THE SOCCER MODEL, STOCHASTIC ORDERING AND MARTINGALE TRANSPORT

GAOYUE GUO, NICOLAS JUILLET, AND WENPIN TANG

Abstract. Tournaments are competitions between a number of teams, the outcome of which determines the relative strength or rank of each team. In many cases, the strength of a team in the tournament is given by a score. Perhaps, the most striking mathematical result on the tournament is Moon's theorem, which provides a necessary and sufficient condition for a feasible score sequence via majorization. To give a probabilistic interpretation of Moon's result, Aldous and Kolesnik introduced the soccer model, the existence of which gives a short proof of Moon's theorem. However, the existence proof of Aldous and Kolesnik is nonconstructive, leading to the question of a "canonical" construction of the soccer model. The purpose of this paper is to provide explicit constructions of the soccer model with an additional stochastic ordering constraint, which can be formulated by martingale transport. Two solutions are given: one is by solving an entropy optimization problem via Sinkhorn's algorithm, and the other relies on the idea of shadow couplings. It turns out that both constructions yield the property of strong stochastic transitivity. The nontransitive situations of the soccer model are also considered.

Key words : Entropy optimization, martingale transport, pairwise comparison, score sequence, Sinkhorn's algorithm, shadow coupling, soccer model, stochastic ordering, strong stochastic transitivity, tournament.

AMS 2010 Mathematics Subject Classification: 60E15, 62F07, 60G42, 49Q22.

1. Introduction

A tournament refers to pairwise competitions between a number of teams (or players), all participating in a sport or game, in order to determine the winner or to produce a ranking of the teams. There has been a long history of research in tournaments and their rankings, including psychology [56, 57], game and economic theory [5, 31, 32], combinatorial and graph theory [22, 23, 39], applied probability [1, 2, 55], statistics [18, 29, 36], and more recently, large language models via direct preference optimization [16, 47]. In everyday language, tournament often means a *single-elimination* or *knockout tournament.* This paper focuses on the n-team *round-robin tournament*, where each team competes against every other team. A common metric for evaluating the strengths of the teams in the tournament is the *score sequence*, or simply, the *score* [30, 35]. We will consider the soccer model [3], which is an alternative to the famous Bradley–Terry model [14]. The main contribution of this work is to provide two explicit constructions of the soccer model that satisfies a stochastic ordering constraint, exploiting various probabilistic techniques such as stochastic ordering, martingale transport and entropy optimization. This answers a question of Aldous and Kolesnik (see Subsection 1.5 for details).

Date: March 8, 2025.

1.1. **Generalized tournament matrices and Moon's theorem.** To provide context, let us first introduce the notion of *tournament matrices* and *scores*. Consider an n-team tournament where each team competes N times against every other opponent. For $i \neq j$, let $p_{ij} \in [0,1]$ be a number representing the relative strength between the teams i and j. It can be interpreted as:

- the proportion of the number of wins of team i over team j in N games;
- the probability that team i wins over its opponent j in a random game.

Definition 1.1 (Generalized tournament matrices and scores)**.** *Let $n \geq 1$.*

(1) Denote by $\mathcal{G}_n$ the set of matrices $P = (p_{ij})_{1\leq i\neq j\leq n}$, where $p_{ij} \in [0,1]$ and $p_{ij}+p_{ji} = 1$ for each $i \neq j$. Such matrices are called the generalized tournament matrices of dimension n. (Note that the diagonals are undetermined.) Moreover, if $p_{ij} \in \{0,1\}$ for each $i \neq j$, then P is a tournament matrix.

(2) Denote by $\mathcal{G}'_n$ the set of matrices $P = (p_{ij})_{1\leq i,j\leq n}$, where $p_{ii} = 1/2$ for each $i = 1,\dots,n$ and $(p_{ij})_{1\leq i\neq j\leq n} \in \mathcal{G}_n$.

(3) For $\mathbf{x} = (x_1,\dots,x_n) \in [0,+\infty)^n$, the subset $\mathcal{G}_n(\mathbf{x}) \subset \mathcal{G}_n$ (resp. $\mathcal{G}'_n(\mathbf{x}) \subset \mathcal{G}'_n$) denotes the collection of matrices P such that

$$\sum_{j\neq i} p_{ij} = x_i \quad \left(\text{resp. } \sum_{j=1}^{n} p_{ij} = x_i\right),$$

for each $i = 1,\dots,n$. The vector $\mathbf{x}$ is called the (generalized tournament) score.

In the probabilistic setting, for every $P \in \mathcal{G}_n(\mathbf{x})$, $\mathbf{x}$ can be interpreted as the vector of expected number of wins because $x_i = \sum_{j\neq i} p_{ij}$ for each i. By convention, for $P \in \mathcal{G}'_n(\mathbf{x})$, a half win is artificially added by setting $p_{ii} = 1/2$ (this would correspond to a match of a team against a copy of itself). In the deterministic case, Nx_i is the number of games that team i wins. A classical theorem of Moon [38] provides a necessary and sufficient condition for the score $\mathbf{x}$ so that $\mathcal{G}_n(\mathbf{x})$ (or $\mathcal{G}'_n(\mathbf{x})$) is not empty. To state Moon's result, we need the notion of *majorization*.

Definition 1.2 (Majorization)**.** *Let $n \geq 1$. On the set of vectors $\mathbf{x} = (x_1,\dots,x_n)$ with increasing coordinates (i.e., $x_1 \leq \cdots \leq x_n$), the partial order of majorization is defined by*

$$\mathbf{x} \preceq \mathbf{y} \quad \text{if and only if} \quad \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \text{ and } \sum_{i=1}^{k} x_i \geq \sum_{i=1}^{k} y_i, \text{ for each } k < n. \tag{1.1}$$

This order[1] extends to $\mathbb{R}^n$ by $\mathbf{x} \preceq \mathbf{y}$ if and only if $\bar{\mathbf{x}} \preceq \bar{\mathbf{y}}$, where the vectors $\bar{\mathbf{x}}, \bar{\mathbf{y}}$ are the rearranged versions of $\mathbf{x}, \mathbf{y}$ (i.e., $\bar{\mathbf{x}} = (x_{\sigma(1)},\dots,x_{\sigma(n)})$, $\bar{\mathbf{y}} = (y_{\sigma'(1)},\dots,y_{\sigma'(n)})$ for some permutations σ, σ') such that the coordinates of $\bar{\mathbf{x}}, \bar{\mathbf{y}}$ are nondecreasing.

Theorem 1.3 (Moon's theorem [38])**.** *Let $\mathbf{x} \in \mathbb{R}^n$.*

(1) $\mathcal{G}_n(\mathbf{x})$ is nonempty if and only if $\mathbf{x} \preceq (0,1,\dots,n-1)$.

(2) $\mathcal{G}'_n(\mathbf{x})$ is nonempty if and only if $\mathbf{x} \preceq (\frac{1}{2},\frac{3}{2},\dots,n-\frac{1}{2})$.

[1] On $\mathbb{R}^n$, it is only a preorder: the reflexivity fails. See Appendix A for reminders.

Note that a result similar to Theorem 1.3 for tournament matrices (see Definition 1.1 (1)) and $\mathbf{x} \in \mathbb{Z}^n$ was proved by H. G. Landau [30]. See also [6, 17, 20, 52] for recent development on the enumeration of score sequences.

1.2. **Zermelo(-Bradley-Terry) model.** Definition 1.1 is quite general. Several parametric models have been developed to provide further structures to p_{ij}'s in pairwise comparisons. The most popular and well-studied example is the *Zermelo model* [60], which is widely known as the *Bradley-Terry model* [14].

Zermelo [60] is arguably the first to consider the inference problem in tournaments. In his model, each team i's strength is specified by a positive number u_i, which is called the "force" (*Spieltärke* in German), with $\sum_{i=1}^n u_i = 1$. For some $k \ge 1$, the outcome of the k^{th} game between the teams i and j can be represented by a Bernoulli variable B_{ij}^k with parameter $\frac{u_i}{u_i+u_j}$, where $\{B_{ij}^k = 1\}$ means that team i beats team j in their k^{th} game, and $\{B_{ij}^k = 0\}$ indicates the other way around. Hence, the generalized tournament matrix is specified by $p_{ij} := \frac{u_i}{u_i+u_j}$. Assuming that the random variables $(B_{ij}^k)_{i,j,k}$ are all independent, the maximum likelihood estimate (MLE) is used to infer the vector parameter $\mathbf{u} = (u_1, \dots, u_n)$. Under an irreducibility condition (see Definition 3.1 below), the MLE is uniquely determined by a system of n equations[2]

$$x_i = \sum_{j \ne i} \frac{u_i}{u_i + u_j}, \qquad \text{for } 1 \le i \le n. \tag{1.2}$$

The Zermelo model was rediscovered in [14, 21]. Since then, there have been various extensions such as the Plackett-Luce model [34, 46] and the Mallows model [36]. See [13, 51] for historical notes and further references.

The Zermelo model can be reparametrized by $v_i = \log u_i$, which yields the (generalized) linear model [37, Section 7.5]:[3]

$$p_{ij} = \frac{u_i}{u_i + u_j} = \frac{1}{1 + e^{v_j - v_i}}, \quad \text{for } 1 \le i \ne j \le n. \tag{1.3}$$

It is easy to check that the model (1.3) enjoys the property of strong stochastic transitivity (SST):

$$p_{ij} \ge 1/2 \text{ and } p_{jk} \ge 1/2 \quad \Longrightarrow \quad p_{ik} \ge \max(p_{ij}, p_{jk}), \tag{1.4}$$

because the left side of (1.4) is equivalent to $v_i \ge v_j \ge v_k$. This property will be the center of our study in Sections 2 and 5.

[2]Though the likelihood is homogeneous of degree zero, the uniqueness is guaranteed by the constraint $\sum_{i=1}^n u_i = 1$.

[3]Here we provide some explanations of the reparametrization $v_i = \log u_i$. The idea, following Thurston and Mosteller [41, 56, 57], is to associate each player i with an independent random variable X_i, and the generalized tournament matrix is specified by $p_{ij} = \mathbb{P}(X_i \ge X_j)$ for $1 \le i, j \le n$. One choice is that $X_i = v_i + V$, where v_i is the indicator of strength, and V is the noise. Hence, $p_{ij} = \mathbb{P}(W \le v_i - v_j)$, where W is distributed as $V - V'$, with V' an independent copy of V. In practice, W is often only required to be a (symmetric) random variable. Specializing to the case where W has the cumulative distribution function $F(t) = \frac{1}{1+e^{-t}}$ recovers the model (1.3). Alternatively, we can take X_i to be an independent exponential random variable with parameter u_i^{-1} to recover the model (1.3).

1.3. **The soccer model.** Recently, Aldous and Kolesnik [3] introduced a new parametric model for tournaments – the soccer model. The idea is in the same spirit to footnote 3 by associating each team with a random variable, or equivalently a probability distribution, for paired comparisons, which is described as follows.

The model is parametrized by

$$\Theta_n := \left\{ (\mu_1, \dots, \mu_n) \in \mathcal{P}(\{0, \dots, n-1\})^n : \sum_{i=1}^n \mu_i = \sum_{k=0}^{n-1} \delta_k \right\}, \tag{1.5}$$

where $\mathcal{P}(\{0, \dots, n-1\})$ denotes the set of probability measures on $\{0, \dots, n-1\}$, and δ_k is the Dirac measure on k. Each team i is assigned a probability measure μ_i, with the constraint $\sum_{i=1}^n \mu_i = \sum_{k=0}^{n-1} \delta_k$. For every $k \geq 1$, the outcome of the k^{th} game between the teams i and j is determined as follows: Let $X_i^k \sim \mu_i$ and $X_j^k \sim \mu_j$ be independent.

- If $X_i^k > X_j^k$, then team i beats team j.
- If $X_i^k < X_j^k$, then team j beats team i.
- If $X_i^k = X_j^k$, then each team is granted a probability $\frac{1}{2}$ (by external randomization) to win the game.

Here the random variables X_i^k and X_j^k can be interpreted as the number of goals scored by the team i and j, which explains the name "soccer". When the scores are equal, a tie-breaking method such that penalty shoot-out or coin tossing will take place and decide which team wins the game with equal probability. Specifically, let Z_{ij}^k be a Bernoulli variable with parameter $\frac{1}{2}$, independent of (X_i^k, X_j^k). Set

$$B_{ij}^k := \mathbb{1}_{\{X_i^k > X_j^k\}} + \mathbb{1}_{\{X_i^k = X_j^k,\, Z_{ij}^k = 1\}},$$

so $\{B_{ij}^k = 1\}$ means that team i beats team j in their k^{th} game, and $\{B_{ij}^k = 0\}$ indicates the other way around. The corresponding tournament matrix is specified by

$$p_{ij} := \mathbb{P}(X_i > X_j) + \frac{1}{2}\mathbb{P}(X_i = X_j), \tag{1.6}$$

where $X_i \sim \mu_i$ for $1 \leq i \leq n$, and $(X_i)_{1 \leq i \leq n}$ are pairwise independent.

