STATS 50: Mathematics of Sport — Spring 2019

Week 8 – Fairness in sport

Lecturer: Maxime Cauchois

Warning: *these notes may contain factual errors*

1 The challenge of fairness

Generally, in most sports, rules and referees guarantee equity, integrity and fair treatment between all contestants, so as to ensure that the winner is indeed the most deserving competitor. In addition to respecting a long-time ideal of what sport should be, this equality of treatment generally makes competitions more appealing and tight, and thus are generally wished not only by the participants but also by the fans.

There are several desiderata which should drive the establishment of new rules, at least in theory. One would hope to put everyone on the same starting line, but it is not clear at all what it means in particular cases. For instance, in athletics, a new law limiting the testosterone level in the body in order to compete in the women category has produced very intense debates between officials, scientists and athletes themselves, and there does not seem to be a general consensus. Making sure that the new rule is well understood and well received is thus another desiderata, somehow linked to the first one, since even a fair rule needs to be approved and understood if one intents to apply it "in the right way".

Now, there are some cases in which mathematics turn to be an effective tool, since they allow to quantify and estimate statistically the impact of any rule. A trivial example is the usual coin toss which precedes most games in order to decide which team should start to attack first, or which player should serve first. As the theory predicts a 50% probability of win for each team, it appears that the method is somehow perfectly fair (i.e. it gives the same odds to both parts).

In the following, we'll study a few cases in which empirical and theoretical analysis can help decide whether or not the laws of the game are actually fair.

2 Tiebreak in tennis

In tennis, the tiebreak was instated in 1965 to reduce the length of each game, and consists in a twelve point game with mandatory two point difference to decide which player gets to win the set. In short, players both try to be the first one to reach 7 points with a 2 point difference. Now, the question was how to distribute serves equitably so that there is no player favored over the other.

In particular, suppose that player A and player B face each other. If the rule was to switch servers between each point (or two), then one of the two would be always first to serve, and might benefit from this situation. On the contrary, the rule which was established and is still used today consists in alternating serves in the following way: A first serves during one point, then B has two serves, then A and so on. The pattern can be called ABBA, by opposition with ABAB.

Now, in [2], they empirically study this pattern in a large sample of different games so as to establish whether or not one of both players gains a statistically significant upper edge from this rule. In order to do so, they model each tiebreak as an independent pair of variable (x_i, y_i), where y_i describes the outcome of the said tiebreak, and x_i includes different explanatory variables susceptible to predict the winner of the tiebreak, including the first player to serve. It is very important to include in x_i not only this piece of information but also additional insight about both players, as one wouldn't want to get a biased analysis. Indeed, suppose that, for some reason (and this is actually debatable), the favorite of the game always decides to serve first when winning the coin toss, while the outsider prefers to return. In this case, you can see that you would observe in practice that the first player to serve wins the tiebreak more often than not, but this wouldn't be because of an unfair treatment between both players, but rather because the first player to serve is positively correlated with the best player on the court! By regressing on other variables, one would hope to reduce such bias. In economics, those are called "control" variables.

The model can then be described as a simple logistic regression:

$$y_i \sim \text{Bernoulli}(\sigma(x_i^\top \theta))$$

where $\sigma : x \mapsto 1/(1 + e^{-x})$ is the sigmoid function.

We want to test whether θ_j, the coefficient associated with the first server is 0 or not. To do so, we fit two different models, one with this piece of information and one without, and compare both fits. If the former is significantly better than the latter, we can reject the hypothesis that $\theta_j = 0$. In the opposite case, we can simply say that our data is not inconsistent with the hypothesis $\theta_j = 0$, or in other terms that there is not enough evidence to reject it.

Table 3
The effect of serving first on the probability to win the tiebreak.

	Logit Regression			*LPM*			*Oster*	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Men's tournaments								
Player A Serves First	0.051**	0.027	0.027	0.051**	0.028	0.027	0.019	0.012
	(0.024)	(0.024)	(0.024)	(0.024)	(0.024)	(0.024)	(0.024)	(0.025)
Number of obs.	1701	1701	1701	1701	1701	1701	1701	1701
Panel B: Women's tournaments								
Player A Serves First	−0.002	−0.007	−0.008	−0.002	−0.007	−0.008	−0.010	−0.007
	(0.033)	(0.033)	(0.033)	(0.033)	(0.033)	(0.033)	(0.033)	(0.032)
Number of obs.	920	920	920	920	920	920	920	920
Player A prob. to win based on betting odds	N	N	Y	N	N	Y	Y	Y
Basic controls	N	Y	Y	N	Y	Y	Y	Y

