Journal of Humanistic Mathematics

Volume 9 | Issue 2 July 2019

What are the Odds at the Russia 2018 FIFA World Cup?

Gunhan Caglayan
New Jersey City University

Recommended Citation

Gunhan Caglayan, "What are the Odds at the Russia 2018 FIFA World Cup?," *Journal of Humanistic Mathematics*, Volume 9 Issue 2 (July 2019), pages 258-280. DOI: 10.5642/jhummath.201902.18. Available at: https://scholarship.claremont.edu/jhm/vol9/iss2/18

What are the Odds at the Russia 2018 FIFA World Cup?

Günhan Caglayan

Department of Mathematics, New Jersey City University, Jersey City, USA
GCaglayan@njcu.edu

Synopsis

In this note we explore the group stage competition format in a standard FIFA World Cup Soccer Championship group phase. The group stage involves thirty-two teams divided into eight groups of four teams each, based on a draw that takes the national teams' seeding and geographical location into consideration. Each of the four teams in a given group is scheduled to play one match against every other team in the same group. Upon the completion of six games in each of the eight groups (for a total of 48 games), the top two highest scoring teams (the winner and the runner-up) advance to the knockout stage. In this note we focus on the forty different ways (sequential configurations or states) that a group stage in an arbitrary group can result at the end of the group stage upon the completion of the six games in a typical World Cup Championship. We generate simulations for these configurations on spreadsheets. We use the Russia 2018 FIFA World Cup as an example, along with other relevant historical data, to compare and contrast theoretical versus actual configurations and their probabilities.

Keywords: *spreadsheets*; *sequential notation*; *sequential configuration*; *number of ways*; *most probable states*; *simulations*; *FIFA World Cup*; *group stage*.

1. FIFA World Cup Group Phase

Thirty-two teams take part in the FIFA (Fédération Internationale de Football Association) World Cup Soccer Championship group phase. They are divided into eight groups $\{A, B, C, D, E, F, G, H\}$ of four teams $\{1, 2, 3, 4\}$ each, based on a draw that takes the national teams' seeding (country success coefficient / FIFA world ranking) and geographical location into consideration.

For instance, no more than two teams from Europe are permitted to be placed in the same group. The host country by default is seeded $A1$ (Table 1). The remaining teams in the eight groups are designated in the remaining spots during the draw (Figure 1).

Groups							
A	**B**	**C**	**D**	**E**	**F**	**G**	**H**
A1 (Russia)	B1	C1	D1	E1	F1	G1	H1
A2	B2	C2	D2	E2	F2	G2	H2
A3	B3	C3	D3	E3	F3	G3	H3
A4	B4	C4	D4	E4	F4	G4	H4

Table 1: Russia 2018 FIFA World Cup Soccer Championship Group Stage Draw

Figure 1: Thirty-two teams are divided into eight groups according to FIFA World Rankings. Here we see the 2018 assignments.

The league format is used during the group stage, in which each of the four teams in a given group is scheduled to play one match against every other team in the same group. Three points are awarded for a win, one point is awarded for a draw, and teams get zero points for a defeat. Upon the completion of six games in each of the eight groups (for a total of 48 games), the top two highest scoring teams advance to the knockout stage.

In the event of equality of points, other criteria (e.g., goal difference, number of goals, head-to-head points/goals, fair play points, etc.) become decisive in determining the final ranking of each team in each group. In this article we focus on the forty different ways (sequential configurations / states) that a group stage in an arbitrary group can result at the end of the group stage upon the completion of the six games.

2. Brazil 2014 and Russia 2018

Because we are studying the range of sequential configurations resulting from each group at the end of the group phase, we first look at some familiar results from 2014 FIFA World Cup and 2018 FIFA World Cup, which were held in Brazil and Russia, respectively.[1] At the end of 48 games, the group standings were recorded; some select records are depicted in Table 2 (for Brazil 2014) and Table 3 (for Russia 2018) below.

A *configuration* (or *state*) in this note is defined as the sequence of four non-negative integers (in a descending order) that corresponds to the final outcome of four-team group stage in a typical FIFA World Cup Group Stage competition upon the completion of six matches. As shown in Table 2, for instance, the configuration of Group G at the Brazil 2014 FIFA World Cup was 7441 because Germany led the group with 7 points, followed by the USA and Portugal with 4 points each, and finally Ghana with 1 point.

Team	Wins	Draws	Losses	Points
Germany	2	1	0	7
USA	1	1	1	4
Portugal	1	1	1	4
Ghana	0	1	2	1

Table 2: Group G results at the Brazil 2014 FIFA World Cup; also see Figure 10.

In the next section we explore the following question: How many different configurations (states) are available (accessible) in a typical FIFA World Cup four-team group stage where the league format is used in which each of the four teams in the same group plays one match against every other team in the same group?

[1] The first draft of this article was submitted to the *Journal of Humanistic Mathematics* on May 20, 2018, prior to the 2018 FIFA World Cup. A revised version was submitted on July 2, 2018, upon the completion of the group stage games.

Group	Result
A	9630
B	5541
C	7531
D	9431
E	7531
F	6633
G	9630
H	6443

Table 3: Group results for the Russia 2018 FIFA World Cup presented in sequential notation.

3. How Many Configurations are Possible?

We use spreadsheets to simulate configurations one million times. We begin by defining a *name box* labeled **game** in cell **P1** which contains **WDL** standing for Win, Draw, Lost, respectively. The cell **P1**, in a sense, can be thought of as an urn from which we pick the outcome for each game (W or D or L).[2] To clarify the process of simulating the group phase for an arbitrary group, we use Group D (Argentina, Iceland, Croatia, Nigeria) from the 2018 FIFA World Cup as an example (Figure 2).

Figure 2: Group D of the 2018 FIFA World Cup. See Appendix A for the full draw.

As shown in Table 4, the outcome of each game will be recorded in the first row of the Spreadsheet as follows: Game 1 result (W or D or L) will be recorded in A1; Game 2 result (W or D or L) in B1; Game 3 result (W or D or L) in C1; Game 4 result (W or D or L) in D1; Game 5 result (W or D or L) in E1; Game 6 result (W or D or L) in F1.

[2]Although in reality teams are placed in groups according to their FIFA World Rankings, we assume for most of this note that all outcomes are equally likely. Also see Sections 5.2-5.3.

Game	Team 1	Team 2	Cell
1	Argentina	Iceland	A1
2	Croatia	Nigeria	B1
3	Argentina	Croatia	C1
4	Nigeria	Iceland	D1
5	Nigeria	Argentia	E1
6	Iceland	Croatia	F1

Table 4: Setting the stage towards simulation of Group D games.

3.1. Group D – Group Phase: Recording the Game Results

We begin by choosing a random outcome for the result of the first game by putting

= MID(game, 1 + INT(RAND() * LEN(game)), 1)

in cell A1. The same syntax is used for cells B1, C1, D1, E1, F1. This way, when executed, each cell from A1 to F1 will result in a W or a D or an L.

3.2. Group D – Group Phase: Recording Team Scores

After recording the results of the six games in cells **A1:F1**, we next use cells **G1:J1** to store each team's score based on the initial Group D – Group Stage Draw in order of appearance: Argentina, Iceland, Croatia, Nigeria, respectively in G1, H1, I1, J1.

We can determine the total score for Argentina via the syntax

= 3*(COUNTIF(A1,"W") + COUNTIF(C1,"W") + COUNTIF(E1,"L")) + COUNTIF(A1,"D") + COUNTIF(C1,"D") + COUNTIF(E1,"D")

placed in cell G1.