1.4. **Proof of Moon's theorem via the soccer model.** As pointed in [3], a remarkable property of this model is that for $(\mu_1, \dots, \mu_n) \in \Theta_n$, the score $x_i = \sum_{j \neq i} p_{ij}$ of team i after exactly one game against every other opponent is the expected number of goals (or points) $\mathbb{E}X_i = \int x \, d\mu_i(x)$. To see this, let $\chi(\mu_i, \mu_j)$ be the probability for team i to win over team j. The map $\chi(\cdot, \cdot)$ can be extended into a linear map for signed measures with compact support. In fact, $\chi(\alpha, \beta) := \int f(y - x) d(\alpha \otimes \beta)(x, y)$, where $f(z) := \mathbb{1}_{\{z > 0\}} + \frac{1}{2}\mathbb{1}_{\{z = 0\}}$. Denoting by $\lambda := \sum_{k=0}^{n-1} \delta_k$, we have

$$\chi(\mu_i, \lambda) = \sum_{k=0}^{n} \chi(\mu_i, \delta_k) = \frac{1}{2} + \mathbb{E}X_i$$

on the one hand, and

$$\chi(\mu_i, \lambda) = \chi(\mu_i, \mu_i) + \sum_{j \neq i} \chi(\mu_i, \mu_j) = \frac{1}{2} + \sum_{j \neq i} p_{ij},$$

on the other hand.

As a consequence, the more difficult implication in Moon's theorem (Theorem 1.3) can easily be proved by using a famous result commonly attributed to Strassen [53] for the set of probability measures with a finite first moment. Earlier versions for discrete measures with finite support such as in Muirhead's inequality (see e.g., [24]) are sufficient for this purpose.

Theorem 1.4 ([15, 24, 53])**.** *Let ρ and μ be two probability measures on $\mathbb{R}^d$ having a finite first moment. Then the following two conditions are equivalent:*

(1) The measures are in convex order $\rho \preceq_C \mu$, i.e., for any convex function $\varphi : \mathbb{R}^d \to \mathbb{R}$, we have $\int \varphi d\rho \leq \int \varphi d\mu$.

(2) There exists a pair of random variables (X, Y) such that $X \sim \rho$, $Y \sim \mu$, and $\mathbb{E}(Y|X = x) = x$ for ρ-almost every x.

Moreover, for $d = 1$, if μ and ν are uniform measures on $\{x_1 \leq \cdots \leq x_n\}$ and $\{y_1 \leq \cdots \leq y_n\}$ respectively, the conditions (1) and (2) are satisfied if and only if $\mathbf{x} \preceq \mathbf{y}$. In this case, the condition (2) is translated as follows: For $1 \leq i \leq n$, let μ_i be the conditional distribution of Y given $X = x_i$. We have $\sum_{i=1}^n \mu_i = \sum_{j=1}^n \delta_{y_j}$ and $\int y \, d\mu_i(y) = x_i$ for each i.

The following proof of Theorem 1.3 by Aldous and Kolesnik is hard to beat.

Proof of Theorem 1.3. The implication $\mathcal{G}_n(\mathbf{x})$ is nonempty $\Rightarrow \mathbf{x} \preceq (0, \dots, n-1)$ follows from the fact that the expected number of wins of the k-weakest teams must be no less than $1 + \cdots + (k-1) = \binom{k}{2}$.

For the more difficult implication, assume that $\mathbf{x} \preceq (0, \dots, n-1)$. By Theorem 1.4, there exists $(\mu_1, \dots, \mu_n) \in \Theta_n$ such that $\int y \, d\mu_i(y) = x_i$ for each i. In the soccer model, $\int y \, d\mu_i(y)$ is the score of team i. So the matrix $P = (p_{ij})_{1 \leq i \neq j \leq n}$ defined by (1.6) is a generalized tournament matrix in $\mathcal{G}_n(x)$. $\square$

1.5. **Motivation and guideline.** The above short proof relies on the observation that for $\mathbf{x} \preceq (0, \dots, n-1)$, the set

$$\Theta_n(\mathbf{x}) := \left\{ (\mu_1, \dots, \mu_n) \in \Theta_n : \int y \, d\mu_i(y) = x_i \text{ for } 1 \leq i \leq n \right\}, \tag{1.7}$$

is nonempty. This fact follows from a soft argument by applying Theorem 1.4 that is non-constructive. It is natural to call for an explicit construction of $(\mu_1, \dots, \mu_n) \in \Theta_n(\mathbf{x})$, which was also asked by the authors of [3]. We quote them twice. First from the abstract:

> In particular, our proof of Moon's theorem on mean score sequences seems more constructive than previous proofs. This provides a comparatively concrete introduction to a longstanding mystery, the lack of a canonical construction for a joint distribution in the representation theorem for convex order.

Quotation from the final "Discussion" section [3]:

> To us, the most interesting part of the bigger picture surrounding convex order is that there is apparently no "canonical" choice of joint distribution in (3), (8): proofs may be constructive but they involve rather arbitrary choices and the resulting joint distributions are not easily described. Recent literature

on peacocks [7][4] studies continuous-parameter processes increasing in convex order, via many different constructions, and ideas from that literature might be relevant in our context.

Also note that the generic dimension of $\Theta_n(\mathbf{x})$ is $(n-1)(n-2)$, which is larger than the dimension $\frac{1}{2}(n-1)(n-2)$ of $\mathcal{G}_n(\mathbf{x})$, and $n-1$ of the Zermelo model. The fact that the soccer model has more degrees of freedom is an advantage because it allows one more flexibility in modeling, e.g., to fit nontransitive situations. It is also a weakness because the system of equations that identifies $\Theta_n(\mathbf{x})$ is underdetermined, which is not the case for the Zermelo model.

The main objective of this paper is to provide explicit constructions of $(\mu_1, \dots, \mu_n) \in \Theta_n(\mathbf{x})$ with the SST as an additional constraint. This approach provides "canonical" representations of the soccer model, thereby responding to the invitation mentioned above. Our approach is based on martingale optimal transport, which is a topic closely related to the one of peacocks, and two different algorithmic solutions are given: one is obtained by solving an entropic martingale transport problem via Sinkhorn's algorithm [44, 49], and the other is related to the concept of *shadow* that was introduced to define a class of interesting martingale transport plans [10, 11] and [8].

Organization of the paper: The remainder of the paper is organized as follows. In Section 2, we prove (in an abstract way) that $\Theta_n(\mathbf{x})$ contains an element that yields the SST. Two explicit constructions of $(\mu_1, \dots, \mu_n) \in \Theta_n(\mathbf{x})$ with the stochastic ordering constraint are presented in Sections 3 and 4. In an opposite direction, we illustrate in Section 5 that $\Theta_n(\mathbf{x})$ also has nontransitive solutions. Finally, general theoretic clarifications on the partial orders induced by a generalized tournament matrix are given in Appendix A.

2. Existence of an SST solution in the soccer model

We have seen in Subsection 1.2 that the Zermelo model (1.3) enjoys the property of SST (1.4). Here we show that the SST also appears in the soccer model. In fact for $\mathbf{x} \preceq (0, 1, \dots, n-1)$, $\Theta_n(\mathbf{x})$ contains an element that satisfies an additional constraint $\mu_1 \preceq_{\text{sto}} \cdots \preceq_{\text{sto}} \mu_n$ (see Proposition 2.2, and the definition of $\preceq_{\text{sto}}$ is recalled at the end of this introduction). With this intermediate constraint, the corresponding generalized tournament matrix enjoys the SST (see Proposition 2.4). The proof of Proposition 2.2 relies on a result of Müller and Rüschendorf [42] (see also [43, Section 2.6]) concerning the existence of conditional martingale kernels that are increasing in stochastic order. This result is equivalent to the existence of 1-Lipschitz martingale transport plans (see e.g., [9] before Lemma 3.3), which is well known in the field of peacocks and martingale transport, and has been discovered several times independently. For instance, Kellerer [28] proved the existence in a nonconstructive way based on Choquet theory; Lowther [33] is based on Hobson's approach to the Skorokhod embedding problem (SEP); Beiglböck, Huesmann and Stebbeg [9] relies on Root's solution that is also not constructive; and Beiglböck and Juillet's sunset coupling [11] is also related to the SEP, and to Kellerer's solution for which it gives a more explicit construction.

Here two novel approaches are developed in the context of the soccer model. The first method is on entropy optimization in the space of martingale transport plans, which will

[4]It is the book by Hirsch, Profeta, Roynette and Yor referred to as [25] in the present paper.

be detailed in Section 3. This method was proposed by Joe [26], but was applied to $P = (p_{ij})_{1\le i\neq j\le n}$ in $\mathcal{G}_n(\mathbf{x})$ instead of $\Theta_n(\mathbf{x})$.[5] The second method is a direct construction based on the shadow embedding, and will be explained in Section 4.

Recall that two probability measures μ and μ' on $\mathbb{R}$ are in stochastic order $\mu \preceq_{\rm sto} \mu'$ if and only if their cumulative distribution functions satisfy $F_\mu \ge F_{\mu'}$. There are many equivalent criterions, e.g., there exists a coupling (X, X') such that $X \sim \mu$, $X' \sim \mu'$ and $\mathbb{P}(X \le X') = 1$ (see [48, Chapter 1]).

Lemma 2.1. *Let μ and ν be two probability measures on $\mathbb{R}$ such that $\mu \preceq_{\rm sto} \nu$. Then for $X \sim \mu$ and $Y \sim \nu$ independent,*

$$\mathbb{P}(X > Y) + \frac{1}{2}\mathbb{P}(X = Y) \ge 1/2.$$

Moreover, if $\mu \neq \nu$, then the equality is strict.

Proof. Consider a coupling (X, X') such that $X \sim \mu$, $X' \sim \nu$ and $\mathbb{P}(X' \ge X) = 1$. Let $Y \sim \nu$ be independent of (X, X') so that $(X, Y) \sim \mu \times \nu$. Since $(Y, X') \sim \nu \times \nu$, we get

$$\frac{1}{2} = \mathbb{P}(Y > X') + \frac{1}{2}\mathbb{P}(Y = X') = \mathbb{E}[f(Y - X')],$$

where

$$f(z) := \mathbb{1}_{\{z>0\}} + \frac{1}{2}\mathbb{1}_{\{z=0\}}. \tag{2.1}$$

Observing that $Y - X' \le Y - X$ almost surely, we have:

$$\frac{1}{2} = \mathbb{E}[f(Y - X')] \le \mathbb{E}[f(Y - X)] = \mathbb{P}(Y > X) + \frac{1}{2}\mathbb{P}(Y = X). \tag{2.2}$$

Furthermore, if $\mu \neq \nu$, the event $\{X < X'\}$ has nonzero probability. Note that Y is independent from (X, X'), and has the same law as X'. Therefore, $\mathbb{P}(X < Y \le X') > 0$. It follows $\mathbb{P}(f(Y - X') \le 1/2,\ f(Y - X) = 1) > 0$, and the inequality in (2.2) is strict. □

The following proposition is an improved version of Strassen's theorem, following [42].

Proposition 2.2. *Let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ be such that $x_1 \le \cdots \le x_n$ and $y_1 \le \cdots \le y_n$ and assume $\mathbf{x} \preceq \mathbf{y}$. There exists $(\mu_1, \dots, \mu_n)$ such that $\sum_{i=1}^n \mu_i = \sum_{k=1}^n \delta_{y_k}$, $\int x d\mu_i = x_i$ for each $1 \le i \le n$ and $(\mu_i)_{1\le i\le n}$ is increasing in $\preceq_{\rm sto}$.*
If $\mathbf{y} = (0, 1, \dots, n-1)$, it can be formulated that there exists $(\mu_1, \dots, \mu_n) \in \Theta_n(\mathbf{x})$ such that $(\mu_i)_{1\le i\le n}$ is increasing in $\preceq_{\rm sto}$.

Proof. As seen in the introduction, Theorem 1.4 has an improved version when ρ, μ are supported on $\mathbb{R}$. In addition to $\mathbb{E}(Y|X = x) = x$ for ρ-almost every x, it was proved in [42] that there exists a family $(\mu_x)_{x\in\mathbb{R}}$ of (regular) conditional laws (i.e., $\mathbb{E}(f(Y)|\ X = x) = \int f(y)\ d\mu_x(y)$ for every positive f) that is increasing in stochastic order: $x \le x'$ implies $\mu_x \preceq_{\rm sto} \mu_{x'}$. Exactly as in the second part of Theorem 1.4, this result translates in the discrete setting into the statement of the proposition. □

Remark 2.3. *In Proposition 2.2, if $\mu_i \preceq_{\rm sto} \mu_{i+1}$, then $\mu_i = \mu_{i+1}$ is equivalent to $x_i = x_{i+1}$. This can be seen by the coupling (X_i, X_{i+1}) such that $X_i \sim \mu_i$, $X_{i+1} \sim \mu_{i+1}$ and $X_i \le X_{i+1}$ almost surely. Then $\mathbb{E}X_i = \mathbb{E}X_{i+1}$ is equivalent to $X_i = X_{i+1}$ almost surely.*

[5]Aldous and Kolesnik [3] also promoted Joe's approach. One main motivation of [3] is to make a connection between Joe's result and Moon's theorem.

