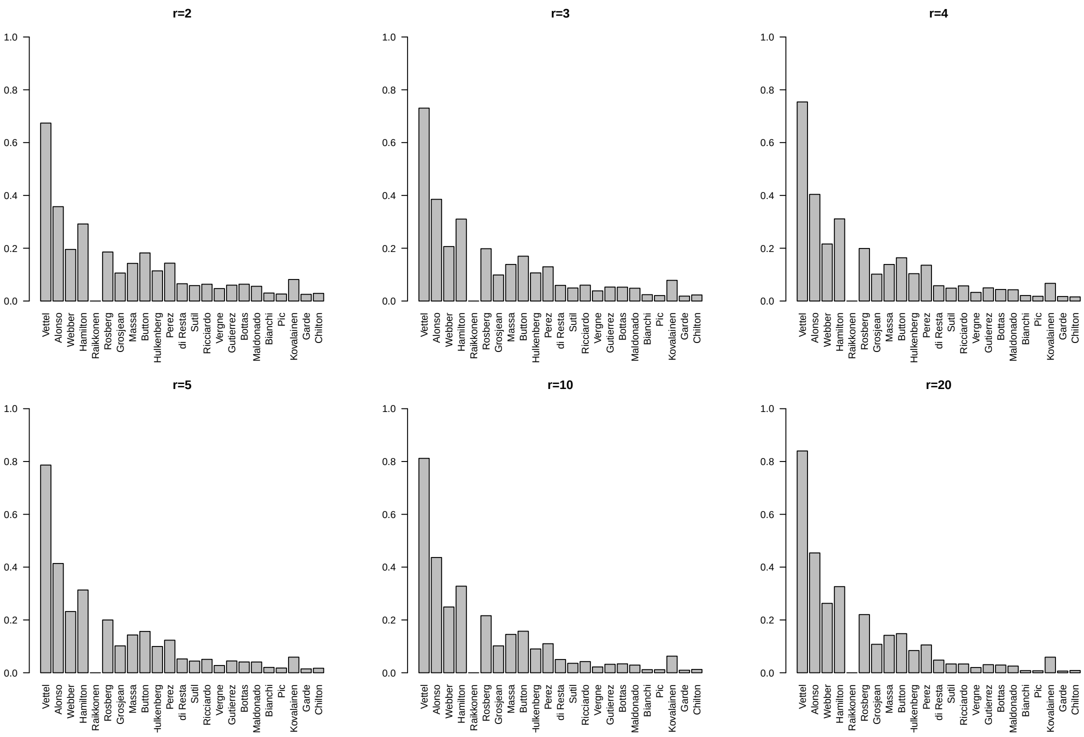


Figure 5.4: Probability of each driver finishing in a podium position in the final race in the 2013 Formula One season based on data on the previous 18 races, and Gamma ranking models with different shape parameters (r)

fact that as the shape parameter increases, the probability of the better drivers finishing in a podium position increases, and the probability of the worse drivers finishing in a podium position decreases. It is interesting to see that Kovalainen's (Driver 21) plot seems to stay fairly constant, even though all other drivers with similar overall finishing positions decrease as the shape parameter increases. This is even more evidence that Kovalainen seemed to be a better driver than the Formula One points system decided he was. Even though no drivers were given 0% chance of finishing in a podium position, some drivers were given very very small probabilties.

Next we will look at the probability of each driver finishing in a points position for the final race of the Formula One 2013 season. We see from Figure 5.5 that again as shape parameter increases, the probability of the better drivers finishing in the points position increases whereas the probability of the worse drivers decrease. Again Kovalainen's probability is much larger than his overall finishing position would suggest. There are again other factors that our model does not take into consideration, such as the fact that Bottas (Driver 17) had finished 8th (his season highest) in the previous race, so would have been full of confidence that he could retain his place in the points positions. Confidence is a big factor in any competitive sport and would have surely given Bottas a boost going into this final race.

