STA531 Final Project

Eric Su

2019-04-26

Abstract

The performance level in men's professional tennis is key in evaluating abilities of players and predicting outcomes of matches. Currently, the most widely used metrics to assess performance are the rankings, number of titles and win percentages. However, these measurements are heavily influenced by a few major tournaments, and are also affected by the proportion of different court surfaces on the ATP tour. The goal of this project is to develop a reliable method that quantifies performance levels of professional tennis players without the aforementioned problems. We utilize a Bayesian network model with performance level as a hidden variable. Analyses are mainly focused on comparing performance levels of four dominant players in the past 10 years, the so-called *Big 4* of men's professional tennis, namely *Roger Federer, Rafael Nadal, Novak Djokovic* and *Andy Murray.*

Introduction

What influence the performance of a professional tennis player? Internal factors such as talent, age, mental strength and external factors such as tournament level, surface properties, sets played might all be essential. However, internal factors are often unobservable and difficult to utilize in practice. We want to find a method to quantify performance that is solely based on external factors. Such model would allow us to estimate effects of changing specific external variables.

Having the ability to estimate perfomance level of players given a set of conditions enables us to answer questions such as "What is the probability of a player winning a particular match?" or "How does a player's performance level change with respect to different tennis surfaces?". This could be useful for predicting match outcomes and comparing performance levels in hypothetical situations.

The most widely considered metrics in evaluating player performance include rankings, points, win percentages and number of major titles. These measurements are all related-rankings of players are determined by the amount of points they earned and dfferent levels of tournaments award different amount of points. For example, an ATP 500 series gives the champion 500 points, a Masters 1000 tournament gives the champion 1000 points, while each Grand Slam (the highest tournament level) awards the champion 2000 points.

Since the amount of points a player gains heavily depend on outcome of certain matches (Grand slam matches in particular), differences in points may be the result of random variations. Win percentage also has the problem of not taking the performance levels of opponents into account. Obviously, winning against tougher opponents is less likely than winning over weaker opponents, but neither points nor win percentage reflect this.

Consequently, we want a more reliable method to quantify players' performance levels. Ideally, the method should put more weight on matches against tougher opponents and it should also give us a number that can be used to compare players. In addition, we want our model to give us information on how each external factor affects a player's performance. The model we use is a Bayesian network model with performance level as a hidden variable. It quantifies performance level into a number and allows us to predict win probabilities of players winning.

The dataset we use is the official men's tennis ATP tour data from 2000 to 2016. Important variables for our project include Series (Tournament level), Court (Indoor or outdoor), Surface (Hard, Clay or Grass), Round (1^{st} round, 2^{nd} round, . . . , semifinal, finals), player names and match outcomes. The dataset can be found on the Kaggle website: https://www.kaggle.com/jordangoblet/atp-tour-20002016

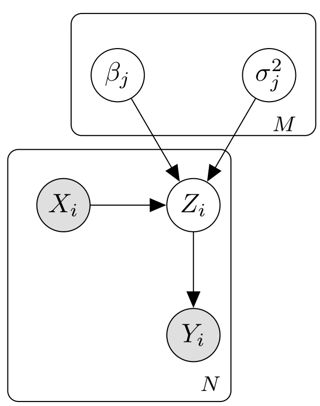
Previous work

Little academic work has been done on evaluating the performance of professional tennis players. Shang-Min Ma et al. (2013) analyzed variables related to tennis performance using a logistic regression model. However, their study included match statistics such as number of aces and percentage of 1st serves won. These variables can only be observed after the match and thus are not useful for prediction. Other researchers have tried to quantify player performance using accumulated matches played and won at various tournament levels (Asmita Chitnisa and Omkarprasad Vaidya, 2014) or using direct match statistics (Filipcic Ales et al., 2015). The result of these studies can not be used to compare potential performance level when particular external factors are altered, since they did not seperate the external factors and match specific factors.