Combining Lemma 2.1, Proposition 2.2 and Remark 2.3 yields that for each $\mathbf{x} \preceq (0, \ldots, n-1)$, there exists $(\mu_1, \ldots, \mu_n) \in \Theta_n(\mathbf{x})$ such that the corresponding generalized tournament matrix P satisfies:

- $p_{ij} \geq 1/2$ if and only if $x_i \geq x_j$.
- $p_{ij} = 1/2$ if and only if $x_i = x_j$.

The next proposition proves the SST (1.4).

Proposition 2.4. *For $\mathbf{x} \preceq (0, \ldots, n-1)$, let $(\mu_1, \ldots, \mu_n) \in \Theta_n(\mathbf{x})$ be specified as in Proposition 2.2. Then $(\mu_i)_{1 \leq i \leq n}$ is increasing in $\preceq_{\text{sto}}$, and hence, the generalized tournament matrix defined by* (1.6) *satisfies the SST.*

Proof. Take $(\mu_1, \ldots, \mu_n) \in \Theta_n(\mathbf{x})$ as in Proposition 2.2, and assume that $p_{kj} \geq 1/2$ and $p_{ji} \geq 1/2$. As explained just above, we have $x_i \leq x_j \leq x_k$. Since $\mu_i \preceq_{\text{sto}} \mu_j$, there exist $X_i \sim \mu_i$, $X_j \sim \mu_j$ and $X_k \sim \mu_k$ such that $X_i \leq X_j$ almost surely and X_k is independent of (X_i, X_j). With the function f defined in (2.1), we have $f(X_k - X_i) \geq f(X_k - X_j)$ almost surely. Taking the expectation on both sides we get $p_{ki} \geq p_{kj}$. Similarly, $\mu_j \preceq_{\text{sto}} \mu_k$ yields $p_{ki} \geq p_{ji}$. Thus, $p_{ki} \geq \max(p_{ji}, p_{kj})$. □

3. The entropic construction

In this section, we provide a construction of $(\mu_1, \ldots, \mu_n) \in \Theta_n(\mathbf{x})$ satisfying the stochastic ordering constraint $\mu_1 \preceq_{\text{sto}} \cdots \preceq_{\text{sto}} \mu_n$ by solving an entropy optimization problem, where an iterative algorithm is given. For ease of presentation, we identify Θ_n and $\Theta_n(\mathbf{x})$, whose elements are $(\mu_1, \ldots, \mu_n) \in \mathcal{P}(\{0, \ldots, n-1\})^n$, with the set of matrices $M = (m_{ij})_{1 \leq i,j \leq n}$ defined by $m_{ij} := \mu_i(\{j-1\})$. More precisely,

$$\Theta_n = \{M \in \mathcal{M}_n(\mathbb{R}) : \ M \text{ is a doubly stochastic matrix}\},$$
$$\Theta_n(\mathbf{x}) = \left\{ M \in \Theta_n : \sum_{j=1}^{n} m_{ij}(j-1) = x_i \ \text{ for } 1 \leq i \leq n \right\},$$

where the measure μ_i is encoded as a vector of n nonnegative coefficients corresponding to the i^{th} row of M by $\mu_i = \sum_{j=1}^{n} m_{ij} \delta_{j-1}$ for $1 \leq i \leq n$.

To proceed further, we need the following irreducibility definition, which is implicit behind the results established in [21, 40, 60].

Definition 3.1 (Irreducibility condition)**.** *Let $P \in \mathcal{G}_n(\mathbf{x})$, and $\bar{\mathbf{x}}$ be the rearranged version of $\mathbf{x}$. The following conditions are equivalent:*

(1) For each (nontrivial) partition $I \cup I^c$ of $\{1, \ldots, n\}$, there exist $i \in I$ and $j \in I^c$ such that $p_{ij} > 0$.

(2) For each $i, j \in \{1, \ldots, n\}$, there exist $r \leq n$ and a chain of coefficients $(i_\ell)_{\ell=1,\ldots,r}$ such that $i_1 = i$, $i_r = j$, and $p_{i_\ell i_{\ell+1}} > 0$ for each $\ell < r$.

(3) There exists $r \geq 1$ such that all the entries of $P^r = \underbrace{P \cdots P}_{r \text{ times}}$ are strictly positive.

(4) $\sum_{i=1}^{n} \bar{x}_i = \frac{n(n-1)}{2}$ and $\sum_{i=1}^{k} \bar{x}_i < \frac{k(k-1)}{2}$ for $k < n$.

If one of the conditions (1)-(3) is satisfied, we say that P is an irreducible generalized tournament matrix. If the condition (4) is satisfied, $\mathbf{x}$ is called an irreducible score.
The following two conditions for reducible generalized tournament matrices and scores are equivalent:

(1) There exists a partition $I \cup I^c$ such that $p_{ij} = 0$ for each $(i,j) \in I \times I^c$.
(2) There exists $1 \leq k < n$ such that $(\bar{x}_1, \dots, \bar{x}_k) \preceq (0, \dots, k-1)$ and $(\bar{x}_{k+1}, \dots, \bar{x}_n) \preceq (k, \dots, n-1)$.

In particular, if $P \in \mathcal{G}_n(\mathbf{x})$ is irreducible (resp. non-irreducible), then so is every $Q \in \mathcal{G}_n(\mathbf{x})$.

For $\mathbf{x} \preceq (0, \dots, n-1)$, define the entropy by

$$H : M \in \Theta_n(\mathbf{x}) \mapsto \sum_{i,j=1}^{n} m_{ij} \log(m_{ij}). \tag{3.1}$$

First, we show in Subsection 3.1 that if $\mathbf{x} \preceq (0, \dots, n-1)$ is an irreducible score, then H has a unique minimizer whose entries are strictly positive. In Subsection 3.2, we prove that $P \in \mathcal{G}_n(\mathbf{x})$ corresponding to this unique minimizer satisfies the SST. Next in Subsection 3.3, we provide an algorithm and prove its convergence to the minimizer of H. Finally, we extend the results to the non-irreducible case in Subsection 3.4.

3.1. **Minimizer of H.** Recall the definition of irreducible scores from Definition 3.1. The following proposition is useful in proving that H has a unique minimizer, as well as the convergence of Sinkhorn's algorithm in Subsection 3.3.

Proposition 3.2. *Let $\mathbf{x} \preceq (0, \dots, n-1)$ be an irreducible score (so that $\mathcal{G}_n(\mathbf{x})$ and $\Theta_n(\mathbf{x})$ are nonempty). Then there exists $M \in \Theta_n(\mathbf{x})$ such that $m_{ij} > 0$ for all (i,j).*

Proof. We consider the partial order on $\Theta_n(\mathbf{x})$, for which $M = (m_{ij})_{ij}$ dominates $N = (n_{ij})_{ij}$ if and only if $m_{ij} > 0 \Rightarrow n_{ij} > 0$ for all (i,j). Let M be maximal for this order. Here we gather a few observations on M:

- For each row i, the set of indices j with $m_{ij} > 0$ is a nonempty (discrete) interval. Assume by contradiction that $m_{ij} = 0$, $m_{ij_0} > 0$ and $m_{ij_1} > 0$ with $j_0 < j < j_1$. Then there exists $i' \neq i$ such that $m_{i'j} > 0$. For $\lambda \in (0,1)$ satisfying $\lambda y_{j_1} + (1-\lambda) y_{j_0} = y_j$, consider $M + hM'$, where M' is a matrix with all entries zero except the six following $m'_{ij_0} = -m'_{i'j_0} = \lambda$, $m'_{ij_1} = m'_{i'j_1} = 1 - \lambda$ and $m'_{i'j} = -m'_{ij} = 1$. For $h < \min(m_{ij}, m_{i'j_0}/\lambda, m_{i'j_1}/(1-\lambda))$, we have $M + hM' \in \Theta_n(\mathbf{x})$, which yields a contradiction to the maximality of M.
- If the four entries m_{ij}, $m_{ij'}$, $m_{i'j}$ and $m_{i'j'}$ are positive, then for the two rows i and i' we have $m_{ik} > 0 \Leftrightarrow m_{i'k} > 0$ (the positive entries are the same for the two rows). The proof follows the same argument as in the first point.

For row i, let J_i be the set of indices j such that $m_{ij} > 0$. With the two observations above and the fact that $\sum_j m_{ij} y_j = x_i$, we see that M is a block matrix with the blocks going from left to right. Assume by contradiction that there are at least two blocks. Then $J_1 = \cdots = J_k = \{1, \dots, \ell\}$ for some $k < n$, and $J_{k+1} \neq J_k$. In fact, $J_{k+1} \cap J_k = \emptyset$ or $\{\ell\}$. Consider two cases $k \geq \ell$ and $k \leq \ell - 1$: For $k \geq \ell$, note that $(i \leq k \text{ and } j > \ell) \Rightarrow m_{ij} = 0$. Thus, the sum of all the entries is larger than that of all entries m_{ij} with $i \leq k$ or $j \geq \ell + 1$, which is $k + (n - \ell)$. (Here k corresponds to the sum over the k first rows, and $n - \ell$ is the sum over the $n - \ell$ last columns.) Note that the sum of all the entries of Θ_n is n. By analyzing the equality case, we get $k = \ell$ and $J_{k+1} \cap J_k = \emptyset$, which contradicts the fact that $\mathbf{x}$ is irreducible because the probability that a team of the rows $0, \ldots, k$ defeats a team indexed by $k+1, \ldots, n$ would be zero (see the second part of Definition 3.1). For $k \leq \ell - 1$, we consider the lower left of M: $(i > k$ and $j < \ell) \Rightarrow m_{ij} = 0$. The sum of all the entries is larger or equal to $(n-k) + (\ell - 1) \geq n$, which corresponds to the $n - k$ last rows plus the $\ell - 1$ first columns. Again this sum is n, so $m_{1\ell} = \cdots = m_{k\ell} = 0$ which contradicts the fact that $J_1 = \ldots = J_k = \{1, \ldots, \ell\}$. □

The following result shows that if $\mathbf{x}$ is irreducible, the entropy function H on $\Theta_n(\mathbf{x})$ has a unique minimum point.

Proposition 3.3. *Let $\mathbf{x} \preceq (0, \ldots, n-1)$ be an irreducible score. Then the function H defined by* (3.1) *has a unique minimizer, whose coefficients m_{ij}'s are all strictly positive.*

Proof. The function H is continuous on the compact set $\Theta_n(\mathbf{x})$ (with the convention $0 \times \log 0 = 0$). Moreover, it is strictly convex so H has a unique minimizer denoted by $M^{(0)}$.

Suppose by contradiction that one of the entries of $M^{(0)}$ is zero. By Proposition 3.2, let $M^{(1)} \in \Theta_n(\mathbf{x})$ with all strictly positive entries. For each $0 \leq \lambda \leq 1$, $M^{(\lambda)} := \lambda M^{(1)} + (1 - \lambda) M^{(0)} \in \Theta_n(\mathbf{x})$. Because all the coefficients of $M^{(1)}$ are strictly positive, the derivative of $\lambda \in [0,1] \mapsto H(M^{(\lambda)})$ at $\lambda = 0^+$ is $-\infty$. This contradicts the fact that $M^{(0)}$ minimizes H. □

Remark 3.4.

(1) *Minimizing $P \in \mathcal{G}_n(\mathbf{x}) \mapsto \sum_{ij} p_{ij} \log(p_{ij})$ (instead of $H : M \mapsto \sum_{ij} m_{ij} \log(m_{ij})$) is used by Joe* [26] *to construct a generalized tournament matrix of the Zermelo-Bradley-Terry model. Similar to Proposition 3.2, we can show that if $\mathbf{x}$ is irreducible, then $\mathcal{G}_n(\mathbf{x})$ contains some P with all strictly positive entries. The argument is as follows: let $P \in \mathcal{G}_n(\mathbf{x})$ with the maximal number of nonzero entries. Assume by contradiction that $p_{i_1 i_0} = 0$, so $p_{i_0 i_1} = 1$. By Definition 3.1, there exists a chain $p_{i_1 i_2}, \ldots, p_{i_{N-1} i_N}, p_{i_N, i_0} > 0$. Now we operate as follows to get $p_{i_1 i_0} > 0$ and guarantee P stays in $\mathcal{G}_n(\mathbf{x})$: for sufficiently small h, replace $p_{i_j i_{j+1}}$ with $p_{i_j i_{j+1}} - h$, with the convention $i_{N+1} = i_0$. Hence, $p_{i_{j+1} i_j}$ is replaced with $p_{i_{j+1} i_j} + h$. It suffices to use the infinite derivative of $p \mapsto p \log p$ at $p = 0^+$ as in Proposition 3.3 to conclude. Note that this argument also works for other functions f, provided that the derivative of f at 0^+ is $-\infty$.*

(2) *Proposition 3.3 can be easily generalized to irreducible $\mathbf{x}$ such that $\mathbf{x} \preceq \mathbf{y}$, where $\mathbf{y} = (k, k+1, \ldots, \ell-1, \ell)$ different from $(0, \ldots, n-1)$. In this case, $\Theta_n(\mathbf{x})$ is replaced by the set of double stochastic matrices with $\sum_{j=1}^{n'} m_{ij} y_j = x_i$ for each $1 \leq i \leq n'$ (with $n' = \ell - k + 1$).*

3.2. **Property of SST.** The following theorem shows that for an irreducible $\mathbf{x}$, the minimizer of H on $\Theta_n(\mathbf{x})$ satisfies the stochastic ordering constraint, and hence, the corresponding generalized tournament matrix enjoys the property of SST.