Likewise, we compute Iceland's points with

= 3*(COUNTIF(A1,"L") + COUNTIF(D1,"L") + COUNTIF(F1,"W")) + COUNTIF(A1,"D") + COUNTIF(D1,"D") + COUNTIF(F1,"D").

in H1. In a similar manner, Croatia's uses

= **3*(COUNTIF(B1,"W") + COUNTIF(C1,"L") + COUNTIF(F1,"L")) + COUNTIF(B1,"D") + COUNTIF(C1,"D") + COUNTIF(F1,"D")**

in I1, and Nigeria uses

= **3*(COUNTIF(B1,"L") + COUNTIF(D1:E1,"W")) + COUNTIF(B1,"D") + COUNTIF(D1:E1,"D").**

in J1.

3.3. Group D – Group Phase: Obtaining the Configuration for One Trial

Once all four team scores are stored in cells **G1:J1** in the first trial, we can put these four numbers in a descending order to obtain the configuration of the trial. We can find the sorted values using the **LARGE** command in a spreadsheet.

Furthermore, we can take the scores and use them as digits in a decimal number. Consider putting

= **1000 * LARGE(G1:J1,1) + 100 * LARGE(G1:J1,2) + 10 * LARGE(G1:J1,3) + 1 * LARGE(G1:J1,4)**

into cell $K1$.

As an example, suppose that one trial resulted in the string **WDDLDW**. This means the first game is won by the "home" side (Argentina). The second, the third, and the fifth games all resulted in a draw. The fourth game is won by the "visitor" side (Iceland), and the sixth game is won by the "home" side (Iceland).

This way, in order of appearance, Argentina scores 5 points, Iceland scores 6 points, Croatia scores 2 points, and Nigeria scores 2 points. Because the configuration must be a sequence of numbers in a descending order, the numbers 5, 6, 2, 2 obtained in cells **G1:J1** must be introduced again in a descending order in cell K1. Figure 3 depicts the above discussion for 120 trials.

	A	B	C	D	E	F	G	H	I	J	K
1	L	D	W	W	L	W	6	6	1	4	6641
2	L	D	L	W	W	L	1	9	1	6	9611
3	W	L	L	L	W	L	3	3	6	6	6633
4	L	W	D	W	W	L	4	9	0	4	9440
5	W	W	W	L	W	W	9	3	6	0	9630
6	D	L	D	W	L	W	2	4	6	4	6442
7	W	L	D	L	D	L	4	1	6	5	6541
8	D	W	W	W	L	L	7	4	0	6	7640
9	L	D	W	W	D	L	4	7	1	4	7441
10	D	D	D	W	D	W	3	5	4	2	5432
11	L	L	W	D	D	L	3	5	4	4	5443
12	W	L	W	D	D	L	6	2	4	4	6442
13	L	D	D	W	D	W	2	7	4	2	7422
14	L	L	L	D	L	D	0	4	5	7	7540
15	D	D	W	W	L	W	5	4	4	3	5443
16	D	D	D	W	L	W	3	4	4	4	4443
17	W	D	L	L	W	L	4	3	4	6	6443
18	W	L	L	D	W	W	3	4	7	3	7433
19	L	L	W	W	W	D	3	9	4	1	9431
20	W	W	L	L	W	L	6	3	3	6	6633
21	L	L	L	D	W	W	0	7	7	3	7730
22	L	L	D	W	W	L	1	9	3	4	9431
23	W	W	L	L	L	D	6	0	4	7	7640
24	D	D	D	W	D	L	3	5	1	5	5531
25	D	D	L	L	W	W	2	4	7	3	7432
26	W	W	W	W	D	L	9	4	0	4	9440
27	L	D	W	W	D	W	4	7	4	1	7441
28	D	W	W	L	D	D	7	2	4	2	7422
29	L	L	D	W	D	D	1	7	4	3	7431
30	D	D	W	D	D	W	5	3	5	1	5531
31	D	D	D	L	D	W	3	2	7	2	7322
32	W	D	W	L	W	L	7	3	4	3	7433
33	L	W	D	L	W	L	4	6	3	4	6443
34	L	W	D	D	D	D	4	5	2	3	5432
35	D	L	L	L	W	W	1	4	9	3	9431
36	W	W	L	W	W	L	6	6	0	6	6660
37	L	W	W	W	L	W	6	6	3	3	6633
38	D	D	L	L	L	L	2	1	4	9	9421
39	L	W	W	D	L	D	6	4	2	4	6442
40	D	L	D	L	L	L	2	1	6	7	7621
41	L	D	L	D	W	L	1	7	2	6	7621
42	L	L	W	L	L	D	3	3	7	4	7433
43	D	D	W	W	W	L	5	7	1	3	7531
44	W	W	D	D	W	D	7	4	2	2	7422
45	W	L	L	L	D	W	3	1	9	4	9431
46	W	W	L	W	L	W	6	3	3	6	6633
47	L	D	D	D	L	L	2	4	2	7	7422
48	L	W	D	W	W	D	4	9	1	2	9421
49	L	D	W	D	D	W	4	5	5	1	5541
50	W	L	L	D	W	W	3	4	7	3	7433
51	L	L	W	D	L	D	3	4	5	4	5443
52	D	L	W	D	W	W	4	5	7	0	7540
53	D	W	D	L	L	W	5	1	6	4	6541
54	D	L	L	D	W	L	1	5	4	6	6541
55	D	L	D	D	D	D	2	3	5	3	5332
56	D	W	D	L	L	L	5	1	3	7	7531
57	L	D	W	W	D	L	4	7	1	4	7441
58	L	W	W	D	D	D	6	5	2	2	6522
59	L	D	D	W	W	W	2	9	4	1	9421
60	D	D	L	W	W	L	2	7	1	6	7621
61	W	L	W	D	W	L	6	4	4	3	6443
62	D	W	W	L	L	L	7	1	3	6	7631
63	W	L	W	L	W	W	6	3	9	0	9630
64	W	D	D	D	L	W	5	1	5	4	5541
65	D	L	L	W	D	L	1	5	3	7	7531
66	W	L	W	L	W	L	6	3	6	3	6633
67	W	D	L	D	L	D	4	1	3	7	7431
68	L	W	D	D	L	W	4	4	4	4	4444
69	D	L	D	L	W	D	2	4	7	2	7422
70	D	W	W	W	D	L	7	5	0	4	7540
71	L	L	W	W	W	W	3	9	6	0	9630
72	D	W	L	L	D	L	4	2	3	7	7432
73	D	L	L	L	L	D	1	1	7	7	7711
74	W	D	L	L	W	D	4	3	5	4	5443
75	W	W	D	L	W	W	7	3	6	1	7631
76	W	L	L	W	L	D	3	3	4	7	7433
77	D	L	W	W	W	D	4	7	4	1	7441
78	L	W	L	D	W	L	3	7	1	6	7631
79	D	L	D	D	L	D	2	2	5	5	5522
80	D	L	L	W	W	L	1	7	3	6	7631
81	L	L	D	L	D	L	1	4	6	5	6541
82	D	D	D	W	W	L	3	7	1	4	7431
83	W	W	L	W	D	D	6	4	1	5	6541
84	L	L	L	D	L	W	0	4	7	6	7640
85	W	L	D	W	D	W	4	4	6	2	6442
86	W	W	W	L	L	L	9	0	3	6	9630
87	D	W	W	W	L	L	7	4	0	6	7640
88	D	D	D	W	W	L	3	7	1	4	7431
89	W	D	L	D	W	L	4	4	2	6	6442
90	D	L	L	L	W	W	1	4	9	3	9431
91	W	L	W	W	L	W	6	3	6	3	6633
92	D	W	W	L	D	D	7	2	4	2	7422
93	W	L	L	D	D	L	3	2	4	7	7432
94	W	L	L	D	W	L	3	4	4	6	6443
95	D	W	D	W	D	W	5	5	3	2	5532
96	D	D	D	L	D	L	3	2	4	5	5432
97	D	D	W	D	L	D	5	2	3	4	5432
98	W	L	D	L	L	D	4	0	7	5	7540
99	W	L	W	W	L	D	6	3	4	4	6443
100	W	D	L	D	W	L	4	4	2	6	6442
101	W	L	L	L	L	L	3	0	6	9	9630
102	D	D	W	L	D	W	5	2	7	1	7521
103	W	D	W	L	D	L	7	1	4	4	7441
104	W	W	D	D	W	D	7	4	2	2	7422
105	W	W	D	D	D	L	7	2	1	5	7521
106	L	D	L	D	L	W	1	4	5	6	6541
107	D	W	L	W	W	L	4	7	0	6	7640
108	D	D	D	D	D	L	3	3	2	5	5332
109	D	L	D	D	D	L	2	3	4	5	5432
110	W	W	L	L	D	D	6	1	4	5	6541
111	W	L	W	D	W	D	6	4	5	1	6541
112	W	D	D	D	D	W	5	2	5	2	5522
113	W	W	D	D	W	W	7	4	4	1	7441
114	W	L	L	D	W	L	3	4	4	6	6443
115	W	L	W	W	L	D	6	3	4	4	6443
116	L	L	L	L	W	W	0	6	9	3	9630
117	L	L	W	D	L	W	3	4	7	3	7433
118	L	W	W	W	L	D	6	6	1	4	6641
119	L	W	D	L	D	D	4	4	4	3	4443
120	L	D	W	W	L	W	4	6	4	3	6443

