A CENTRAL LIMIT THEOREM FOR INTRANSITIVE DICE

LUIS G. COELHO, TERTULIANO FRANCO, LAEL V. LIMA, JOÃO P. C. DE PAULA, JOÃO V. A. PIMENTA, GUILHERME L. F. SILVA, AND DANIEL UNGARETTI

Abstract. Intransive dice $D^{(1)},\dots,D^{(\ell)}$ are dice such that $D^{(1)}$ has advantage with respect to $D^{(2)}$, dice $D^{(2)}$ has advantage with respect to $D^{(3)}$ and so on, up to $D^{(\ell)}$, which has advantage over $D^{(1)}$. In this twofold work, we present: first, (deterministic) results on existence of general intransitive dice. Second and mainly, a central limit theorem for the vector of normalized victories of a die against the next one in the list when the faces of a die are i.i.d. random variables and all dice are independent, but different dice may have distinct distributions associated to, as well as they may have distinct number of faces. From this central limit theorem we derive a criteria to assure that the asymptotic probability of observing intransitive dice is null, which applies for many cases, including all continuous distributions and many discrete ones.

Contents

1. Introduction

Intransitivity is an inherent facet of nature, it is part of the equilibrium in evolutionary dynamics, where different relations between predators and prey create the balance for common existence. This phenomenon is noted for instance in the eighteenth century Condorcet's paradox, in which three candidates are intransitive in the sense that the candidate A wins when running against B, the candidate B wins when running against candidate C and the candidate C wins when running against candidate A. It is worthy mentioning that the Condorcet's paradox is intrinsically related to the classical Arrow's Theorem. Intransitivity also manifests itself in sports leagues, network relations, interactions between different medications, and an ever-expanding array of scenarios.

While being a fundamental mathematical concept, intransitivity can lead to intriguing outcomes even in simple models. Consider a basic dice game as an example: there are two players, each toss a (possibly different) die and the one with the highest outcome wins. It is possible to construct three dice, A, B and C, for which A is better than B (in the sense that the player with die A has a higher chance of winning against the player with die B), B is better than C and C is better than A? What about constructing an intransitive chain of more than three dice? And dice with a very large number of faces?

The answer about existence of such dice is positive, and to the best of our knowledge, the intransitivity of dice was first addressed in Martin Gadner's column [5]. However, it is worthy mentioning that the particular intransitive dice therein cited was previously mentioned in the sixties by Bradley Effron, which are

$$\begin{aligned} A &= (0,0,4,4,4,4), \quad & B &= (3,3,3,3,3,3) \\ C &= (2,2,2,2,6,6), \quad & D &= (1,1,1,5,5,5). \end{aligned}$$

For those dice it holds that the probability of A beats B, B beats C, C beats D and D beats A are all equal to $2/3$. Intransitive dice are also natural examples of the *Steinhaus and Trybuła's paradox* (see [11]), consisting on the existence of independent random variables X, Y, and Z such that $\mathbb{P}(X > Y) > 1/2$, $\mathbb{P}(Y > Z) > 1/2$, and $\mathbb{P}(Z > X) > 1/2$ (related to Steinhaus and Trybuła's paradox, see [7]). The property of intransitivity can be found in various domains, such as Statistics (see [1]) and voting systems (see [6]).

From a probabilistic point of view, there has been a recent upraise of interest in intransitive dice phenomena. In part, such recent trend started with a discussion by Conrey, Gabbard, Grant, Liu and Morrisson in [2], where the authors considered a model of random dice where the n faces of a given random die are given by uniformly choosing n entries among positive integers conditioned to sum to $n(n+1)/2$, which they call *a balanced model*. For instance, for $n = 4$, the faces of a die are chosen by picking uniformly one of the multisets below:

$$(1,1,4,4), \quad (1,2,3,4), \quad (1,3,3,3), \quad (2,2,2,4), \quad (2,2,3,3)\,.$$

In [2], supported by computational evidence, it has been posed two conjectures for the set of three dice: first, that the asymptotic probability of ties is zero, and second, that the asymptotic probability of picking a set intransitive dice is $1/4$ (see also [9] which evaluates some exact probabilities for three and four dice). These two conjectures were proved later by the Polymath group (see [10]).

In our present work, we consider a general scenario where we do not impose any constraints regarding the sum of their entries, but consider intransitivity phenomenon both from deterministic and random perspectives, in particular allowing for arbitrary number of dice, and also in an asymptotic regime where the number of faces of each die grows large in a proportional (but not necessarily equal) way.

The first part of this paper is devoted to study existence of intransitive dice: in a deterministic setup that does not allow for ties among different dice faces, we are able to characterize completely when an ordered collection of intransitive dice exist, in terms of the size of the dice and the number of different entries used for the faces. In short terms, we essentially prove that there are always intransitive collections with arbitrarily large number of faces, provided each die has at least 3 faces. Naturally, one is faced with the question of how many of these collections there are. We are able to show that the proportion of ordered collections of intransitive dice among all possible collections (not necessarily intransitive) decays with the number of faces of the dice, and we perform numerical experiments on the decay rate.

For both the previously mentioned existence results and decay of the proportion of intransitive dice, the key observation is a bijection between collections of dice, not necessarily intransitive, and words with appropriate number of letters. We explore this connection to construct, from a given collection of intransitive dice, a new collection with a larger number of faces of each die or with a larger number of dice, while preserving the intransitivity. This construction is algorithmic, and as we mentioned it is based on the connection between intransitive dice and words with a particular combinatorial property which may be of independent interest.

The second and main part of this paper deals with models of random dice, where the numbers on the faces of a given dice are independent random variables, but the distributions of different dice may vary. Our main interest lies in determining the chance that a finite collection of random dice is intransitive, when the number of faces of each die grows large. To do so, we prove a central limit theorem for the vector of number of victories of the faces of die against the faces of the next die in the list (whose entries are strongly correlated). The proof of this CLT is based on the moment method, where the crucial step consists of a careful estimate of the moments via an identification with a combinatorial problem in graph theory. This is much inspired by the now classical moment method used in the proof of Wigner's semicircle law in random matrix theory.