Methods

To evaluate performance levels of different players, we must estimate player specific parameters such as the effect of playing on clay for different players. Since there are over 1200 players in the dataset and many of them only played in a few matches, we run into the issue of not having enough data for these players. As a result, we filtered out all tennis players who played in less than 100 matches and group them into a single catagory called "Other". This catagory will be used as a benchmark group which represents the average players.

We use a Bayesian network model with perfomance level as a hidden variable to represent how variables affect one another. Our hypothesis is that each player exhibits a certain level of performance in a match and the the winner of the match is the one that has a higher performance level. The performance levels are hypothethized to be normally distributed, with mean being a linear combination of external factors and a baseline. The baseline indicates the performance level of a player in a first-round, indoor hard court, ATP 250 tournament. This model can be interpreted as a combination of two probit regression models for each players performance level, and the outcome of any particular match is determined by comparing the results from the two probit regressions. The model is illustrated below as a graph.



where matches are represented using the index $i = 1, ..., N$ and players are represented using the index $j = 1, ..., M$ with variables

X_i : Vector of external conditions (Series, Court, Surface, Round, Opponent rank)

Y_i : Match outcome

$Z_{i,1:2}$: Performance level of the two players

β_j : Vector of regression coefficients

σ_j^2 : Variance parameter for performance

The outcome of each match will be modelled using the distribution:

$$Y_i = \begin{cases} 1 & \text{if } Z_{i,1} \geq Z_{i,2} \\ 0 & \text{if } Z_{i,1} < Z_{i,2} \end{cases}$$

For player j, his performance level in match i will be a linear combination of X_i(external factors) as shown below.

$$Z^{(j)}_{i,1 \text{ or } 2} = X_i^T \beta_j + \epsilon_{i,j} \quad \epsilon_{i,j} \sim N(0, \sigma_j^2)$$

The prior distributions for this model are

$$\beta_j \sim N(\mu_0, \Sigma_0)$$
$$\sigma_j^2 \sim \text{inverse-gamma}(\nu_0/2, \nu_0\sigma_0^2/2)$$

with

$$\mu_0 = \mathbf{0}$$
$$\Sigma_0 = 1000\mathbf{I}$$
$$\nu_0 = 1$$
$$\sigma_0^2 = 1000$$

The parameters of our model would be estimated using a Gibbs sampler with full conditionals as follows. We use Z_j and X_j to indicate all (matrix) performace levels and external conditions for player j and n_j to represent the number of matches player j is involved.

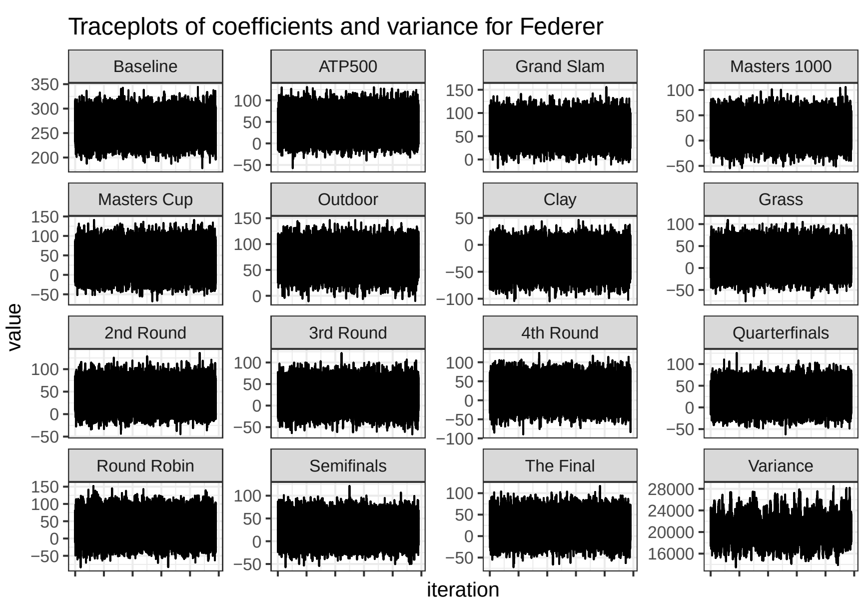
$$p(Z^{(j)}_{i,1} \mid X_i, Y_i, Z_{i,2}, \beta_j, \sigma_j^2) = \begin{cases} \text{Truncated Normal}(X_i\beta_j, \sigma_j^2, Z_{i,2}, \infty) & \text{if } Y_i = 1 \\ \text{Truncated Normal}(X_i\beta_j, \sigma_j^2, -\infty, Z_{i,2}) & \text{if } Y_i = 0 \end{cases}$$