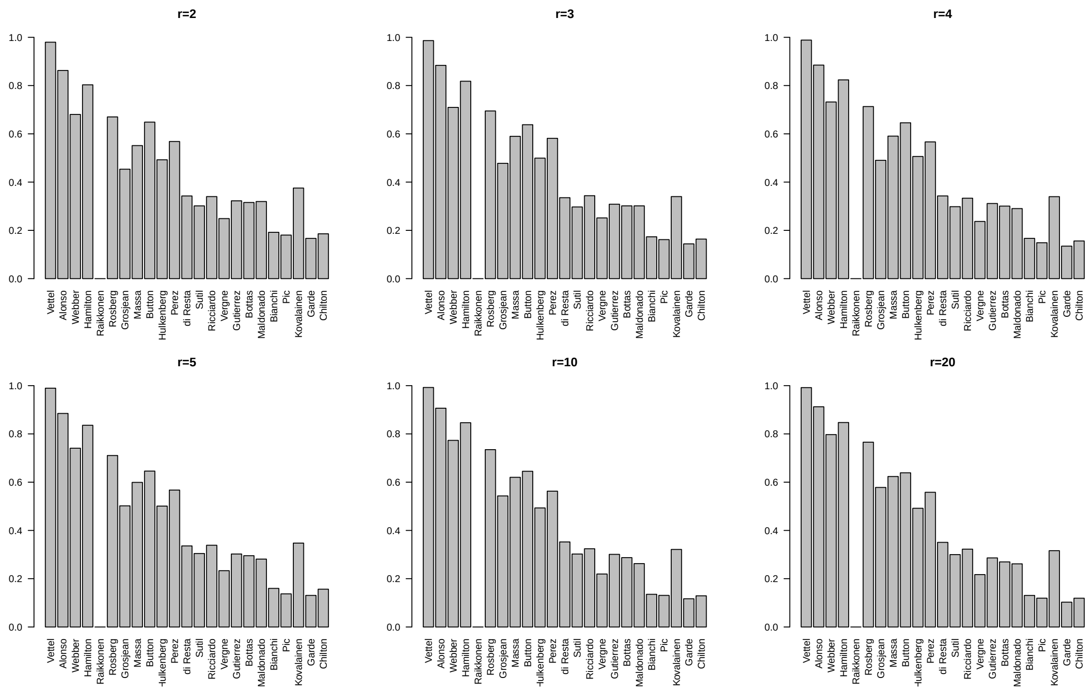


Figure 5.5: Probability of each driver finishing in a points position in the final race in the 2013 Formula One season based on data on the previous 18 races, and Gamma ranking models with different shape parameters (r)

	Plackett-Luce	Truncated	Gamma					
			r=2	r=3	r=4	r=5	r=10	r=20
P(Win)	0.1928	0.2767	0.3397	0.3839	0.4291	0.4396	0.4985	0.5271
P(Podium)	0.0017	0.0058	0.0041	0.0073	0.0066	0.0095	0.0125	0.0138
P(Points)	2.16×10^{-10}	2.99×10^{-09}						
P(Overall)	9.00×10^{-20}	1.92×10^{-21}						

Table 5.1: Comparison of predictive probabilities

5.4 Prediction

Using predictive simulations, we shall look at how each of the different models predicted the winner, the podium positions, the points positions and the entire 22 positions for the final race of the 2013 Season. From Table 5.1 we can see that the probability for correctly predicting the winner of the final race is greater for larger values of r. It seems that as r increases, the probability of correctly predicting the winner is directed correlated. The podium finish follows a similar pattern, with one exception where for r=4 has a lower probability for r=3, but if we ran for a larger number of iterations, we suspect that would change. We note that the truncated model is better at predicting the podium finish than the gamma model with r=2, but worse than predicting the winner.

Unfortunately we can not get a realistic prediction for the gamma ranking models for predicting the points positions or for every position in the final race. As there is such a large number of permutations for both of these (22!/12! and 22!, respectively) we unfortunately do not have the resources to run our rjags model for as many iterations as we would need to get an accurate probability.

5.5 Summary

The gamma ranking models seem to give us similar information as the truncated model. It seems that as we increase the shape parameter, r, the model distinguishes more between the better and the worse drivers, As we increase r, our model increases the probability of the good drivers winning whereas it decreases the chances. This is similar to what the truncated model did for us, however the gamma ranking models does it to a larger extent, for example giving us a 52.71% chance of correctly predicting the winner of the last race compared to the truncated model's probability of 27.67%.