Figure 3: Recording configurations in Column K.

3.4. Getting Ready to Count Configurations

Before we count the configurations in Column K, we set the stage for recording the number of each configuration to use later. There are altogether forty different configurations: one configuration starting with 3, two configurations starting with 4, nine configurations starting with 5, seven configurations starting with 6, fourteen configurations starting with 7, and seven configurations starting with 9, as depicted in Figure 4.

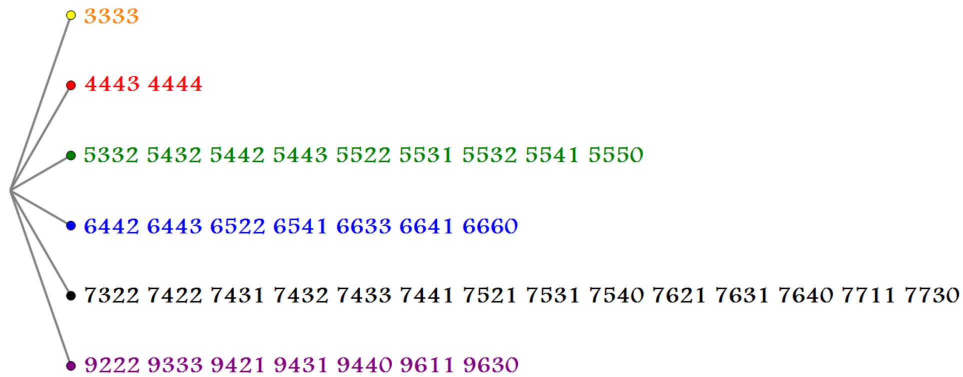

Figure 4: Forty possible configurations.

3.5. Using COUNTIF to Determine the Frequency of Each Configuration

We list the forty configurations depicted in Figure 4 within the range **M4:M43** on the spreadsheet in order to record their frequencies in cells **N4:N43**. We use the **COUNTIF** command to determine the frequency of each configuration via the syntax **N4:** $=$ **COUNTIF(K:K,M4)**, which is then dragged all the way down to Cell N43. We also calculate the total frequency of the configurations starting with (3) in Cell 04, (4) in Cell O6, (5) in Cell O15, (6) in Cell O22, (7) in Cell O36, and (9) in Cell O43. The experimental probabilities of these are also recorded in the appropriate cells Q4, Q6, Q15, Q22, Q36, and Q43 respectively (see Figure 5).

3.6. Frequency Histogram

To support the data in Figure 5, we also plot a frequency histogram for the forty accessible configurations for further analysis (Figure 6).

We make the following preliminary observations:

- 3333 is the least likely configuration.
- 5550 and 9222 each appear with probability about 0.0055.
- 4444 and 7711 appear with probability 0.008.
- 4443, 6660, and 9333 appear with probability about 0.011.
- 5332, 5522, 5531, 5532, 6522, 7322, 7730, 9440, and 9611 each have probability about 0.0163.

	M	N	O	P	Q
3	Scenario	Frequency			Exp. Prob.
4	3333	1359	1359	1	0.001359
5	4443	11085			
6	4444	8136	19221	2	0.019221
7	5332	16332			
8	5432	32672			
9	5442	33363			
10	5443	33004			
11	5522	16200			
12	5531	16562			
13	5532	16465			
14	5541	33107			
15	5550	5441	203146	9	0.203146
16	6442	32936			
17	6443	49430			
18	6522	16531			
19	6541	33082			
20	6633	32748			
21	6641	32913			
22	6660	11041	208681	7	0.208681
23	7322	16334			
24	7422	33077			
25	7431	32698			
26	7432	32831			
27	7433	32838			
28	7441	49474			
29	7521	32782			
30	7531	33100			
31	7540	32746			
32	7621	32755			
33	7631	32973			
34	7640	32876			
35	7711	8175			
36	7730	16371	419030	14	0.41903
37	9222	5465			
38	9333	10919			
39	9421	33102			
40	9431	32999			
41	9440	16580			
42	9611	16441			
43	9630	33057	148563	7	0.148563
44	TOTAL	1000000	1000000	40	1

Figure 5: Recording the frequency of each configuration for 1,000,000 trials.

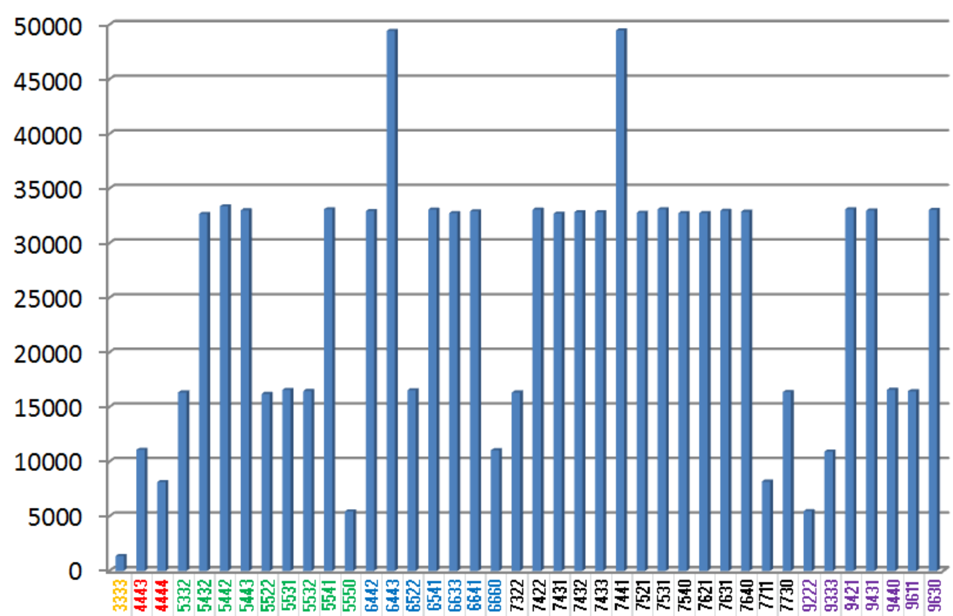

Figure 6: Frequency histogram of the distinct configurations.

