The Canadian Journal of Statistics
Vol. 00, No. 00, 2023, Pages 000–000
La revue canadienne de statistique

Football group draw probabilities and corrections

Gareth O. ROBERTS[1] and Jeffrey S. ROSENTHAL[2*]

[1] *Department of Statistics, University of Warwick, Coventry, CV4 7AL, UK*
[2] *Department of Statistics, University of Toronto, Toronto, ON, M5S 3G3, Canada*

Key words and phrases: FIFA World Cup group draw; Monte Carlo simulation; sequential Monte Carlo; rejection sampler; sequential sampler; uniform draw.

MSC 2020: Primary 62-08; secondary 62P99.

Abstract: This article considers the challenge of designing football group draw mechanisms, which have a uniform distribution over all valid draw assignments, but are also entertaining, practical and transparent. Although this problem is trivial in completely symmetric problems, it becomes challenging when there are draw constraints that are not exchangeable across each of the competing teams, so that symmetry breaks down. We explain how to simulate the FIFA sequential draw method and compute the nonuniformity of its draws by comparison with a uniform rejection sampler. We then propose several practical methods of achieving the uniform distribution while still using balls and bowls in a way which is suitable for a televised draw. The solutions can also be carried out interactively. The general methodology we provide can readily be transported to different competition draws and is not restricted to football events.

Résumé: Dans cette étude, les auteurs se penchent sur le défi de concevoir des mécanismes de tirage au sort pour les groupes de football, visant à assurer une distribution uniforme des affectations valides, tout en étant divertissants, pratiques et transparents. Alors que ce problème semble anodin dans le contexte de problèmes complètement symétriques, il acquiert une complexité certaine lorsque des contraintes de tirage ne sont pas interchangeables entre les équipes concurrentes, entraînant ainsi une rupture de la symétrie. Les auteurs exposent dans ce travail la méthode de simulation du tirage séquentiel de la FIFA et évaluent la non-uniformité de ses tirages en les comparant à un échantillonneur par rejet uniforme. Par la suite, ils proposent plusieurs approches pratiques visant à parvenir à une distribution uniforme, tout en conservant l'utilisation de balles et de bols, adaptées à une diffusion télévisée. De plus, ces solutions peuvent être mises en œuvre de manière interactive. La méthodologie générale qu'ils présentent peut aisément être adaptée à d'autres tirages au sort de compétitions, sans aucune restriction quant à leur lien avec le football.

1. INTRODUCTION

Major football (soccer) tournaments such as the FIFA World Cup, European Championships and UEFA Champions League hold public draws to decide who plays whom. It is customary to employ one or more celebrities to draw balls from bowls to sequentially construct the draw, to add excitement and increase interest in the competition. However, such mechanisms can affect the draw probabilities in unexpected ways.

This article will focus on the group draw for the FIFA World Cup, although similar ideas could be applied to other competitions. This event involves teams from 32 different nations. The 2022 event was of particular interest to Canada, who qualified for the first time since 1986. Its group

* *Corresponding author: jeff@math.toronto.edu*

draw had various restrictions, based upon geographical constraints, about which assignments were permissible, leading to a complicated space of potential draws. In addition, the draw needed to take place sequentially, to allow for public interest and transparency. The statistical challenge, then, was to simulate from the uniform distribution on a nonsymmetric high-dimensional space in a way which is also sequential and entertaining. We shall present several potential solutions that we have developed to address this challenge; they are available for interactive use (Roberts and Rosenthal, 2022), and have been reported on in the media (Rosenthal, 2022b; Jones, 2022). See also their simple, informal descriptions in Section 7 below.

The methodology we develop is quite general and is readily transportable to different football competition draws, and indeed to other sports that carry out public draws of this kind. We will focus on the FIFA World Cup given its global appeal.

Sequential simulation of complex probability distributions has a long history of use in statistics (e.g., particle filtering, nested sampling, rare event simulation, simulated annealing, etc.). This article considers a rather different application of these techniques, which nevertheless require statistical thinking, design and analyses. We will thus utilize key state of the art techniques from the computational statistics literature such as stratified resampling for sequential Monte Carlo (Kitagawa, 1996), and retrospective simulation for the simulation of distributions with intractable probabilities, see, for example, Beskos et al. (2006).

1.1. The 2022 FIFA World Cup Group Draw

The 2022 FIFA World Cup took place in November/December 2022, in Qatar. It involved 32 national teams—31 who qualified through competition, plus the host team Qatar who qualified automatically. These 32 teams needed to be partitioned into 8 groups of 4 teams each, who would all play each other in the group stage of the competition.

The group assignments were subject to various restrictions as follows. The 32 teams were first partitioned into 4 seeded pots, with Pot 1 consisting of the host team Qatar plus the seven most highly ranked teams (according to the official FIFA national team rankings), Pot 2 consisting of the next 8 highest ranked teams, and so on (with the proviso that at the time there were three places still to be decided via various playoffs, and each of these undetermined slots were placed in Pot 4). In addition, each team is affiliated to one of the six continental federation regions: UEFA (Europe; 13 qualifying teams; henceforth **Eu**), CONMEBOL (South America; 4 or 5 teams; henceforth **SA**), CONCACAF (North and Central America; 3 or 4 teams; henceforth **NA**), AFC (Asia; 5 or 6 teams; henceforth **As**), CAF (Africa; 5 teams; henceforth **Af**) and OFC (Oceania; 0 or 1 team; henceforth **Oc**). The assigned pots were as follows:

Pot 1: Qatar[**As**], Belgium[**Eu**], Brazil[**SA**], France[**Eu**], Argentina[**SA**], England[**Eu**], Portugal[**Eu**], Spain[**Eu**].

Pot 2: Denmark[**Eu**], Netherlands[**Eu**], Germany[**Eu**], Switzerland[**Eu**], Croatia[**Eu**], Mexico[**NA**], USA[**NA**], Uruguay[**SA**].

Pot 3: Iran[**As**], Serbia[**Eu**], Japan[**As**], Senegal[**Af**], Tunisia[**Af**], Poland[**Eu**], KoreaRep[**As**], Morocco[**Af**].

Pot 4: Wales/Scot/Ukr[**Eu**], Peru/UAE/Au[**SA**,**As**], CostaRica/NZ[**NA**,**Oc**], Saudi Arabia[**As**], Cameroon[**Af**], Ecuador[**SA**], Canada[**NA**], Ghana[**Af**].

The reason for the uncertainty in three of the team names in Pot 4 is that, due to delays caused by Covid-19 and the war in Ukraine, not all teams had been finalized by the time of the draw, so placeholders were used. Two of the placeholders corresponded to two different potential regions, so their placements had to satisfy the geographical constraints for *both* of the corresponding regions.

 DOI: 10.1002/cjs.11798

In terms of these specifications, the restrictions on group formations were that each group needed to include one team from each of the four pots, and furthermore include either one or two teams from **Eu** plus either 0 or 1 teams from each of the other regions.

The FIFA group draw, held on 1 April 2022, then proceeded as follows (FIFA, 2022). First, the host team Qatar was automatically placed in Group A. Then, the remaining teams from Pot 1 were selected one at a time, uniformly at random, and each placed into the next group from B through H. Then, the teams from Pot 2 were selected one at a time, uniformly at random, and assigned to the next available group, that is, the first group that would not cause a conflict with any of the geographical restrictions (either immediately in the group where they were placed, or subsequently by making it impossible to validly fill in the remainder of the draw). This procedure was then repeated with Pot 3, and then with Pot 4 to conclude the draw. Each random selection was performed by a celebrity footballer, who picked a ball uniformly at random from a round bowl, and opened it to reveal the chosen team. (Each team was also randomly assigned a "position" within their group, to determine the order in which the matches would be played, but we do not consider that issue here.)

Without the geographical constraints, just drawing the teams in random order from the four pots in sequence, there would be $7! \times (8!)^3 \doteq 3.3 \times 10^{17}$ possible draws that could be produced. We shall see below that about 1 in 560 of these unconstrained draws satisfy the geographical constraints. Hence, the number of valid draws is approximately 5.9×10^{14}. A *uniform* draw is one for which all valid draws have an equal chance of materializing. This article will explore the nonuniformity in the FIFA 2022 draw procedure and also propose various methods of refining the draw to achieve complete uniformity.