The vector of victories is actually connected to intransitivity, which is simple to illustrate when there are no ties and that each die has n faces: in this particular case, intransitity of the list of dice is equivalent to the fact that each entry of the vector of victories is larger than $n^2/2$. This can be properly generalized to account for possible ties among entries, and when combined with the CLT obtained, we are able to deduce a criteria to ensure that the probability of finding intransitive collections of dice decays to zero when the number of faces grows. Such result is obtained under rather mild conditions on the distribution of the random variables determining the faces. *Grosso modo*, the two conditions are a bound from above on the number of ties and a bound from below on the variance of victories of a die against another one in the list, avoiding degeneracy in the central limit setting. These mild conditions cover many situations. For instance, it includes the scenarios where all dice have same distribution (including all continuous distributions and many discrete ones), and also situations where the underlying distributions of faces depend on scaling parameters. And also the case where dice have different distributions in some cases. We also provide a way of constructing asymptotic intransitive dice (not satisfying the previous conditions, of course), which is argued via a concentration inequality.

We now move forward to the discussion of our main findings.

2. Statement of results

We split the discussion of these major results into two subsections, first for deterministic dice and then for random dice. As we hope to convey with this text, simple models of dice display rather interesting and rich aspects worth investigating more deeply. However, many interesting phenomena may depend on somewhat subtle specific features of the model considered. Nevertheless, questions surrounding intransitive dice phenomena are rather simple to state. For the latter reason, we mostly introduce new terminology and notation along the text, reserving formal definitions solely for more technical assumptions needed. For convenience, such definitions along the text are highlighted in bold.

2.1. Main results for deterministic models of dice.

An **n-sided die** is a pair (D, X), where $D = (D_1, \dots, D_n)$ is a real-valued vector where each D_k represents the number on the k-th face, and X is a random variable taking values on $[n] := \{1, 2, \dots, n\}$ that represents the label of the face in the outcome of a toss. The number n is the number of faces, or simply size, of the die D. The die is said to roll the face k with probability $\mathbb{P}(X = k)$ and results D_k. If this probability equals $1/n$ for every k, the die is **honest** or **fair**. Otherwise the die is **unfair** or **biased**. If there is no ambiguity, the die will be denoted as D, and in that case, it is useful to denote the random result of D in a roll by D_X. Thus, in general, the entries of D need not be integer-valued, nor even positive. We reserve capital letters A, B, C etc. to represent dice, and lower indices A_i, B_i etc. to represent a entries of the dice A, B. It is also useful to distinguish different dice with an upper index, writing for instance $D^{(1)}$, $D^{(2)}$ etc, and the corresponding entries by $D_i^{(1)}$, $D_i^{(2)}$ etc.

A die A is said to be **better than** a die B, and it is denoted by $A \triangleright B$ if the probability of A rolling a higher value than B is greater than the probability of B rolling a higher value than A. To the same extent, the die B is said to be **worse than** A, and it is denoted by $B \triangleleft A$.

In mathematical terms, one way to verify whether a fair die A is better than a fair die B is by counting against how many faces of B a given face of A wins, summing the result over all possible faces of A, and comparing with the count we obtain when we do the same interchanging the roles of A and B. In other words, $A \triangleright B$ if, and only if, the inequality

$$\sum_{A_i > B_j} 1 \; > \; \sum_{B_j > A_i} 1$$

is satisfied. With n_A and n_B being the number of faces of A and B, respectively, there are in total $n_A n_B$ pairs of faces from A and B to compare, and $A \triangleright B$ if, and only if,

$$\sum_{A_i > B_j} 1 \; > \; \frac{1}{2} n_A n_B - \frac{1}{2} \sum_{A_i = B_j} 1 \,. \tag{2.1}$$

An ordered collection of dice $\mathbf{D} = (D^{(1)}, \dots, D^{(\ell)})$ is said to be **intransitive** if $D^{(1)} \triangleright \cdots \triangleright D^{(\ell)} \triangleright D^{(1)}$. Note that while $\triangleright$ is an asymmetric relation, it is not necessarily transitive, so it does not define an order relation. When computing whether a given collection $\mathbf{D} = (D^{(1)}, \dots, D^{(\ell)})$ of dice is intransitive, the ordering of the entries does matter, and it is possible that $\mathbf{D}$ is not intransitive, but for some permutation σ of length ℓ a reordering $(D^{(\sigma(1))}, \dots, D^{(\sigma(\ell))})$ is intransitive. We will either be interested in existence results for deterministic collections $\mathbf{D}$, or in asymptotic results when the distributions of the entries of each die are rather arbitrary, so this ordering will not be relevant in any essential way.

By a **no-tie collection** of dice we mean that no pair of faces, either from the same die or from different dice, share the same number.

Our first two results concern intransitive families of deterministic dice. The first result deals with the existence of intransitive, fair dice.

Theorem 1. *Consider dice whose face entries are positive integers. For every $\ell \geq 3$ and $n \geq 3$ there exists a no-tie collection of ℓ honest n-sided dice which is intransitive. Furthermore, for any $\ell \geq 3$ there does not exist a no-tie family of ℓ honest 2-sided dice which is intransitive.*

The notion of a die A being better than B is not a relation on the specific numbers on their faces, but rather among relative ordering of these numbers. For instance, the die $A = (2, 4, 9, 10, 11)$ is better than the die $B = (1, 5, 7, 9, 10)$. Now, increase, say, the first entry of A to a new die $\widetilde{A} = (x, 4, 9, 10, 11)$ with any choice $x = 3, 4$. Then when we choose either one of dice A or $\widetilde{A}$, the chance of winning against a roll of die B is the same. So, in terms of *chance of winning* against B, the dice A and $\widetilde{A}$ are indistinguishable.

In that sense, when we talk about comparison of ℓ non-tie dice with n faces each, it suffices to distribute the numbers in the set $[\ell n]$ among the faces of the dice, without repetition. In fact, the proof of the existence claim in Theorem 1 is inductive/constructive, and shows that such ℓ honest n-sided dice can always be chosen with distinct entries in $[\ell n]$.