Theorem 3.5. *Let* $\mathbf{x} \preceq (0, \dots, n-1)$ *be an irreducible score, and* M *be the (unique) minimizer of* H *on* $\Theta_n(\mathbf{x})$*. Then* $(\mu_i)_{1 \le i \le n}$ *is increasing in stochastic order, and the generalized tournament matrix corresponding to* M *(defined by* (1.6)*) satisfies the SST.*

Proof. Since $\mathbf{x}$ is irreducible, H has a unique minimizer on $\Theta_n(\mathbf{x})$ by Proposition 3.3. Moreover, the minimizer is in the interior (of the affine linear space spanned by $\Theta_n(\mathbf{x})$). By the Karush-Kuhn-Tucker theorem, we can neglect the constraints $m_{ij} \ge 0$, and only consider those on the rows with multipliers a_i's, on the columns with multipliers b_j's, and on the barycenters with multipliers c_i's. Differentiating the Lagrangian, we get:

$$(1 + \log(m_{ij})) + a_i + b_j + c_i y_j = 0, \quad \text{where } y_j = j - 1.$$

So we have $m_{ij} = e^{-a_i - 1} e^{-b_j} e^{-c_i y_j}$, and the probability vector $\mu_i = (m_{i1}, \dots, m_{in})$ is specified by the vector $(e^{-b_1}, \dots, e^{-b_n})$ modulated by entrywise product with the shape vector $(e^{-c_i y_1}, \dots, e^{-c_i y_n})$ and normalized by the constant e^{-a_i}. For ease of presentation, we fix $(b_1, \dots, b_n)$, let $\bar{\mu} = (e^{-b_1}, \dots, e^{-b_n})$, and define $\bar{\mu}(s)$ by

$$\bar{\mu}(s) = (\bar{\mu}_1(s), \dots, \bar{\mu}_n(s)) = (\bar{\mu}_1 e^{s y_1}, \dots, \bar{\mu}_n e^{s y_n}).$$

(Note that $\bar{\mu}(0) = \bar{\mu}$, and recall $y_j = j - 1$.) Let $C(s) = \sum_{j=1}^n \bar{\mu}_j e^{s y_j}$, so that $\mu_i = C(-c_i)^{-1} \bar{\mu}(-c_i)$ for $1 \le i \le n$. Now for $s < t$ and $1 \le k < n$, we have:

$$\frac{\sum_{j=1}^k \bar{\mu}_j(t)}{\sum_{j=1}^k \bar{\mu}_j(s)} \le e^{(t-s)y_k} < e^{(t-s)y_{k+1}} \le \frac{\sum_{j=k+1}^n \bar{\mu}_j(t)}{\sum_{j=k+1}^n \bar{\mu}_j(s)}.$$

Therefore, $\frac{\sum_{j=1}^k \bar{\mu}_j(t)}{\sum_{j=k+1}^n \bar{\mu}_j(t)} < \frac{\sum_{j=1}^k \bar{\mu}_j(s)}{\sum_{j=k+1}^n \bar{\mu}_j(s)}$ so that

$$g_k : s \mapsto C(s)^{-1} \sum_{j=1}^k \bar{\mu}_j(s) = \frac{\sum_{j=1}^k \bar{\mu}_j(s)}{\sum_{j=1}^k \bar{\mu}_j(s) + \sum_{j=k+1}^n \bar{\mu}_j(s)} = \left(1 + \frac{\sum_{j=k+1}^n \bar{\mu}_j(s)}{\sum_{j=1}^k \bar{\mu}_j(s)}\right)^{-1}$$

is strictly increasing. Comparing the cumulative distribution functions of

$$\mu(s) = \sum_{j=1}^n C(s)^{-1} \bar{\mu}_j(s) \delta_{y_j} \quad \text{and} \quad \mu(t) = \sum_{j=1}^n C(t)^{-1} \bar{\mu}_j(t) \delta_{y_j}, \tag{3.2}$$

we see that $\mu(s) \preceq_{\text{sto}} \mu(t)$. Recall that $\mu_i = C(-c_i)^{-1} \mu(-c_i)$ for some $c_i \in \mathbb{R}$. As a result, the probability measures $\mu_1, \dots, \mu_n$ are totally ordered (in stochastic order), and are in the same order as their barycenters $x_i = \int y \, d\mu_i(y)$. Hence, $(\mu_i)_{1 \le i \le n}$ is increasing in stochastic order. The property of SST follows readily from Proposition 2.4. □

3.3. **Sinkhorn's algorithm and convergence.** In this subsection, we provide a numerical scheme to approximate the minimizer of the entropy in the soccer model with the stochastic ordering constraint (and the SST) as in Theorem 3.5. This scheme is closely related to the one for the entropic optimal transport with discrete marginals, which was initiated in the pioneering work of Benamou, Carlier, Cuturi, Nenna and Peyré [12]. Our situation is different due to the martingale constraint. Thus, we need to adapt the existing theory carefully, depending on whether $\mathbf{x}$ is irreducible or not (see Subsection 3.4).[6] Our presentation closely follows the one by Bauschke and Lewis in [7].

Set $C := C_1 \cap C_2 \cap C_3 \in \mathbb{R}^{n\times n}$, where

$$C_1 := \left\{ M \in \mathbb{R}_+^{n\times n} : \sum_{j=1}^n m_{ij} = 1, \text{ for all } i = 1, \ldots, n \right\}, \tag{3.3}$$

$$C_2 := \left\{ M \in \mathbb{R}_+^{n\times n} : \sum_{i=1}^n m_{ij} = 1, \text{ for all } j = 1, \ldots, n \right\}, \tag{3.4}$$

$$C_3 := \left\{ M \in \mathbb{R}_+^{n\times n} : \sum_{j=1}^n (j-1) m_{ij} = x_i, \text{ for all } i = 1, \ldots, n \right\}. \tag{3.5}$$

By Proposition 3.2, $C \cap (0,\infty)^{n\times n} \neq \emptyset$.

Let $h : \mathbb{R} \to (-\infty, \infty]$ be defined by $h(x) = x \log x$ for $x \geq 0$, and ∞ otherwise. The goal is to solve the convex optimization problem:

$$\min_{M\in C} \left\{ H(M) := \sum_{i,j=1}^n h(m_{ij}) \right\}. \tag{3.6}$$

To this end, we use Bregman's (pseudo-)distance $D : \mathbb{R}^{n\times n} \times \mathrm{idom}(H) \to [0,\infty]$ defined by

$$\begin{aligned} D(L,M) :&= H(L) - H(M) - \nabla H(M) \cdot (L - M) \\ &= \sum_{i,j=1}^n \Big[l_{ij} \log l_{ij} - m_{ij} \log m_{ij} - \big(1 + \log m_{ij}\big)(l_{ij} - m_{ij}) \Big], \end{aligned} \tag{3.7}$$

where $\mathrm{dom}(H) = \mathbb{R}_+^{n\times n}$ denotes the domain of H, and its interior and boundary are denoted by $\mathrm{idom}(H) = (0,\infty)^{n\times n}$ and $\mathrm{bdom}(H) = \{M \in \mathbb{R}_+^{n\times n} : m_{ij} = 0 \text{ for some } (i,j)\}$, respectively.

Set $M^0 := (m_{ij}^0 \equiv 1) \in \mathrm{idom}(H)$,[7] so $D(L, M^0) = H(L) - \sum_{i,j=1}^n l_{ij} + K$, where $K \in \mathbb{R}$ depends only on M^0. Because $\sum_{i,j=1}^n l_{ij} = n$ for $L \in C$, we have:

$$\arg\min_{L\in C} H(L) = \arg\min_{L\in C} D(L, M^0) \quad \text{and} \quad \min_{L\in C} H(L) = \min_{L\in C} D(L, M^0) + n - K$$

For $k \geq 4$, let $C_k := C_{k \bmod 3}$. For $k \geq 1$, define M_k as the Bregman projection of M_{k-1} on C_k:

$$M^k := \arg\min_{L\in C_k} D(L, M^{k-1}), \tag{3.8}$$

where M^k is uniquely determined. The following theorem establishes the convergence of the sequence $(M^k)_{k\geq 0}$, where M^0 is given previously and M^k is uniquely defined. Finally, the convergence of our scheme is summarized as below.

[6] We point out that some similar but different algorithms have been considered for martingale optimization problems satisfying either marginal or expectation constraints [4, 19].

[7] The initialization $M^0 := (m_{ij}^0 \equiv 1) \in \mathrm{idom}(H)$ is chosen so that the entropy optimization problem $\min_{L\in C} H(L)$ is equivalent to the problem $\min_{L\in C} D(L, M^0)$, which can be provably solved by Sinkhorn's algorithm.

Theorem 3.6. *Let* $\mathbf{x} \preceq (0, \ldots, n-1)$ *be an irreducible score, so that* $C \cap (0,\infty)^{n\times n} \neq \emptyset$. *Then*

$$\lim_{k\to\infty} M^k = \arg\min_{L\in C} D(L, M^0) = \arg\min_{L\in C} H(L). \tag{3.9}$$

Proof. Note that $\mathrm{dom}(h) = \mathbb{R}_+$, $\mathrm{idom}(h) = (0,\infty)$ and $\mathrm{bdom}(h) = \{0\}$. It is easy to check:

(1) h is a proper convex function. Moreover, h is closed because $\{x \in \mathrm{dom}(h) : h(x) \leq \alpha\}$ is closed for all $\alpha \in \mathbb{R}$.

(2) h is Legendre because
- h is differentiable on $\mathrm{idom}(h)$;
- $\lim_{t\to 0+} h'(x + t(y - x))(y - x) = -\infty$ for all $x \in \mathrm{bdom}(f)$ and $y \in \mathrm{idom}(h)$;
- h is strictly convex on $\mathrm{idom}(h)$.

(3) h is co-finite because $\lim_{t\to\infty} \frac{h(tx)}{t} = \infty$ for all $x \neq 0$.

(4) h is very strictly convex because $h''(x) > 0$ for all $x \in \mathrm{idom}(h)$.

Since C_1, C_2, C_3 are all affine subsets, it suffices to apply [7, Theorem 4.3] to conclude. □

See also [45] for recent development on the convergence rate of Bregman's iteration under further technical assumptions, which we do not pursue here.

Next we propose a computational scheme inspired by Theorem 3.6. The key is to compute numerically, for each $M \in (0,\infty)^{n\times n}$, its Bregman projection on C_k. We distinguish three cases:

- $k = 1$: We introduce the Lagrange multiplier $\lambda = (\lambda_1, \ldots, \lambda_n) \in \mathbb{R}^n$, and set $\Phi(L,\lambda) := D(L, M) + \sum_{i=1}^n \lambda_i \left(\sum_{j=1}^n l_{ij} - 1\right)$. Differentiating Φ with respect to m_{ij} and λ_i yields $\partial_{l_{ij}}\Phi = \log(l_{ij}) - \log(m_{ij}) + \lambda_i$ and $\partial_{\lambda_i}\Phi = \sum_{j=1}^n l_{ij} - 1$. By setting these to zero, we get
$$\arg\min_{L\in C_1} D(L, M) = \left(\frac{m_{ij}}{\sum_{k=1}^n m_{ik}}\right)_{1\leq i,j\leq n}. \tag{3.10}$$
- $k = 2$: The same reasoning as in the previous case yields:
$$\arg\min_{L\in C_2} D(L, M) = \left(\frac{m_{ij}}{\sum_{k=1}^n m_{kj}}\right)_{1\leq i,j\leq n}. \tag{3.11}$$
- $k = 3$: Define $\Phi(L,\lambda) := D(L, M) + \sum_{i=1}^n \lambda_i \left(\sum_{j=1}^n (j-1) l_{ij} - x_i\right)$, and differentiate Φ with respect to m_{ij} and λ_i yields $\partial_{l_{ij}}\Phi = \log(l_{ij}) - \log(m_{ij}) + \lambda_i(j-1)$ and $\partial_{\lambda_i}\Phi = \sum_{j=1}^n (j-1) l_{ij} - x_i$. Setting these to zero yields
$$\arg\min_{L\in C_3} D(L, M) = \left(m_{ij} r_{ij}^{j-1}\right)_{1\leq i,j\leq n}, \tag{3.12}$$
where r_{ij} is the unique positive root to the polynomial equation $\sum_{j=1}^n (j-1) m_{ij} r^{j-1} - x_i = 0$. Let $f(r) := \sum_{j=1}^n (j-1) m_{ij} r^{j-1} - x_i$. It is easy to see that f is strictly increasing on $[0,\infty)$ with $f(0) \leq 0$ and $\lim_{r\to\infty} f(r) = +\infty$. Thus, it is easy (and quick) to find a numerical root of f on $[0,\infty)$ by Newton's method.