1.2. Previous Literature

Unfairness of sporting rules is currently an active area of research in the operations research literature, see, for example, Kendall and Lenten (2017), Lenten and Kendall (2021), and Csató (2021b). Hidden biases in draw mechanisms provide good examples of this. It has long been known that sequential draws such as those adopted by FIFA and UEFA led to nonuniform probabilities, see, for example, Jones (1990), Rathgeber and Rathgeber (2007), Klößner and Becker (2013), Guyon (2015), and Csató (2021a). Various papers have looked at different mechanisms for carrying out different sorts of sequential draws; see, for example, Csató (2021a). Much of the literature has focused on how to obtain a *balanced* draw rather than a uniform one, that is, trying to make each group roughly equal in strength, see, for example, Guyon (2015), Guyon (2018), Csató (2023b), Cea et al. (2020), and Laliena and López (2019). This literature has been successful in influencing rules for football draws. Indeed, balance is what inspired FIFA to create the seeded pots based on world rankings rather than continental affiliation, for both the 2018 and 2022 World Cup group draws. However, seeking balance by adjusting the constraints within the draw is a separate question from the question of obtaining uniformity without changing constraints, which is the focus of this article.

Other papers consider other related issues, such as how teams' incentives to perform well are affected by tournament designs (Guyon, 2022) and by group draws (Csató, 2022).

One might wonder whether nonuniformity really matters in practice, or whether it is merely of theoretical interest. However, a study based on the UEFA Champions League (Klößner and Becker, 2013, p. 262) shows that even small probability differences can translate into quite substantial financial differences in expected revenues for different clubs. For the 2018 FIFA World Cup, Csató (2021a) finds that small biases can also significantly affect nations' progression probabilities beyond the initial group stage.

Various papers have proposed solutions to the nonuniformity problem. Boczon and Wilson (2022) show that the UEFA mechanism is close to constrained-best. Guyon (2014) (see also Guyon (2015)) proposes three innovative solutions, each of which is transparent using only balls and bowls (with no multiple bowls) and very promising, and in one case was actually adopted by FIFA. However, these solutions have certain acknowledged limitations: their Suggestion 1 does not lead to a uniform draw; their Suggestion 2 is uniform but requires intractable preliminary computation to list all possible continental distributions; and their Suggestion 3 adds an additional "S-curve-type" constraint, which increases group balance but significantly decreases the number of admissible draws, so it produced a tractable transparent draw, which is uniform but only on a different (much smaller) draw set. Also, Klößner and Becker (2013) briefly postulate a possible Markov chain procedure (specifically an exclusive Gibbs sampler) for the UEFA Champions League draw. However, none of these methods respect the desired sequential nature of the draw in a practically feasible way while also achieving perfect uniformity over all draws satisfying the specified FIFA constraints, which is our goal here.

1.3. Summary of Main Contributions and Structure of Paper

Our main contributions in this article are as follows:

1. We explicitly calculate biases inherent in the 2022 FIFA World Cup draw procedure.
2. We propose three practically implementable solutions to completely remove bias. All of these methods could potentially be used by FIFA (and other sporting bodies) for improved future draws, no matter what (hard) constraints are imposed, since they involve considering different potential draws and then checking to see whether or not they are valid.
3. We give prototype software to illustrate our solutions.
4. We provide a user's guide to sporting bodies to suggest how these solutions could be implemented in ways which are fairly similar to current procedures.

In Section 2, we investigate the extent of the bias in the 2022 FIFA draw, comparing a simulation of the FIFA draw mechanism with uniform draws, which are achieved by rejection sampling. In Section 3, we attempt to demystify this bias created by the FIFA mechanism via a simplified example and investigate how multiple-ball procedures could potentially be utilized to circumvent the problem. The next three sections are devoted to our solutions: Section 4 describes the Metropolis (swap) solution; Section 5 introduces our multiple-ball solution, while a multiple-rejection solution is provided in Section 6. Section 7 provides informal, nontechnical descriptions of our two most practical methods. We conclude with a discussion of various related issues in Section 8.

2. COMPARING UNIFORM AND FIFA PROBABILITIES

Before proposing alternative solutions, we investigate the extent to which the FIFA sequential algorithm (described in Section 1.1 above) is nonuniform.

A preliminary look illustrates the nature of the problem. For example, consider the question of whether the United States is assigned to Group A with Qatar. (This is an important question, since Qatar is weaker than the other teams in Pot 1, so Group A is the most desirable placement.) Under the FIFA method, any of the eight teams in Pot 2 is equally likely to be selected first, and none of them have any regional conflict with Qatar, so the United States has probability exactly 1/8 or 12.5% of being placed in Group A. But under the uniform distribution, the United States should be *less* likely to land in Group A, because that leaves fewer ways for the numerous **Eu** teams to be placed in the other groups. (Indeed, we shall see below that the uniform distribution gives a probability of approximately 9.0% to the United States being in Group A, which is significantly smaller.) These calculations reinforce the findings of Csató (2023b, Figure 1), which illustrated that the average strength of Group A is much smaller compared with the average strength of the other seven groups.

To compute the differences between the two distributions, we shall use a Monte Carlo approach. Thus, we will repeatedly simulate from both the uniform distribution **U** and the distribution **P** of draws created by the FIFA sequential method. We consider each of these problems in turn in Sections 2.1 and 2.2 below.

Note that refinements on our basic Monte Carlo approach are possible, for instance, we could utilize the exchangeability between teams in each pot from the *same* continental confederation. Thus, for instance, we know that (for both **U** and **P**) the probabilities of Mexico being placed in Group 1 with Qatar is identical to that for the United States being placed in Group 1 with Qatar. Using these symmetries can improve the Monte Carlo accuracy. However, in practice we found that we could achieve high levels of accuracy without needing these symmetries.

2.1. Uniform Simulation by Rejection Sampling

To obtain draws with equal probabilities, the simplest way is to use a rejection sampler. This algorithm proceeds as follows:

1. First, assign a draw completely randomly, without regard to the geographical constraints. That is, select uniformly at random from the $7! \times (8!)^3 \doteq 3.3 \times 10^{17}$ possible unconstrained draws discussed in Section 1.1 (which is straightforward).
2. Then, compute the number of teams from each region in each group, and check whether the geographic constraints are satisfied (i.e., 1 or 2 **Eu** teams plus 0 or 1 from each other region in each group).
3. If the constraints are satisfied, then output the chosen draw.
4. If the constraints are *not* satisfied, then reject the chosen draw, and start again at Step 1 above with a fresh unconstrained draw.

It is well known (e.g., Devroye, 1986, Section II.3; see also Guyon, 2014) that this rejection sampler produces a draw which is uniform over all possible valid draws, that is, which is distributed as the uniform distribution **U**. Then, by sampling many times and averaging the results (Rosenthal, 2022a), we can get a good estimate (by the Law of Large Numbers) of the expected value according to **U** of any functional, or the probability of any particular event.

In addition, our simulations (Rosenthal, 2022a) show that about 1 in 560 of the proposed draws were accepted, that is, satisfied the geographical constraints. This means that the total number of valid draws is approximately $7! \times (8!)^3 \big/ 560 \doteq 5.9 \times 10^{14}$.

2.2. Simulating the FIFA Sequential Draw

Although the FIFA sequential draw method is easy to describe (see Section 1.1), and can usually be implemented without difficulty, it is surprisingly challenging to simulate with a computer program (and neither FIFA nor UEFA provides an algorithm for this). This is because potential group assignments have to be skipped whenever they would conflict with the geographical constraints.

Immediate conflicts are simple to detect. Indeed, if adding a team to a group would create a third **Eu** team, or a second team from any other region, then obviously that group must be skipped. However, subsequent conflicts are much more complicated. It may be that assigning a certain team to a certain group would then make it impossible to fill in the rest of the draw in a valid way, either due to being forced to later add a third **Eu** team or second team from another region to a group, or failing to add at least one **Eu** team to a group. These subsequent conflicts are usually relatively simple to resolve (and indeed, they did not occur at all in the actual FIFA 2022 draw), but this is not guaranteed.