The gamma ranking model predicts the podium places correctly with the highest probability too, however it is unfortunate we can not predict the probability of a points finish and the entire 22 positions using these models. We would expect the gamma models to be better at predicting the points finish than our original Plackett-Luce model however worse at predicting the entire 22 positions as the gamma models, especially for large r, give the worse drivers very very small probabilities so it would be hard to distinguish between these worse drivers. This is one of the major advantages of the Plackett-Luce model that we can get accurate predictive probabilities for these events, whereas this is difficult for the gamma models using rjags.

Chapter 6

Conclusions

When starting this report one of our main objectives was trying to analyse data from the Formula One 2013 season to find who was the best driver. We have used many different models and we find for every case that Vettel was a deserved champion. The Plackett-Luce model, truncated model and the Gamma ranking models with different shape parameters, all predict Vettel as the strongest driver.

There were some other interesting results though; take Kovalainen for example. The Formula One points system ranked him as 21st position overall, however he only participated in the last two races, and then all of our models used throughout this report had different rankings for him. The truncated model also ranked him quite low down, as the 15th best driver, whereas the Plackett-Luce model ranked him as the 11th best driver. The gamma ranking models give him the 12th largest probability of winning, coming in a podium position or finishing in a points position out of all the drivers. The worst any of our models predicts Kovalainen is the 15th best driver, however he was not given a chance to drive in the Formula One 2014 season, but drivers such as Chilton (23rd position in 2013), Bianchi (19th position in 2013), Maldonado (18th position in 2013), Gutierrez (16th position in 2013) and Vergne (15th position in 2013) were all given the opportunity to participate in the 2014 Formula One season. If we were to just look at data for the 2013 season, we would have expected Kovalainen to have been given a chance in the 2014 season.

Another driver with interesting results is Webber, who finished 3rd overall in the 2013 Formula One season. However none of our models predict him to finish so high, with the Plackett-Luce predicting him as the 6th best driver, the truncated Plackett-Luce predicting him as 4th best driver and the Gamma ranking models giving him the 4th highest probability of winning, finishing in a podium position and finishing in a points position out of all the drivers. So it seems here that the Plackett-Luce model ranks Webber lower than the other models and lower than his actual finishing position. Webber had 4 races where he finished outside the points positions and he finished very low, this may be down to driving more safely to protect himself or the car, or because his team mate Vettel was already winning the race. Whatever the reason, this is an example of a disadvantage of using the Plackett-Luce model, as a couple of poor results can change the overall results dramatically, especially for the better drivers.

There is no clear way of deciding which model is the best and there are advantages and disadvantages of all the models. Each model is useful for different aspects. For example if you wanted to predict the outcome of an entire race, then the Plackett-Luce model would most likely be the most appropriate model to choose as it takes into account all 22 drivers in each race. However if you were trying to calculate who will win or finish in a podium position then it seems using the gamma ranking models with a larger shape parameter would be more appropriate. If you wanted to find out the probability of a points finish, then the truncated Plackett-Luce model seems like the most ideal model as you can easily determined such a probability unlike the gamma ranking models, but it gives a larger probability of correctly predicting the points places than the original Plackett-Luce model.

6.1 Future Work

Formula One has added a new points system for the 2014 season where the final race of each season will be worth double points. This could be interesting to go over past seasons and see how this could have effected previous years' results. It certainly wouldn't have changed the fact that Vettel won the 2013 season, in fact it would have made his win by a larger margin, but in seasons previous to this it would have in fact changed the driver who won the season.

It would also be interesting to look over data for more than one season. Earlier we spoke about Kovalainen, suggesting that he deserved a place in the 2014 season from his results in the 2013 season, but when you look at his results from the previous three seasons before 2013, you can understand why he wasn't given the opportunity to participate in the 2014 season. So for looking over data for more than one season, we could look at information for who is the greatest driver in recent history? How does age affect a driver's performance and at what age do drivers generally hit their peak?

There are also other models we could have looked into. In a paper by Graves et al (2003), they suggest that when calculating the likelihood for drivers, that we do so in reverse order. In our example from Chapter 1 where 4 drivers participated in a race finished in the order 3,1,2,4; instead of calculating

$$\Pr(\rho_1 \mid \lambda) = \frac{\lambda_3}{\lambda_3+\lambda_1+\lambda_2+\lambda_4} \times \frac{\lambda_1}{\lambda_1+\lambda_2+\lambda_4} \times \frac{\lambda_2}{\lambda_2+\lambda_4}$$

we would calculate

$$\Pr(\rho_1 \mid \lambda) = \frac{\lambda_4}{\lambda_4+\lambda_2+\lambda_1+\lambda_3} \times \frac{\lambda_2}{\lambda_2+\lambda_1+\lambda_3} \times \frac{\lambda_1}{\lambda_1+\lambda_3}$$

which Graves et al (2003) claim distinguishes more between the better and worse drivers, and drivers' scores don't get as affected by one or two poor results.