To summarize, we normalize the rows ($k = 1$), the columns ($k = 2$) and the barycenters of the rows ($k = 3$) sequentially. The two first operations correspond to the standard steps in Sinkhorn's algorithm. The third one is not exactly a barycenter normalization of the rows μ_i, since the total mass of every row changes from 1 to another value in (3.12). From a theoretical viewpoint, we can merge the steps 1 and 3 to a minimization problem in two variables. It can be checked that the minimum is specified by the proper value of s for $\mu(s)$ in (3.2). Nevertheless, the current presentation is more tractable from a numerical viewpoint.

3.4. **The non-irreducible case.** Without loss of generality, we assume $\mathbf{x} = \bar{\mathbf{x}}$, i.e., $x_1 \le \cdots \le x_n$. The non-irreducible case can be treated similarly on each irreducible components, i.e., for score subsequences $(x_{k_r+1}, \ldots, x_{k_{r+1}}) \preceq (k_r, k_r+1, \ldots, k_{r+1}-1)$, where all inequalities in (1.1) are strict except the equality of the two total sums. More precisely, put $k_0 := 0$, and define recursively k_{r+1} whenever $k_r < n$ by

$$k_{r+1} := \min\left\{k \in \mathbb{N} : k_r < k \le n \text{ and } \sum_{i=1}^{k} x_i = \frac{k(k-1)}{2}\right\}.$$

Let $R \in \mathbb{N}$ be such that $k_R = n$. Obviously, $R \ge 2$. We see from Definition 3.1 that for the non-irreducible case, there are different leagues with the probability for a team in a lower league to defeat a team of an upper league being zero. As a result, an element $M \in \Theta_n(\mathbf{x})$ must take the form:

$$M = \begin{pmatrix} M_1 & 0 & \cdots & 0 \\ 0 & M_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & M_R \end{pmatrix},$$

where M_r has $k_r - k_{r-1}$ rows, the number of teams in the r^{th} league. Multiplying M on the left with the row vectors $(1, \ldots, 1, 0, \ldots, 0)$ with k_r entries 1, and on the right with the column vector $(y_1, \ldots, y_n) = (0, 1, \ldots, n-1)$, permits us to prove that the M_r's are square matrices. Conversely, every block diagonal matrix made of R doubly stochastic matrices that satisfy

$$\sum_{j=1}^{k_r - k_{r-1}} m_{ij}(k_{r-1} + j - 1) = x_{k_{r-1}+1} \quad \text{for each } i \le k_r - k_{r-1},$$

is clearly an element of $\Theta_n(\mathbf{x})$. Also note that the constraints in the definition of $\Theta_n(\mathrm{x})$ are separable.[8]

Now we see that in the minimization problem, the function $h : m \mapsto m \log m$ is zero on the off-diagonal blocks, and $H(M) = \sum_{r=1}^{R} H_r(M_r)$, where each H_r is the entropy function defined on $\mathcal{M}_{k_r - k_{r-1}}(\mathbb{R})$. Since the constraints on each block are separable, we have R separate problems, each of which has a unique minimizer.[9] Numerically, we can just solve the R problems separately, which requires us to adapt the step 3 properly by replacing $j-1$ by the proper y_j. However, the R problems can be merged into one with the same steps as

[8]The set $\Theta_n(\mathbf{x})$ is the Cartesian product of factors $\Theta_{k_r - k_{r-1}}(\mathbf{x}_r, \mathbf{y}_r)$ with $\mathbf{x}_k = (x_{k_{r-1}+1}, \ldots, x_{k_r})$ and $\mathbf{y}_k = (k_{r-1}+1, \ldots, k_r)$, and the appropriate definition concerning the constraints.

[9]This is a slight extension of Proposition 3.3, which is mentioned in Remark 3.4. The difference is that the vector $\mathbf{y} = (0, 1, \ldots, n-1)$ is replaced by $(k_{r-1}+1, \ldots, k_r)$.

in the irreducible case. It suffices to set the initial matrix with ones on the diagonal blocks, and zeros on the off-diagonal blocks.

4. The shadow construction

In this section, we provide another construction of $M \in \Theta_n(\mathbf{x})$ for $\mathbf{x} \preceq (0, \dots, n-1)$ with the stochastic ordering constraint. We call it the shadow solution and present it in the form of a martingale coupling, i.e., a probability measure π^* on $\mathbb{R}^2$ with marginals the uniform measures $\{x_1, \dots, x_n\}$ and $\{0, \dots, n-1\}$, and a martingale constraint corresponding to (3.12). Then $M = (\mu_1, \dots, \mu_n) \in \Theta_n(\mathbf{x})$ are defined by $\mu_i(A) = n\pi^*(\{x_i\} \times A)$ for each $1 \le i \le n$ and Borel $A \subset \mathbb{R}$, or $m_{ij} = \pi^*(\{(x_i\} \times \{j-1\}\})$ for each $1 \le i, j \le n$.

We start by recalling the terminology in Beiglböck and Juillet [10]. For any closed set $E \subset \mathbb{R}$, denote by $\mathfrak{M}(E)$ the set of measures on E. Next we generalize the convex order (see Theorem 1.4) to the *extended convex order*, denoted by $\preceq_E$. For $\mu, \nu \in \mathfrak{M}(\mathbb{R})$, we say $\mu \preceq_E \nu$ if and only if for any nonnegative convex function $f : \mathbb{R} \to \mathbb{R}$,

$$\int_{\mathbb{R}} f(x)\mu(dx) \le \int_{\mathbb{R}} f(x)\nu(dx).$$

Furthermore, for $\mu, \nu \in \mathfrak{M}(\mathbb{R})$, we say $\mu \preceq_+ \nu$ if and only if for any nonnegative function $f : \mathbb{R} \to \mathbb{R}$,

$$\int_{\mathbb{R}} f(x)\mu(dx) \le \int_{\mathbb{R}} f(x)\nu(dx).$$

Define the shadow embedding: Let $\mu, \nu \in \mathfrak{M}(\mathbb{R})$ such that $\mu \preceq_E \nu$. Then there exists a measure $S^\nu(\mu)$, called the *shadow of* μ *in* ν, such that

(1) $S^\nu(\mu) \preceq_+ \nu$.
(2) $\mu \preceq_C S^\nu(\mu)$.
(3) If η is another measure satisfying (1) and (2), then we have $S^\nu(\mu) \preceq_C \eta$.

Now we use the shadow embedding to construct a martingale coupling π^*. Denoting $\frac{1}{k}\sum_{i=1}^k \delta_{z_i}$ by $U_{(z_1,\dots,z_k)}$ we first fix

$$\mu := U_{(x_1,\dots,x_n)}, \quad \nu := U_{(0,\dots,n-1)}, \tag{4.1}$$

and for each permutation σ we denote $(x_{\sigma(k)})_{1 \le k \le n}$ by $(x_k^\sigma)_{1 \le k \le n}$. By [10, Lemma 4.13 and Example 4.20], we can carry out the following algorithm because we have $\frac{1}{n}\delta_{x_k^\sigma} \le_E \nu_{k-1}$ at each step k.

(1) Let $\eta_0^\sigma := 0$ and $\nu_0^\sigma := \nu$ (we have $\eta_0^\sigma + \nu_0^\sigma = \nu$, and $\eta_k^\sigma + \nu_k^\sigma = \nu$ for each $1 \le k \le n$ with $\eta_0^\sigma \le \eta_1^\sigma \le \cdots$ and $\nu_0^\sigma \ge \nu_1^\sigma \ge \cdots$).
(2) For $k = 1, \dots, n$, define recursively:

$$\eta_k^\sigma := \eta_{k-1}^\sigma + S^{\nu_{k-1}^\sigma}\left(\frac{1}{n}\delta_{x_k^\sigma}\right), \quad \nu_k^\sigma := \nu_{k-1}^\sigma - S^{\nu_{k-1}^\sigma}\left(\frac{1}{n}\delta_{x_k^\sigma}\right). \tag{4.2}$$

(In particular, $\eta_k^\sigma + \nu_k^\sigma = \nu$ for each $1 \le k \le n$.)
(3) Define $\pi^\sigma \in \mathcal{P}(\mathbb{R}^2)$ (the set of probability measures on $\mathbb{R}^2$) by

$$\pi^\sigma(dx, dy) := \sum_{k=1}^n \delta_{x_k^\sigma}(dx) \otimes S^{\nu_{k-1}^\sigma}\left(\frac{1}{n}\delta_{x_k^\sigma}\right)(dy). \tag{4.3}$$

(4) Define $\pi^* \in \mathcal{P}(\mathbb{R}^2)$ by

$$\pi^* := \frac{1}{n!} \sum_{\sigma \in \mathfrak{S}(n)} \pi^\sigma. \tag{4.4}$$

Recall that the output measure π^* can be used to construct an element of $\Theta_n(\mathbf{x})$ by defining $\mu_i(A) = n\pi^*(\{x_i\} \times A)$, or equivalently, $m_{ij} = n\pi^*(\{x_i\} \times \{j-1\})$. Here we give a probabilistic interpretation of the above algorithm. If we want to know how many goals are scored by team i, we pick uniformly at random a sequence $x_{\sigma(1)}, x_{\sigma(2)}, \dots$ until $\sigma(k) = i$. During this process, the measure ν is filled with

$$\eta^\sigma_{k-1} = S^\nu\left(\frac{1}{n}\delta_{x_{\sigma(1)}}\right) + S^{\nu^\sigma_1}\left(\frac{1}{n}\delta_{x_{\sigma(2)}}\right) + \cdots + S^{\nu^\sigma_{k-2}}\left(\frac{1}{n}\delta_{x_{\sigma(k-1)}}\right) \preceq_+ \nu, \tag{4.5}$$

and it remains ν^σ_{k-1}. Now we embed $\frac{1}{n}\delta_{x_{\sigma(k)}}$ in ν^σ_{k-1}, and obtain (up to a factor of $1/n$) a probability distribution that underlies the random number of goals. Finally, note that this conditional distribution is obtained by conditioning on $\sigma(1), \dots, \sigma(k)$ rather than the entire σ.

Theorem 4.1. *For $\mathbf{x} \preceq (0, \dots, n-1)$, let $\pi^* \in \mathcal{P}(\mathbb{R}^2)$ be defined by the above algorithm. The marginals of π^* are the uniform measures μ and ν defined as in* (4.1)*. Define*

$$\mu_i := n\pi^*(\{x_i\} \times \cdot) \text{ for } 1 \le i \le n, \quad \text{and} \quad m_{ij} := n\pi^*(\{x_i\} \times \{j-1\}) \text{ for } 1 \le i, j \le n.$$

Then $M = (\mu_1, \dots, \mu_n) \in \Theta_n(\mathbf{x})$. Moreover, $(\mu_i)_{1 \le i \le n}$ is increasing in stochastic order, and the generalized tournament matrix corresponding to M (defined by (1.6)*) satisfies the SST.*

Proof. The first part of the theorem concerning the marginals and the score is in fact satisfied by π^σ for each permutation σ, and hence by π^*. For π^σ, this has been proved in [10, Lemma 4.13 and Example 4.20]. Note that we can check directly from (4.3) that the first marginal of π^* is $U_{(x_1,\dots,x_n)}$. We also see from this equation that $\mu_i = nS^{\nu^\sigma_{\sigma^{-1}(i)-1}}\left(\frac{1}{n}\delta_{x_i}\right)$, so $\frac{1}{n}\delta_{x_i} \preceq_C \frac{1}{n}\mu_i$ and $\delta_{x_i} \preceq_C \mu_i$, which implies $\int y d\mu_i(x) = x_i$. (This is a basic fact in martingale transport theory, but it can also be checked by integrating $\varphi : y \mapsto \pm y$ as in Theorem 1.4 (1).) The fact that the second marginal of π^σ is $U_{(0,\dots,n-1)}$ is less direct, but is still a consequence of [10, Section 4]. This second marginal is also $\sum_{k=1}^n S^{\nu^\sigma_{k-1}}\left(\frac{1}{n}\delta_{\sigma(k)}\right)$ where each of the summands is a measure of mass $\frac{1}{n}$, and is a shadow embedded in the measure $\nu^\sigma_{k-1} \preceq_+ \nu$, whose mass is $n-(k-1)$. The measure ν^σ_k is the new measure of mass $n-k$ obtained by the formula $\nu^\sigma_k = \nu^\sigma_{k-1} - S^{\nu^\sigma_{k-1}}\left(\frac{1}{n}\delta_{\sigma(k)}\right)$ after embedding.