$$p(Z^{(j)}_{i,2} \mid X_i, Y_i, Z_{i,1}, \beta_j, \sigma_j^2) = \begin{cases} \text{Truncated Normal}(X_i\beta_j, \sigma_j^2, -\infty, Z_{i,1}) & \text{if } Y_i = 1 \\ \text{Truncated Normal}(X_i\beta_j, \sigma_j^2, Z_{i,1}, \infty) & \text{if } Y_i = 0 \end{cases}$$

$$p(\beta_j \mid X_j, Z_j, \mu, \Sigma) = N((\Sigma^{-1} + X_j^T X_j/\sigma_j^2)^{-1}(\Sigma^{-1}\mu + X_j^T Z_j/\sigma_j^2), (\Sigma^{-1} + X_j^T X_j/\sigma_j^2)^{-1})$$

$$p(\sigma_j^2 \mid \beta_j, X_j) = \text{inverse-gamma}((\nu_0 + n_j)/2, [\nu_0\sigma_0^2 + \sum_{i=1}^{n_j}(Z_{i,j} - \beta_j^T X_{i,j})^2]/2)$$

We use a burn-in of 2000 samples with no thinning. The effective sample size of the coefficients are all around 15-20% of the total sample size. Since there are 244 unique players, we cannot show traceplots for all players. Below we show traceplots of coefficients for the player *Roger Federer*. No apparent convergent issue can be found in these plots.

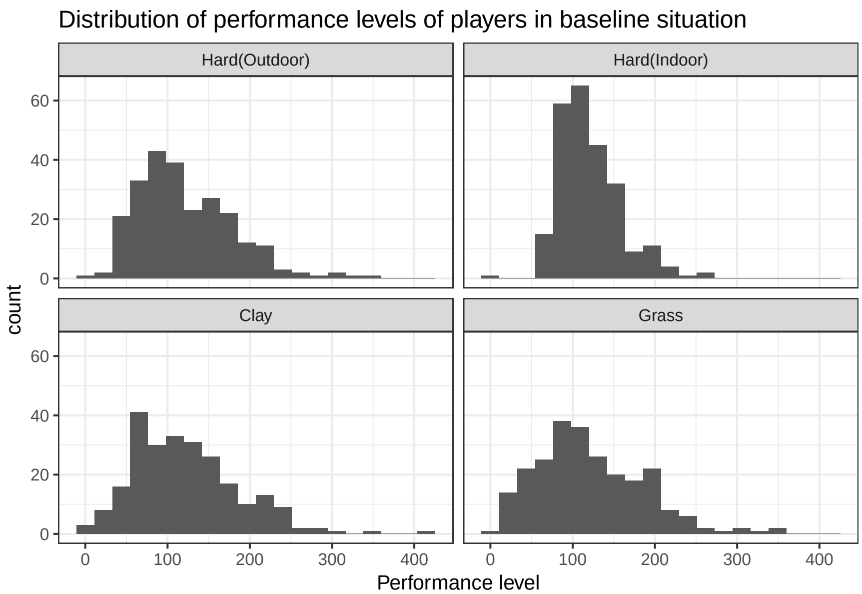


We also use Geweke's convergence diagnostics to test whether the parameters have convergent problems. None of the tests are significant, suggesting that no evidence for unequal means exists in the MCMC chain.