We could also have looked at assigning the shape parameter r for the gamma ranking models a prior distribution and sampling from its posterior in rjags, to see which value of r gives the optimum probability of predicting the winner, points and podium positions of races.

References

Brooks, S. P, Gelman, A., Jones, G. L. and Meng, X.-L. (2011) *Handbook of Markov Chain Monte Carlo*. Chapman & Hall/CRC.

Caron, F. and Doucet, A. (2012) Efficient Bayesian inference for generalized Bradley-Terry models. *Journal of Computational and Graphical Statistics*, **21**, 174–196.

Gelman, A., Carlin, J. B., Stern, H. S. and Rubin, D. B. (2004) *Bayesian Data Analysis*. Second Edition. Chapman & Hall/CRC.

Graves, T., Reese, C. S. and Fitzgerald, M. (2003) Hierarchical models for permutations: analysis of auto racing results. *Journal of the American Statistical Association*, **98**, 282–291.

Luce, R. D. (1959) *Individual Choice Behavior: A Theoretical Analysis*. New York: Wiley & Sons.

Plackett, R. L. (1975) The analysis of permutations. *Applied Statistics*, **24**, 193–202.

Plummer, M. (2014) *rjags: Bayesian graphical models using MCMC.* `http://CRAN.R-project.org/package=rjags`

R Development Core Team (2012) *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing. `http://www.R-project.org`

Stern, H. S. (1990) Models for distributions on permutations. *Journal of the American Statistical Association*, **85**, 558–564.

Appendix A

Formula One 2013 Season

* Results

Pos	Driver	AUS	MAL	CHN	BHR	ESP	MON	CAN	GBR	GER	HUN	BEL	ITA	SIN	KOR	JPN	IND	ABU	USA	BRA	Points
1	Sebastian Vettel	3	1	4	1	4	2	1	21	1	3	1	1	1	1	1	1	1	1	1	397
2	Fernando Alonso	2	22	1	8	1	7	2	3	4	5	2	2	2	6	4	11	5	5	3	242
3	Mark Webber	6	2	20	7	5	3	4	2	7	4	5	3	15	21	2	20	2	3	2	199
4	Lewis Hamilton	5	3	3	5	12	4	3	4	5	1	3	9	5	5	20	6	7	4	9	189
5	Kimi Raikkonen	1	7	2	2	2	10	9	5	2	2	21	11	3	2	5	7	22			183
6	Nico Rosberg	20	4	19	9	6	1	5	1	9	19	4	6	4	7	8	2	3	9	5	171
7	Romain Grosjean	10	6	9	3	22	17	13	19	3	6	8	8	21	3	3	3	4	2	22	132
8	Felipe Massa	4	5	6	15	3	21	8	6	22	8	7	4	6	9	10	4	8	12	7	112
9	Jenson Button	9	17	5	10	8	6	12	13	6	7	6	10	7	8	9	14	12	10	4	73
10	Nico Hulkenberg	DNS	8	10	12	15	11	21	10	10	11	13	5	9	4	6	19	14	6	8	51
11	Sergio Perez	11	9	11	6	9	16	11	20	8	9	11	12	8	10	15	5	9	7	6	49
12	Paul di Resta	8	21	8	4	7	9	7	9	11	18	20	22	20	22	11	8	6	15	11	48
13	Adrian Sutil	7	20	21	13	13	5	10	7	13	22	9	16	10	20	14	9	10	22	13	29
14	Daniel Ricciardo	19	18	7	16	10	18	15	8	12	13	10	7	22	19	13	10	16	11	10	20
15	Jean-Eric Vergne	12	10	12	22	20	8	6	22	20	12	12	21	14	18	12	13	17	16	15	13
16	Esteban Gutierrez	13	12	22	18	11	13	20	14	14	20	14	13	12	11	7	15	13	13	12	6
17	Valtteri Bottas	14	11	13	14	16	12	14	12	16	21	15	15	13	12	17	16	15	8	21	4
18	Pastor Maldonado	21	19	14	11	14	20	16	11	15	10	17	14	11	13	16	12	11	17	16	1
19	Jules Bianchi	15	13	15	19	18	19	17	16	21	16	18	19	18	16	22	18	20	18	17	0
20	Charles Pic	16	14	16	17	17	22	18	15	17	15	22	17	19	14	18	21	19	20	20	0
21	Heikki Kovalainen																		14	14	0
22	Giedo van der Garde	18	15	18	21	21	15	22	18	18	14	16	18	16	15	21	22	18	19	18	0
23	Max Chilton	17	16	17	20	19	14	19	17	19	17	19	20	17	17	19	17	21	21	19	0