The second part on the fact that $(\mu_i)_{1 \le i \le n}$ is increasing in stochastic order is purely a property of π^* obtained by symmetrization. Let $1 \le i < i' \le n$. We need to prove $\mu_i \preceq_{\text{sto}} \mu_{i'}$, i.e.,

$$\frac{1}{n!}\sum_{\sigma \in \mathfrak{S}(n)} S^{\nu^\sigma_{\sigma^{-1}(i)-1}}\left(\frac{1}{n}\delta_{x_i}\right) \preceq_{\text{sto}} \frac{1}{n!}\sum_{\sigma \in \mathfrak{S}(n)} S^{\nu^\sigma_{\sigma^{-1}(i')-1}}\left(\frac{1}{n}\delta_{x_{i'}}\right). \tag{4.6}$$

Now we explain how to derive (4.6) from (4.3) and (4.4). For the sum on the left side, we embed x_i (in fact the atomic measure $\frac{1}{n}\delta_{x_i}$) in $n!$ different ways depending on σ. Here $\sigma^{-1}(i)-1$ is the number of goals $x_{\sigma(a)}$ (in fact the atomic measures $\frac{1}{n}\delta_{x_{\sigma(a)}}$) that are embedded in ν before embedding x_i. The remaining part of ν, where $\frac{1}{n}\delta_{x_i}$ is embedded, is the measure $\nu^\sigma_{\sigma^{-1}(i)-1}$. The same is done for $x_{i'}$ for the sum on the right side. Since stochastic ordering is preserved by addition, it suffices to prove (4.6) for the sum of two terms σ and $\sigma' = (i,i')\circ\sigma$, where (i,i') is the transposition of i and i' Assume without loss of generality that $k < k'$, with $k = \sigma^{-1}(i)$ and $k' = \sigma^{-1}(i')$. Also assume for simplicity $k = 1$ (we will explain how to relax this assumption later). Let

$$\alpha = \frac{1}{n}\delta_{x_i}, \quad \beta = \frac{1}{n}\sum_{a=2}^{k'-1}\delta_{x_{\sigma(a)}}, \quad \gamma = \frac{1}{n}\delta_{x_{i'}}.$$

Note that $\alpha \preceq_{\mathrm{sto}} \gamma$. By Lemma 4.2 (whose proof is postponed), we have:

$$S^\nu\left(\frac{1}{n}\delta_{x_i}\right) + S^{\nu - S^\nu(\sum_{a=1}^{k'-1}\frac{1}{n}\delta_{x_{\sigma(a)}})}\left(\frac{1}{n}\delta_{x_{i'}}\right) \preceq_{\mathrm{sto}} S^\nu\left(\frac{1}{n}\delta_{x_{i'}}\right) + S^{\nu - S^\nu(\sum_{a=2}^{k'}\frac{1}{n}\delta_{x_{\sigma(a)}})}\left(\frac{1}{n}\delta_{x_{i'}}\right). \tag{4.7}$$

We recognize (without the factor $1/n!$) the contribution of σ and σ' to the two sums defining μ_i and $\mu_{i'}$ (one part for $k = 1$ and the other for k'). By associativity of shadows from the theory in [10], $S^\nu(\sum_{a=1}^{k'-1}\frac{1}{n}\delta_{x_{\sigma(a)}})$ and $S^\nu(\sum_{a=2}^{k'}\frac{1}{n}\delta_{x_{\sigma(a)}})$ are the measure $\eta^\sigma_{k'-1}$ and $\eta^{\sigma'}_{k'-1}$ respectively, so we recognize in (4.7) the measures $\nu^\sigma_k = \nu - \eta^\sigma_{k-1}$ and $\nu^\sigma_{k'} = \nu - \eta^\sigma_{k'-1}$ appearing in (4.6). It remains the case $1 < k < k'$. Let $\xi = \sum_{a=1}^{k-1}\delta_{x_a}$. The two Dirac masses $\alpha = \delta_{x_{\sigma(k)}} = \delta_{x_{\sigma'(k')}}$ and $\gamma = \delta_{x_{\sigma(k')}} = \delta_{x_{\sigma(k')}}$ are no longer embedded in ν and $\nu - S^\nu(\gamma+\beta)$, and ν and $\nu - S^\nu(\alpha+\beta)$ respectively, but in $\nu' := \nu - S^\nu(\xi)$ and $\nu - S^\nu(\xi+\gamma+\beta) = \nu' - S^{\nu'}(\gamma+\beta)$, and ν' and $\nu - S^\nu(\xi+\alpha+\beta) = \nu' - S^{\nu'}(\alpha+\beta)$ respectively. It suffices to apply Lemma 4.2 to ν' instead of ν to conclude. □

We conclude this section with the following lemma for the shadow embedding.

Lemma 4.2. *If $\alpha + \beta + \gamma \preceq_E \nu$ and $\alpha \preceq_{\mathrm{sto}} \gamma$, then*

$$S^\nu(\alpha) + S^{\nu - S^\nu(\gamma+\beta)}(\alpha) \preceq_{\mathrm{sto}} S^\nu(\gamma) + S^{\nu - S^\nu(\alpha+\beta)}(\gamma)$$

Proof. Since $\alpha \preceq_{\mathrm{sto}} \gamma$, it follows from [27] that $S^\nu(\alpha) \preceq_{\mathrm{sto}} S^\nu(\gamma)$. Since $\alpha+\beta \preceq_{\mathrm{sto}} \gamma+\beta$, the same argument shows $S^\nu(\alpha+\beta) \preceq_{\mathrm{sto}} S^\nu(\gamma+\beta)$. By associativity of the shadow projection (with $\alpha+\beta+\gamma \preceq_{\mathrm{sto}} \nu$), we have:

$$S^{\nu - S^\nu(\gamma+\beta)}(\alpha) = S^\nu(\alpha+\beta+\gamma) - S^\nu(\gamma+\beta) \text{ and } S^{\nu - S^\nu(\alpha+\beta)}(\gamma) = S^\nu(\alpha+\beta+\gamma) - S^\nu(\alpha+\beta).$$

Thus, $S^{\nu - S^\nu(\gamma+\beta)}(\alpha) \preceq_{\mathrm{sto}} S^{\nu - S^\nu(\alpha+\beta)}(\gamma)$, which yields the desired result. □

Remark 4.3. *Computing the distribution μ_i that gives the number of goals scored by team i seems to be difficult for large n, because there are $n!$ terms in the sum. Nevertheless, it is relatively easy to simulate the random number of goals from the probabilistic interpretation (4.5) of the algorithm. More precisely, the shadow $S^{\nu'}(m\delta_x)$ is the unique measure of mass m and center of mass x that takes the form $(F^{-1}_{\nu'})_\#\mathrm{Leb}_{]\alpha,\beta]}$. Here $F^{-1}_{\nu'}$ is any quantile function of ν', i.e., any nondecreasing function such that for every $\beta \in]0,1]$, the measure $\nu'_\beta := (F^{-1}_{\nu'})_\#\mathrm{Leb}_{[0,\beta]}$ satisfies $\nu'_\beta \preceq_+ \nu'$ and is of mass β. One such example is the inverse of the cumulative distribution function $F^{-1}_{\nu'}(\gamma) = \inf\{x \in \mathbb{R} : \nu'(]-\infty,x]) \geq \gamma\}$.*

5. Nontransitivity in the soccer model

In this section, we consider (possible) nontransitive situation in the soccer model. The idea comes from the work of Steinhaus and Trybuła [50, 58], in which they characterize the triples $(\xi, \eta, \zeta) \in [0,1]^2$ that correspond to the probabilities $\xi = \mathbb{P}(X < Y)$, $\eta = \mathbb{P}(Y < Z)$ and $\zeta = \mathbb{P}(Z < X)$ for three independent random variables X, Y, Z. However, their result was only established under the assumption that $\mathbb{P}(X = Y) = \mathbb{P}(Y = Z) = \mathbb{P}(Z = X) = 0$, which does not necessarily hold for the soccer model. Here we adapt Steinhaus and Trybuła's result to the soccer model, where ties are allowed.

Theorem 5.1 ([50, 58])**.** *Let*

$$A := \left\{ (\mathbb{P}(X < Y), \mathbb{P}(Y < Z), \mathbb{P}(Z < X)) \in [0,1]^3 : \begin{array}{l} (X, Y, Z) \textit{ are independent and} \\ \mathbb{P}(X = Y) = \mathbb{P}(Y = Z) = \mathbb{P}(Z = X) = 0 \end{array} \right\},$$

and $\alpha : [0,1]^2 \to \mathbb{R}_+$ *be defined by*

$$\alpha(\xi, \eta) = \begin{cases} \max\left(\frac{1-\xi}{\eta}, \frac{1-\eta}{\xi}, 1 - \xi\eta\right) & \textit{for } \xi + \eta > 1, \\ 1 & \textit{for } \xi + \eta \le 1. \end{cases} \tag{5.1}$$

Then $A = D$*, where*

$$D = \left\{ (\xi, \eta, \zeta) \in [0,1]^3 : 1 - \alpha(1 - \xi, 1 - \eta) \le \zeta \le \alpha(\xi, \eta) \right\}. \tag{5.2}$$

More symmetric descriptions of D are given in [54, 59]. Steinhaus and Trybuła observed a nontransitive phenomenon (which they called a "paradox"): $(\xi, \eta, \zeta) \in D$ can have all its three coordinates larger that $1/2$. For instance, the point $\left(\frac{\sqrt{5}-1}{2}, \frac{\sqrt{5}-1}{2}, \frac{\sqrt{5}-1}{2}\right)$ lies on the boundary of D, with $\frac{\sqrt{5}-1}{2} \approx 0,618 > \frac{1}{2}$.

To apply Theorem 5.1 to the soccer model, we introduce a set A' that is a priori larger than A: Similar to (1.6) it is the set of triples (ξ', η', ζ') with

$$\xi' = \mathbb{P}(X < Y) + \frac{1}{2}\mathbb{P}(X = Y),\ \eta' = \mathbb{P}(Y < Z) + \frac{1}{2}\mathbb{P}(Y = Z),\ \zeta' = \mathbb{P}(Z < X) + \frac{1}{2}\mathbb{P}(Z = X), \tag{5.3}$$

and (X, Y, Z) independent. Here we do not assume $\mathbb{P}(X = Y) = \mathbb{P}(Y = Z) = \mathbb{P}(Z = X) = 0$, but with this additional assumption, we have $(\xi', \eta', \zeta') = (\mathbb{P}(X < Y), \mathbb{P}(Y < Z), \mathbb{P}(Y < Z)) \in A$. So $A \subset A'$.

Proposition 5.2. *With the notation above, we have* $A' = D$.

Proof. By Theorem 5.1, we have $D = A$. Since $A \subset A'$. It remains to show that $A' \subset A = D$. Let $(\xi', \eta', \zeta') \in A'$, and (X, Y, Z) be a triple of random variables satisfying (5.3). Let U_X, U_Y and U_Z be three independent random variables uniform on $[-\frac{1}{2}, \frac{1}{2}]$, and also independent of (X, Y, Z). Define

$$X_n = X + \frac{1}{n}U_X, \quad Y_n = Y + \frac{1}{n}U_Y, \quad Z_n = Z + \frac{1}{n}U_Z.$$

First we claim that $\mathbb{P}(X_n = Y_n) = \mathbb{P}(Y_n = Z_n) = \mathbb{P}(Z_n = X_n) = 0$. To see this, we observe $\{X_n = Y_n\} = \{U_Y - U_X = n(X - Y)\}$, where both sides of the equality are independent. Since $U_Y - U_X$ is diffuse, we get $\mathbb{P}(U_Y - U_X = z) = 0$ for each $z \in \mathbb{R}$. Integrating in z with respect to the distribution of $n(X-Y)$ yields $\mathbb{P}(X_n = Y_n) = 0$. The other two equalities follow the same. Thus, $(\mathbb{P}(X_n < Y_n), \mathbb{P}(Y_n < Z_n), \mathbb{P}(Z_n < X_n)) \in A = D$ by Theorem 5.1.

Next we prove the following limit:

$$\mathbb{P}(X_n < Y_n) \longrightarrow_{n\to\infty} \frac{\mathbb{P}(X < Y) + \mathbb{P}(X \leq Y)}{2} = \mathbb{P}(X < Y) + \frac{1}{2}\mathbb{P}(X = Y). \tag{5.4}$$

Note that

$$\begin{aligned}\mathbb{P}(X_n < Y_n) =& \mathbb{P}(X = Y)\,\mathbb{P}(X_n < Y_n \,|\, X = Y) \\ &+ \mathbb{P}(X_n < Y_n \text{ and } X < Y) + \mathbb{P}(X_n < Y_n \text{ and } X > Y).\end{aligned}$$

Clearly, $\mathbb{P}(X_n < Y_n \,|\, X = Y) = \mathbb{P}(U_X < U_Y) = 1/2$. Since $\mathbb{P}(\frac{1}{n}|U_X - U_Y| \leq \frac{1}{n}) = 1$, we also have:

$$\mathbb{P}\left(X - Y < -\frac{1}{n}\right) \leq \mathbb{P}(X_n < Y_n \text{ and } X < Y) \leq \mathbb{P}(X - Y < 0).$$

By the dominated convergence theorem and the squeeze theorem, we have $\lim_{n\to\infty} \mathbb{P}(X_n < Y_n \text{ and } X < Y) = \mathbb{P}(X < Y)$. Similarly, $\lim_{n\to\infty} \mathbb{P}(X_n < Y_n \text{ and } X > Y) = 0$. Similar to (5.4), η' and ζ' can be obtained as limits in terms of (X_n, Y_n, Z_n). Finally, since D is closed, we get $(\xi', \eta', \zeta') \in A$ by sending $n \to \infty$. □

Now we see that the soccer model permits nontransitivity for some parameters in the limits of Proposition 5.2.

