Hierarchical models

Žiga Trojer 63200440

I. Introduction

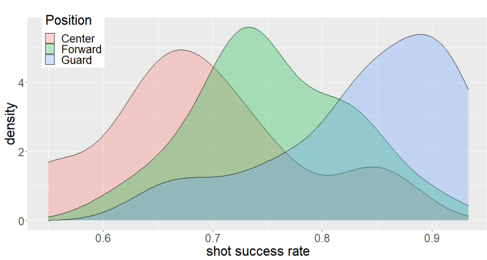
In this homework, we developed a hierarchical model that is appropriate for analyzing the provided basketball dataset. We were attempting to answer two main questions about basketball players' performance based on their position:

1) With what certainty can we claim that the best guard (Player #20) is on average a better free throw shooter than the best forward (Player #3)?
2) We take a random guard and a random forward, each of them shoots one free throw. What are the probabilities of the guard winning (guard makes the shot, forward misses the shot), the forward winning (guard misses the shot, forward makes the shot), and a tie (both players either miss the shot or hit the shot). What are the probabilities for other combinations?

II. Data and model description

We were given 5400 records describing basketball free shots from various player positions (guard, forward, center) and basketball rims. We filtered out data describing shots on the special rim size because we were only interested in free throw shots made on the traditional rim size, leaving us with 2700 records for 45 players (19 guard, 19 forward and 7 center players). We also discarded variables, describing the angle of a shot and the consecutive number of the shot out. There is no empty values. Figure 1 shows the distribution of shot success rate per player's position. We can observe that guard position players have a highest success rate of shots on average and center position players the lowest success rate.

Figure 1: Distribution of shot success rate per position.



We used a two-level hierarchical model. A Bernoulli distribution was used at the bottom level, as our target variable is binary. At the second level, a reparameterized Beta distribution was used, as it suits the nature of the problem. The formal description of the model is as following:

Hierarchical model

$$y_{player,k}|\theta_{player} \sim \text{Bernoulli}(\theta_{player}),$$

$$\theta_{player}|\mu,\tau \sim \text{Beta}(p,q),$$

where $p = \tau\mu$, $q = \tau(1-\mu)$. Reparametrization was done in the following way: $\mu = \frac{\alpha}{\alpha+\beta}$, $\tau = \alpha + \beta$, which stands for mean and sample size. Default priors were used.

We fit the model where we used 1000 iterations for the warm-up phase and 1000 for the sampling phase - for each chain. We performed posterior checks to ensure that the model sufficiently converged. First, we observed the $\hat{R}$ and the sample size for all θ's. They indicated that sampling was successful, which was also confirmed when observing the trace plots.

III. Results

We begin by presenting the results that support the answer to Section I's question 1). We calculated the probability that the player #20 (best guard) is on average a better free throw shooter than he player #3 (best forward): $P(\theta_{20} \geq \theta_3) = 0.897 \pm 0.005$[1]. This probability implies that the best guard is on average a better free throw shooter than the best forward player.

To be able to answer the question 2) from the Section I, we needed to conduct the following experiment: we chose two players at random for each of the two positions. For each, we randomly extracted probabilities of a shot being scored. Using those probabilities, we first calculated each of the following probabilities:

- A player from Group 1 scores, but a player from Group 2 fails to score.
- A player from Group 1 fails to score, but a player from Group 2 succeeds.
- Either both score or both miss.

We saved those probabilities and repeated the same experiment 1000 times. Results are shown in Table I. We then averaged the result and estimated the uncertainty.

Position 1	Position 2	First win	Tie	Second win
Center	Forward	0.168 ± 0.001	0.595 ± 0.001	0.236 ± 0.001
Center	Guard	0.111 ± 0.001	0.627 ± 0.002	0.262 ± 0.002
Forward	Guard	0.120 ± 0.001	0.674 ± 0.001	0.206 ± 0.002

Table I: Probabilities for each position combination: player from Position 1 wins, player from Position 2 wins, and a tie occurs. Standard error is provided.

IV. Conclusion

During the analysis, we found that Figure 1, which depicts the distribution of shot success rate per position, provided very similar insights. We'd venture to say that based on the distribution graph, the best player guard is almost certainly better than the best player forward on average. In terms of the second question, such conclusions would not be possible in the absence of modeling. When we compare the probabilities for a tie result, we see that the center-forward comparison have the lowest probability. Based on Figure 1, we would expect this to be the lowest in a center-guard comparison. It'd be interesting to see how much more information we'd get if we added another top level to the hierarchical model. We did not test this for the purposes of this homework, because the model had already produced quite nice results. As a result, we learned where to use hierarchical models and how useful they are.

[1]SE