* Format of data in R and rjags

```
Driver<-c("Vettel","Alonso","Webber","Hamilton","Raikkonen","Rosberg","Grosjean","Massa","Button","Hulkenberg","Perez","di Resta",
"Sutil","Ricciardo","Vergne","Gutierrez","Bottas","Maldonado","Bianchi","Pic","Kovalainen","Garde","Chilton")
AUS<-c(3,2,6,5,1,20,10,4,9,0,11,8,7,19,12,13,14,21,15,16,0,18,17);rho.1<-order(AUS)[-c(1,2)]
MAL<-c(1,22,2,3,7,4,6,5,17,8,9,21,20,18,10,12,11,19,13,14,0,15,16);rho.2<-order(MAL)[-1]
CHN<-c(4,1,20,3,2,19,9,6,5,10,11,8,21,7,12,22,13,14,15,16,0,18,17);rho.3<-order(CHN)[-1]
BHR<-c(1,8,7,5,2,9,3,15,10,12,6,4,13,16,22,18,14,11,19,17,0,21,20);rho.4<-order(BHR)[-1]
ESP<-c(4,1,5,12,2,6,22,3,8,15,9,7,13,10,20,11,16,14,18,17,0,21,19);rho.5<-order(ESP)[-1]
MON<-c(2,7,3,4,10,1,17,21,6,11,16,9,5,18,8,13,12,20,19,22,0,15,14);rho.6<-order(MON)[-1]
CAN<-c(1,2,4,3,9,5,13,8,12,21,11,7,10,15,6,20,14,16,17,18,0,22,19);rho.7<-order(CAN)[-1]
GBR<-c(21,3,2,4,5,1,19,6,13,10,20,9,7,8,22,14,12,11,16,15,0,18,17);rho.8<-order(GBR)[-1]
GER<-c(1,4,7,5,2,9,3,22,6,10,8,11,13,12,20,14,16,15,21,17,0,18,19);rho.9<-order(GER)[-1]
HUN<-c(3,5,4,1,2,19,6,8,7,11,9,18,22,13,12,20,21,10,16,15,0,14,17);rho.10<-order(HUN)[-1]
BEL<-c(1,2,5,3,21,4,8,7,6,13,11,20,9,10,12,14,15,17,18,22,0,16,19);rho.11<-order(BEL)[-1]
ITA<-c(1,2,3,9,11,6,8,4,10,5,12,22,16,7,21,13,15,14,19,17,0,18,20);rho.12<-order(ITA)[-1]
SIN<-c(1,2,15,5,3,4,21,6,7,9,8,20,10,22,14,12,13,11,18,19,0,16,17);rho.13<-order(SIN)[-1]
KOR<-c(1,6,21,5,2,7,3,9,8,4,10,22,20,19,18,11,12,13,16,14,0,15,17);rho.14<-order(KOR)[-1]
JPN<-c(1,4,2,20,5,8,3,10,9,6,15,11,14,13,12,7,17,16,22,18,0,21,19);rho.15<-order(JPN)[-1]
IND<-c(1,11,20,6,7,2,3,4,14,19,5,8,9,10,13,15,16,12,18,21,0,22,17);rho.16<-order(IND)[-1]
ABU<-c(1,5,2,7,22,3,4,8,12,14,9,6,10,16,17,13,15,11,20,19,0,18,21);rho.17<-order(ABU)[-1]
USA<-c(1,5,3,4,0,9,2,12,10,6,7,15,22,11,16,13,8,17,18,20,14,19,21);rho.18<-order(USA)[-1]
BRA<-c(1,3,2,9,0,5,22,7,4,8,6,11,13,10,15,12,21,16,17,20,14,18,19);rho.19<-order(BRA)[-1]
rho<-list()
rho[[1]]<-rho.1;rho[[2]]<-rho.2;rho[[3]]<-rho.3;rho[[4]]<-rho.4;rho[[5]]<-rho.5;rho[[6]]<-rho.6;rho[[7]]<-rho.7;
rho[[8]]<-rho.8;rho[[9]]<-rho.9;rho[[10]]<-rho.10;rho[[11]]<-rho.11;rho[[12]]<-rho.12;rho[[13]]<-rho.13;
rho[[14]]<-rho.14;rho[[15]]<-rho.15;rho[[16]]<-rho.16;rho[[17]]<-rho.17;rho[[18]]<-rho.18;rho[[19]]<-rho.19
```