Theorem 5.3. *For every $n \geq 6$ and each $(\xi, \eta, \zeta) \in D$, there exists $(\mu_1, \ldots, \mu_n) \in \Theta_n$ such that the corresponding generalized tournament matrix (defined by* (1.6)*) satisfies $(p_{12}, p_{23}, p_{31}) = (\xi, \eta, \zeta)$.*

Proof. Let $(\xi, \eta, \zeta) \in [0,1]^3$ be a element of D (we call it "cyclic" as in [59]). We look for $S \subset \mathbb{R}$ of cardinal 6, and three S-valued random variables (X, Y, Z) with disjoint values such that $(\mathbb{P}(X < Y), \mathbb{P}(Y < Z), \mathbb{P}(Z < X)) = (\xi, \eta, \zeta)$. Permutations of the random variables or replacement by their opposite permits us to see that D is invariant by permutation of the coordinates, and by replacement with the complementary to 1. Therefore, we can assume $\xi \leq \eta \leq \zeta$, and also $\xi \leq 1 - \zeta$. By the alternative characterization in [59, Lemma 2.4], $(\xi, \eta, \zeta) \in D$ implies that $(\xi/(1-\zeta), \eta, 0)$ is also cyclic. Thus, a symmetric construction to the one just after (21) in [58] permits to introduce three disjoint laws μ_{X_0}, μ_Y, μ_Z on a set of five points with $(\mathbb{P}(X_0 < Y), \mathbb{P}(Y < Z), \mathbb{P}(Z < X_0)) = (\xi/(1-\zeta), \eta, 0)$. Replacing μ_{X_0} by a Dirac measure $\mu_{X_1} = \delta_x$ for $x \in \mathbb{R}$ larger than the 5 previous points, one creates the cyclic vector $(0, \eta, 1)$ attached to (X_1, Y, Z). Finally, (ξ, η, ζ) is obtained as an element of D for the triple (X, Y, Z), where $X \sim \zeta\mu_{X_1} + (1-\zeta)\mu_{X_0}$ (comparing with [58, p. 328], where the interpolation is not linear). Note that Z is supported on one point, X on two points and Y on three points. An increasing homeomorphism φ now maps S on $\{-5, -4, \ldots, 0\}$. Let μ_1, μ_2, μ_3 be the distributions of $-\varphi(X), -\varphi(Y), -\varphi(Z)$ respectively. Since the supports of these measures are disjoint, we can complete (μ_1, μ_2, μ_3) with $(\mu_4, \ldots, \mu_n)$ to obtain an admissible $(\mu_1, \ldots, \mu_n) \in \Theta_n$. □

Appendix A. Partial orders via the generalized tournament matrix

In this part, we provide further thoughts on partial orders induced by the generalized tournament matrix, scrutinizing various results in [3, 26]. As discussed in the introduction, there are two ways to compare the teams $\mathcal{T}_n := (T_i)_{1\le i\le n}$ based on different levels of information:

- We compare T_i and T_j by looking at p_{ij} from the generalized tournament matrix $P \in \mathcal{G}_n$ or $P \in \mathcal{G}'_n$.
- We compare T_i and T_j by looking at x_i and x_j from the score $\mathbf{x} = (x_1, \dots, x_n)$, where $x_i := \sum_{j\neq i}^n p_{ij}$ or $x_i := \sum_{j=1}^n p_{ij}$.

Given P and $\mathbf{x}$, there are two binary relations naturally associated with $\mathcal{T}_n$:

(1) *Score-based relation*: $(T_i \ \le_x \ T_j)$ if and only if $x_i \le x_j$.
(2) *Results-based relation*: $(T_i \ \le_P \ T_j)$ if and only if $p_{ij} \le 1/2$ (if p_{ij} exists).

We can also define the strict relations: $(T_i \ <_x \ T_j)$ if $x_i < x_j$, and $(T_i \ <_P \ T_j)$ if $p_{ij} < 1/2$. Finally, $(T_i \ =_x \ T_j)$ means $(T_i \ \le_x \ T_j)$ and $(T_i \ \ge_x \ T_j)$, i.e., $x_i = x_j$; and $(T_i \ =_P \ T_j)$ means $(T_i \ \le_P \ T_j)$ and $(T_i \ \ge_P \ T_j)$, i.e., $p_{ij} = 1/2$.

Now we recall the definition of a partial order $\preceq$ on a set $\mathcal{T}$:

(a) Reflexivity: $T \ \preceq \ T$.
(b) Transitivity: $(T \ \preceq \ T')$ and $(T' \ \preceq \ T'')$ implies $(T \ \preceq \ T'')$.
(c) Antisymmetry: $(T \ \preceq \ T')$ and $(T' \ \preceq \ T)$ implies $(T = T')$.

Moreover, if all the elements are comparable, i.e., $(T \ \preceq \ T')$ or $(T' \ \preceq \ T)$ for all $T, T' \in \mathcal{T}$, the partial order is called a *total order*. If all the properties except (c) are satisfied, the relation is called a *preorder* or a *total preorder*, respectively. It is easy to see that $\le_x$ is a total preorder. In the remaining of this section, we focus on the relation $\le_P$, and the question whether it is a total preorder.

For a general $P \in \mathcal{G}_n$ or $P \in \mathcal{G}'_n$, the relation $\le_P$ may be far from being a total preorder. The definition of $\mathcal{G}'_n$ (by setting $p_{ii} = \frac{1}{2}$) ensures the reflexivity and totality (which are not satisfied by $\mathcal{G}_n$). Here our goal is to characterize $\le_P$ as a total preorder on $\mathcal{G}'_n$, so it remains to consider whether $\le_P$ is transitive. We will review and elaborate some results evoked in the literature, especially in [3, 26], and present a new observation on SST in Proposition A.2. It was explained in in [26, p.917] that even when $\le_P$ is transitive, this preorder can be different from $\le_x$. Nevertheless, the notion of SST defined in (1.4) and studied in Proposition 2.4 and Theorem 3.5 allows us to establish the equivalence between $\le_x$ and $\le_P$. We can reformulate it as follows: for all $1 \le i, j, k \le n$,

$$(T_i \ \le_P \ T_j) \text{ and } (T_j \ \le_P \ T_k) \quad \Longrightarrow \quad (T_i \ \le_P \ T_k) \text{ and } p_{ik} \le \min(p_{ij}, p_{jk}).$$

Note that the SST is not a property of the binary relation $\le_P$ alone, but rather of the generalized tournament matrix P. Moreover, the property of SST already contains the transitivity of $\le_P$ in its formulation: if $P \in \mathcal{G}'_n$, the SST implies that $(\mathcal{T}, \le_P)$ is a total preorder. On the other hand, the SST can a priori be satisfied by $P \in \mathcal{G}_n$ even though $T_i \le_P T_i$ does not hold. It is easy to see that the transitivity of $\le_P$ for $P \in \mathcal{G}_n$ implies that for $P \in \mathcal{G}'_n$. The contrary is false because if $p_{ij} = 1/2$ for some $i \neq j$, both relations $(T_i \le_P T_j)$ and $(T_j \le_P T_i)$ should enforce $(T_i \le_P T_i)$ and $(T_j \le_P T_j)$, but these relations do not apply for $\mathcal{G}_n$, whose matrices have undetermined diagonal. It seems to us that most authors only consider $\le_P$ for $P \in \mathcal{G}_n$.

In [26, Theorem 2.1], Joe stated that $\leq_P$ is equivalent to $\leq_x$ under the SST. However, the one-line proof is not entirely convincing. We restate this result in Corollary A.3. That is a consequence of Proposition A.2 where the SST is compared with the almost equivalent notion of monotonicity.

Definition A.1 (Monotonic matrix P)**.** *Let $P = (p_{ij})_{1\leq i\neq j\leq n} \in \mathcal{G}_n$, or $P = (p_{ij})_{1\leq i,j\leq n} \in \mathcal{G}'_n$. The matrix P is called* monotonic *if P is decreasing along rows, i.e., if for every $1 \leq i,j,k \leq n$, inequality $j < k$ implies $p_{ij} \geq p_{ik}$, provided p_{ij} and p_{ik} are defined.*
Equivalently, the matrix P is monotonic if P is increasing down columns, i.e., for every $1 \leq i,j,k \leq n$, inequality $j < k$ implies $p_{ji} \leq p_{ki}$, provided p_{ji} and p_{ki} are defined.

In [3, p.849], Aldous and Kolesnik recalled that if the SST is satisfied, the matrix P is monotonic. They also showed that these two properties are not equivalent, though the same terminology is often used interchangeably for the two close concepts. Their counterexample is based on a matrix P with some entries having value one half.[10] Example A.4 below provides another counterexample, where the monotonicity does not imply the SST for a matrix $P \in \mathcal{G}_n$ with all the entries different from $\frac{1}{2}$.

The following proposition shows the equivalence between the SST and the monotonicity.

Proposition A.2. *Let $P = (p_{ij})_{1\leq i\neq j\leq n} \in \mathcal{G}_n$. The following statements are equivalent:*

(i) The relation $\leq_P$ satisfies the property of SST (1.4) *except that we don't require $p_{ii} = 1/2$ and $p_{jj} = 1/2$ when $p_{ij} = p_{ji} = 1/2$ for some $i \neq j$.*

(ii) The relation $\leq_P$ satisfies the property of SST (1.4) *for P seen as a matrix in $\mathcal{G}'_n$.*

(iii) Both conditions are satisfied:
- *There exists a relabelling of P by a permutation matrix $M \in \mathcal{M}_n(\mathbb{R})$ such that MPM^T is monotonic in $\mathcal{G}'_n$.*
- *For every i, j, $p_{ij} = \frac{1}{2}$ implies $p_{ik} = p_{jk}$ for every $1 \leq k \leq n$.*

Moreover, a relabelling in (iii) is admissible if and only if $i < j \Rightarrow (T_i \ \leq_P \ T_j)$.

Proof. $(i) \Leftrightarrow (ii)$. It is straightforward that (ii) implies (i). Conversely, if (i) is satisfied, extending P to $\mathcal{G}'_n$ permits us to recover the full condition (1.4).
$(ii) \Rightarrow (iii)$. Assume that $P \in \mathcal{G}'_n$ satisfies the SST. We relabel the indices in such a way that $T_1 \ \leq_P \ T_2 \ \leq_P \cdots \leq_P \ T_n$. (The reason why this is possible is straightforward, see [3] for a graph theoretic proof.) Let i, j, k be three different indices with $j < k$. We want to prove $p_{ij} \geq p_{ik}$. Recall that $p_{jk} \leq 1/2$. If $i < j < k$, then we have $p_{ij} \leq 1/2$, and the SST implies $p_{ik} \leq p_{ij}$. The case $j < k < i$ is similar. Finally, if $j < i < k$, we have $p_{ik} \leq 1/2 \leq p_{ij}$. Now we prove the second part of (ii). Assume $p_{ij} = 1/2$, and in particular $p_{ji} \leq 1/2$. If $p_{ik} \leq 1/2$, then we have $p_{jk} \leq p_{ik} \leq 1/2$. Since $p_{ij} \leq 1/2$, we obtain $p_{ik} \leq p_{jk} \leq 1/2$. Thus, $p_{ik} = p_{jk}$. The case $p_{ki} > 1/2$ can be treated similarly.
$(iii) \Rightarrow (ii)$. Assume that there exists a relabelling such that $(p_{ij})_{i,j} \in \mathcal{G}'_n$ is monotonic and $p_{ij} = \frac{1}{2}$ implies $p_{ik} = p_{jk}$. The first property implies $p_{ij} \leq p_{ii} = \frac{1}{2}$ for every pair (i,j) such that $i < j$. Conversely, if $p_{ij} < 1/2 = p_{ii}$, we must have $i < j$. Assume now $T_i \ \leq_P \ T_j \ \leq_P \ T_k$. If these relations are from $p_{ij} < 1/2$ and $p_{jk} < 1/2$, we have $i < j < k$, and $p_{ik} \leq \min(p_{ij}, p_{jk})$ follows from the monotonicity of $P \in \mathcal{G}'_n$. Otherwise, $\max(p_{ij}, p_{jk}) = 1/2$, and the second property permits us to prove $p_{ik} = \min(p_{ij}, p_{jk})$.

[10] Their example is $p_{21} = p_{23} = 1/2$ and $p_{13} < 1/2$.

To conclude, we make it clear when a given relabelling is admissible. We have already proved in part $(ii) \Rightarrow (iii)$ that if the labelling satisfies $T_1 \leq_P \cdots \leq_P T_n$, the matrix is admissible for (iii). Conversely, assume that after relabelling the matrix P satisfies (iii). By monotonicity, the entries on the upper right part of P are larger or equal to $1/2$ because it is the value on the diagonal. Therefore, $i < j$ implies $p_{ij} \leq 1/2$, i.e., $T_i \leq_P T_j$. □

Corollary A.3. *Assume that $\leq_P$ satisfies the property of SST for $P \in \mathcal{G}'_n$. Then for all $i, j \leq n$,*

$$(T_i \leq_P T_j) \Longleftrightarrow (T_i \leq_x T_j).$$

Proof. With the SST both $\leq_P$ and $\leq_x$ are total preorder on the set of teams. Therefore, it suffices to prove for every i, j, the relation $T_i <_P T_j$ implies $T_i <_x T_j$, and $T_i =_P T_j$ implies $T_i =_x T_j$. In fact, $T_i =_P T_j$ means that $p_{ij} = 1/2$. By the SST, it implies $p_{ik} = p_{jk}$ for every $k \leq n$. Hence, $T_i =_x T_j$. Now assume $T_i <_P T_j$. Up to a proper relabelling corresponding to (iii) in Proposition A.2, it implies $i < j$. Thus, the elements of the i^{th} row are entrywise smaller or equal to those of the j^{th} row. But they are not all equal because $p_{ij} < 1/2 = p_{jj}$, Thus, we get $T_i <_x T_j$. □

The following example shows that if $P = (p_{ij})_{1 \leq i \neq j \leq n} \in \mathcal{G}_n$ is only monotonic, the relation $\leq_P$ may fail to satisfy the SST, and be equivalent to $\leq_x$.