For a given choice of positive integers $n_1, \dots, n_\ell$, let $\mathcal{D}(n_1, \dots, n_\ell)$ be the set of collections of dice $\mathbf{D} = (D^{(1)}, \dots, D^{(\ell)})$ for which $D^{(j)}$ has exactly n_j faces, and where each number in $[n_1 + \cdots + n_\ell]$ appears exactly once in the faces in $\mathbf{D}$. In other words, the dice are filled with numbers in $[n_1 + \cdots + n_\ell]$, without repetition. Observe that with this definition, in $\mathcal{D}(n_1, \dots, n_\ell)$ we do not distinguish between the ordering of faces in each die. Or, alternatively, dice in $\mathcal{D}(n_1, \dots, n_\ell)$ are always viewed in increasing order, so that for instance the dice $(1, 2, 3)$ and $(2, 1, 3)$ are the same and are always represented by $(1, 2, 3)$. But we do distinguish between orderings within a collection, so that the collections $\mathbf{D} = ((1, 2, 4), (3, 5, 6)), \widehat{\mathbf{D}} = ((3, 5, 6), (1, 2, 4))$ are distinct elements of $\mathcal{D}(3, 3)$. In other words,

$$
\begin{aligned}
\mathcal{D}(n_1, \dots, n_\ell) \;&:=\; \Big\{ \mathbf{D} = (D^{(1)}, \dots, D^{(\ell)}) : D^{(j)} = (D^{(j)}_1, \dots D^{(j)}_{n_j}) \in \mathbb{Z}^{n_j}, \\
0 < D^{(j)}_1 < \cdots < D^{(j)}_{n_j} &\text{ for } j = 1, \dots, \ell, D^{(j_1)}_{i_1} \neq D^{(j_2)}_{i_2} \text{ for } j_1 \neq j_2, \{D^{(j)}_i\}_{i,j} = [n_1 + \cdots + n_\ell] \Big\}.
\end{aligned}
$$

We denote by $\mathcal{D}_{\triangleright}(n_1, \dots, n_\ell)$ the subset of $\mathcal{D}(n_1, \dots, n_\ell)$ that consists of intransitive dice, that is,

$$
\mathcal{D}_{\triangleright}(n_1, \dots, n_\ell) \;:=\; \big\{ \mathbf{D} \in \mathcal{D}(n_1, \dots, n_\ell) : D^{(1)} \triangleright \cdots \triangleright D^{(\ell)} \triangleright D^{(1)} \big\},
$$

and additionally also set

$$
\mathcal{D}_\ell(n) \;:=\; \mathcal{D}(\underbrace{n, \dots, n}_{\ell \text{ times}}), \quad \mathcal{D}_{\triangleright, \ell}(n) \;:=\; \mathcal{D}_{\triangleright}(\underbrace{n, \dots, n}_{\ell \text{ times}}) \,. \tag{2.2}
$$

Exploring a connection between non-tie dice with integer entries and the set of words in a given alphabet which is explained in Section 4.1, we will be able to estimate the size of $\mathcal{D}_{\triangleright}(n, n, n)$.

Theorem 2. *For each $\ell \geq 3$, there exists a constant $L(\ell) \geq 0$ for which*

$$
|\mathcal{D}_{\triangleright, \ell}(n)| \;=\; \mathrm{e}^{nL(\ell) + o(n)} \quad \textit{as } n \to \infty \,.
$$

For any $n \geq 1$, a simple combinatorial argument shows that $|\mathcal{D}_\ell(n)| = (\ell n)!/(n!)^\ell$, and by combining Theorem 2 with Stirling's approximation we see that

$$\frac{|\mathcal{D}_{\triangleright,\ell}(n)|}{|\mathcal{D}_\ell(n)|} \;=\; \frac{\mathrm{e}^{-n(\ell\log\ell - L(\ell))+o(n)}}{(2\pi n\ell)^{(\ell-1)/2}}\,. \tag{2.3}$$

Equipping $\mathcal{D}(\ell)$ with the uniform distribution, one may view the quantity $\frac{|\mathcal{D}_{\triangleright,\ell}(n)|}{|\mathcal{D}_\ell(n)|}$ as the probability of selecting an ℓ-uple of intransitive dice from this distribution. As a consequence of Theorem 6 to be seen in a moment, applied to random dice with uniform law on $[0,1]$ we can infer that

$$\lim_{n\to\infty}\frac{|\mathcal{D}_{\triangleright,\ell}(n)|}{|\mathcal{D}_\ell(n)|} \;=\; 0\,.$$

When $\ell = 3$, we will explain in Subsection 4.5 how to obtain the estimate $2.445 < L(3) \leq 3\log 3$, and also display some numerical experiments that indicate that $L(3) = 3\log 3$. It may be natural to expect that $L(\ell) = \ell\log\ell$ for any $\ell \geq 3$, but besides the case $\ell = 3$ we do not have numerical evidence to support this conjecture.

We remark that the definition of the sets $\mathcal{D}_\ell(n)$ and $\mathcal{D}_{\triangleright,\ell}(n)$ as above do not account for possible permutations of dice when checking intransitivity, that is, the list of dice has a fixed order. We stick to this convention as well as in the central limit theorem to be stated in the sequel.

2.2. **Main results for random models of dice.**

When each D_i is a random variable, we say that the corresponding die $D = (D_1, \ldots, D_n)$ is a **random die**. Whenever we say that the **law** of a die D is $\mathcal{L}^D$, we mean that the entries D_i are all i.i.d. random variables with law $\mathcal{L}^D$. We say that the dice in a collection $\mathbf{D} = (D^{(1)}, \ldots, D^{(\ell)})$ are **independent** if the family of random variables $\{D_i^{(k)}\}$ are all mutually independent. We stress that for independent dice the laws $\mathcal{L}^{(1)} := \mathcal{L}^{D^{(1)}}, \ldots, \mathcal{L}^{(\ell)} := \mathcal{L}^{D^{(\ell)}}$ need not coincide, but entries within the same die are i.i.d. random variables.