For an extreme example, suppose a draw in progress has teams from the various regions as follows:

Group	A	B	C	D	E	F	G	H	To go
Pot 1	**Af**	**NA**	**NA**	**NA**	**NA**	**NA**	**NA**	**NA**	
Pot 2	**As**?								**SA**,**SA**,**SA**,**SA**,**SA**,**SA**,**SA**
Pot 3									**Eu**,**Eu**,**Eu**,**Eu**,**Eu**,**Eu**,**NA**,**SA**
Pot 4									**Oc**,**Oc**,**Oc**,**Oc**,**Oc**,**Oc**,**Oc**,**Oc**

Then placing the **As** team from Pot 2 into Group A already creates a subsequent conflict, since the seven **SA** teams from Pot 2 must then be placed in Groups B through H, and then in Pot 3 the **NA** and **SA** teams cannot both be placed in valid groups without violating their respective geographic constraints. Instead, the **As** team in Pot 2 must be placed in some other group besides Group A, thus leaving Group A for the **NA** team from Pot 3, and the other group for the **SA** team from Pot 3. However, this subsequent conflict might not become apparent until the end of the Pot 3 placements.

This example illustrates that a general program to simulate the FIFA method must be robust enough to detect subsequent conflicts many steps later. Our approach (Rosenthal, 2022a) was to program this recursively (as also discussed in, e.g., Guyon, 2017b). We created a subroutine `placerest()`, which, given a partial draw, randomly selects a team to place next, and then attempts to place that team in the next available slot. It then recursively calls itself with that one additional placement, to see whether the rest of the draw can then be filled in successfully. If it can, then the draw is complete. If it cannot, then it instead attempts to place that team in the subsequent available slot. Continuing in this way, it eventually successfully places every team. The recursive nature of the program ensures that any subsequent conflicts will be dealt with, no matter how much further along they occur; for details see the computer program (Rosenthal, 2022a). Related backtracking algorithms are presented in Guyon (2015), and in Csató (2021a), which also makes connections to the well-known computer science problem of generating all permutations of a given sequence.

Thus, this algorithm produces a draw with probabilities as in the current FIFA sequential method, that is, which is distributed as the FIFA distribution **P**. Then, by sampling many times and averaging the results (Rosenthal, 2022a), we can get a good estimate (by the Law of Large Numbers) of the expected value according to **P** of any functional, or the probability of any event.

2.3. Specific Probability Comparisons

Armed with our computer program (Rosenthal, 2022a), we can now compare the probabilities of various events according to the uniform distribution **U** and the FIFA sequential method **P**. We proceed by running each method one million times, and taking the fraction of successes as our point estimate. We also compute corresponding 95% confidence intervals, using the confidence interval for a proportion formula $p \pm 1.96\sqrt{p(1-p)/n}$ for sample size n with success proportion p.

For example, the probability that England is in the same group as Germany should be 10.53% (95%CI=[10.47%,10.59%]) under **U**, but increases to 11.78% (95%CI=[11.72%,11.85%]) under **P**. The probability that Germany is in the same group as Qatar should be 13.74% (95%CI=[13.67%,13.80%]) under **U**, but decreases to 12.50% (exact) under **P**. The probability that Canada is with Qatar should be 15.53% (95%CI=[15.46%,15.60%]) under **U**, but increases to 16.51% (95%CI=[16.44%,16.58%]) under **P**. And the probability that the United States is with Qatar should be just 9.06% (95%CI=[9.00%,9.11%]) under **U**, but increases to 12.50% (exact) under **P** (a relative increase of 38%). Probabilities of other events can similarly be computed (Rosenthal, 2022a).

These probability differences are not huge, but they are large enough to illustrate that the FIFA method is significantly different from a uniform draw and could lead to significantly different group assignments. We now consider how to fix this problem, to achieve uniform group draws while still preserving the interest and excitement of the FIFA method.

3. A MOTIVATING EXAMPLE

To see more clearly why FIFA's sequential procedure described in Section 1.1 above fails to achieve uniformity, and how we might correct that, consider a simplified setup with only six teams to be allocated to three groups of two teams each. Suppose there are two seeded pots (a subset of the actual World Cup 2022 pots):

Pot 1: Qatar (Q) [**Af**], France (F) [**Eu**], Brazil (B) [**SA**]
Pot 2: Mexico (M) [**NA**], Switzerland (S) [**Eu**], Uruguay (U) [**SA**]

Qatar, as host, were automatically pre-assigned to Group A. Without loss of generality, assume that France is placed in Group B and Brazil in Group C. Then the two **SA** teams, Brazil in Pot 1 and Uruguay in Pot 2, must be kept apart, so Uruguay cannot be placed in Group C. There are thus four possible draws for assigning the three different groups:

D_1: QM, FU, BS.
D_2: QS, FU, BM.
D_3: QU, FM, BS.
D_4: QU, FS, BM.

Let **P** be the probability measure resulting from FIFA's sequential procedure described above but adapted to this simplified setting, and let **U** be the probability measure that assigns equal probability to each valid draw. Thus, $\mathbf{U}(D_i) = 1/4$ for $i = 1, 2, 3, 4$. However, a simple calculation gives that $\mathbf{P}(D_1) = \mathbf{P}(D_2) = 1/3$ while $\mathbf{P}(D_3) = \mathbf{P}(D_4) = 1/6$. In particular, if QU is the event that Qatar is paired with Uruguay, then $\mathbf{P}(QU) = 1/3$, whereas $\mathbf{U}(QU) = 1/2$. This clearly illustrates the potential nonuniformity arising from FIFA's sequential procedure, as also noted elsewhere, for example, Guyon (2014, Section 3).

3.1. In Search of Debiasing: Random Order Sequential Procedures

It is natural to ask how to fix the biases of the above example while retaining the sequential nature of the draw. Perhaps the most obvious approach is to randomize the order in which opponents for the Pot 1 teams are found. In certain cases, judicious choices of ordering and randomness can reduce bias (e.g., see Csató (2023a), which also includes an overview of group draws in sports tournaments under constraints), but here we shall investigate whether such a procedure can provide generic solutions to the problem.

In the usual FIFA method, the first drawn team in Pot 2 is first attempted to be placed in Group A with Qatar. To emphasize that we start with Qatar, write $\mathbf{P}_Q$ for the probabilities induced by this method (so $\mathbf{P}_Q$ is the same as the FIFA probabilities **P**). If instead the first drawn team in Pot 2 is attempted to be placed in Group B with France, then this leads to different probabilities $\mathbf{P}_F$, and similarly $\mathbf{P}_B$ if the first team is attempted to be placed in Group C with Brazil. Alternatively, if the first drawn team in Pot 2 is attempted to be placed in Group A or B or C with probability 1/3 each, then such a procedure could be called a *random order sequential draw*, with resulting probabilities given by $\mathbf{P}_{rand} = \frac{1}{3}(\mathbf{P}_Q + \mathbf{P}_F + \mathbf{P}_B)$. In principle, we could also specify that the *second* team drawn from Pot 2 should have their attempted group randomized among the two remaining teams, but this example is sufficiently simple that we can ignore that.

In this example, it can be checked that $\mathbf{P}_Q$ and $\mathbf{P}_F$ are both nonuniform, but $\mathbf{P}_B$ happens to be uniform, i.e., $\mathbf{P}_B = \mathbf{U}$. Moreover, a simple calculation yields that

$$\mathbf{P}_{rand} = \frac{1}{3}(\mathbf{P}_Q + \mathbf{P}_F + \mathbf{P}_B) = \mathbf{U},$$

i.e., the random order sequential draw is uniform in this case. This leads to the question of whether random order sequential draws are always uniform.