The estimated coefficients for the intercept (baseline performance) are all centered to make the baseline performance of the "Other" player group to be 0. Our model have an accuracy around 70% for predicting match outcomes.

Results

We first look at the overall distribution of the players' performance level at the baseline situation, which are matches played in the first-round of ATP 250 tournaments.



All the distributions are skewed to the left, with most players centered around the 100 mark and a handful of exceptionally good players. In fact, members of the *Big 4* are always in the top 5 players of all surfaces, showing that their performance levels are consistantly higher than others regardless of the condition. In addition, the variance of performance level on indoor hard court is much smaller than the variance on other types of surfaces.

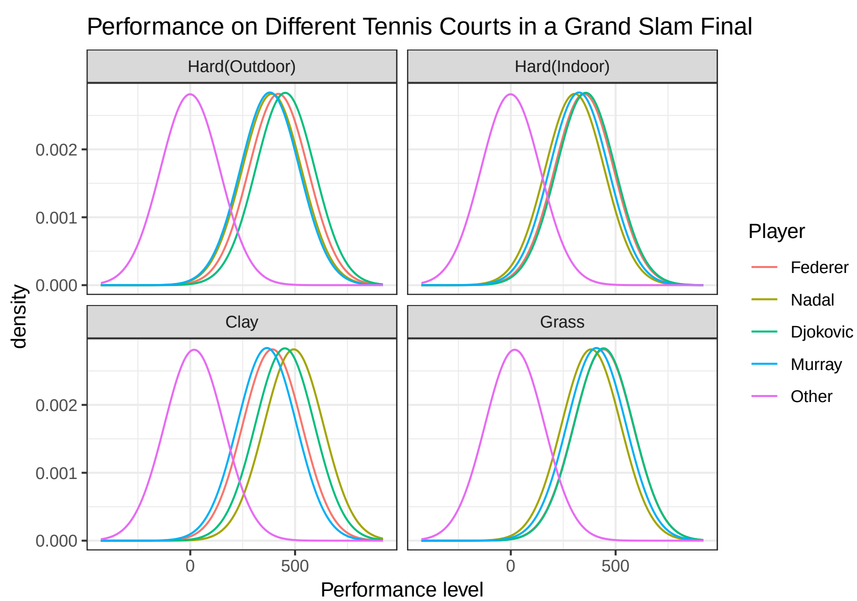
Next let us turn our attention to results related to the *Big 4* and the "Other" player catagory in more detail. Below is a summary table (Table 1) showing the estimated coefficients and the respective highest posterior density (HPD) intervals.

The baseline performance coefficients of all the *Big 4* players are significantly higher than that of the "Other" players. We also see that both *Federer* and *Murray* performs best on grass courts, *Nadal* performs best on clay courts, and *Djokovic*'s level is highest on outdoor hard courts, which is consistant with the general perception of these players. The *Big 4* all have better performance at Grand Slam tournaments. Also note that the estimated standard deviation for performance levels of all the players are relatively the same, this is also the case for all the players not presented in the table, which is against our general perception of some players being more consistant than others.