Our main goal is to determine whether a sequence of random dice may be intransitive when they grow in size. Fix an integer $\ell \geq 3$ and consider a sequence $\{\mathbf{D}_m\}_m$ of collections $\mathbf{D}_m = (D^{(1)}, \ldots, D^{(\ell)})$ of random independent dice. Each die $D^{(k)} = D^{(k)}(m)$ depends on the index m of the sequence $\{\mathbf{D}_m\}_m$, but to lighten notation we mostly omit this dependence. We assume each die $D^{(k)} = D^{(k)}(m)$ has $n_k = n_k(m) \leq m$ faces, which may vary with m, and we set

$$f_k = f_k(m) \;:=\; \frac{n_k}{m} \quad \text{so that } D^{(k)} \text{ has size } n_k \;=\; f_k m, \quad k = 1, \ldots, \ell. \tag{2.4}$$

The assumption $n_k = n_k(m) \leq m$ is made solely for convenience as in this case $f_k \leq 1$. We can always re-index the sequence to accomplish this restriction. Although there are no further relations imposed between the sizes n_k and m, it is instructive to think about m as essentially given the size of the die with largest number of faces. As we already mentioned before, we always assume that different entries of the same die $D^{(k)}$ are independent random variables with the same law $\mathcal{L}^{(k)} = \mathcal{L}^{D^{(k)}}$ which may now vary with m, and we write $\mathcal{L}_m^{(k)}$ when we want to emphasize this dependence.

The main question we investigate is the probability of intransitivity, namely

$$\mathbb{P}\big(D^{(1)} \triangleright D^{(2)} \triangleright \cdots \triangleright D^{(\ell)} \triangleright D^{(1)}\big) \tag{2.5}$$

as the numbers of faces of our dice go to infinity, which we measure by sending $m \to \infty$. The intransitivity event in (2.5) is the intersection of $D^{(k)} \triangleright D^{(k+1)}$ for $1 \leq k \leq \ell$, where we convention for the rest of the paper that $D^{(0)} = D^{(\ell)}$ and $D^{(\ell+1)} = D^{(1)}$. Such events are intimately connected to the values of the random variables

$$N_k \;\coloneqq\; \sum_{i=1}^{n_k}\sum_{j=1}^{n_{k+1}} \mathbb{1}_{D_i^{(k)}>D_j^{(k+1)}}, \quad k=1,\dots,\ell \tag{2.6}$$

and

$$E_k \;\coloneqq\; \sum_{i=1}^{n_k}\sum_{j=1}^{n_{k+1}} \mathbb{1}_{D_i^{(k)}=D_j^{(k+1)}}, \quad k=1,\dots,\ell. \tag{2.7}$$

From the inequality (2.1) we learn that $D^{(k)} \triangleright D^{(k+1)}$ if, and only if, the inequality

$$N_k \;>\; \frac{1}{2}n_k n_{k+1} - \frac{1}{2}E_k$$

is satisfied, and therefore

$$\mathbb{P}\left(D^{(1)} \triangleright \dots \triangleright D^{(\ell)} \triangleright D^{(1)}\right) \;=\; \mathbb{P}\Big(N_k > \frac{1}{2}n_k n_{k+1} - \frac{1}{2}E_k\,,\; k=1,\dots,\ell\Big), \tag{2.8}$$

which will be at the core of our method to analyze (2.5), and shows the relevance of N_k and E_k. One should view the N_k as the *relative strength* of the die $D^{(k)}$ against $D^{(k+1)}$. Observe that for dice coming from a sequence $\{\mathbf{D}_m\}_m$, the random variables $N_k = N_k(m)$ and $E_k = E_k(m)$ also depend on m, and N_k, N_{k+1}, E_k and E_{k+1} are all pairwise strongly correlated.

We will analyze (2.8) in the limit $m\to\infty$ via a Central Limit Theorem (CLT) for the vector $(N_1,\dots,N_\ell)$. For this CLT some probabilities associated to the underlying laws of the dice are of utmost importance. By

$$\mathbf{p}_k \;=\; \mathbf{p}(\mathcal{L}^{(k)},\mathcal{L}^{(k+1)}) \;\coloneqq\; \mathbb{P}\left(D_1^{(k)} > D_1^{(k+1)}\right) \;=\; \mathbb{E}\left(\mathbb{1}_{D_1^{(k)}>D_1^{(k+1)}}\right) \tag{2.9}$$

we denote the probability that a given face of the k-th die beats a given face of the $(k+1)$-th die. By

$$\mathbf{q}_k \;=\; \mathbf{q}(\mathcal{L}^{(k)},\mathcal{L}^{(k+1)}) \;\coloneqq\; \mathbb{P}\left(D_1^{(k)} > D_1^{(k+1)}, D_2^{(k)} > D_1^{(k+1)}\right) \tag{2.10}$$

we denote the probability that two given faces of the k-th die beat a given face of the $(k+1)$-th die. By

$$\mathbf{r}_k \;=\; \mathbf{r}(\mathcal{L}^{(k)},\mathcal{L}^{(k+1)}) \;\coloneqq\; \mathbb{P}\left(D_1^{(k)} > D_1^{(k+1)}, D_1^{(k)} > D_2^{(k+1)}\right) \tag{2.11}$$

we denote the probability that a given face of the k-th die beats two given faces of the $(k+1)$-th die. Finally, also set

$$\mathbf{s}_k \;=\; \mathbf{s}(\mathcal{L}^{(k-1)},\mathcal{L}^{(k)},\mathcal{L}^{(k+1)}) \;\coloneqq\; \mathbb{P}\left(D_1^{(k-1)} > D_1^{(k)} > D_1^{(k+1)}\right), \tag{2.12}$$

which is the probability that a given face of $D^{(k-1)}$ beats a given face of $D^{(k)}$ at the same time that the latter beats a given face of $D^{(k+1)}$. As we will see in a moment, these quantities will play a role in understanding the covariance between different dice. We use cyclic notation for these quantities, so that $\mathbf{p}_{\ell+1} \coloneqq \mathbf{p}_1$, $\mathbf{q}_{\ell+1} \coloneqq \mathbf{q}_1$ and so forth.