Appendix B

Gibbs sampler for Plackett-Luce model

```
PL.gibbs <- function(rho,a=1,b=1,its=1000,lambda.init=rgamma(max(as.vector(unlist(rho))),a,b))
{
  ## number of individuals
  K <- max(as.vector(unlist(rho)))
  ## number of multiple comparisons
  n <- length(rho)
  ## number of individuals in each multiple comparison
  p <- rep(0,n)
  for(i in 1:n){
    p[i] <- length(rho[[i]])
  }
  ## identify individual in last place
  ## and count up times for each individual
  last <- rep(0,n)
  participated <- rep(0,K)
  wlast <- rep(0,K)
  for(i in 1:n){
    participated[rho[[i]]] <- participated[rho[[i]]]+1
    last[i] <- rho[[i]][p[i]]
    wlast[last[i]] <- wlast[last[i]]+1
  }
  w <- participated-wlast
  ## compute delta
  delta <- array(0,c(n,K-1,K))
  for(k in 1:K){
    for(i in 1:n){
      for(j in 1:(p[i]-1)){
        delta[i,j,k] <- sum(k==rho[[i]][j:p[i]])
      }
    }
  }
  ## storage for samples
  Z.res <- array(0,c(its,n,K-1))
  lambda.res <- matrix(0,nrow=its,ncol=K)
  ll.res <- rep(0,its)

 ## initial values
  lambda.curr <- lambda.init
  Z.curr <- matrix(0,nrow=n,ncol=(K-1))
  ## Markov chain sampling
  for(t in 1:its){
    print(t)
    ## sample latent variables | everything else
    for(i in 1:n){
      for(j in 1:(p[i]-1)){
        Z.curr[i,j] <- rexp(1,sum(lambda.curr[rho[[i]]][j:p[i]]))
      }
    }
    ## sample parameters | everything else
    for(k in 1:K){
      lambda.curr[k] <- rgamma(1,a+w[k],b+sum(delta[,,k]*Z.curr))
    }
    ## normalize and rescale
    Lambda.curr <- rgamma(1,K*a,b)
    lambda.curr <- Lambda.curr*lambda.curr/sum(lambda.curr)
    ## store sampled values
    lambda.res[t,] <- lambda.curr
    Z.res[t,,] <- Z.curr
    ## compute observed data loglikelihood
    ll.res[t] <- loglike.pl(rho,lambda.res[t,])
    print(ll.res[t])
  }
  return(list(Z=Z.res,lambda=lambda.res,ll=ll.res))
}
```