- The twenty-one configurations 5432, 5422, 5443, 5541, 6442, 6541, 6633, 6641, 7422, 7431, 7432, 7433, 7521, 7531, 7540, 7621, 7631, 7640, 942, 9431, and 9630 all have probability about 0.033.
- Finally, we have 6443 and 7441 each with probability about 0.05.

3.7. Equally Accessible Configurations (EAC)

The total frequency of these seven groups of "equally accessible configurations" (from highest to least) are also calculated in Cells **T4:T10** for further analysis (Figure 7).

		S	T
	3	**Eq. Acc. Conf.**	**Freq.**
5422 5443 5541 6442 6541 6633 6641 7422 7431 7432 7433 7521 7531 7540 7621 7631 7640 9421 9431 9630	4	21	691026
5332, 5522, 5531, 5532, 6522, 7322, 7730, 9440, 9611	5	9	148580
4443, 6660, 9333	6	3	32996
6443, 7441	7	2	98472
4444, 7711	8	2	16400
5550, 9222	9	2	11100
3333	10	1	1426
	11	**Total**	1000000

Figure 7: Equally accessible configurations.

3.8. Comparing Experimental and Theoretical Probabilities

To compare the experimental and theoretical probabilities, we use **V4:V43** and **W4:W43**, respectively (Figure 8). We also calculate the total probabilities of the configurations starting with (3) in Cell X4, (4) in Cell X6, (5) in Cell X15, (6) in Cell X22, (7) in Cell X36, and (9) in Cell X43.

Further inferences are possible. For instance, upon the completion of the group stage:

1. It is impossible to score a total of 8 points – a team can only score a total of 0, 1, 2, 3, 4, 5, 6, 7, 9 points.
2. Consider the team that wins the group.
   (a) This team has 7 points with $306/729 \approx 42\%$ probability (the most likely event).
   (b) This team has 6 points with $152/729 \approx 20.9\%$ probability.
   (c) This team has 5 points with $148/729 \approx 20.3\%$ probability.
   (d) This team has 9 points with $108/729 \approx 14.8\%$ probability.
   (e) This team has 3 or 4 points each with $15/729 \approx 2\%$ probability.
3. The most likely configurations are 6443 and 7441, each with a $36/729 \approx 5\%$ probability.

	U	V	W	X
2		Exp. Prob.	Th. Prob.	
3		Out of 729	Out of 729	Out of 729
4	3333	1.026432	1	1
5	4443	8.14293	8	
6	4444	6.01425	6	14
7	5332	12.154617	12	
8	5432	24.050439	24	
9	5442	23.987016	24	
10	5443	23.88204	24	
11	5522	11.99934	12	
12	5531	11.798136	12	
13	5532	12.051099	12	
14	5541	23.96223	24	
15	5550	3.919104	4	148
16	6442	23.825178	24	
17	6443	36.105183	36	
18	6522	11.825109	12	
19	6541	23.989203	24	
20	6633	23.992119	24	
21	6641	23.957127	24	
22	6660	7.935165	8	152

	U	V	W	X
		Exp. Prob.	Th. Prob.	
		Out of 729	Out of 729	Out of 729
23	7322	12.014649	12	
24	7422	23.900994	24	
25	7431	23.755194	24	
26	7432	24.072309	24	
27	7433	23.847777	24	
28	7441	35.948448	36	
29	7521	24.18822	24	
30	7531	24.121881	24	
31	7540	24.226128	24	
32	7621	23.955669	24	
33	7631	24.106572	24	
34	7640	23.784354	24	
35	7711	6.036849	6	
36	7730	11.994966	12	306
37	9222	3.971592	4	
38	9333	8.053263	8	
39	9421	23.927238	24	
40	9431	24.031485	24	
41	9440	11.941749	12	
42	9611	12.011004	12	
43	9630	23.945463	24	108
44	TOTAL	729	729	729

Figure 8: Comparing experimental and theoretical probabilities.

4. When equally accessible configurations (EAC) are considered:

   (a) The group

$$5432, 5422, 5443, 5541, 6442, 6541, 6633,$$
$$6641, 7422, 7431, 7432, 7433, 7521, 7531,$$
$$7540, 7621, 7631, 7640, 9421, 9431, 9630$$

   of 21 configurations stands out as the most accessible EAC with $21 \cdot 24/729 \approx 69.14\%$ total probability.

   (b) Next we have the cluster

$$5332, 5522, 5531, 5532, 6522, 7322, 7730, 9440, 9611$$

   of 9 configurations with $9 \cdot 12/729 \approx 14.81\%$ total probability.

   (c) Then the set $\{6443, 7441\}$ with $2 \cdot 36/729 \approx 9.88\%$ total probability.

3.9. Conditional Probability Inferences

It is possible to come up with some conditional probability inferences as well. Continuing our list of inferences above, perhaps we can add the following:

For a team to advance to the knockout stage, it is required to finish the group in the first or second position. Therefore, according to the probability data outlined in Figure 8, the conditional probability that a national team advances to the knockout stage is

- 100% given that it scores 9 points;
- 100% given that it scores 7 points;
- $(252 + (2/3)(8))/260 \approx 98.97\%$ given that it scores 6 points.
- $(2562 + (2/3)(4))/256 \approx 99.48\%$ given that it scores 5 points.
- $((16/3) + 249)/458 \approx 55.53\%$ given that it scores 4 points.
- $((1/2) + 18 + (8/3))/352 \approx 6\%$ given that it scores 3 points.

Therefore, it is highly unlikely to advance to the knockout stage with 3 points (that is, 3D or 1W-2L combination); while collecting 4 points (that is, 1W-1D-1L combination) does not seem to guarantee a spot at the knockout stage, either. Whereas collecting 9 or 7 points (that is, 3W or 2W-1D combination) brings advancement to the knockout stage with 100% certainty, collecting 5 or 6 points (that is, 1W-2D combination or 2W-1L combination) will highly likely be enough to move the team to the next stage.

3.10. Inferences about the Team Finishing Third

A national team will finish the group stage in the third position (hence will fail to advance to the knockout stage in the World Cup) having scored:

- 1 point with $18/729 \approx 2.47\%$ probability.
- 2 points with $136/729 \approx 18.66\%$ probability.
- 3 points with $273/729 \approx 37.45\%$ probability.
- 4 points with $290/729 \approx 39.78\%$ probability (the most likely event).
- 5 points with $4/729 \approx 0.55\%$ probability (the least likely event).
- 6 points with $8/729 \approx 1.1\%$ probability.

It is impossible to end the group in the third position with no points.

4. So What? Exploring Conditional Probability Further

What good are these simulations, or knowing the number of possible configurations, or the probability of each configuration? This knowledge might have implications in an actual group stage, in determining the odds of qualification (right before the fifth or the sixth game, for instance, given the outcome of the first four or five games). For that purpose, we simulate again one million times in order to get a feel of the available configurations (aka the "so far" configurations) after five matches. Using a similar procedure as described above, we obtain the spreadsheet snapshot in Figure 9; note that Column F corresponding to the sixth game is left blank.