Our main tool to analyze the probability (2.5) is a CLT for the correlated random variables $N_1,\dots,N_\ell$, so it is natural to introduce their normalized version

$$\widetilde{N}_k \;\coloneqq\; \frac{N_k - \mathbb{E}(N_k)}{\sqrt{\operatorname{Var}(N_k)}}\,. \tag{2.13}$$

Let

$$\sigma_k \;=\; \sigma_k(\mathbf{p}_k, \mathbf{q}_k, \mathbf{r}_k, \mathbf{s}_k) \;\coloneqq\; \Big[f_k f_{k+1}\left(f_k(\mathbf{q}_k - \mathbf{p}_k^2) + f_{k+1}(\mathbf{r}_k - \mathbf{p}_k^2)\right)\Big]^{1/2} \tag{2.14}$$

and

$$\gamma_k \;\coloneqq\; \frac{1}{\sigma_{k-1}\sigma_k} f_{k-1} f_k f_{k+1}(\mathbf{s}_k - \mathbf{p}_{k-1}\mathbf{p}_k)\,. \tag{2.15}$$

A straightforward calculation (see Lemma 14 below) shows that, as $m \to \infty$,

$$\begin{aligned} \mathbb{E}(N_k) &= f_k f_{k+1} m^2 \mathbf{p}_k\,, \\ \mathrm{Var}\,(N_k) &= \sigma_k^2 m^3 + o(m^3)\,, \quad \text{and} \\ \mathrm{Corr}\,(N_{k-1}, N_k) &= \gamma_k + o(1)\,. \end{aligned} \tag{2.16}$$

We stress that the values $\sigma_k = \sigma_k(m)$ and $\gamma_k = \gamma_k(m)$ depend explicitly on probabilities associated to the laws $\mathcal{L}_m^{(k-1)}, \mathcal{L}_m^{(k)}$ and $\mathcal{L}_m^{(k+1)}$. Moreover, they are $O(1)$ as $m \to \infty$, and they do not depend on regularity features of these laws, such as existence of moments or tail behavior.

Since we are considering a sequence $\{\mathbf{D}_m\}_m$ of collections of independent dice, all the quantities we just introduced depend on m, and when needed to stress such dependence we write $\mathbf{p}_k = \mathbf{p}_k(m), \sigma_k = \sigma_k(m), \gamma_k = \gamma_k(m)$ etc. Our main working assumptions are the following.

Assumption 3. *Fix $\ell \geq 3$. We assume that the sequence $\{\mathbf{D}_m\}_m$ is a collection of ℓ independent random dice, each with number of faces $n_k = f_k m$ as in (2.4), and satisfying the following conditions:*

(i) For $k = 1, \dots, \ell$, the relative sizes $f_k = f_k(m)$ satisfy

$$f_k(m) \;\to\; f_k(\infty) \in (0, 1]\,, \quad \textit{as } m \to \infty.$$

(ii) For $k = 1, \dots, \ell$, the rate of growth of the mean and variance of N_k, and covariance between N_{k-1} and N_k satisfy

$$\begin{aligned} \mathbf{p}_k(m) &\to \mathbf{p}_k(\infty) \in (0, 1], \\ \sigma_k(m) &\to \sigma_k(\infty) \in (0, \infty)\,, \\ \gamma_k(m) &\to \gamma_k(\infty) \in [-1, 1]\,, \end{aligned}$$

as $m \to \infty$.

We insist that the values $\mathbf{p}_k(m)$ and $\sigma_k = \sigma_k(m)$ depend only on probabilities associated to the underlying laws rather than on qualitative features of them. In particular, Assumption 3–(i) is solely a non-degeneracy condition, which ensures that the number of faces of the dice are all growing, with the same speed m but possibly different rates. With (2.9) in mind, condition (ii) on $\mathbf{p}_k$ is essentially saying that the limiting laws do not generate to a deterministic situation when intransitivity does not occur by degeneration. Also, as we said earlier, under Assumption 3–(i) the values $\sigma_k = \sigma_k(m)$ are bounded functions of m. Thus, with (2.16) in mind, the second condition in (ii) says that the variance in the relative strength of consecutive dice is growing at true speed m^3 and not slower. The quantities $\gamma_k(m)$ are correlation coefficients, so they are always bounded, and the third convergence condition in (ii) can always be achieved with a replacement of the original sequence of dice $\{\mathbf{D}_m\}_m$ by a subsequence of it.

Theorem 4. *Fix $\ell \geq 3$ and for each m let $\mathbf{D}_m = (D^{(1)}(m), \dots, D^{(\ell)}(m))$ be a collection of random independent dice, for which $\{\mathbf{D}_m\}_m$ satisfies Assumption 3, and let $(\widetilde{N}_1(m), \dots, \widetilde{N}_\ell(m))$ be the corresponding variables from (2.13).*

Then, as $m \to \infty$, the random vector $(\widetilde{N}_1(m), \cdots, \widetilde{N}_\ell(m))$ converges in distribution to a centered Gaussian vector $(X_1, \dots, X_\ell)$ whose covariance matrix is given by

$$\Sigma \;=\; \begin{pmatrix} 1 & \gamma_2(\infty) & 0 & \cdots & 0 & \gamma_1(\infty) \\ \gamma_2(\infty) & 1 & \gamma_3(\infty) & \cdots & 0 & 0 \\ 0 & \gamma_3(\infty) & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & \gamma_\ell(\infty) \\ \gamma_1(\infty) & 0 & 0 & \cdots & \gamma_\ell(\infty) & 1 \end{pmatrix}, \tag{2.17}$$

where the coefficients $\gamma_k(\infty)$ are the ones in Assumption 3–(ii).

As said, Theorem 4 will be central in our understanding of intransitivity as $m \to \infty$. In general, the very definition of $\mathbf{p}$ in (2.9) would say that

$$1 \;=\; \mathbb{P}\big(D_1^{(k)} > D_1^{(k+1)}\big) + \mathbb{P}\big(D_1^{(k)} < D_1^{(k+1)}\big) + \mathbb{P}\big(D_1^{(k)} = D_1^{(k+1)}\big) \tag{2.18}$$

$$= \; \mathbf{p}(\mathcal{L}^{(k)}, \mathcal{L}^{(k+1)}) + \mathbf{p}(\mathcal{L}^{(k+1)}, \mathcal{L}^{(k)}) + \mathbb{P}\big(D_1^{(k)} = D_1^{(k+1)}\big)\,. \tag{2.19}$$