To answer this question, consider a modification of our simple example. Suppose that we also forbid pairings between two **Eu** teams, thus disallowing D_4 (which pairs France with Switzerland). In that case,

$$\mathbf{U}(D_1) = \mathbf{U}(D_2) = \mathbf{U}(D_3) = 1/3.$$

For this modified example, $\mathrm{P} = \mathbf{P}_Q$ is easily seen to be uniform, but $\mathbf{P}_F$ and $\mathbf{P}_B$ both turn out to be nonuniform. Moreover, for the random order sequential draw $\mathbf{P}_{rand} = (1/3)(\mathbf{P}_Q + \mathbf{P}_F + \mathbf{P}_B)$, we calculate that

$$\mathbf{P}_{rand}(D_1) = 5/18; \quad \mathbf{P}_{rand}(D_2) = 13/36; \quad \mathbf{P}_{rand}(D_3) = 13/36\ .$$

So, in this case, FIFA's fixed order draw P_Q is uniform, while the randomized order draw P_{rand} is nonuniform. This shows that randomized order draws do not solve the nonuniformity problem. In fact, they can sometimes make the nonuniformity worse, or even introduce nonuniformity in cases where the original FIFA algorithm happened to be uniform. These findings concur with the findings of Guyon (2017a) and Csató (2021a) that the order of pots can have a non-negligible effect on the size of the bias in the draw mechanism, and that randomizing the order cannot generically solve the bias problem.

3.2. In Search of Debiasing: Multiple-Ball Procedures

We now return to the original motivating example (at the beginning of Section 3), where the only constraint is on the two **SA** teams. Then we see from Section 3.1 that two of the possible draws (D_3 and D_4) put Uruguay in Group A with Qatar, while only one (D_1) puts Mexico in Group A, and one (D_2) puts Switzerland in Group A. This suggests that, to achieve the uniform distribution $\mathbf{U}$, when selecting the team from Pot 2 to put in Group A, we could use a bowl with *two*Uruguay balls, and only one from each of Mexico and Switzerland. The next team in the draw is then selected uniformly at random from those four balls, just as before. This simple "multiple balls solution" thus achieves the correct conditional probability (in terms of $\mathbf{U}$) for the team from Pot 2 to be placed in Group A.

More formally, this solution could be described as follows. When selecting the team from Pot 2 to put in Group A, we first count, for each team in Pot 2, the number of valid draws which have that team in that position while also keeping all of the previously selected teams in their previously selected positions. In our motivating example, we have $n_U = 2$ for Uruguay, and $n_M = 1$ for Mexico, and $n_S = 1$ for Switzerland. We then place n_U balls for Uruguay, n_M balls for Mexico and n_S balls for Switzerland, all into a bowl, and then sample one of the balls uniformly at random. In this way, we select Uruguay with probability $n_U/(n_U + n_D + n_S)$, and so on. This ensures the correct conditional probability for the next position. Hence, if this procedure is repeated for each new position, then it ensures the correct full uniform probability $\mathbf{U}$ for the entire draw.

Can this procedure be extended to larger draws? In principle, yes. However, the number of balls would soon get completely out of hand. For instance, in the World Cup 2022 draw, when choosing the Pot 2 team to play Qatar, we would need to put about $(8!)^3/560 \doteq 10^{11}$ balls into the bowl, clearly impossible in practice. Such pure multiple-ball draws were already considered in Guyon (2015), which also noted their impracticality in larger draws. Nevertheless, we shall see in Section 5 below that we can exhibit a practical and completely uniform multiple balls solution to this problem using internal random computer simulations, with far smaller total numbers of balls.

4. A METROPOLIS (SWAP) ALGORITHM SOLUTION

The rejection sampler algorithm of Section 2.1 provides a perfectly uniform group draw distributed according to **U**, so in some sense, it completely solves the problem. However, the use of this algorithm would require the public to "trust" the computer to sample correctly, and would not provide any drama or entertainment value during the draw. So, we next consider ways to achieve a more interesting and entertaining and transparent draw while still preserving the uniform probabilities **U**.

One solution is as follows. Begin with (say) a rejection sampler uniform sample as above. Then, repeatedly choose two teams at random from the same pot, and "swap" their group assignments provided that swap does not violate any of the geographical constraints. (If the swap would violate any constraint, then the group assignments are left unchanged.) The use of a Metropolis algorithm like this was first suggested but not implemented in the context of the UEFA Champions League knockout draw by Klößner and Becker (2013). Unlike the other methods, it does not draw teams sequentially, but rather continually modifies the entire draw until settling on a final version.

The validity of this Metropolis solution follows from:

Proposition 1. *If we begin with a valid draw chosen from* **U** *(e.g., using a rejection sampler as in Section 2.1), and then repeatedly perform a fixed number of swap moves as described above, then the distribution of the draw remains equal to the uniform distribution* **U**.

Proof. See the Appendix. ∎

The swap moves used by this method can easily be performed manually, by choosing pairs of teams by drawing balls from bowls, and checking directly if any geographic constraints would be violated by swapping them. Unlike the FIFA sequential draw of Section 2.2, there are no subsequent conflicts, just immediate conflicts, so they can be easily checked.

As more swaps are performed, the overall group assignment continues to change randomly, in unexpected and entertaining ways which could make for an exciting spectacle. A somewhat related exclusive Gibbs sampler for the UEFA Champions League draw was proposed in Klößner and Becker (2013). If desired, a large number of initial swaps could be performed quickly by a computer. Then, a certain fixed number of final swaps could be performed manually, by physically selecting balls from bowls to determine which two teams are selected next for possible swap. The final assignment would then be whatever configuration remains after the final manual swap has been performed.

An interactive simulation of this method applied to the 2022 FIFA World Cup is available at Roberts and Rosenthal (2022).

5. A MULTIPLE-BALL SOLUTION

We now investigate how to generalize the multiple-ball method introduced in Section 3.1 to a full World Cup draw.

This method fills in the groups one team at a time. At each step, the computer generates a collection of balls corresponding to all the teams who could potentially occupy the next spot.

One of those balls is then chosen uniformly at random, and that team is placed in the next spot. Once all spots are filled, it provides a complete group draw.

Recall that, for this method to generate the uniform distribution **U**, the next team needs to be selected from its correct conditional probability according to **U**. That is, given a partial draw, if n_j is the number of ways of completing the draw which put team j in the next position, then we should select team j with probability proportional to n_j.

This approach immediately presents several challenges. How could we compute the n_j values, or at least the corresponding probabilities $p_j = n_j/S$ where $S = \sum_i n_i$? And even if we knew the values of the p_j, how could we sample with probabilities proportional to them, in a practical way without needing a massive collection of S different balls? We first consider the second problem, of sampling with probabilities p_j (Section 5.1). We then explain how it suffices to use *estimates* of the p_j, so that they do not need to be computed analytically (Section 5.2). This leads to a practical algorithm for conducting the draws (Section 5.3).

5.1. Discrete Random Rational Simulation

Consider the problem of simulating a discrete random event from J possible values with given non-negative probabilities $p_1, \ldots, p_J$ summing to 1. Call this distribution **D**. We wish to sample from **D** using a *rational simulation algorithm*, meaning we select some small non-negative integer number m_j of balls representing each value j, and then draw each ball with probability $1/M$ where $M = \sum_{j=1}^{J} m_j$.

Now, if the values of the p_j were small integer multiples of each other, then this would be easy. For example, if $p_1 = 1/2$, $p_2 = 1/3$, and $p_3 = 1/6$, then we could choose $m_1 = 3$, $m_2 = 2$ and $m_3 = 1$, and then drawing uniformly at random from the $3 + 2 + 1 = 6$ balls would accomplish our task. However, we do not wish to assume that the p_j are rationally related, and we want the total number of balls M to remain moderate even if the ratios of the p_j have no simple fractional form. We therefore propose the following algorithm. Our solution uses the stratified sampling strategy common in sequential Monte Carlo (Kitagawa, 1996; Fearnhead, 1998). Other resampling strategies could also be used; see Douc and Cappé (2005) for a discussion of options.