	A	B	C	D	E	F	G	H	I	J
1	L	D	L	D	D		1	4	4	3
2	W	D	D	L	W		5	3	4	1
3	L	D	L	L	L		1	3	4	6
4	D	D	W	W	W		5	7	1	0
5	L	W	L	W	L		3	6	0	6
6	W	D	W	L	W		7	3	4	0
7	L	W	D	W	L		4	6	0	4
8	W	W	L	L	D		6	1	3	4
9	L	L	D	D	D		1	5	4	2
10	W	L	L	L	L		3	0	6	6
11	W	L	W	W	W		6	6	3	0
12	L	D	D	L	L		2	3	4	4
13	L	D	L	W	W		1	9	1	3
14	W	W	W	D	W		9	4	1	0
15	D	L	D	L	W		2	4	6	1
16	L	D	D	W	L		2	6	1	4
17	D	D	L	D	L		2	2	2	6
18	W	W	L	L	D		6	1	3	4
19	L	W	D	W	L		4	6	0	4
20	L	W	D	L	W		4	6	3	1
21	L	L	L	W	L		0	6	3	6
22	L	D	L	W	W		1	9	1	3
23	W	D	D	W	D		5	4	1	2
24	W	W	D	D	D		7	2	1	2
25	D	D	D	L	L		3	1	4	4
26	L	L	L	D	D		0	5	4	4
27	L	D	D	W	L		2	6	1	4
28	D	L	W	W	L		4	4	3	3
29	L	W	W	L	D		6	4	3	1
30	L	D	D	L	W		2	6	4	1
31	L	W	L	W	D		3	7	0	4
32	D	D	W	D	L		5	2	2	3
33	W	W	D	W	W		7	6	0	1
34	D	D	W	L	D		5	2	4	1
35	L	L	L	W	D		0	7	3	4
36	W	L	W	D	L		6	1	4	3
37	W	W	D	W	L		7	3	0	4
38	D	L	W	D	D		4	3	4	1
39	W	W	L	W	W		6	6	0	3
40	L	W	L	W	W		3	9	0	3
41	D	L	D	W	W		2	7	3	1
42	D	L	L	D	W		1	5	4	3
43	L	D	D	L	W		2	6	4	1
44	L	D	W	D	D		4	5	2	1
45	D	W	D	L	L		5	1	3	4
46	D	L	W	W	L		4	4	3	3
47	L	D	L	D	L		1	4	2	6
48	W	W	D	W	L		7	3	0	4
49	W	W	W	D	W		9	4	1	0
50	W	D	W	D	W		7	4	2	0
51	L	W	D	D	W		4	7	1	1
52	D	D	L	W	D		2	5	1	4
53	D	D	L	W	W		2	7	1	3
54	L	L	D	D	L		1	4	4	4
55	D	D	W	L	D		5	2	4	1
56	W	W	D	W	L		7	3	0	4
57	W	D	L	W	W		4	6	1	3
58	L	D	L	W	D		1	7	1	4
59	D	L	L	W	D		1	5	3	4
60	W	L	L	L	D		3	1	6	4
61	D	W	D	L	L		5	1	3	4
62	W	D	L	L	L		4	0	4	6
63	L	L	L	W	L		0	6	3	6
64	D	W	D	D	W		5	5	1	1
65	W	L	W	W	D		6	4	3	1
66	L	W	L	W	L		3	6	0	6
67	L	W	D	L	D		4	4	3	2
68	W	L	D	W	D		4	4	3	2
69	W	L	W	W	D		6	4	3	1
70	D	W	L	W	W		4	7	0	3
71	W	L	L	W	L		3	3	3	6
72	W	D	W	W	L		7	3	1	3
73	W	W	L	D	W		6	4	1	3
74	D	D	L	W	D		2	5	1	4
75	L	L	D	L	L		1	3	6	4
76	D	W	D	L	W		5	4	3	1
77	D	W	D	W	D		5	5	0	2
78	D	D	D	W	L		3	4	1	4
79	D	L	W	D	W		4	5	4	0
80	W	W	L	D	D		6	2	1	4
81	L	L	W	D	W		3	7	4	0
82	D	W	W	W	L		7	4	0	3
83	D	L	W	W	D		4	5	3	1
84	W	L	L	L	D		3	1	6	4
85	W	W	D	W	D		7	4	0	2
86	W	D	D	D	L		5	1	2	4
87	W	D	L	D	L		4	1	2	6
88	W	D	W	W	D		7	4	1	1
89	D	W	D	L	D		5	2	3	2
90	D	W	L	W	D		4	5	0	4
91	D	W	D	L	W		5	4	3	1
92	D	L	L	L	D		1	2	6	4
93	W	W	D	W	L		7	3	0	4
94	W	L	D	W	W		4	6	3	1
95	L	W	L	L	W		3	6	3	3
96	L	L	L	L	D		0	4	6	4
97	L	L	D	D	D		1	5	4	2
98	W	W	W	L	D		9	1	3	1
99	L	W	W	L	W		6	6	3	0
100	L	W	D	L	W		4	6	3	1

Figure 9: The so far configurations for a hundred random trials.

4.1. From "So What?" to "So Far" Configurations

Once again using the same example above from Table 4, this time we record the result of the first five games in cells **A1:E1**, while we still use cells **G1:J1** to store each team's score based on the initial Group D – Group Stage Draw in order of appearance: Argentina, Iceland, Croatia, Nigeria, respectively in G1, H1, I1, J1. Because the sixth game is between Iceland and Croatia, and the configuration "so far" will be based on the first five games, there are exactly three possible configurations (out of the forty distinct available) upon the completion of the sixth game.

For instance, suppose that the first five games resulted in the string WDDLD. This means the first game is won by the "home" side (Argentina). The second, the third, and the fifth games all resulted in a draw. The fourth game is won by the "visitor" side (Iceland). This way, in order of appearance, "so far" Argentina scores G1=5 points; Iceland scores H1=3 points;

Croatia scores I1=2 points; and Nigeria scores J1=2 points. However, G1 and J1 are temporary scores subject to change because the sixth game, that is, the Iceland-Croatia game, is still pending. Thus, the "so far" configuration 5322 could become either 5622 (if Iceland wins), or 5352 (if Croatia wins), or 5432 (if it is a draw). Equivalently, we would obtain the revised (in descending order) final configuration 6522, 5532, or 5432, respectively. From Figure 8, we know these configurations have respective probabilities of $12/729, 12/729, 24/729$. This means that the "so far" configuration 5322 will be finalized as 6522, 5532, or 5432, with respective (conditional) probabilities $12/(12+12+24) = 25\%, 12/(12+12+24) = 25\%, 24/(12+12+24) = 50\%$. Therefore, the "so far" configuration 5322 will more likely become 5432. Equivalently, given the five game results described above, it is more likely for the sixth game between Iceland and Croatia to end in a draw than either side winning the game. Figure 9 depicts the above discussion for 100 trials.