Example A.4. *Take $p_{12} = 0.3,\ p_{13} = 0,\ p_{23} = 0.6$, and symmetrically $p_{21} = 0.7,\ p_{31} = 1,\ p_{32} = 0.4$. This matrix is monotonic in $\mathcal{G}_3$ (but not in $\mathcal{G}'_3$). The coefficients $p_{23},\ p_{31},\ p_{21}$ are larger than $1/2$, but $p_{23} = 0.6 \geq \max(p_{21}, p_{13})$ is not satisfied. So $\leq_P$ does not satisfy the property of SST. Furthermore, $\leq_P$ is transitive with $T_1 \leq_P T_3 \leq_P T_2$, which is different from $T_1 \leq_x T_2 \leq_x T_3$. Note also that there is no relabelling that make P monotonic in $\mathcal{G}'_3$.*

The following example show that even if $P = (p_{ij})_{i \neq j} \in \mathcal{G}_n$ is both strongly transitive and monotonic, the order of the indices can be different from $\leq_P$ and $\leq_x$.

Example A.5. *For $n \in 2\mathbb{N}_{>0}$, take $p_{2k,2k-1} = 0.4$, $p_{2k-1,2k} = 0.6$ and the other entries 0 or 1 with $p_{ij} = 0$ if $i < j$. The order $\leq_P$ ranks the teams as follows:*

$$(T_2 \leq_P T_1) \leq_P (T_3 \leq_P T_4) \leq_P \cdots \leq_P (T_{2k} \leq_P T_{2k-1}) \leq_P \cdots \leq_P (T_n \leq_P T_{n-1}).$$

Still $(p_{ij})_{i \neq j} \in \mathcal{G}_n$ is monotonic.

Acknowledgments. We thank David Aldous, Brett Kolesnik and Erhan Bayraktar for helpful discussions. Juillet thanks Tom Muller and Franck-kabrel Ngoupe for numerical experiments.

Funding. This research was funded, in whole or in part, by the Agence Nationale de la Recherche (ANR), Grant ANR-23-CE40-0017. Guo is supported by Cai Yuanpei Découverte 2024 and Bourse *"Systemic Robustness and Systemic Failure"* provided by the Institut Europlace de Finance. Tang acknowledges financial support by NSF grant DMS-2206038, the Columbia Innovation Hub grant, and the Tang Family Assistant Professorship.

References

[1] I. Adler, Y. Cao, R. Karp, E. A. Peköz, and S. M. Ross. Random knockout tournaments. *Oper. Res.*, 65(6):1589–1596, 2017.

[2] D. J. Aldous. A prediction tournament paradox. *Amer. Statist.*, 75(3):243–248, 2021.

[3] D. J. Aldous and B. Kolesnik. To stay discovered: on tournament mean score sequences and the Bradley-Terry model. *Stochastic Process. Appl.*, 150:844–852, 2022.

[4] M. Avellaneda. Minimum entropy calibration of asset pricing models, internat. *J. Theoret. Appl. Finance*, 1:447472, 1998.

[5] R. Axelrod and W. D. Hamilton. The evolution of cooperation. *Science*, 211(4489):1390–1396, 1981.

[6] M. Bassan, S. Donderwinkel, and B. Kolesnik. Tournament score sequences, Erdös-Ginzburg-Ziv numbers, and the Lévy-Khintchine method. 2024. arXiv:2407.01441.

[7] H. H. Bauschke and A. S. Lewis. Dykstra's algorithm with Bregman projections: a convergence proof. *Optimization*, 48(4):409–427, 2000.

[8] E. Bayraktar, S. Deng, and D. Norgilas. A potential-based construction of the increasing supermartingale coupling. *Ann. Appl. Probab.*, 33(5):3803–3834, 2023.

[9] M. Beiglböck, M. Huesmann, and F. Stebegg. Root to Kellerer. In *Séminaire de Probabilités XLVIII*, volume 2168 of *Lecture Notes in Math.*, pages 1–12. Springer, Cham, 2016.

[10] M. Beiglböck and N. Juillet. On a problem of optimal transport under marginal martingale constraints. *Ann. Probab.*, 44(1):42–106, 2016.

[11] M. Beiglböck and N. Juillet. Shadow couplings. *Trans. Amer. Math. Soc.*, 374(7):4973–5002, 2021.

[12] J.-D. Benamou, G. Carlier, M. Cuturi, L. Nenna, and G. Peyré. Iterative Bregman projections for regularized transportation problems. *SIAM J. Sci. Comput.*, 37(2):A1111–A1138, 2015.

[13] R. A. Bradley. Science, statistics, and paired comparisons. *Biometrics*, 32(2):213–232, 1976.

[14] R. A. Bradley and M. E. Terry. Rank analysis of incomplete block designs: I. the method of paired comparisons. *Biometrika*, 39(3/4):324–345, 1952.

[15] P. Cartier, J. M. G. Fell, and P.-A. Meyer. Comparaison des mesures portées par un ensemble convexe compact. *Bull. Soc. Math. France*, 92:435–445, 1964.

[16] H. Chen, H. Zhao, H. Lam, D. Yao, and W. Tang. MallowsPO: Fine-tune your LLM with preference dispersions. In *ICLR*, volume 13, 2025.

[17] A. Claesson, M. Dukes, A. F. Franklín, and S. u. O. Stefánsson. Counting tournament score sequences. *Proc. Amer. Math. Soc.*, 151(9):3691–3704, 2023.

[18] H. A. David. Tournaments and paired comparisons. *Biometrika*, 46(1-2):139–149, 1959.

[19] H. De March and P. Henry-Labordere. Building arbitrage-free implied volatility: Sinkhorn's algorithm and variants. *arXiv preprint arXiv:1902.04456*, 2019.

[20] S. Donderwinkel and B. Kolesnik. Tournaments and random walks. 2024. arXiv:2403.12940.

[21] L. R. Ford, Jr. Solution of a ranking problem from binary comparisons. *Amer. Math. Monthly*, 64(8):28–33, 1957.

[22] R. L. Graham and J. H. Spencer. A constructive solution to a tournament problem. *Canad. Math. Bull.*, 14:45–48, 1971.

[23] F. Harary and L. Moser. The theory of round robin tournaments. *Amer. Math. Monthly*, 73:231–246, 1966.

[24] G. H. Hardy, J. E. Littlewood, and G. Pólya. *Inequalities*. Cambridge, at the University Press,, 1952. 2d ed.

[25] F. Hirsch, C. Profeta, B. Roynette, and M. Yor. *Peacocks and associated martingales, with explicit constructions*, volume 3 of *Bocconi & Springer Series*. Springer, Milan; Bocconi University Press, Milan, 2011.

[26] H. Joe. Majorization, entropy and paired comparisons. *Ann. Statist.*, 16(2):915–925, 1988.

[27] N. Juillet. Stability of the shadow projection and the left-curtain coupling. *Ann. Inst. Henri Poincaré Probab. Stat.*, 52(4):1823–1843, 2016.

[28] H. G. Kellerer. Integraldarstellung von Dilationen. In *Transactions of the Sixth Prague Conference on Information Theory, Statistical Decision Functions, Random Processes (Tech. Univ. Prague, Prague, 1971; dedicated to the memory of Antonín Špaček)*, pages 341–374. Academia [Publishing House of the Czechoslovak Academy of Sciences], Prague, 1973.

[29] M. G. Kendall. Further contributions to the theory of paired comparisons. *Biometrics*, 11(1):43–62, 1955.

[30] H. G. Landau. On dominance relations and the structure of animal societies. III. The condition for a score structure. *Bull. Math. Biophys.*, 15:143–148, 1953.
[31] J.-F. Laslier. *Tournament solutions and majority voting*, volume 7. Springer, 1997.
[32] E. P. Lazear and S. Rosen. Rank-order tournaments as optimum labor contracts. *J. Political Econ.*, 89(5):841–864, 1981.
[33] G. Lowther. Limits of one-dimensional diffusions. *Ann. Probab.*, 37(1):78–106, 2009.
[34] R. D. Luce. *Individual Choice Behavior a Theoretical Analysis.* John Wiley & Sons Inc., 1959.
[35] P. A. MacMahon. An american tournament treated by the calculus of symmetric functions. *Quart. J. Math*, 49:1–36, 1920.
[36] C. L. Mallows. Non-null ranking models. I. *Biometrika*, 44(1/2):114–130, 1957.
[37] P. McCullagh and J. A. Nelder. *Generalized linear models.* Monographs on Statistics and Applied Probability. Chapman & Hall, London, 1983.
[38] J. W. Moon. An extension of Landau's theorem on tournaments. *Pacific J. Math.*, 13:1343–1345, 1963.
[39] J. W. Moon. *Topics on tournaments.* Holt, Rinehart and Winston, New York-Montreal, Que.-London, 1968.
[40] J. W. Moon and N. J. Pullman. On generalized tournament matrices. *SIAM Rev.*, 12:384–399, 1970.
[41] F. Mosteller. Remarks on the method of paired comparisons: III. a test of significance for paired comparisons when equal standard deviations and equal correlations are assumed. *Psychometrika*, 16(2):207–218, 1951.
[42] A. Müller and L. Rüschendorf. On the optimal stopping values induced by general dependence structures. *J. Appl. Probab.*, 38(3):672–684, 2001.
[43] A. Müller and D. Stoyan. *Comparison methods for stochastic models and risks.* Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd., Chichester, 2002.
[44] G. Peyré and M. Cuturi. Computational optimal transport: With applications to data science. *Foundations and Trends® in Machine Learning*, 11(5-6):355–607, 2019.
[45] P. Pinto and N. Pischke. On Dykstra's algorithm with Bregman projections. 2024. Available at `https://www2.mathematik.tu-⊂darmstadt.de/~pinto/articles/B-⊂Dykstra.pdf`.
[46] R. Plackett. The analysis of permutations. *Appl. Stat.*, pages 193–202, 1975.
[47] R. Rafailov, A. Sharma, E. Mitchell, S. Ermon, C. D. Manning, and C. Finn. Direct preference optimization: Your language model is secretly a reward model. In *Neurips*, volume 36, 2023.
[48] M. Shaked and J. G. Shanthikumar. *Stochastic orders.* Springer Series in Statistics. Springer, New York, 2007.
[49] R. Sinkhorn and P. Knopp. Concerning nonnegative matrices and doubly stochastic matrices. *Pacific J. Math.*, 21:343–348, 1967.
[50] H. Steinhaus and S. Trybuła. On a paradox in applied probabilities. *Bull. Acad. Polon. Sci*, 7(67-69):108, 1959.
[51] M. Stob. A supplement to: "A mathematician's guide to popular sports" [Amer. Math. Monthly **90** (1983), no. 4, 246–266; MR0700265 (85i:05122a)] by T. Jech. *Amer. Math. Monthly*, 91(5):277–282, 1984.
[52] P. K. Stockmeyer. Counting various classes of tournament score sequences. *J. Integer Seq.*, 26(5):Art. 23.5.2, 9, 2023.
[53] V. Strassen. The existence of probability measures with given marginals. *Ann. Math. Statist.*, 36:423–439, 1965.
[54] R. Suck. Independent random utility representations. *Math. Soc. Sci.*, 43(3):371–389, 2002.
[55] L. Takács. A Bernoulli excursion and its various applications. *Adv. in Appl. Probab.*, 23(3):557–585, 1991.
[56] L. Thurstone. A law of comparative judgment. *Psychological review*, 34(4):273, 1927.
[57] L. Thurstone. Rank order as a psycho-physical method. *Journal of Experimental Psychology*, 14(3):187, 1931.
[58] S. Trybuła. On the paradox of three random variables. *Zastos. Mat.*, 5:321–332, 1960/61.
[59] P. Vuksanovic and A. J. Hildebrand. On cyclic and nontransitive probabilities. *Involve*, 14(2):327–348, 2021.
[60] E. Zermelo. Die Berechnung der Turnier-Ergebnisse als ein Maximumproblem der Wahrscheinlichkeitsrechnung. *Math. Z.*, 29(1):436–460, 1929.

Université Paris-Saclay CentraleSupélec, Laboratoire MICS and CNRS FR-3487.
Email address: gaoyue.guo@centralesupelec.fr

(1) Université de Haute Alsace, IRIMAS (Institut de Recherche en Informatique, Mathématiques, Automatique et Signal) UR 7499, F-68 100 Mulhouse, France (2) Université de Strasbourg.
Email address: nicolas.juillet@uha.fr

Department of Industrial Engineering and Operations Research, Columbia University, New York, USA.
Email address: wt2319@columbia.edu