1. Let $M = \lceil \max\{(1/p_j) : p_j > 0\} \rceil$ be the ceiling of reciprocals of the nonzero p_j values.
2. Set $r_j = M\, p_j$, for $1 \le j \le J$. (Note that our choice of M is the smallest one which ensures that $r_j \ge 1$ for each j with $p_j > 0$, which guarantees at least one ball of type j. Other values of M, either smaller or larger, could also be used.)
3. For each $1 \le j \le J$, place $a_j := \lfloor r_j \rfloor$ balls of type j into the bowl.
4. Set $u_j = r_j - a_j$, and $v_j = \sum_{\ell=1}^{j} u_\ell$ for $1 \le j \le J$, with $v_0 = 0$. Also set $K = v_J$. Thus, $K = \sum_j u_j = \sum_j r_j - \sum_j a_j = M - \sum_j a_j$ which is a non-negative integer.
5. Simulate independent uniform random variables $U_1, \ldots, U_K$ with $U_i \sim \text{Uniform}[i-1, i)$.
6. For each $1 \le j \le J$, let $b_j = \#\{i : U_i \in [v_{j-1}, v_j)\}$ be the number of random variables U_i which lie in the interval $[v_{j-1}, v_j)$, and add b_j additional balls of type j to the bowl. (Note that we must have $0 \le b_j \le 2$, and $\sum_j b_j = K = M - \sum_j a_j$. Furthermore, the total number of balls of type j is $m_j = a_j + b_j$, so the total number of balls in the bowl is $\sum_j m_j = \sum_j a_j + \sum_j b_j = \sum_j a_j + (M - \sum_j a_j) = M$.)
7. Select a ball uniformly at random from the M balls in the bowl.

Proposition 2. *Given a collection $p_1, \ldots, p_J$ of non-negative probabilities summing to 1, the above procedure selects a ball of type j with probability p_j.*

Proof. See the Appendix. ■

5.2. Estimating the Probabilities

To use the previous algorithm for group draws would require that we know the conditional probabilities $p_j = n_j/S$ for the next team to be chosen in a partially completed group draw, where n_j is the number of ways of completing the draw with team j in the next position, and $S = \sum_i n_i$.

Now, it might perhaps be possible to compute $\{n_j\}$ directly. We first observe that the number of possible completions depends only on the geographic region match-ups of the various teams, not on the actual team names (cf. Guyon, 2015). So, in the 2022 FIFA World Cup as described in Section 1.1, Pot 1 can be distributed arbitrarily without affecting the subsequent n_j. Then, all we need to know about the Pot 2 teams is how many **Eu** teams were put in the same group as an **Eu** team from Pot 1, whether the two **NA** teams were put with **Eu** or **As** or **SA** teams, and so on. This leads to cascading combinatorial counts for the number of ways of putting different regions into different positions in the draw. Then, multiplying by corresponding factorials gives the numbers of ways of placing actual teams into the draw.

Such combinatorial problems can eventually be solved. However, they quickly become rather complicated and messy. Furthermore, they have to be re-computed for each possible partial draw, and the calculations are entirely different for different regional distributions of the teams in each pot. So, this does not appear to be a feasible way of proceeding.

Fortunately, there is a practical alternative. Proposition 2 remains true if the algorithm in Section 5.1 instead uses values $\hat{p}_j$ which are unbiased estimates of the p_j. That is:

Proposition 3. *Given a collection $\hat{p}_1, \ldots, \hat{p}_J$ of non-negative unbiased estimators summing to 1, with $\mathrm{E}(\hat{p}_j) = p_j$, if we replace p_j by $\hat{p}_j$ throughout the algorithm in Section 5.1, then it will still select a ball of type j with probability p_j.*

Proof. See the Appendix. ■

Remark 1. The approach of using an estimator $\hat{p}_j$ for simulation, without knowing the true value p_j, is the defining feature of a collection of simulation techniques known as retrospective sampling, see Beskos et al. (2006). However, our use of these methods here is rather different from those in the literature, which have focused largely on Bayesian inference and exact diffusion simulation problems.

The advantage of Proposition 3 is that there is a natural way to estimate the values of the p_j in an unbiased way: classical Monte Carlo samples. That is, given a partial draw, we can generate a large number N of valid draw completions using a rejection sampler similar to Section 2.1 above. In fact, it is even easier, since part of the draw is already chosen and does not need to be sampled. Then, we can estimate p_j by the fraction of those valid draw completions which have team j in the next position. This gives a good unbiased estimator $\hat{p}_j$ of the true conditional probability p_j that team j would be placed in the next position according to the uniform distribution **U**. That provides the final piece of the puzzle for us to produce an effective multiple balls uniform draw sample, which we now present.

5.3. Generating Draws with Multiple Balls

Combining the previous ideas together gives the following algorithm for choosing the team in the next position of a uniform group draw with distribution **U**, by drawing uniformly at random from a modest number of balls, as follows:

1. Select a positive integer valued algorithm parameter N.

2. Simulate N different uniformly distributed completions of the current partial draw, using a rejection sampler.
3. For each team j, let n_j be the number of such completions which have team j in the next position, and set $\hat{p}_j = n_j/N$.
4. Let $M = \lceil \max\{(1/\hat{p}_j) : \hat{p}_j > 0\} \rceil$ be the ceiling of the reciprocals of the nonzero estimates $\hat{p}_j$.
5. Set $r_j = M\ \hat{p}_j$, for $1 \le j \le J$. (So, $r_j \ge 1$ whenever $n_j > 0$.)
6. For each $1 \le j \le J$, place $a_j := \lfloor r_j \rfloor$ balls of type j into the bowl.
7. Set $u_j = r_j - a_j$, and $v_j = \sum_{\ell=1}^{j} u_\ell$ for $1 \le j \le J$, with $v_0 = 0$. Also set $K = v_J$.
8. Simulate independent uniform random variables $U_1, \ldots, U_K$ with $U_i \sim \text{Uniform}[i-1, i)$.
9. For each $1 \le j \le J$, let $b_j = \#\{i : U_i \in [v_{j-1}, v_j)\}$ be the number of random variables U_i which lie in the interval $[v_{j-1}, v_j)$, and add b_j additional balls of type j to the bowl (so that the total number of balls of type j is $m_j = a_j + b_j$).
10. Select a ball uniformly at random from the M balls in the bowl.

Consider a simple illustration based on the motivating example in Section 3. Although in this example we can easily compute the true p_i probabilities, we shall assume that we cannot, so that we still need to carry out a Monte Carlo experiment to estimate them. Assume that we have allocated Qatar to Group 1, France to Group 2 and Brazil to Group 3. Now to be consistent with the notation above, we (arbitrarily) number the Pot 2 teams: (Mexico, Switzerland, Uruguay) = $(1, 2, 3)$. Then, when choosing which Pot 2 team to allocate to Group A, we have

$$p_1 = 1/4; \quad p_2 = 1/4; \quad p_3 = 1/2.$$

Suppose we pick $N = 100$, and then carry out 100 simulations from a multinomial with these probabilities, obtaining $(24, 29, 47)$ observations of each of the three types. Then $(\hat{p}_1, \hat{p}_2, \hat{p}_3) = (0.24, 0.29, 0.47)$, and the largest value of $1/\hat{p}_j$ is $1/0.24 = 4.167$. The algorithm therefore dictates that we choose $M = \lceil 4.167 \rceil = 5$. So we shall have 5 balls, and next need to choose what type they should be. Here $M(\hat{p}_1, \hat{p}_2, \hat{p}_3) = (\mathbf{1}.2, \mathbf{1}.45, \mathbf{2}.35)$, so we initially allocate $(1, 1, 2)$ balls of types 1,2,3 respectively. We then assign the final ball using the remainder probabilities $(u_1, u_2, u_3) = (0.2, 0.45, 0.35)$, with $K = \sum_i u_i = 1$.

Returning to the method in general, the validity of this method is ensured through the following result. It follows from Proposition 3 that:

Proposition 4. *If the above procedure is used sequentially for each position of a group draw, then the final group draw will have distribution* $\mathbf{U}$, *i.e., will be equally likely to be any of the potential valid draws.*

An interactive simulation of this group draw generation method for the 2022 FIFA World Cup is available at Roberts and Rosenthal (2022).