We give another example to clarify the procedure for calculating the conditional probabilities and making inferences. Suppose that the first five games resulted in the string LWWLW as in trial 99 of Figure 9, corresponding to the "so far" configuration 6630. This "so far" configuration 6630 could become either 6930 (if Iceland wins), or 6660 (if Croatia wins), or 6740 (if it is a draw). Equivalently, we would have either of the final configurations 9630, or 6660, or 7640, revised in descending order, respectively. As before, we go back to Figure 7 to retrieve the probability of these three configurations: $24/729, 8/729, 24/729$ respectively. This suggests that the "so far" configuration 6630 will become 9630, or 6660, or 7640 with respective (conditional) probabilities $24/(24+8+24) = 3/7, 8/(24+8+24) = 1/7, 24/(24+8+24) = 3/7$. Thus, given the first five game results in the way recounted above, it is more likely for the last game between Iceland and Croatia to not end in a draw. Equivalently, it it is more likely that Iceland or Croatia wins the sixth game. Although it is possible to make a list of all the "so far" available states and then to calculate the corresponding conditional probabilities, we recommend using only the states needed, as illustrated in the two examples in the context of the 2018 FIFA World Cup Group D above.

4.2. What Happened in 2014 FIFA World Cup – Group G?

To delve further into the properties of "so far" configurations, we look at an actual scenario from Brazil 2014 FIFA World Cup – Group G, also called the Group of Death, where the winner and the runner-up were unknown until the last moment. We looked at this group briefly in Table 2. More details, including the scores and timing of all six game scores, are provided in Figure 10.

Day 1

16 JUN 2014 - 13:00 Local time, GROUP G, Arena Fonte Nova, Salvador	GERMANY	FULL-TIME 4-0	PORTUGAL
16 JUN 2014 - 19:00 Local time, GROUP G, Estadio das Dunas, Natal	GHANA	FULL-TIME 1-2	USA

Day 2

21 JUN 2014 - 16:00 Local time, GROUP G, Estadio Castelao, Fortaleza	GERMANY	FULL-TIME 2-2	GHANA
22 JUN 2014 - 18:00 Local time, GROUP G, Arena Amazonia, Manaus	USA	FULL-TIME 2-2	PORTUGAL

Day 3

26 JUN 2014 - 13:00 Local time, GROUP G, Arena Pernambuco, Recife	USA	FULL-TIME 0-1	GERMANY
26 JUN 2014 - 13:00 Local time, GROUP G, Estadio Nacional, Brasilia	PORTUGAL	FULL-TIME 2-1	GHANA

GROUP G

TEAMS	MP	W	D	L	GF	GA	+/-	Pts
GERMANY	3	2	1	0	7	2	5	7
USA	3	1	1	1	4	4	0	4
PORTUGAL	3	1	1	1	4	7	-3	4
GHANA	3	0	1	2	4	6	-2	1

Figure 10: Analyzing Group G (the Group of Death) from the Brazil 2014 FIFA World Cup.
.

Let us now explore the "so far" configuration (SFC) that is based on the first five games in this actual context, assuming that we did not know the outcome of the last game between Portugal and Ghana. The SFC corresponding to the first five games is 7114.We proceed as before: The SFC 7114 could become either 7414 (if Portugal wins – *which is what actually happened*), or 7144 (if Ghana wins), or 7224 (if the last game is a draw). Equivalently, we would have either 7441 or 7422 as the revised descending order final configuration (unlike the previous examples where we had three possibilities). The probability of these two configurations are 36/729 and 24/729, respectively.

(Note actually that we know 7441 as one of the two most accessible configurations.) Therefore the SFC 7114 will turn into 7441 or 6660 with respective (conditional) probabilities $36/(36+24) = 60\%$ and $24/(36+24) = 40\%$.

As a final example, suppose that the USA-Germany game was postponed and thus became the last game instead of the Portugal-Ghana game. The SFC would then be 4414, which in turn would become either 4417 (if USA wins), or 7414 (if Germany wins – *which is what actually happened*), or 5415 (if it is a draw); suggesting that the SFC turns into 7441 or 5541, with respective (conditional) probabilities $36/(36+24) = 60\%$, $24/(36+24) = 40\%$ (same conditional probabilities as in the previous example).

5. Implications and Modifications

The "3 points for a win" scoring system has been in regulation six times in the FIFA World Cup Championships (1994, 1998, 2002, 2006, 2010, 2014, 2018). The historical data reveals the configurations as listed in Table 5.

Year	Actual configurations
1994	**4444, 6443, 6660, 6660, 7521, 7531**
1998	**5531, 6541, 6541, 7322, 7631, 7730, 9421, 9630**
2002	**5541, 7433, 7433, 7521, 7531, 7540, 9440, 9440**
2006	**7531, 7540, 7631, 7730, 9421, 9421, 9611, 9630**
2010	**5432, 5541, 6443, 6641, 7441, 7540, 9431, 9630**
2014	**7441, 7631, 7640, 7730, 9421, 9431, 9431, 9630**
2018	**5541, 6443, 6633, 7531, 7531, 9431, 9630, 9630**

Table 5: Actual configuration data from six FIFA World Cup Championships.

5.1. HST vs. EAC Categorization

Out of 54 configurations that actually happened, we note the following:

- Configuraton 9630 occured most often (6 times),
- followed by 7531 (5 times),
- 9421, 9431 (4 times),
- 5541, 6443, 7540, 7631, 7730 (3 times),
- 6541, 6660, 7433, 7441, 7521, 9440 (twice),
- 4444, 5432, 5531, 6641, 6633, 7322, 7640, 9611 (once).

In addition,

- The highest scoring team won the group with

  7 points $22/54 \approx 40.74\%$ of the time (compare with $\approx 42\%$ theoretical probability),

  6 points $9/54 \approx 16.67\%$ of the time (compare with $\approx 20.9\%$ theoretical probability),

  5 points $5/54 \approx 9.26\%$ of the time (compare with $\approx 20.3\%$ theoretical probability),

  9 points $17/54 \approx 31.48\%$ of the time (compare with $\approx 14.8\%$ theoretical probability).

- The theoretically most likely configurations 6443 and 7441 each appeared $3/54 \approx 5.56\%$ and $2/54 \approx 3.70\%$ of the time (compare with $\approx 4.94\%$ theoretical probability).

In other words the actual probability turned out to be very close to the theoretical one in the case of "*The group winner with 7 points*" perspective; see the fourth row of Table 6.

HST	**Actual Prob.**	**Theoretical Prob.**	**Error**
4444	$1/54 \approx 1.85\%$	$14/729 \approx 1.92\%$	**0.00068587105**
5432 5531 5541 5541 5541	$5/54 \approx 9.26\%$	$148/729 \approx 20.3\%$	0.1104
6443 6443 6443 6541 6541 6641 6660 6660 6633	$9/54 \approx 16.67\%$	$152/729 \approx 20.9\%$	0.0418
7322 7433 7433 7441 7441 7521 7521 7531 7531 7531 7531 7531 7540 7540 7540 7631 7631 7631 7640 7730 7730 7730	$22/54 \approx 40.74\%$	$306/729 \approx 42\%$	**0.01234567901**
9421 9421 9421 9421 9431 9431 9431 9431 9440 9440 9611 9630 9630 9630 9630 9630 9630	$17/54 \approx 31.48\%$	$108/729 \approx 14.8\%$	0.1667

Table 6: Configurations data – Highest Scoring Team (HST) perspective.

When equally accessible configurations (EAC) are considered, we see the actual probabilities to be much closer to the theoretical probabilities in all categories; we display the relevant data in Table 7. The EAC categorization seems to "balance out" the error discrepancies of Table 6.

