Sacramento Kings Case Competition: Fouling Out

Dani Chu, Lucas Wu, Matthew Reyers, Conor Doyle
Simon Fraser University

April 30, 2018

1 Introduction

In today's NBA, star players are the engines that drive their respective teams. The impact that these types of players can have on a game can not be overstated. So when a star player gets into foul trouble the coach is faced with a difficult decision. Coaches need to decide how to best utilize this valuable resource. Should they sit their best player and save him for crucial end-of-game match ups? Or risk the player fouling out by leaving them in the game? A popular rule-of-thumb that coaches use to inform this decision is the "Q+1" methodology detailed in Maymin and Shen's research on early foul trouble [3]. The "Q+1" rule is an easy way to determine when a coach should sit a player in foul trouble. Simply stated, if a player has more fouls than the current quarter they should be benched. For example, the "Q+1" rule asserts that a player with 3 fouls in the second quarter should sit.

Despite its popularity the "Q+1" assessment isn't always the best approach. This is one of the main points in Ben Falk's article "The Trouble with Foul Trouble" [2]. In his article Falk recounts a scenario that unfolded in the first round of the 2018 NBA playoffs between the Cleveland Cavaliers and the Indiana Pacers. After wining the first game of the series the Pacers are looking to double-up on the Cavaliers in game two. However, Victor Oladipo, the sole all-star and first option for the Pacers picks up 3 early fouls, 2 in the 1st quarter and 1 in the 2nd. Following the "Q+1" rule, Indiana's coach, Nate McMillan, sits Oladipo, who would ultimately only play 8 minutes in the first half of the game. In the second half of the game Oladipo played 20 minutes without picking up another foul and the Pacers lost the game by 3 points. Clearly, sacrificing Oladipo's minutes to the bench in favor of the flexibility to play him in the final moments of the game hurt the Pacers. This begs the question of whether the "Q+1" rule is the best strategy for foul management. Additionally, Falk points out that no foul management strategy can save a player minutes only limit them. The decision to bench a player due to fouls is a trade off:" the player will play less in exchange for the ability to control when he plays."

More and more NBA analysts and enthusiasts are seeing the shortcomings of the "Q+1" ideology and are looking for better alternatives. In her study [1] the foul rates of NBA centres, Katherine Evans suggests using a survival model to estimate the time until a player commits his next foul. Moreover, Evans considers how emotions affect the rate at which players pick up fouls, which she calls "tilting". Evan's shows that fouling rates are not independent of the number of fouls and that emotional players often tilt, but a players tilt can be reduced by a quick substitution. Evans approach to the problem of foul trouble is refreshing and offers a new way to help coaches manage players who are in or approaching foul trouble.

We are tasked with creatively outlining objective guidelines to present to a coaching staff about how to handle-in-game foul management. To do so, we take into account the findings of the relevant literature and suggest an improvement to the current "Q+1" foul management strategy identified by Maymin. We use Evan's process of fitting survival curves to different foul levels and extend it by fitting exponential distributions to the curves and combining them to gather a time to foul out distribution instead of simply a time to next foul distribution. We take into account Falk's suggestions and identify foul trouble on a player by player basis, are more lenient earlier on, and help coaches rearrange minutes instead of limiting them. Additionally, we agree that there are in-game adjustments that can be used to limit a player's foul rate. Finally, we place our trust in our coaching staff while providing them with more information than they've ever had previously to make a more informed decision regarding foul trouble.

2 Model

Survival analysis is a set of statistics methods for analyzing data where the outcome variable is the time until the occurrence of an event. The event can be death, customer churn or the occurrence of a disease etc. In the fouling out analysis setting, we are interested in the event - time to next foul.

The standard classification or regression methods may not be the best choice for this problem as they cannot effectively handle the censoring of observations, where the exact event time of some observations are unknown. For instance, a player who never picked up his sixth foul before the game ends is right censored by the end of that game.