Remark 2. The above procedure can be improved in a few minor ways. For example, the true conditional probabilities p_j must be the same for all teams in the same geographical region, so it is possible after step 3 to replace each $\hat{p}_j$ by the average of the $\hat{p}_i$ over all teams in the same region as team j. Also, if it happens after step 9 that the final numbers of balls m_j have a nontrivial common factor, i.e., $gcd\{m_j\} > 1$, then each m_j can be divided by this common factor to produce a simpler draw that still maintains the same probabilities.

5.4. Managing M

From a practical point of view, it is important to know how large the number of balls M could be. The sporting authority could be severely embarrassed if they ran out of balls for a particular team! Indeed, the practical feasibility of the draw depends on M not being too large, which could be a concern since the algorithm in Section 5.3 computes M randomly.

To illustrate this, return to the motivating example of Section 3. We recall that we wanted to carry out a Monte Carlo simulation of $N = 100$ draws from a trivariate random variable with probabilities

$$p_1 = 1/4; \quad p_2 = 1/4; \quad p_3 = 1/2.$$

Suppose we obtained (unusual) counts such as $(3, 36, 61)$. This would lead to a choice of $M = \lceil \max\{(1/\hat{p}_j) : \hat{p}_j > 0\} \rceil = \max\{1/(3/100), 1/(36/100), 1/(61/100)\} = 34$. In fact it is easy to see that the maximum possible M value is 100 in this case, and equals N in general. Fortunately this is vanishingly unlikely, and indeed by taking N sufficiently large, we can ensure that M is small with high probability. For instance, for $N = 100$, we can directly compute that $\mathbf{P}(M \geq 6)$ is about 0.042, while for $N = 1000$ that probability decays to less than 10^{-11}.

Unfortunately, the required N for making the probability that M is large depends crucially on the probabilities $\{p_i\}$. For instance, suppose p_1 was the smallest probability at 0.1. Then M will usually have to be at least 10 and we require $N = 1000$ to ensure that $\mathbf{P}(M \geq 15)$ is around 10^{-4}.

In the case of the FIFA World Cup draw, we have the additional complication that we do not actually know the p_is so we cannot predict the size of N required to make $\mathbf{P}(M \geq k)$ sufficiently small. Empirically we have observed in simulations that the maximum M value is extremely rarely above 16, and for any future draw we could carry out a similar empirical study. However, this does not provide any guarantees.

There is, however, a modification of the algorithm in Section 5.3, which allows us to specify a maximum possible M, say $M_{\max}$. We simply replace Step 4 there by

4.′ Let $M = \min\{M_{\max}, \lceil \max\{(1/\hat{p}_j) : \hat{p}_j > 0\} \rceil\}$.

We then still set $r_j = M\, \hat{p}_j$ for each j as before. This guarantees that we will never use a total of more than $M_{\max}$ balls, and the method remains completely valid. The only disadvantage of this method is that now it is possible to get $r_j = 0$ balls even though $n_j > 0$, though this will only occur if $1/\hat{p}_j > M_{\max}$, which is quite unlikely for reasonable choices of $M_{\max}$.

One special case of this algorithm is when $M_{\max} = 1$. In this case, only a single ball is produced, and all the random simulation takes place within the computer, which is clearly undesirable. By contrast, as $M_{\max}$ gets larger, more of the potential teams will get at least one ball. In general, a larger M means a larger percentage of the balls are from the deterministic a_j integer part, with a smaller percentage from the random b_j simulation part, leading to less randomness and more balanced ball counts and thus a more transparent-seeming draw. In practice, the user should choose $M_{\max}$ judiciously: not too large for reasons of practicality, and not too small for transparency and to avoid the $r_j = 0$ case, such as $M_{\max} = 20$.

6. A MULTIPLE-REJECTION SOLUTION

Finally, we present a somewhat different solution, which still selects the teams sequentially one position at a time, still uses uniform draws from bowls with modest sized numbers of balls representing the different teams, and still produces a completely uniform valid draw having distribution $\mathbf{U}$, but with details that are somewhat different as we now describe.

Given a partial draw, suppose we know the number n_j of valid ways of completing the draw with team j in the next position. In terms of this, let $\mathcal{T}$ be the set of all teams with $n_j > 0$, and set $n_{\max} = \max_{j\in\mathcal{T}} n_j$. Then, for each team $j \in \mathcal{T}$, we sample a geometric random variable $G_j \sim$ Geometric$(n_j/n_{\max})$. (The G_j could be produced automatically by a computer in advance.) Thus, $G_j \geq 1$. Also, if $n_j = n_{\max}$, then $G_j = 1$. Furthermore, if the n_j are all roughly equal, then most (if not all) of the G_j will be equal to 1.

Given these G_j values, the selection process proceeds as follows:

1. Choose one of the teams $j \in \mathcal{T}$, uniformly at random (e.g., from balls in a bowl).
2. If team j has now been chosen a total of G_j times, then select team j for the next position in the draw.
3. Otherwise, return to step 1 and again choose a team uniformly at random.

This procedure produces a "race" in which the different teams are each trying to be the first to be chosen G_j times. Eventually one team will be chosen G_j times, and will then be selected for the next position in the draw.

The usefulness of this procedure is given by:

Proposition 5. *If the above procedure is used sequentially for each position of a group draw, then the final group draw will have distribution* **U**, *i.e., will be equally likely to be any of the potential valid draws.*

Proof. See the Appendix. ■

Although this multiple-rejection solution produces a valid uniform draw, it requires knowledge of the actual n_j values, which might be difficult in practice (as discussed in Section 5.2 above). Moreover, allocating a team to a particular slot takes multiple draws and is thus more time-consuming, and therefore less practical. Hence, for actual group draws, we believe that the methods of Sections 4 and 5 are preferable.

7. INFORMAL DESCRIPTIONS

In this section, we provide less formal/technical descriptions of our proposed draw methods, to explain to less technical readers how they would work in actual implementation. Although we have presented a number of possible methods (including rejection sampler, multiple rejections, etc.), here we focus on our two main practical methods, namely the Metropolis (swap) method of Section 4, and the Multiple Balls method of Section 5. Both methods are also demonstrated interactively at Roberts and Rosenthal (2022).

7.1. The Metropolis (Swap) Method

This method begins with some specific valid initial assignment, that is, by placing all teams in some group in a valid way which satisfies all of the restrictions. This initial assignment could be done using a rejection sampler on a computer to remove all bias. Alternatively, it could be done mechanically by any valid method; the initial assignment is not very important since it will be significantly modified in any case.

Then, the method does some prespecified number of "swap" moves. Each individual swap move proceeds as follows:

- First, choose one of the four pots, e.g., Pot 2. This choice could be done systematically, by choosing first Pot 1, then Pot 2, etc.
- Then, select two teams from that pot, e.g., Germany and Mexico. This choice could be done by, for example, placing one ball in a bowl for each of the eight teams in that pot,

and then randomly selecting two of the balls. (As an exception, if Pot 1 is chosen, then the host team should *not* be considered since they should always remain in Group A.)

- Then, check if it is possible to swap the two selected teams, for example, if the draw would still be valid if Germany and Mexico switch to each other's group. If it is possible, then the swap should be made. If not, then no change should be made.

A reasonable number of these swap moves should be performed. We recommend first performing a large number (e.g., 50) on a computer, which can be done quickly and easily. Then, the final number (e.g., 20) could be performed by celebrities selecting balls from bowls as above, under the watchful gaze of an excited general public.

Once the prespecified total number of swap moves has been performed, then the resulting draw shall be the final draw for the tournament. (For an interactive demonstration, click the "Swap speed" buttons at Roberts and Rosenthal (2022).)

7.2. The Multiple Balls Method

This method fills in the spots on the draw sequentially. The Pot 1 team in Group A is always the host team, as usual. Then, a team from Pot 1 is chosen for Group B, then one for Group C and so on. Once the Pot 1 teams are all assigned, then a team from Pot 2 is chosen for Group A, then one for Group B and so on. Then, a team from Pot 3 is chosen for Group A, then Group B and so on. Finally, a team from Pot 4 is chosen for Group A, then Group B and so on.