In order for the die $D^{(k)}$ to be sufficiently stronger than the die $D^{(k+1)}$, we would expect that $\mathbf{p}(\mathcal{L}^{(k)}, \mathcal{L}^{(k+1)}) > \mathbf{p}(\mathcal{L}^{(k+1)}, \mathcal{L}^{(k)})$, and in such a case we would expect $D^{(k)} \triangleright D^{(k+1)}$ with high probability. Likewise, if $\mathbf{p}(\mathcal{L}^{(k+1)}, \mathcal{L}^{(k)}) < \mathbf{p}(\mathcal{L}^{(k)}, \mathcal{L}^{(k+1)})$ then we would instead expect $D^{(k+1)} \triangleright D^{(k)}$ with high probability. Hence, intransitivity becomes a nontrivial question precisely when $\mathbf{p}(\mathcal{L}^{(k)}, \mathcal{L}^{(k+1)}) \approx \mathbf{p}(\mathcal{L}^{(k+1)}, \mathcal{L}^{(k)})$ asymptotically as $m \to \infty$, in which case the equality above becomes

$$\mathbf{p}(\mathcal{L}^{(k+1)}, \mathcal{L}^{(k)}) + \frac{1}{2}\mathbb{P}\big(D_1^{(k)} = D_1^{(k+1)}\big) = \mathbf{p}_k + \frac{1}{2}\mathbb{P}\big(D_1^{(k)} = D_1^{(k+1)}\big) \;\approx\; \frac{1}{2}\,,$$

and our next result gives a rate of decay of such approximation under which we can use the CLT to estimate the probability of intransitivity in the large-dice limit $m \to \infty$.

Theorem 5. *Fix $\ell \geq 3$ and for each m let $\mathbf{D}_m = (D^{(1)}(m), \dots, D^{(\ell)}(m))$ be a collection of random independent dice, for which $\{\mathbf{D}_m\}_m$ satisfies Assumption 3, and let $(X_1, \dots, X_\ell)$ be a Gaussian vector with covariance matrix (2.17). Suppose that there exists $\delta > 0$ and a function $r(m)$ with $\lim_{m\to\infty} r(m) = +\infty$, for which*

$$\frac{1}{2} - \mathbf{p}_k - \frac{1}{2}\mathbb{P}(D_1^{(k)} = D_1^{(k+1)}) \;\geq\; -\frac{\delta}{m^{1/2} r(m)}\,, \quad k = 1, \dots, \ell, \tag{2.20}$$

for every m sufficiently large, and in addition

$$\lim_{m\to\infty} \mathbb{P}\left(D_1^{(k)}(m) = D_1^{(k+1)}(m)\right) \;=\; 0 \quad \text{for } k = 1, \dots, \ell. \tag{2.21}$$

Then

$$\limsup_{m\to\infty} \mathbb{P}\left(D^{(1)} \triangleright \cdots \triangleright D^{(\ell)} \triangleright D^{(1)}\right) \;\leq\; \mathbb{P}\left(X_j \geq 0, j = 1, \dots, \ell\right). \tag{2.22}$$

In the particular case when all the laws are the same $\mathcal{L}^{(1)} = \cdots = \mathcal{L}^{(\ell)}$, then (2.18) yields that

$$\frac{1}{2} - \mathbf{p}_k - \frac{1}{2}\mathbb{P}\left(D_1^{(k)} = D_1^{(k+1)}\right) = 0, \quad k = 1, \dots, \ell,$$

and (2.20) always holds true.

Note that (2.21) says that there are no ties between different dice in the asymptotic limit. Similarly as for N_k, the mean and variance of the E_k's are given in terms of probabilities associated to the underlying laws. For arbitrary underlying laws of the entries of the dice, they satisfy the rough bound

$$\mathbb{E}(E_k) = O(m^2) \quad \text{and} \quad \operatorname{Var}(E_k) = O(m^3) \quad \text{as } m \to \infty, \tag{2.23}$$

see Lemma 15 below. These quantities have the same order as the corresponding quantities for N_k (compare (2.16) with (2.23)). Lemma 16 below shows that (2.21) implies

$$\mathbb{E}(E_k) = o(m^2), \quad \operatorname{Var}(E_k) = o(m^3) \quad \text{as } m \to \infty, \quad \text{for } k = 1, \dots, \ell. \tag{2.24}$$

Thus, condition (2.21) in Theorem 5 may also be interpreted as saying that whenever the E_k's grow slightly slower than N_k's, either in their mean or in their variance, then the intransitive dice problem can be bounded from above by the Gaussian probability (2.22). Exploring this limit, we conclude our next result.

Theorem 6. *Let $\{\mathbf{D}_m\}_m$ be a sequence of random independent dice satisfying the conditions of Theorem 5. Then*

$$\lim_{m\to\infty} \mathbb{P}\left(D^{(1)} \rhd \cdots \rhd D^{(\ell)} \rhd D^{(1)}\right) = 0 .$$

In words, under the assumptions of Theorem 6 above, the proportion of intransitive random independent dice becomes negligible as the number of faces grows.

The proof of Theorem 4 is the most involved proof in this paper, and is based on the moment method. We look at the moment generating function of $\widetilde{N}_k$'s, which always exist because these are rescaled Bernoulli random variables.

The proof of Theorem 5 is based on Theorem 4. We look at the probability in (2.22), and with the help of Chebyshev's inequality we condition on the event of no ties, reducing the right-hand side of (2.8) to a probability that involves only the N_k's plus an additional term which is small in virtue of the variance control (2.24). The right-hand side then naturally arises when taking the large n limit. In virtue of a particular structure of the coefficients $\gamma_k(\infty)$ in the covariance matrix (2.17), we are able to show that the probability on the right-hand side of (2.22) vanishes, and Theorem 6 follows.

2.3. **Organization of the remainder of the paper.**

The remainder of the paper is structured as follows. In Section 3 we discuss examples of random dice, in particular when our core Assumption 3 and the variance control (2.24) are satisfied, allowing to apply our main result. We also provide a sequence of random independent dice that do not satisfy these conditions and for which intransitivity survives in the limit. In Section 4 we discuss intransitivity in deterministic contexts, and in particular we explore a connection between intransitive dice and combinatorics of words in order to construct intransitive dice. We then turn to the context of random dice. In Section 5 we briefly discuss the counting functions N_k and E_k from (2.6)–(2.7), which correspond to victories and ties, respectively, and which play a central role in the connection between our CLT and intransitivity. In our CLT, Gaussian vectors with a covariance matrix of a particular structure appear (see (2.17)), and in Section 5 we also collect several properties of them in a form suitable for our needs. In Section 6, we assume Theorem 4, which is our central limit theorem, and we use it to prove Theorems 5 and 6, which are tests of asymptotic intransitivity. Finally, at Section 7, we prove the Theorem 4.