A popular regression model for the analysis of survival data is Cox Proportional Hazards Regression, which is used to relate predictors or covariates to survival time. In a Cox proportional hazards regression model, the measure of effect is the hazard rate, which is the risk of failure (i.e., the probability of picking up the next foul), given that the participant has survived up to a specific time. We consider 3 covariates, player position (Guard, Wing and Big), number of fouls committed (foul from 0 to 5) and season experience (rookies who played 0-3 years vs. senior who played 3+ years in the league) in our Cox model, where

$$\lambda(t|Z) = \lambda_0(t)\exp(\beta_1 \cdot Position + \beta_2 \cdot FoulFrom + \beta_3 \cdot SeasonExperience)$$

Considering the work of CausalKathy [1], players tend to have differing foul rates than other players as well as different foul rates dependent on what foul number they are currently at. This remains consistent with what we found in our work.

The coefficients of our Cox model is shown in Table 1. However, the proportional hazards assumption states that the hazard of covariates must be constant over time. After we checked the assumption of proportionality, we found that some of the effects of our covariates are changing throughout the game, which is a deviation from Cox's model. If time permitted, we would like to investigate an extended version of Cox model which allows time-dependent covariates as well as incorporating variables, such as point differentials, minutes left during the game and so on.

Making note of these differences, we designed a model that could provide more informative results when we include the number of fouls a player currently has. This is first done by incorporating the survival curves shown in Figure 1. Each survival curve is based on the first two years of data we had available to us (2014-15 and 2015-16, [1]. We use some of her data and merged with the original dataset), meaning that player did not incur any more fouls for the remainder of the game and picked up the corresponding number of seconds of playtime at that foul level.

Though the curves are stepwise functions, they bare close resemblance to the exponential distribution. Previous work has used this similarity to justify the assumption that foul rates follow an exponential distributions [1]. We will make the same claim.

In considering each curve as the result of an exponential distribution, we can fit the appropriate rate parameter λ using numerical integration. Details of this process can be found at Eureka Statistics [4]. For athlete i we fit an exponential distribution to the jth foul level $(0 \ldots 5)$ and obtain λ_{ij}, the foul rate for that player at that foul level. The time to incur the next foul, t_{ij}, follows $exp(\lambda = \lambda_{ij})$. In the event that a player is missing data for foul rates or simply has insufficient data to create the desired curves at each foul level, their overall foul rate is used independently of the number of fouls they currently have, such that $\lambda = \lambda_i$. This primarily occurs for a small subset of bench and reserve players, often totaling three or fewer one-rate players per team. Given player i's foul level j, we can then use these exponential distributions to estimate the quantiles of the distribution of time to foul out by:

- Generating 5000 samples from each exponential distribution with parameters λ_{ij}
- For each of the 5000 simulations per player, calculate $\sum_j^5 t_{ij}$ as the time to foul out from foul j for player i
- Use bootstrap sampling of size 1000 to estimate the 25th, 50th, and 75th quantiles of the time to foul out distribution

For the aforementioned simulations, the truncated exponential distributions are used. Truncation is done based on the maximum playing time observed for a given player at a given foul rate. This is done to provide more reasonable estimates in the context of Basketball being a sport with a finite end time. Truncation is especially useful when considering the final foul rate as many players are artificially cut short in playing time due to the end of the game or a benching. This foul rate is significantly different than the other foul rates and greatly impacts playing time estimates when not truncated. Further, as players continue to play and accumulate data, the truncation bounds can be adjusted as to represent their new maximum playtime.

After creating the foul simulation and identifying the bounds on expected time to foul out for a given player and foul number, we are then able to validate the model against the test set of 2016-17 data that was excluded from the model building process. With validated results on out of sample data we outline how we proceed to present our results for use by the coaching staff.

3 Coaching Staff Explanation

For practical in game use we present the Time to Foul Out Tool. This tool is easy to use. Simply input the number of fouls each player has and you will be provided with estimates for the amount of playing time that a player should be able to survive at minimum 75%, 50%, and 25% of the time. This way you can make an informed decision based on the situation of the game that is catered to the tendencies of each specific player on your team.