Appendix C

Gibbs sampler for truncated Plackett-Luce model

```
PL.gibbs.10 <- function(rho,a=1,b=1,its=1000,lambda.init=rgamma(max(as.vector(unlist(rho))),a,b))
 {
   ## number of individuals
   K <- max(as.vector(unlist(rho)))
   ## number of multiple comparisons
   n <- length(rho)
   ## number of individuals in each multiple comparison
   p <- rep(0,n)
   for(i in 1:n){
     p[i] <- length(rho[[i]])
   }
   top10 <- array(0,c(n,10))
   w <- rep(0,max(as.vector(unlist(rho))));
   for(i in 1:n){
     for(j in 1:10){
       top10[i,j]<- rho[[i]][j]
       w[top10[i,j]] <- w[top10[i,j]]+1
     }
   }
   w
  ## compute delta
   delta <- array(0,c(n,K-1,K))
   for(k in 1:K){
     for(i in 1:n){
       for(j in 1:(p[i]-1)){
         delta[i,j,k] <- sum(k==rho[[i]][j:p[i]])
       }
     }
   }
   ## storage for samples
   Z.res <- array(0,c(its,n,K-1))
   lambda.res <- matrix(0,nrow=its,ncol=K)
   ll.res <- rep(0,its)

  ## initial values
   lambda.curr <- lambda.init
   Z.curr <- matrix(0,nrow=n,ncol=(K-1))
   ## Markov chain sampling
   for(t in 1:its){
     print(t)
     ## sample latent variables | everything else
     for(i in 1:n){
       for(j in 1:10){
         Z.curr[i,j] <- rexp(1,sum(lambda.curr[rho[[i]]][j:p[i]]))
       }
     }
     ## sample parameters | everything else
     for(k in 1:K){
       lambda.curr[k] <- rgamma(1,a+w[k],b+sum(delta[,1:10,k]*Z.curr[,1:10]))
     }
     ## normalize and rescale
     Lambda.curr <- rgamma(1,K*a,b)
     lambda.curr <- Lambda.curr*lambda.curr/sum(lambda.curr)
     ## store sampled values
     lambda.res[t,] <- lambda.curr
     Z.res[t,,] <- Z.curr
     ## compute observed data loglikelihood
     ll.res[t] <- loglike.pl(rho,lambda.res[t,])
     print(ll.res[t])
   }
   return(list(Z=Z.res,lambda=lambda.res,ll=ll.res))
 }
```

Appendix D

Gamma Ranking model

```
library(rjags)
K <- 23
x <- matrix(NA,nrow=n,ncol=22)
for(i in 1:n){
  x[i,1:length(rho[[i]])] <- rho[[i]]
}

y <- ifelse(is.na(x)==1,NA,1)

ind <- matrix(NA,nrow=n,ncol=22)
p <- rep(0,n)
for(i in 1:n){
  ind[i,1:length(rho[[i]])] <- 1:length(rho[[i]])
  p[i] <- length(rho[[i]])
}

#Can change r for different Gamma models
r <- 2
a <- 1
b <- 1
lambda.init1 <- rgamma(K,a,b)
Z.init1 <- matrix(NA,ncol=22,nrow=n)
lambda.init2 <- rgamma(K,a,b)
Z.init2 <- matrix(NA,ncol=22,nrow=n)
lambda.init3 <- rgamma(K,a,b)
Z.init3 <- matrix(NA,ncol=22,nrow=n)
for(i in 1:n){
  Z.init1[i,1:p[i]] <- sort(rgamma(p[i],r,lambda.init1))
Z.init2[i,1:p[i]] <- sort(rgamma(p[i],r,lambda.init2))
Z.init3[i,1:p[i]] <- sort(rgamma(p[i],r,lambda.init3))
}

data=list(x=x,y=y,n=n,p=p,K=K,ind=ind)
hyper=list(a=1,b=1,r=1)
init=list(list(lambda=lambda.init1,Z=Z.init1),list(lambda=lambda.init2,Z=Z.init2),list(lambda=lambda.init3,Z=Z.init3))

modelstring="
model{
  for(i in 1:n){
    for(j in 1:p[i]){
      lims[i,j,1] <- equals(j,1)*0 +inprod(Z[i,1:p[i]], equals(ind[i,1:p[i]],j-1))
      lims[i,j,2] <- equals(j,p[i])*(10e6) +inprod(Z[i,1:p[i]], equals(ind[i,1:p[i]],j+1))
      y[i,j]~dinterval(Z[i,j],lims[i,j,])
      Z[i,j]~dgamma(r,lambda[x[i,j]])
    }
  }
  for(k in 1:K){
    lambda[k]~dgamma(a,b)
  }
  Lambda ~ dgamma(K*a,b)
  for(k in 1:K){
    lambda.star[k] <- Lambda*lambda[k]/sum(lambda)
  }
}
"

hyper=list(a=a,b=b,r=r)
model=jags.model(textConnection(modelstring), data=append(data,hyper), inits=init,n.chains=3)
update(model,n.iter=10000)
output=coda.samples(model=model,variable.names=c("lambda.star"),n.iter=100000,thin=10)
plot(output)
autocorr.plot(output)
summary(output)
effectiveSize(output)
HPDinterval(output)
crosscorr(output)
```