Acknowledgments. T.F. acknowledges support by the National Council for Scientific and Technological Development (CNPq) via a Universal Grant (number 406001/2021-9) and a Bolsa de Produtividade (number 311894/2021-6).

G.S. acknowledges support by São Paulo Research Foundation (FAPESP) under Grants # 2019/16062-1 and # 2020/02506-2, and by Brazilian National Council for Scientific and Technological Development (CNPq) under Grant # 315256/2020-6.

J.P. acknowledges support by São Paulo Research Foundation (FAPESP) under Grant # 2023/02674-0.

J.Pa. acknowledges support by Brazilian National Council for Scientific and Technological Development (CNPq) under Grant # 118536/2023-0

L.C. acknowledges support by São Paulo Research Foundation (FAPESP) under Grant # 2023/02240-0.

L.L. acknowledges support by São Paulo Research Foundation (FAPESP) under Grant # 2023/02397-7.

Part of this work was carried out during the undergraduate research program "Jornadas de Pesquisa em Matemática do ICMC 2023" held at the Instituto de Ciências Matemáticas e de Computação (ICMC) - Universidade de São Paulo (USP), and which was partially supported by the Centro de Ciências Matemáticas Aplicadas à Indústria (CeMEAI - CEPID) under FAPESP Grant # 2013/07375-0. Research carried out using the computational resources of the Center for Mathematical Sciences Applied to Industry (CeMEAI) funded by FAPESP (grant 2013/07375-0). We thank the hospitality of ICMC-USP during the program.

3. Examples

In this section we describe examples of random dice such that the probability of observing intransitivity is asymptotic null by applying the Theorem 6 and we also illustrate some cases of asymptotically intransitive random dice.

We start by pointing out that condition (2.20) is satisfied, by symmetry, whenever all the dice have same law, which is the case in what follows, except in the last example.

The first example has been already commented below Theorem 4: assuming that each die has the same number of faces, and those faces are i.i.d. random variables with the same continuous (but not necessarily absolutely continuous) law $\mathcal{L}$, the probability that the random dice $(D^{(1)},\dots,D^{(\ell)})$ are intransitive goes to zero as $m\to\infty$. Theorem 6 straightforwardly extends this to a more general situation, as we explain in the next paragraph.

If the law of any die is given by a same continuous law $\mathcal{L}$, there will be no ties, so (2.21) holds trivially. Moreover, $\mathbf{p}_k(m),\sigma_k(m)$ and $\gamma_k(m)$ do not depend on m neither k, hence it is trivial to check Assumption 3-(ii). Assuming that the quantity of faces in the kth die is given by $n_k=f_k m$, where f_k, $k=1,\dots,\ell$, are positive constants, we verify Assumption 3-(i). Putting these conditions together leads to conclude by Theorem 6 that the sequence of dice in this way constructed has probability asymptotically null of being intransitive. That is, under a continuous law, intransitivity is not achievable regardless of the quantities of faces in each die, provided these quantities are proportional to the scaling parameter m.

Let us see now a discrete example. Assume that all ℓ dice have same law $\mathcal{L}_m$, the law of a geometric random variable of parameter p. Since

$$\mathbb{P}\left(D_1^{(k)}(m)=D_1^{(k+1)}(m)\right) \;=\; \sum_{i=1}^{\infty}(1-p)^{2(i-1)}p^2 \;=\; \frac{p}{2-p}\,,$$

in order to assure condition (2.21) on ties, it is necessary to impose that $p = p(m) \to 0$ as $m \to \infty$. A long but elementary calculation yields that

$$\begin{aligned} \mathbf{p}_k(m) &= \frac{1-p}{2-p}, \\ \mathbf{q}_k(m) &= r_k(m) = \frac{(1-p)^2}{3-3p-p^2}, \\ \mathbf{s}_k(m) &= \frac{(1-p)^3}{(2-p)(2-2p-p^2)}. \end{aligned}$$

Recalling the formulas (2.14) and (2.15) for σ_k and γ_k, respectively, gives us that, as $m \to \infty$,

$$\begin{aligned} \mathbf{p}_k(m) &\longrightarrow \frac{1}{2} \in (0,1], \\ \sigma_k(m) &\longrightarrow \sigma(\infty) = \sqrt{\frac{f_k f_{k+1}(f_k + f_{k+1})}{6}} \in (0,\infty), \\ \gamma_k(m) = \frac{f_{k-1} f_k f_{k+1}(\mathbf{s}_k - \mathbf{p}_{k-1}\mathbf{p}_k)}{\sigma_{k-1}(m)\sigma_k(m)} &\longrightarrow 0 \in [-1,1], \end{aligned}$$

so all assumptions of Theorem 6 have been checked, hence the probability of observing a sequence of intransitive dice is asymptotically null. A similar example may be constructed choosing the Poisson distribution with parameter $\lambda = \lambda(m) \to \infty$ as $m \to \infty$.

The previous examples may give us the impression that asymptotic intransitivity with positive probability is never attainable when using dice with i.i.d. faces. This is not true and the idea to construct such sequence of random dice is indeed simple: starting from a list of deterministic dice that is intransitive, we chose the distribution of the faces of each random die according to that list, and concentration inequality will then assure that the sequence of random dice in this way constructed is asymptotically intransitive. This is explained in the next proposition.

Proposition 7. *Let $A^{(k)} = (a_1^{(k)}, \ldots, a_m^{(k)})$ for $k \in [\ell]$ be a set of ℓ deterministic honest dice with m faces that is known to be intransitive: $A^{(k)} \triangleright A^{(k+1)}$ for every k. Consider random dice $(B^{(k)} : k \in [\ell])$, each with n faces, where the faces of die $B^{(k)}$ are independently chosen with law*

$$B_j^{(k)} \sim \mathrm{U}\{a_i^{(k)} : i \in [m]\},$$

that is, uniformly over the faces of die $A^{(k)}$. Then, there is a constant $c > 0$, depending only on the set of dice $A^{(k)}$, such that

$$\mathbb{P}\big(B^{(1)} \triangleright B^{(2)} \triangleright \cdots \triangleright B^{(\ell)} \triangleright B^{(1)}\big) = 1 + o(e^{-cn}), \quad \text{as } n \to \infty. \tag{3.1}$$