Each of the individual choices is conducted as follows. The computer will conduct internal random simulations in order to specify which teams are eligible for the next spot, and how many balls each of those teams should get. Then, that collection of balls is placed into a bowl, and a single ball is selected. Whatever team is specified by that ball is then chosen and placed in the corresponding position on the draw.

For example, when selecting the Pot 2 team to go in Group E, the computer might specify two balls for each of Netherland and Uruguay, and one each for Denmark and Mexico, for six in total. Then, those six balls are placed in a bowl and one is selected. If the selected ball says Uruguay, then Uruguay is placed into Group E.

Once each spot has been chosen, that is, a team from each of the four pots has been assigned to each of the eight groups, then the draw is complete, and is taken as the final draw. Since the computer only allowed balls for valid choices of teams, this complete draw will automatically be a valid draw. (For an interactive demonstration, click the "Balls Update" button at Roberts and Rosenthal (2022).)

8. DISCUSSION

This article has considered the challenge of designing football group draw mechanisms which have uniform distribution over all valid draw assignments, but are also entertaining, practical and transparent. We have explained (Section 2) how to simulate the FIFA sequential draw method, to compute the nonuniformity of its draws by comparison to a uniform rejection sampler. We have then proposed two practical methods (Metropolis swaps in Section 4, and multiple balls in Section 5) of achieving the uniform distribution while still using balls and bowls in a way which is suitable for a televised draw. These two solutions can be tried interactively for the World Cup groups at Roberts and Rosenthal (2022).

An important advantage of our methodology is its generality, allowing it to be immediately translated for use with different geographic constraints and other restrictions. The methodology can also be used seamlessly in conjunction with the constraints proposed in Csató (2022) for mitigating the risks of *tanking*, that is, teams intentionally losing games. Indeed, given any definition of what constitutes a "valid" draw, our methods can be used to simulate uniformly from all such valid draws, in ways that are still entertaining, practical and transparent.

We have focused here on methods that are practical, are unbiased, respect the sequential nature which FIFA and other footballing authorities seem to require, and try to ensure that in some sense *most* of the randomness takes place in a transparent way within a live television show. As detailed earlier, other literature has concentrated more on changing draw mechanisms to try to ensure a balanced draw. Clearly there is a tension between ensuring balance and creating an unbiased fair draw, and which aspect is more important will depend on political considerations. However, we note that the flexibility of our methods is readily transferable so that, should the authorities wish to impose further constraints to ensure balance, our methods can be readily translated to that context.

Our approach also has implications for other football draws, which are also biased, and are all amenable to correction using the methods we have presented in this article. For example, the UEFA (European) Champions League group stage draw in August 2022 used a slightly different procedure from FIFA (UEFA, 2022), choosing first a club and then randomizing uniformly over possible valid allocations for that club (as opposed to choosing the lexicographically first available slot). The UEFA Champions League round of 16 draw in December 2022 was different again. In that case, slots were filled in order with only the possible teams being put into a glass ball from which one was drawn uniformly at random. This can be seen as an approximation to a sequential Monte Carlo algorithm (e.g., Doucet et al., 2001) in which all of the nonzero incremental particle weights are assumed to be equal.

Note that this UEFA round of 16 draw is *not* equivalent to the FIFA sequential algorithm. For example, suppose there are four pots with three teams each, as follows:

Pot 1: Qatar[**As**], Belgium[**Eu**], France[**Eu**]
Pot 2: Switzerland[**Eu**], Mexico[**NA**], Germany[**Eu**]
Pot 3: Iran[**As**], Tunisia[**Af**], Senegal[**Af**]
Pot 4: Canada[**NA**], Ecuador[**SA**], Peru[**SA**]

Assume Qatar as host is automatically placed in Group A. Then consider the conditional probability that Iran and Canada are in the same group, given that Mexico is in Group A with Qatar. Under the FIFA sequential algorithm, this conditional probability equals $(2/3)^2 + (1/3)^2 = 5/9$, but under the UEFA round of 16 rules it equals $(1/2)^2 + (1/2)^2 = 1/2 \neq 5/9$. (Note that this difference persists even if Qatar's group assignment is random, although the conditional probabilities change somewhat due to the nature of the FIFA sequential algorithm.)

In a different direction, the 2026 World Cup will have 48 teams instead of 32, leading to different constraints and challenges. We plan to study the UEFA and World Cup 2026 draws in greater detail in subsequent work.

Our three proposed solutions are all practical and unbiased. However they all have disadvantages. Existing draw mechanisms for both FIFA and UEFA use background computer calculations, as do ours. However, unlike existing methods, we carry out random simulations in these calculations. This could be seen as a disadvantage as this randomness is not *transparent*. Although, as suggested by an editor, this transparency could perhaps be improved by agreeing in advance on a specific pseudo-random number generation algorithm and seed value to be used for the random simulations. Existing draw mechanisms also use computer-aided decision-making to check feasibility. Since these decisions can be checked a posteriori, these methods can thus be considered to be more transparent in some sense. On the other hand, such methods are clearly biased, so it seems that the choice is between correctness and some degree of transparency.

Our solutions all balance transparent randomness (i.e., celebrities drawing balls from glass bowls) with hidden randomness (i.e., computational calculations carried out instantaneously in the background to determine, for instance, the number of each type of balls to use, or the number of lives a team possesses). The aim is to carry out as much of the draw with transparent randomness as possible, and all three methods can do this reasonably effectively. However, it will be difficult to explain to a lay audience why some teams have more balls (or lives) than others. Therefore, both the multiple-ball and multiple-rejection solutions intrinsically require nonsymmetry to the transparent randomness part of the draw. On the other hand, the Metropolis (swap) solution has the disadvantage of being very different from current procedures. Viewers might also be worried about apparent biases created by the draw obtained when the hidden computer random iterations have been completed, thinking *we started in the group of death, so it is no surprise that we ended up with a tough draw!*

Note that in the Metropolis solution, we can in some sense decrease the proportion of computer randomness by running the final iterations carried out by balls and glass bowls for more iterations. Markov chain convergence arguments could in principle also be used to quantify this iteration/transparency trade-off.

Although practical, all our solutions are more complex than existing methods. For the multiple-ball draw, when the number of balls required is large, the draw could become complex to implement, though Section 5.4 explains how the required number of balls can be strictly bounded. It should also be emphasized that no specialist computer or statistical expertise is required by those carrying out the draw. It is also not necessary for technical nomenclature such as geometric or binomial distributions or retrospective simulation techniques to be used at all as part of the draw.

In summary, we have proposed three solutions to solve the bias caused by the sequential draw mechanisms used by football authorities. These solutions are flexible and can be used with a wide variety of constraints and structures. We believe our methods could be used as they are, or in modified form, by football authorities to provide fairer draws for future tournaments, and we encourage these organizations to consider them seriously.

ACKNOWLEDGEMENTS

We thank David Firth for helpful discussions of these topics, and thank the anonymous referees for detailed helpful reports. GOR was supported by EPSRC grants Bayes for Health (R018561) and CoSInES (R034710). JSR was supported by NSERC discovery grant RGPIN-2019-04142.

REFERENCES

Beskos, A., Papaspiliopoulos, O., & Roberts, G. O. (2006). Retrospective exact simulation of diffusion sample paths with applications. *Bernoulli*, 12(6), 1077–1098.

Boczon, M. & Wilson, A. J. (2022). Goals, constraints, and public assignment: A field study of the UEFA Champions League. *Management Science*, 1–37.

Brooks, S. P., Gelman, A., Jones, G. L., & Meng, X.-L. (Eds.), (2011). *Handbook of Markov chain Monte Carlo*, New York: Chapman & Hall.

Cea, S., Durán, G., Guajardo, M., Sauré, D., Siebert, J., & Zamorano, G. (2020). An analytics approach to the FIFA ranking procedure and the World Cup final draw. *Annals of Operations Research*, 286(1), 119–146.

Csató, L. (2021a). On the fairness of the restricted group draw in the 2018 FIFA World Cup. arXiv preprint, arXiv:2103.11353.

Csató, L. (2021b). *Tournament Design: How Operations Research Can Improve Sports Rules*, London: Springer Nature.

Csató, L. (2022). Quantifying incentive (in)compatibility: A case study from sports. *European Journal of Operational Research*, 302, 717–726.

