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

Gunhan Caglayan
New Jersey City University

Follow this and additional works at: https://scholarship.claremont.edu/jhm

Part of the Arts and Humanities Commons, and the Mathematics Commons

Recommended Citation
Caglayan, G. "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).

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

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 a 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.

<table>
<thead>
<tr>
<th>Team</th>
<th>Wins</th>
<th>Draws</th>
<th>Losses</th>
<th>Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>Germany</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>USA</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Portugal</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Ghana</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

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.
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).\textsuperscript{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).

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.

\textsuperscript{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.

<table>
<thead>
<tr>
<th>Group</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>9630</td>
</tr>
<tr>
<td>B</td>
<td>5541</td>
</tr>
<tr>
<td>C</td>
<td>7531</td>
</tr>
<tr>
<td>D</td>
<td>9431</td>
</tr>
<tr>
<td>E</td>
<td>7531</td>
</tr>
<tr>
<td>F</td>
<td>6633</td>
</tr>
<tr>
<td>G</td>
<td>9630</td>
</tr>
<tr>
<td>H</td>
<td>6443</td>
</tr>
</tbody>
</table>

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

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

\[
= \text{MID(game, 1 + INT(RAND() * LEN(game)), 1)}
\]

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

\[
= 3*( \text{COUNTIF(A1,"W") + COUNTIF(C1,"W") + COUNTIF(E1,"L") } + \text{COUNTIF(A1,"D") + COUNTIF(C1,"D") } + \text{COUNTIF(E1,"D")})
\]

placed in cell G1.

Likewise, we compute Iceland’s points with

\[
= 3*( \text{COUNTIF(A1,"L") + COUNTIF(D1,"L") + COUNTIF(F1,"W") } + \text{COUNTIF(A1,"D") + COUNTIF(D1,"D") } + \text{COUNTIF(F1,"D")} )
\]
in H1. In a similar manner, Croatia’s uses

\[= 3*( \text{COUNTIF(B1,"W")} + \text{COUNTIF(C1,"L")} + \text{COUNTIF(F1,"L")} ) + \text{COUNTIF(B1,"D")} + \text{COUNTIF(C1,"D")} + \text{COUNTIF(F1,"D")}\]

in I1, and Nigeria uses

\[= 3*( \text{COUNTIF(B1,"L")} + \text{COUNTIF(D1:E1,"W")} ) + \text{COUNTIF(B1,"D")} + \text{COUNTIF(D1:E1,"D")}.\]

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

\[= 1000 * \text{LARGE(G1:J1,1)} + 100 * \text{LARGE(G1:J1,2)} + 10 * \text{LARGE(G1:J1,3)} + 1 * \text{LARGE(G1:J1,4)}\]

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.
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.
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 O4, (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.
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).

![Figure 7: Equally accessible configurations.](image)

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 $\frac{306}{729} \approx 42\%$ probability (the most likely event).
   (b) This team has 6 points with $\frac{152}{729} \approx 20.9\%$ probability.
   (c) This team has 5 points with $\frac{148}{729} \approx 20.3\%$ probability.
   (d) This team has 9 points with $\frac{108}{729} \approx 14.8\%$ probability.
   (e) This team has 3 or 4 points each with $\frac{15}{729} \approx 2\%$ probability.

3. The most likely configurations are 6443 and 7441, each with a $\frac{36}{729} \approx 5\%$ probability.
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.

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.

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 WDLDL. 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 $I_1=2$ points; and Nigeria scores $J_1=2$ points. However, $G_1$ and $J_1$ 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 $\frac{12}{729}$, $\frac{12}{729}$, $\frac{24}{729}$. This means that the “so far” configuration $5322$ will be finalized as $6522$, $5532$, or $5432$, with respective (conditional) probabilities $\frac{12}{12 + 12 + 24} = 25\%$, $\frac{12}{12 + 12 + 24} = 25\%$, $\frac{24}{12 + 12 + 24} = 50\%$. 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: $\frac{24}{729}$, $\frac{8}{729}$, $\frac{24}{729}$ respectively. This suggests that the “so far” configuration $6630$ will become $9630$, or $6660$, or $7640$ with respective (conditional) probabilities $\frac{24}{24 + 8 + 24} = \frac{3}{7}$, $\frac{8}{24 + 8 + 24} = \frac{1}{7}$, $\frac{24}{24 + 8 + 24} = \frac{3}{7}$. 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 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.
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 $\frac{36}{729}$ and $\frac{24}{729}$, 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 \( \frac{36}{36 + 24} = 60\% \) and \( \frac{24}{36 + 24} = 40\% \).

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 \( \frac{36}{36 + 24} = 60\% \), \( \frac{24}{36 + 24} = 40\% \) (same conditional probabilities as in the previous example).

5. Implications and Modifications


<table>
<thead>
<tr>
<th>Year</th>
<th>Actual configurations</th>
</tr>
</thead>
<tbody>
<tr>
<td>1994</td>
<td>4444, 6443, 6660, 6660, 7521, 7531</td>
</tr>
<tr>
<td>1998</td>
<td>5531, 6541, 6541, 7322, 7631, 7730, 9421, 9630</td>
</tr>
<tr>
<td>2002</td>
<td>5541, 7433, 7433, 7521, 7531, 7540, 9440, 9440</td>
</tr>
<tr>
<td>2006</td>
<td>7531, 7540, 7631, 7730, 9421, 9421, 9611, 9630</td>
</tr>
<tr>
<td>2010</td>
<td>5432, 5541, 6443, 6641, 7441, 7540, 9431, 9630</td>
</tr>
<tr>
<td>2014</td>
<td>7441, 7631, 7640, 7730, 9421, 9431, 9431, 9630</td>
</tr>
<tr>
<td>2018</td>
<td>5541, 6443, 6633, 7531, 7531, 9431, 9630, 9630</td>
</tr>
</tbody>
</table>

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:

- Configuration 9630 occurred 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 \( \frac{22}{54} \approx 40.74\% \) of the time (compare with \( \approx 42\% \) theoretical probability),
  6 points \( \frac{9}{54} \approx 16.67\% \) of the time (compare with \( \approx 20.9\% \) theoretical probability),
  5 points \( \frac{5}{54} \approx 9.26\% \) of the time (compare with \( \approx 20.3\% \) theoretical probability),
  9 points \( \frac{17}{54} \approx 31.48\% \) of the time (compare with \( \approx 14.8\% \) theoretical probability).

- The theoretically most likely configurations 6443 and 7441 each appeared
  \( \frac{3}{54} \approx 5.56\% \) and \( \frac{2}{54} \approx 3.70\% \) 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.

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

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

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

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 where 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.
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**, **LL-LLDW**, **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.
For each of A1, B1, C1, D1, E1, F1, the game result (W or D or L) is recorded using

$$=\text{MID(game1,1+INT(RAND() \times \text{LEN(game1)}),1)}$$

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.

![Figure 13: Adjusted Urn Model simulation.](image)

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 approximately 43.22% probability (the most likely scenario);
- 5 points with approximately 29.36% probability (the second most likely scenario);
- 9 points with approximately 13.4% probability, and
- 6 points with approximately 11.85% probability.
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).
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](http://www.fifa.com), last accessed on July 19, 2019.
A. Russia 2018 FIFA World Cup Group Stage Draw