Table 1: Estimated coefficients and standard deviations with HPD intervals

	Other			Federer			Nadal			Djokovic			Murray		
	Est.	Lower	Upper	Est.	Lower	Upper	Est.	Lower	Upper	Est.	Lower	Upper	Est.	Lower	Upper
Baseline	0.00	-16.80	16.06	262.48	225.16	300.75	224.58	183.73	265.90	257.27	215.38	297.58	247.58	208.70	287.24
Series															
ATP500	10.56	-4.75	25.52	45.12	1.27	86.86	23.98	-20.06	69.14	46.62	-1.75	94.19	4.17	-42.16	49.36
Grand Slam	-13.05	-27.87	2.13	65.42	27.21	104.64	69.01	27.04	110.99	66.71	24.04	110.21	51.67	9.50	94.63
Masters 1000	9.97	-8.39	27.35	21.34	-15.00	58.55	50.43	12.16	88.81	52.70	13.04	93.51	24.60	-15.68	64.43
Masters Cup	1.75	-58.38	59.58	39.36	-12.06	88.86	9.50	-48.29	61.22	25.60	-28.30	78.83	-7.24	-63.01	48.85
Court															
Outdoor	0.19	-16.67	17.06	70.81	33.22	108.45	84.23	41.19	125.55	95.03	52.87	137.20	54.67	13.59	94.95
Surface															
Clay	19.19	6.20	31.81	-29.24	-64.96	8.59	104.45	67.13	142.09	-1.54	-42.68	40.50	-14.54	-59.19	27.41
Grass	19.55	2.40	36.26	21.30	-22.49	63.66	-5.79	-53.40	43.00	-10.03	-57.96	42.48	28.25	-17.36	75.34
Round															
2nd Round	-9.89	-22.77	3.05	40.64	-0.37	81.18	24.49	-18.64	69.15	10.78	-33.91	56.05	17.16	-25.06	59.44
3rd Round	12.59	-16.98	42.16	22.09	-21.60	64.75	14.29	-31.51	58.26	18.77	-28.24	64.85	5.49	-39.38	51.08
4th Round	46.54	-4.34	99.88	18.88	-29.46	67.04	18.79	-34.33	69.63	27.56	-25.59	80.19	8.47	-43.21	59.79
Quarterfinals	32.65	10.45	54.77	26.74	-11.97	67.18	-7.93	-48.68	33.97	34.51	-10.17	78.85	9.03	-32.55	53.63
Round Robin	-0.08	-49.74	47.98	34.01	-20.18	87.67	13.83	-43.61	69.05	4.28	-50.27	60.01	6.75	-50.57	61.60
Semifinals	32.63	-0.69	65.09	26.88	-13.90	65.73	8.58	-34.12	51.64	26.95	-16.78	70.44	35.09	-11.75	80.16
The Final	12.01	-34.69	62.18	21.83	-19.29	65.37	11.33	-35.58	55.39	33.88	-12.13	79.46	26.10	-22.82	74.18
Standard deviation															
SD	141.88	131.78	151.80	141.56	128.67	154.03	141.58	127.77	154.57	140.83	126.71	154.26	140.59	126.25	154.60

The relative performance level of these players can be investigated in more detail by ploting the posterior predictive distribution of their performance levels. The plot below shows posterior predictive distributions of performance levels at Grand Slam final matches for different tennis courts.



From the plots, we again see that the performance level of the *Big 4* is considerably higher than the "Other" players. We also see that *Djokovic* is the highest performing player on hard courts (both indoor and outdoor), with *Federer* being the second best player on hard courts. *Nadal* dominates clay court matches and *Federer* shares the crown of the best grass court player with *Djokovic*. Something worth noting is that the differences in clay court performance is much higher than those of other surfaces. This is consistant with the general perception of the clay court being more distinct in the challenges it brings to players.

Our model allows us to estimate probabilities of winning in different situations. Table 2 shows the win probability of *Big 4* and "Other" against each other. The results confirm with the plots above.