If after a foul, a player's expected playing time remaining reaches a level that you are uncomfortable with we recommend the following steps. Firstly, take your player out of the game. This is rationalized by Evans' work where she points out that player's may tilt due to fatigue or emotions and that a quick substitution lowers foul rates. We recommend putting the player back in two to three minutes later. Secondly, we recommend giving the player the most foul efficient defensive assignment in a further attempt to prevent fouls.

The suggestions above, among other behavioural changes that a player can make on the court, will in most cases lower his foul rate.

Following Falks suggestions, we suggest rearranging minutes instead of limiting them. We give 3 estimates of differing levels of certainty so that you can adjust your budget of minutes to rearrange based on the situation. As an underdog, we recommend taking risks and trying to get the most playing time out of your star players. Therefore, use the 25% estimate to budget a player's remaining playing time. On the other hand if you are a favourite, and decide that the game will be close at the end and you value the contributions of your star player highly in end of game situations, we recommend using the 75% estimate to budget the player's remaining playing time. We acknowledge that using this estimate may severely limit a player's playing time and only recommend it in specific circumstances.

Using our method in this manner allows for more flexibility than the classical "Q+1" rule while still evaluating foul trouble. Consider the case of Victor Oladipo in Game 2 of his 2018 NBA playoffs. The Indiana Pacers were heavy underdogs with 2 fouls and 47 minutes left in the game. His 25% estimate is 35 minutes of playing time remaining and his 50% estimate is 27 minutes. With this information Coach McMillan would have known that Oladipo could have sat for 2 minutes and then played his regular rotation of minutes without fear of fouling out. Instead he sat for 10 minutes and his playing time ended up limiting his potential contribution.

4 Summary

Through our literature review we found that the basketball community does not yet have a robust understanding of how long, on average, it will take a player to foul out given a certain number of fouls. We propose a methodology to estimate the distribution of time to fouling out and using the 25th, 50th, and 75th quartiles, improve the ability to make in game decisions. Finally, we provide coaches with a tool that can be easily run during a game to help with in game decision making.

There are limitations to our approach. It is biased by the way coaches already act and by sampling bias for larger foul levels. Perception is often more powerful than reality. If we don't see many players in our dataset playing in "Q+1" foul trouble we cannot predict how other players and our player may react when they perceive our player to be in foul trouble. The player may play more cautiously as noted in the Maymin paper, and other players may attack our player more and increase their foul rate. Additionally, we do not have a clear view on the foul rates at high foul totals for players who do not reach high foul totals often. The foul rates at high foul totals are biased more often as there is less game time left to see what would happen. Finally, we do not discount data from the 4th quarter of games where there is presumably more intentional fouling.

We hope to see extended analyses as behaviours change and more data becomes available with regards to quick substitutions and the role of fatigue on foul rates.