EAC Perspective	Actual Prob.	Theoretical Prob.	Error
5432 5541 5541 5541 6541 6541 6633 6641 7433 7433 7521 7521 7531 7531 7531 7531 7531 7540 7540 7540 7631 7631 7631 7640 9421 9421 9421 9421 9431 9431 9431 9431 9630 9630 9630 9630 9630 9630	$38/54 \approx 70.37\%$	$504/729 \approx 69.14\%$	0.01234567901
5531 7322 7730 7730 7730 9440 9440 9611	$8/54 \approx 14.81\%$	$108/729 \approx 14.81\%$	0.00000
6443 6443 6443 7441 7441	$5/54 \approx 9.26\%$	$72/729 \approx 9.88\%$	0.00617
6660 6660	$2/54 \approx 3.70\%$	$24/729 \approx 3.29\%$	0.00412
4444	$1/54 \approx 1.85\%$	$12/729 \approx 1.65\%$	0.00206

Table 7: Configurations data – Equally Accessible Configurations (EAC) Perspective.

5.2. Suggestions for More Accurate Simulations

In the HST perspective, the theoretical results regarding configurations where the highest scoring team ends up with 6 or 7 points appear to be in closer agreement with the actual results, than those configurations wwhere the highest scoring team ends up with 5 or 9 points. Here we propose some plausible explanations.

In the simulations we used only one "urn" including the equally likely outcomes W or D or L. In reality, this is not the case. In actual World Cup Championships, each four-team group has one team from Pot 1 (strongest teams with highest country coefficients), one team from Pot 2, one team from Pot 3, and one team from Pot 4 (weakest teams when FIFA World Rankings are considered); see Figure 11 for the four pots used for the 2018 FIFA World Cup. So, visiting our FIFA 2018 Group D [D1: Argentina, D2: Iceland, D3: Croatia, D4: Nigeria] example again, we see that Argentina, being drawn from Pot 1, has a higher ranking than the other three teams and hence is the strongest team of the group, at least in theory.

Our spreadsheet simulation may produce "better" results by taking FIFA World Rankings into consideration. For instance, instead of using one urn with WDL, each game could have its own urn from which the simulation results are to be generated. One such "different urn per game" technique is proposed in Figure 12 as an example.

Figure 11: Drawing from the pots for the 2018 FIFA World Cup.

5.3. Simulating the Adjusted Urn Model

Once again we use spreadsheets, this time to simulate the configurations based on the Adjusted Urn Model (AUM) according to Figure 12 one million times. We begin by defining six *name boxes*, respectively labeled as **game1**, **game2**, **game3**, **game4**, **game5**, **game6**, in cells **P1:U1**. These contain the adjusted urns **WWWDDL**, **WWDDDL**, **WWDDDL**, **WDDDLL**, **LLLLDW**, **WDDDDL**, respectively.

To clarify the process of simulating the AUM, we simulate the group phase for the familiar example of Group D (Argentina, Iceland, Croatia, Nigeria) from the 2018 FIFA World Cup. The outcome of each game will be recorded in the first row of the Spreadsheet as before.

Game 1 Argentina vs Iceland → **Urn1=WWWDDL**

Game 2 Croatia vs Nigeria → **Urn2=WWDDDL**

Game 3 Argentina vs Croatia → **Urn3= WWDDDL**

Game 4 Nigeria vs Iceland → **Urn4= WDDDLL**

Game 5 Nigeria vs Argentina → **Urn5=LLLLDW**

Game 6 Iceland vs Croatia → **Urn6=WDDDDL**

Figure 12: Adjusted urns for simulating Group D games.

For each of A1, B1, C1, D1, E1, F1, the game result (W or D or L) is recorded using

=MID(game1,1+INT(RAND()*LEN(game1)),1)

to uniformly draw from the result. Using a procedure similar to the one described in Sections 3.1-3.7, we obtain the desired results; see Figure 13 for a partial screenshot of the final spreadsheet.

	A	B	C	D	E	F	G	H	I	J	K
1	W	W	D	W	D	D	5	1	5	4	5541
2	D	W	L	D	W	D	1	3	7	4	7431
3	W	D	W	D	L	D	9	2	2	2	9222
4	D	L	W	L	L	W	7	7	0	3	7730
5	W	W	L	D	W	D	3	2	7	4	7432
6	D	D	D	L	L	D	5	5	3	1	5531
7	W	L	W	L	L	L	9	3	3	3	9333
8	W	W	W	D	L	D	9	2	4	1	9421
9	D	W	L	D	D	L	2	2	9	2	9222
10	D	W	D	W	L	D	5	2	5	3	5532
11	L	W	L	L	W	W	0	9	6	3	9630
12	L	D	D	D	L	W	4	7	2	2	7422
13	L	D	D	L	W	D	1	7	3	4	7431
14	W	W	W	L	W	D	6	4	4	3	6443
15	D	D	D	W	L	L	5	1	5	4	5541
16	L	D	D	W	L	W	4	6	2	4	6442
17	L	D	W	D	L	D	6	5	2	2	6522
18	W	W	L	L	D	W	4	6	6	1	6641
19	W	L	D	D	D	D	5	2	2	5	5522
20	W	W	D	W	L	D	7	1	5	3	7531
21	W	W	W	L	W	W	6	6	3	3	6633
22	W	W	W	D	W	D	6	2	4	4	6442
23	D	W	W	D	L	W	7	5	3	1	7531
24	W	L	D	W	W	D	4	1	2	9	9421
25	L	D	W	D	D	D	4	5	2	3	5432
26	W	D	W	L	L	D	9	4	2	1	9421
27	W	D	D	D	L	W	7	4	2	2	7422
28	D	W	D	D	L	D	5	3	5	1	5531
29	L	D	D	D	L	D	4	5	3	2	5432
30	D	W	D	L	L	L	5	4	7	0	7540
31	W	W	L	D	L	D	6	2	7	1	7621
32	W	D	D	L	D	D	5	4	3	2	5432
33	D	D	W	L	W	D	4	5	2	4	5442
34	W	W	D	D	L	D	7	2	5	1	7521
35	D	W	D	D	L	D	5	3	5	1	5531
36	W	D	D	D	D	D	5	2	3	3	5332
37	L	W	L	L	D	D	1	7	7	1	7711
38	W	L	D	D	L	D	7	2	2	4	7422
39	W	W	W	D	L	D	9	2	4	1	9421
40	W	D	D	D	L	W	7	4	2	2	7422
41	W	W	W	L	W	L	6	3	6	3	6633
42	W	W	L	W	L	D	6	1	7	3	7631
43	W	W	W	L	L	D	9	4	4	0	9440
44	L	W	D	D	W	W	1	7	4	4	7441
45	W	D	D	W	L	W	7	3	2	4	7432
46	W	D	L	W	W	D	3	1	5	7	7531
47	W	W	W	L	W	D	6	4	4	3	6443
48	W	W	W	D	L	D	9	2	4	1	9421
49	L	L	D	D	L	L	4	4	4	4	4444
50	W	D	D	L	L	D	7	4	3	1	7431
51	W	D	D	D	L	L	7	1	5	2	7521
52	W	D	W	D	L	D	9	2	2	2	9222
53	W	D	L	D	L	D	6	2	5	2	6522
54	W	D	D	D	L	D	7	2	3	2	7322
55	L	L	D	L	L	D	4	7	2	3	7432
56	W	L	W	L	L	D	9	4	1	3	9431
57	W	L	D	D	L	W	7	4	1	4	7441
58	D	W	D	D	D	D	3	3	5	2	5332
59	D	W	D	D	W	D	2	3	5	4	5432
60	W	W	W	D	W	D	6	2	4	4	6442
61	W	L	D	L	L	D	7	4	2	3	7432
62	W	D	D	L	D	D	5	4	3	2	5432
63	D	W	W	D	W	L	4	2	6	4	6442
64	W	D	D	L	L	D	7	4	3	1	7431
65	W	L	L	L	L	L	6	3	6	3	6633
66	D	L	W	D	W	W	4	5	0	7	7540
67	D	W	W	D	L	D	7	3	4	1	7431
68	D	L	D	W	L	D	5	2	2	6	6522
69	W	D	D	D	L	D	7	2	3	2	7322
70	L	L	L	W	L	D	3	4	4	6	6443
71	W	W	W	D	L	W	9	4	3	1	9431
72	D	L	D	L	L	D	5	5	2	3	5532
73	D	D	D	D	L	D	5	3	3	2	5332
74	D	D	D	L	L	D	5	5	3	1	5531
75	W	D	D	L	D	D	5	4	3	2	5432
76	D	W	D	D	D	D	3	3	5	2	5332
77	L	D	D	D	L	W	4	7	2	2	7422
78	W	W	W	D	L	D	9	2	4	1	9421
79	W	W	D	L	L	L	7	3	7	0	7730
80	W	D	W	D	L	W	9	4	1	2	9421
81	L	W	W	L	D	L	4	6	6	1	6641
82	L	W	W	L	L	D	6	7	4	0	7640
83	D	D	D	D	W	D	2	3	3	5	5332
84	W	D	L	W	L	D	6	1	5	4	6541
85	L	D	D	L	L	D	4	7	3	1	7431
86	W	L	D	D	L	W	7	4	1	4	7441
87	D	D	D	L	L	L	5	4	5	1	5541
88	W	D	W	L	L	D	9	4	2	1	9421
89	W	W	D	D	L	D	7	2	5	1	7521
90	W	D	W	D	L	D	9	2	2	2	9222
91	W	W	W	L	W	D	6	4	4	3	6443
92	W	L	D	D	D	D	5	2	2	5	5522
93	D	W	W	D	W	L	4	2	6	4	6442
94	W	D	D	D	D	L	5	1	5	3	5531
95	D	D	D	D	L	L	5	2	5	2	5522
96	D	W	D	D	D	L	3	2	7	2	7322
97	D	W	W	D	W	L	4	2	6	4	6442
98	W	W	L	W	L	L	6	0	9	3	9630
99	L	D	W	L	L	D	6	7	2	1	7621
100	W	W	D	D	D	D	5	2	5	2	5522