Csató, L. (2023a). How to improve the fairness of group draw in sports tournaments. arXiv preprint, arXiv:2109.13785, 2023a.

Csató, L. (2023b). Group draw with unknown qualified teams: A lesson from the 2022 FIFA World Cup draw. *International Journal of Sports Science & Coaching,* 18(2), 539–551.

Devroye, L. (1986). *Non-Uniform Random Variate Generation*, Springer-Verlag, New York http://www.nrbook.com/devroye/.

Douc, R. & Cappé, O. (2005). Comparison of resampling schemes for particle filtering. In *ISPA 2005 – Proceedings of the 4th International Symposium on Image and Signal Processing and Analysis*, Piscataway, New Jersey: IEEE, 64–69.

Doucet, A., De Freitas, N., & Gordon, N. J. (Eds.), (2001). *Sequential Monte Carlo Methods in Practice*, Springer, New York.

Fearnhead, P. (1998). *Sequential Monte Carlo Methods in Filter Theory*. PhD thesis, Oxford, U.K.: Oxford University.

FIFA. (2022). *FIFA World Cup Qatar 2022 Draw Procedures*. https://digitalhub.fifa.com/m/2ef762dcf5f577c6/original/Portrait-Master-Template.pdf.

Guyon, J. (2014). Rethinking the FIFA World Cup final draw (longer preprint version). http://ssrn.com/abstract=2424376.

Guyon, J. (2015). Rethinking the FIFA World Cup final draw. *Journal of Quantitative Analysis in Sports*, 11(3), 169–182.

Guyon, J. (2017a). Ligue des champions: Pourquoi le PSG a presque une chance sur trois de rencontrer Chelsea. *Le Monde* https://www.lemonde.fr/ligue-des-champions/article/2017/12/10/ligue-des-champions-pourquoi-le-psg-a-une-chance-sur-trois-de-rencontrer-chelsea&uscore;5227638&uscore;1616944.html

Guyon, J. (2017b). Tirage au sort de la Coupe du monde: Comment ça marche et quelles probabilités pour la France. *Le Monde* https://www.lemonde.fr/mondial-2018/article/2017/11/30/tirage-au-sort-de-la-coupe-du-monde-comment-ca-marche-et-quelles-probabilites-pour-la-france&uscore;5222713&uscore;5193650.html

Guyon, J. (2018). Pourquoi la coupe du monde est plus équitable cette année. https://theconversation.com/pourquoi-la-coupe-du-monde-est-plus-equitable-cetteannee-97948 .

Guyon, J. (2022). "Choose your opponent": A new knockout design for hybrid tournaments. *Journal of Sports Analytics*, 8(1), 9–29.

Jones, C. (2022). Toronto professor claims FIFA draw contains hidden bias that skews the odds. https://www.cbc.ca/sports/soccer/canada-world-cup-toronto-professors-better-draw-chris-jones-april-1-1.6405610. *CBC News*.

Jones, M. C. (1990). The World Cup draw's flaws. *The Mathematical Gazette*, 74(470), 335–338.

Kendall, G. & Lenten, L. J. (2017). When sports rules go awry. *European Journal of Operational Research*, 257(2), 377–394.

Kitagawa, G. (1996). Monte Carlo filter and smoother for non-Gaussian nonlinear state space models. *Journal of Computational and Graphical Statistics*, 5(1), 1–25.

Klößner, S. & Becker, M. (2013). Odd odds: the UEFA Champions League round of 16 draw. *Journal of Quantitative Analysis in Sports*, 9(3), 249–270.

Laliena, P. & López, F. J. (2019). Fair draws for group rounds in sport tournaments. *International Transactions in Operational Research*, 26(2), 439–457.

Lenten, L. J. & Kendall, G. (2021). Scholarly sports: Influence of social science academe on sports rules and policy. *Journal of the Operational Research Society*, 73(2), 1–11.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6), 1087–1092.

Rathgeber, A. & Rathgeber, H. (2007). Why Germany was supposed to be drawn in the group of death and why it escaped. *Chance*, 20(2), 22–24.

Roberts, G. O. & Rosenthal, J. S. (2022). Football group draw methods and probabilities. http://probability.ca/fdraw/.

Rosenthal, J. S. (2022a). C program to simulate Uniform and FIFA group draws. http://probability.ca/fdraw/drawprobs.c.

Rosenthal, J. S. (2022b). World Cup groups – What are the odds? https://www.thestar.com/opinion/contributors/2022/03/30/for-kicks-we-came-up-with-a-fairer-way-to-determine-the-world-cup-draw.html. *Toronto Star*, A17.

UEFA. (2022). UEFA champions league group stage draw – procedure. https://www.uefa.com/uefachampionsleague/draws/2023/2001673/.

APPENDIX: PROOFS OF PROPOSITIONS

Proof of Proposition 1. First note that the proposed swaps are symmetric, since making the same swap twice is equivalent to not changing at all. Furthermore, swaps are accepted if the swapped draw is still valid, otherwise they are rejected. It follows that the swap moves correspond to a Metropolis algorithm (Metropolis et al., 1953) with stationary distribution $\mathbf{U}$. Hence, $\mathbf{U}$ is a stationary distribution for the Markov Chain corresponding to this method, that is, this method induces a Markov Chain Monte Carlo (MCMC) algorithm (Brooks et al., 2011), which preserves the distribution $\mathbf{U}$ as it runs. Hence, the overall distribution of the group assignment remains uniform no matter how many swap moves are attempted. ■

Proof of Proposition 2. The interval $[v_{j-1}, v_j)$ has length $v_j - v_{j-1} = u_j < 1$. If it lies entirely inside an interval $[i-1, i)$, then $\mathrm{P}(U_i \in [v_{j-1}, v_j)) = u_j$, so $\mathrm{E}(b_j) := \mathrm{E}\left[\#\{i : U_i \in [v_{j-1}, v_j)\}\right] = u_j$. Or, if there is an integer i with $i-2 < v_{j-1} \leq i-1 < v_j \leq i$, then $\mathrm{P}(U_{i-1} \in [v_{j-1}, v_j)) = (i-1) - v_{j-1}$ and $\mathrm{P}(U_i \in [v_{j-1}, v_j)) = v_j - (i-1)$, so we still have $\mathrm{E}(b_j) := \mathrm{E}\left[\#\{i : U_i \in [v_{j-1}, v_j)\}\right] = (i-1) - v_{j-1} + v_j - (i-1) = u_j$. Hence, in any case, $\mathrm{E}(b_j) = u_j$, whence $\mathrm{E}(m_j) = a_j + u_j = a_j + (r_j - a_j) = r_j = M p_j$. That is, the expected number of balls of type j is proportional to p_j. Hence, the probability that a ball drawn uniformly at random will be of type j is also proportional to p_j. Then, since $\sum_j p_j = 1$, this probability is actually equal to p_j, as claimed. ■

Proof of Proposition 3. Conditional on the values of the estimators $\widehat{p}_j$, the proof of Proposition 2 shows that the probability of selecting a ball of type j is equal to $\widehat{p}_j$. Hence, taking expectations over the $\widehat{p}_j$, it follows that the probability that the algorithm using the estimators $\widehat{p}_j$ will select a ball of type j is given by $\mathrm{E}(\widehat{p}_j) = p_j$, as claimed. ■

Proof of Proposition 5. It suffices to show that each new position of the draw is filled with the correct conditional probability according to $\mathbf{U}$. This follows because the procedure is actually a way of carrying out a corresponding rejection sampler. Indeed, consider a rejection sampler with proposal distribution which is uniform on $\mathcal{T}$, with acceptance probability $n_j/n_{\max}$ for each team j. Then each choice of team in step 1 corresponds to one proposal from the rejection sampler. And, G_j corresponds to the number of times that team j must be proposed before it is finally accepted. Hence, team j being the first to be chosen G_j times is precisely equivalent to team j being the team selected by a rejection sampler with target probabilities proportional to $n_j/n_{\max}$, i.e., proportional to n_j. The result follows. ■

Received 26 July 2022
Accepted 17 July 2023