5 Appendix

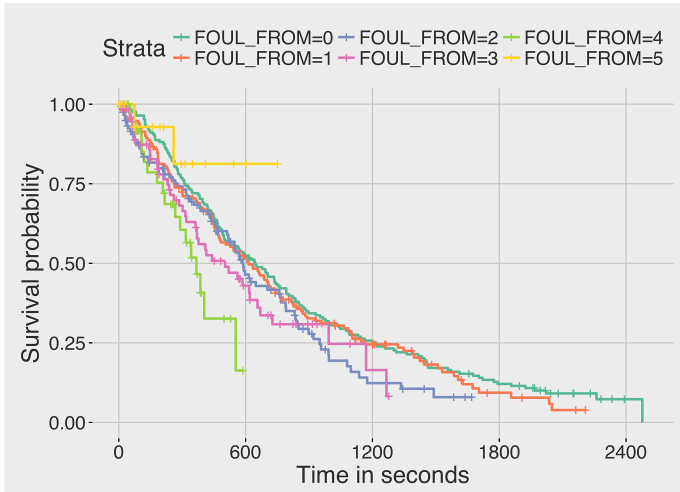
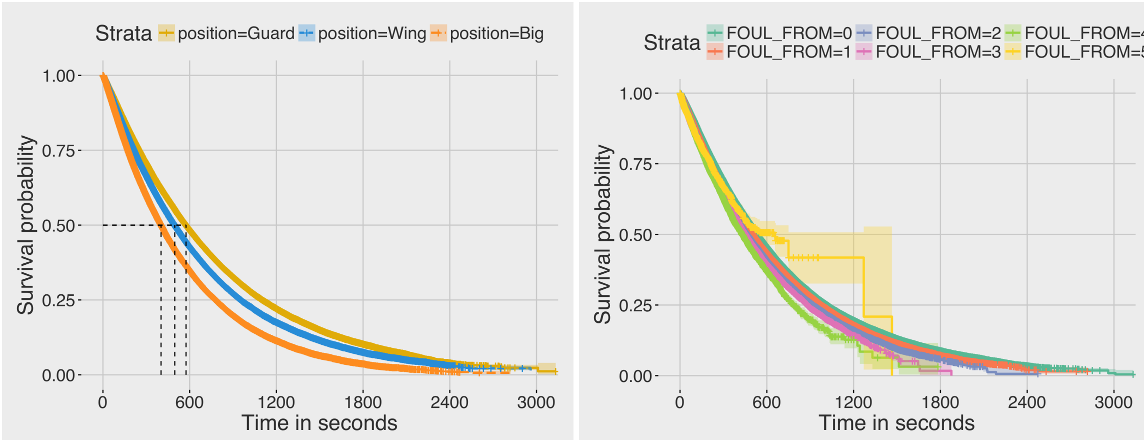


Figure 1: Survival Curves for Rudy Gay by Foul Number from 2014 to 2016

	Name	Foul	Player will survive past this many minutes x% of the time		
			75%	50%	25%
	Darren Collison	3	15	22	29
	Kosta Koufos	1	20	27	33
	DeMarcus Cousins	2	16	23	30
	Arron Afflalo	2	20	26	34
	Rudy Gay	3	12	17	22

Figure 2: Foul Trouble Tool. Can be used live in games to update the time to foul out after a player picks up a foul. Used here on the 16-17 roster of the Sacramento Kings



(a) (b)

Figure 3: (a) Foul Rate by Position (b) Foul Rate by Number of Fouls Committed

Variable	coef	exp(coef)	se(coef)	P-value
FOUL_FROM1	0.0580	1.0598	0.0059	<2e-16
FOUL_FROM2	0.1250	1.1331	0.0072	<2e-16
FOUL_FROM3	0.1569	1.1699	0.0098	<2e-16
FOUL_FROM4	0.2054	1.2280	0.0157	<2e-16
FOUL_FROM5	0.0441	1.0451	0.0359	0.22
season_experience3+	-0.1072	0.8984	0.0050	<2e-16
positionWing	0.1487	1.1603	0.0058	<2e-16
positionBig	0.3625	1.4369	0.0062	<2e-16

Table 1: This table shows the coefficients of our fitted Cox proportional hazards regression. For a player whose "position" = big, the exponential of coefficient is 1.4369 means that when will result in an increased probability (43.69%) of picking up next foul compared to when "position" = Guard, holding other variables constant

References

[1] Evans, K. NBA Fouls - Survival Analysis. `https://causalkathy.com/2017/02/21/nba-fouls-survival-analysis/`, 2017. [Online; accessed April-2018].

[2] Falk, B. The Trouble with Foul Trouble. `https://cleaningtheglass.com/the-trouble-with-foul-trouble/`, 2018. [Online; accessed April-2018].

[3] Allan Maymin, Philip Maymin, and Eugene Shen. How much trouble is early foul trouble? Strategically idling resources in the NBA. *International Journal of Sport Finance*, 7(4):324–339, 11 2012.

[4] Rosenmai, P. Fitting an Exponential Curve to a Stepwise Survival Curve. `https://eurekastatistics.com/fitting-an-exponential-curve-to-a-stepwise-survival-curve/`, 2016. [Online; accessed April-2018].