The list of basic controls includes the ranking indexes of players *A* and *B*, whether each of the players has a home advantage, the height and BMI index of each player, the round of the match in the tournament relative to the total number of rounds, as well as type of tournament- and surface-fixed effects. In Column 7 we report Oster's bias-adjusted treatment effect when the amount of selection on unobservables is recovered from the amount of selection on all observables. In Column 8 we treat the betting odds as part of the identification strategy and thus recover the amount of selection on unobservables from the amount of selection on all the other observed characteristics, where the betting odds is included both in the controlled and uncontrolled regressions. Robust standard errors are presented in parentheses. Standard errors in columns 7 and 8 are obtained from bootstrapping (500 replications) *, **, *** denote significance at the 10%, 5%, 1% level respectively.

Figure 1: Findings of logistic regression for tennis in tiebreak

In figure 1, it appears that when regressing using not only the dummy variable representing the first server but also the betting odds (see column (3)), the first server does not seem to gain any systematic advantage. On other terms, two equally-skilled players would have the same probability of winning the tiebreak, independently of who serves first.

3 Penalty Shootout in Soccer

Penalty shootouts in soccer are almost unanimously described as a mental and physical challenge more than technical, due to the tension and exhaustion that very often characterizes both team when they reach this critical moment. Due to their nature and huge impact on the outcome of a game, one can wonder whether or not one of both teams receive a significant advantage from shooting first. Actually, the ABAB order provides between 53% and 60% (see [1], [3], [4], [5]) to the team shooting first of winning the game, which in every case tends to show a statistically significant departure from the fair value 50%.

The question which arises asks for a better and fairer scheme so as to guarantee the integrity of the game, and minimize the advantage conferred to any of both teams. Several approaches have been proposed, along with probabilistic analysis of the chances for each of the teams to win the shootout. In particular, figure 2 describes four different schemes. The first two have already been summarized in prior sections, so here we rather focus on the Catch-Up and Adjusted Catch-Up rules. The former consists in taking into account past results so as decide which team shoots first on a given round: if team A shooted first during the previous round, then team B will shoot first during the next one (as in ABBA), except if A failed and B succeeded. This is for instance the case after the 8th kick: A was supposed to shoot first, but since it scored and not B, B shooted the 9th kick. In the adjusted case, the rule is the same with a difference at the beginning of the sudden death: one makes sure that the team to shoot the first kick is not the same as the one shooting the 11th one.

Table 1: An illustration of penalty shootout rules

Rule	*ABAB*		*ABBA*		Catch-Up		Adj. Catch-Up	
Team	Red	Blue	Red	Blue	Red	Blue	Red	Blue
1st kick	✔		✔		✔		✔	
2nd		✔		✔		✔		✔
3rd	✗			✗		✗		✗
4th		✗	✗		✗		✗	
5th	✔		✔		✔		✔	
6th		✔		✔		✔		✔
7th	✔			✗		✗		✗
8th		✗	✔		✔		✔	
9th	✗		✗			✔		✔
10th		✔		✔	✗		✗	
11th	✔			✔	✔			✔
12th		✔	✔			✔	✔	
13th	✔		✔			✗	✔	
14th		✗		✗	✔			✗

Table 1 shows an example of how the four rules work. The Red team is the first kicker, ✔ means a successful, and ✗ indicates an unsuccessful penalty. Since the result after five rounds is 3-3, the sudden death stage starts, where the Red team kicks first in the sixth round according to the Catch-Up Rule as the Blue team was the first-mover in the previous round, but the Blue team kicks first in the sixth round if the Adjusted Catch-Up Rule is used because it was disadvantaged in the first round.

Figure 2: Different shootout rules

This rule is supposed to be fairer insofar as shooting first empirically provides shooters with a higher percentage of success (see figure 3), around 75% when shooting first against 70% when shooting second.

We'll suppose in the following that probabilities of scoring only depend on the position of the shooter: the first one will have a probability p of scoring, when the second one instead has a probability $q < p$. In this model, there exists a probability $W(p, q)$ that the team to start the

Table 2: Penalty shootout success rates per round

	First kicker	Second kicker
Round 1	0.79	0.72
Round 2	0.82	0.77
Round 3	0.77	0.64
Round 4	0.74	0.68
Round 5	0.74	0.67

Source: Apesteguia and Palacios-Huerta (2010, p. 2558)

Figure 3: Different shootout probabilities of scoring

shootout actually wins it. The order of shooting with the probability $W(p, q)$ actually closer to 50% is likely to yield the fairer model. In some way, it is even what we can define to be the actual fairer model. Because each order rule can be represented with a different Markov chain with a specific transition matrix, it is possible to compute the actual probability of winning.