Table 2: Estimated win Probability in Grand Slam Finals

	Hard(Outdoor)	Hard(Indoor)	Clay	Grass
Federer				
Nadal	0.56	0.59	0.30	0.61
Djokovic	0.44	0.48	0.38	0.50
Murray	0.58	0.55	0.55	0.57
Other	0.98	0.96	0.97	0.98
Nadal				
Federer	0.44	0.41	0.70	0.39
Djokovic	0.37	0.40	0.58	0.38
Murray	0.52	0.46	0.74	0.45
Other	0.97	0.94	0.99	0.97
Djokovic				
Federer	0.56	0.52	0.62	0.50
Nadal	0.63	0.60	0.42	0.62
Murray	0.64	0.56	0.67	0.57
Other	0.99	0.96	0.98	0.98
Murray				
Federer	0.42	0.45	0.45	0.43
Nadal	0.48	0.54	0.26	0.55
Djokovic	0.36	0.44	0.33	0.43
Other	0.97	0.95	0.96	0.97

Next we will look into some hypothetical scenarios. As previously mentioned, the proportion of each surface is unequal on the ATP tour. The current tour consists of around 60% hard court, 30% clay, and 10% grass court tournaments. Obviously, players who prefer hard courts are more likely to gain more points and in turn get a higher ranking. We have also argued that the hard courts and grass courts are more similar in nature and clay courts requires players use different strategies. We propose two hypothetical scenarios, equal surface proportions (17% indoor hard, 17% outdoor hard, 33% clay, 33% grass) and half clay (50% clay, 25% grass, 12.5% indoor hard, 12.5% outdoor hard). The estimated overall win percentage of Grand Slam finals in these two scenarios are recalculated for the *Big 4*, as shown in table 3 below.

Table 3: Estimated win Probability in hypothetical scenarios

	Against			
	Federer	Nadal	Djokovic	Murray
Equal surface proportion				
Federer		0.50	0.45	0.56
Nadal	0.50		0.45	0.56
Djokovic	0.55	0.55		0.61
Murray	0.44	0.44	0.39	
Half clay				
Federer		0.45	0.43	0.56
Nadal	0.55		0.48	0.60
Djokovic	0.57	0.52		0.63
Murray	0.44	0.40	0.37	

In these two hypothetical scenarios, *Murray* have a win percentage lower than 50% against all the other three, while *Djokovic* will have an advantage over the others. Moreover, *Federer* and *Nadal* would have no overall advantage over one another if the ATP tour has equal surface proportions. Due to the gap between *Federer* and *Nadal*'s clay court performance levels, *Nadal* will gain an upperhand if the proportion of clay courts is increased. Analyses on these hypothetical scenarios suggests that *Djokovic* is the overall strongest player of the four, while *Murray* is the weakest in general.

Conclusions

We developed a method to quantify performance levels of professional men's tennis players by using a Bayesian graphical model. Results from our model shows that performance levels of players are skewed, with a few top players above others. We further identified that the players with top performance levels are indeed the *Big 4* of men's professional tennis. By plotting the posterior predictive distributions and calculating the win probability, we showed that members of the *Big 4* each has particular surfaces that they excel on, while *Djokovic* would be the overall strongest player if the proportion of surfaces is equal.

Potential future work might involve adding a time component to the model, either by including an age term or directly letting the current performance level be dependent on the performance level of previous matches. This would account for the change in performance level over the years and would enable the model to assess performance level more accuratly at a particular given time. Another topic for future work is putting better priors on players, especially for those without sufficient data. Priors can be formulated using intuition or other factors such as service speed and ball spinrate that a player can produce.

References

Shang-Min Ma, Chao-Chin Liu, Yue Tan & Shang-Chun Ma (2013). Winning matches in Grand Slam men's singles: An analysis of player performance-related variables from 1991 to 2008, Journal of Sports Sciences, 31:11, 1147-1155

Asmita Chitnis, Omkarprasad Vaidya (2014). Performance Assessment of Tennis Players: Application of DEA, Procedia - Social and Behavioral Sciences 133, 74-83

Filipcic, Ales; Zecic, Miroslav; Reid, Machar; Crespo, Miguel; Panjan, Andrej; Nejc, Sarabon (2015). Differences in Performance Indicators of Elite Tennis Players in the Period 1991-2010, Journal of Physical Education and Sport 15(4), 671-677