Figure 13: Adjusted Urn Model simulation.

We display the frequency of each configuration in our simulation in Figure 14. Figure 15 presents the frequency histogram of the forty distinct configurations in our AUM simulation. Studying the results carefully we can once again make several observations. In particular our specific choice of AUM as applied to Group D seems to yield the following scenarios:

6660 and 9333 stand out as the least likely configurations.

Then we have 4444 and 3333.

Due to the application of AUM, we no longer have the previous sets of EAC: different sets of EAC seem to emerge.

We observe 5432 and 7431 as the most accessible configurations.

The winner of the group will score:

7 points with ≈ 43.22% probability (the most likely scenario);

5 points with ≈ 29.36% probability (the second most likely scenario),

9 points with ≈ 13.4% probability, and

6 points with ≈ 11.85% probability.

	M	N	O	P	Q
3	Scenario	Frequency			Exp. Prob.
4	3333	4644	4644	1	0.00464
5	4443	13171			
6	4444	3882	17053	2	0.01705
7	5332	49020			
8	5432	63617			
9	5442	31772			
10	5443	19845			
11	5522	25427			
12	5531	38278			
13	5532	20063			
14	5541	35738			
15	5550	9843	293603	9	0.2936
16	6442	24324			
17	6443	17703			
18	6522	26431			
19	6541	27552			
20	6633	6047			
21	6641	13740			
22	6660	2695	118492	7	0.11849
23	7322	53538			
24	7422	44303			
25	7431	63399			
26	7432	25860			
27	7433	11793			
28	7441	32259			
29	7521	54882			
30	7531	28238			
31	7540	36916			
32	7621	32209			
33	7631	13988			
34	7640	17990			
35	7711	7811			
36	7730	9035	432221	14	0.432221
37	9222	19444			
38	9333	2955			
39	9421	52800			
40	9431	21577			
41	9440	14118			
42	9611	12020			
43	9630	11073	133987	7	0.133987
44	TOTAL	1000000	1000000	40	1

Figure 14: Frequency of each configuration for one million trials in the AUM.

6. Concluding Remarks

In the HTS categorization, our simplistic all-outcomes-are-equally-likely assumption seemed to offer reasonable simulation results, which were in agreement with the actual results for certain situations (see, for example, "*The group winner with 7 points*" perspective, and more generally, Table 6).

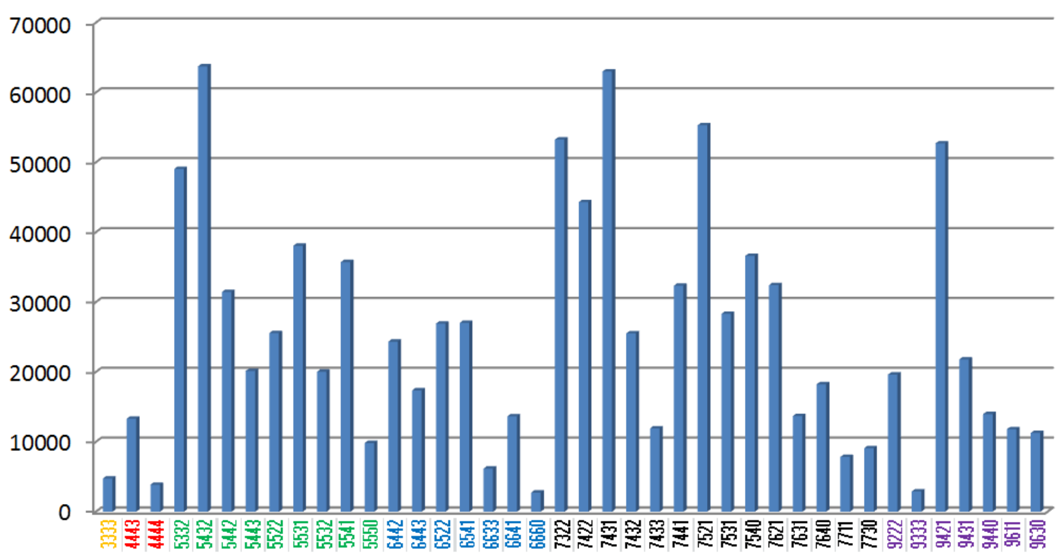

Figure 15: Frequency histogram of the forty distinct configurations in the AUM simulation.

In the EAC perspective, on the other hand, simulation results offered plausible results in all categories (Table 7). Yet, in the previous section, we went beyond that and suggested an adjusted urn model (AUM), which we then used to simulate the six group phase games of one particular group in the 2018 FIFA World Cup a million times. This model provided us with even more accurate results.

The adjusted urn model we proposed is just one model out of many that could be designed. Thus we could conceivably simulate other groups in the same competition or in different championships (e.g., UEFA, CONMEBOL, Copa America, CONCACAF, African Cup of Nations, AFC, OFC, UEFA Champions League, UEFA Europa League, etc.) by taking the success coefficients of each team in a given group into consideration in the group phase. This way, simulations could be expected to offer better results in agreement with the actual results.

Acknowledgments. The data and many of the images for this work were retrieved from http://www.fifa.com, last accessed on July 19, 2019.

A. Russia 2018 FIFA World Cup Group Stage Draw