In the sudden death stage, all three rules ABBA, Catch-Up and Adj Catch-up are the same: if A kicks first during sudden death, its probability of winning won't depend on the chosen rule. One can even check that it is exactly:

$$W_{SD}(p, q) = \frac{1 - p + pq}{2 - p - q + 2pq}$$

To see the latter result, one can derive a recursion by considering the possible four outcomes of the first round.

In soccer, the presence of two distinct phases in penalty shootouts makes the probability depending on the number of rounds played before the sudden death stage. For instance, in figure 4, one can observe that, for five rounds, if $p = 3/4$ and $q = 2/3$, the Adjusted Catch-Up rule gives a probability of success just above 50%, and actually closer to 1/2 than the two other rules. In addition, one can see in figure 5 that the procedure is uniformly better than the two others on a large region of different p and q, which gives a hint about the "stability" of the procedure to a shift in the actual probabilities of scoring.

Table 3: The probability that A wins including sudden death ($p = 3/4$ and $q = 2/3$)

	Catch-Up Rule	Adjusted Catch-Up Rule	Alternating ($ABBA$) Rule
1 Round	0.526	0.526	0.526
2 Rounds	0.516	0.495	0.511
3 Rounds	0.518	0.515	0.519
4 Rounds	0.513	0.501	0.508
5 Rounds	**0.514**	**0.509**	**0.515**
6 Rounds	0.512	0.504	0.507
7 Rounds	0.512	0.507	0.513
8 Rounds	0.511	0.504	0.506

Figure 4: Probabilities of winning with number of rounds

As an aside, it can be noted that both Catch-Up rules yield higher probabilities of reaching the sudden death stage, making the game more exciting and uncertain, which is another benefit for free with the method.

Figure 3: The fixed scoring probabilities in sudden death which guarantee that the Adjusted Catch-Up Rule is fairer than the other penalty shootout designs

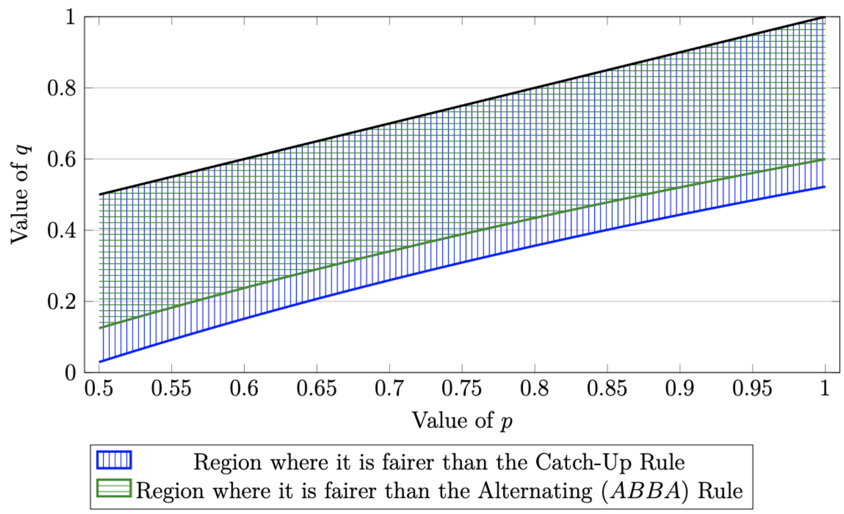

Figure 5: Uniform performance of the Adjusted Catch-up Rule

References

[1] Apesteguia, J. and Palacios-Huerta, I. (2010). Psychological pressure in competitive environments: Evidence from a randomized natural experiment. American Economic Review, 100(5):25482564.

[2] Cohen-Zada, D., Krumer, A. and Shapir, O.M., 2018. Testing the effect of serve order in tennis tiebreak. Journal of Economic Behavior & Organization, 146, pp.106-115.

[3] Kocher, M. G., Lenz, M. V., and Sutter, M. (2012). Psychological pressure in competitive environments: New evidence from randomized natural experiments. Management Science, 58(8):15851591

[4] Palacios-Huerta, I. (2014). Beautiful game theory: How soccer can help economics. Princeton University Press, Princeton, New York.

[5] Da Silva, S., Mioranza, D., and Matsushita, R. (2018). FIFA is right: The penalty shootout should adopt the tennis tiebreak format. Open Access Library Journal, 5(3):123.