Proof. The deterministic dice $A^{(k)}$, $k \in [\ell]$ form an intransitive cycle. This can be translated into the following collection of inequalities:

$$\sum_{\substack{i,j:\\ a_i^{(k)} > a_j^{(k+1)}}} 1 > \sum_{\substack{i,j:\\ a_i^{(k)} < a_j^{(k+1)}}} 1, \quad \text{for every } k \in [\ell]. \tag{3.2}$$

Let $N_{k,i}$ be the random variable that counts the number of appearances of face $a_i^{(k)}$ at dice $B^{(k)}$. As discussed in Section 4, the quantity $N_{k,i}$ represents the weight of face $a_i^{(k)}$ in die $B^{(k)}$. Hence, we can write event $B^{(k)} \rhd B^{(k+1)}$ as a function of $N_{k,i}$ and $N_{k+1,j}$:

$$\sum_{\substack{i,j:\\ a_i^{(k)} > a_j^{(k+1)}}} N_{k,i}N_{k+1,j} \;>\; \sum_{\substack{i,j:\\ a_i^{(k)} < a_j^{(k+1)}}} N_{k,i}N_{k+1,j}\,, \quad \text{for every } k \in [\ell]. \tag{3.3}$$

It is clear that $N_{k,i}$ has a binomial distribution with parameters n and $\frac{1}{m}$, and by Hoeffding's inequality

$$\mathbb{P}\Big(\Big|N_{k,i} - \frac{n}{m}\Big| > \varepsilon n\Big) \;\leq\; 2e^{-2\varepsilon^2 n}\,. \tag{3.4}$$

Define the event $G := \bigcap_{k,i}\big\{\big|N_{k,i} - \frac{n}{m}\big| \leq \varepsilon n\big\}$. By union bound,

$$\mathbb{P}(G^c) \;=\; \mathbb{P}\Big(\bigcup_{k,i}\big\{\big|N_{k,i} - \frac{n}{m}\big| > \varepsilon n\big\}\Big) \;\leq\; 2\ell m\, e^{-\varepsilon^2 n}\,.$$

Notice that on event G we have that

$$n^2\Big(\frac{1}{m} - \varepsilon\Big)^2 \;<\; N_{k,i}N_{k+1,j} \;<\; n^2\Big(\frac{1}{m} + \varepsilon\Big)^2. \tag{3.5}$$

From (3.2) it is clear that

$$\sum_{\substack{i,j:\\ a_i^{(k)} > a_j^{(k+1)}}} 1 \;-\; \sum_{\substack{i,j:\\ a_i^{(k)} < a_j^{(k+1)}}} 1 \;\geq\; 1\,, \quad \text{for every } k \in [\ell].$$

By continuity, one can choose $\varepsilon > 0$ such that

$$\Big(\frac{1}{m} - \varepsilon\Big)^2 \sum_{\substack{i,j:\\ a_i^{(k)} > a_j^{(k+1)}}} 1 \;-\; \Big(\frac{1}{m} + \varepsilon\Big)^2 \sum_{\substack{i,j:\\ a_i^{(k)} < a_j^{(k+1)}}} 1 \;>\; \frac{1}{2m^2} \;>\; 0\,, \quad \text{for every } k \in [\ell].$$

Apply the upper estimate of (3.5) for pairs i,j with $a_i^{(k)} < a_j^{(k+1)}$ and the lower estimate for pairs with $a_i^{(k)} > a_j^{(k+1)}$. Then, on the event G we have

$$\sum_{\substack{i,j:\\ a_i^{(k)} < a_j^{(k+1)}}} N_{k,i}N_{k+1,j} \;<\; n^2\Big(\frac{1}{m} + \varepsilon\Big)^2 \sum_{\substack{i,j:\\ a_i^{(k)} < a_j^{(k+1)}}} 1 \;<\; n^2\Big(\frac{1}{m} - \varepsilon\Big)^2 \sum_{\substack{i,j:\\ a_i^{(k)} > a_j^{(k+1)}}} 1 \;<\; \sum_{\substack{i,j:\\ a_i^{(k)} > a_j^{(k+1)}}} N_{k,i}N_{k+1,j}\,,$$

implying that on G we have $B^{(k)} \rhd B^{(k+1)}$ for every $k \in [\ell]$. □

As an application of above, take for instance the Effron's dice, which are given by

$$\begin{aligned} A &= (0,0,4,4,4,4), \quad & B &= (3,3,3,3,3,3)\\ C &= (2,2,2,2,6,6), \quad & D &= (1,1,1,5,5,5), \end{aligned}$$

as mentioned in the Introduction. In this case, the laws associated to each die would be:

$$\mathcal{L}_A = \frac{1}{3}\delta_0 + \frac{2}{3}\delta_4\,, \quad \mathcal{L}_B = \delta_3\,, \quad \mathcal{L}_C = \frac{2}{3}\delta_2 + \frac{1}{3}\delta_6\,, \quad \text{and} \quad \mathcal{L}_D = \frac{1}{2}\delta_1 + \frac{1}{2}\delta_5\,.$$

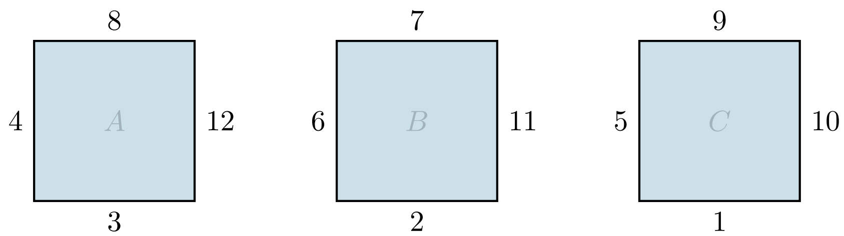
Of course, since the corresponding sequence of dice with the above laws is asymptotically intransitive, it cannot fulfill the assumptions of Theorem 6. Note that there are no ties in Effron's dice example, so condition (2.21) is trivially satisfied. The assumptions of Theorem 6 are not satisfied because the quantity of victories of a die against another die becomes deterministic (in view of the law of large numbers) as the number of faces in each die increases, and its variance degenerates to zero.

4. On deterministic intransitive dice

The main goal of this section is to prove Theorems 1 and 2. We keep using the notation introduced in Section 2.1, but in this section every dice considered is deterministic, that is, the entries $D_i^{(j)}$ of the die $D^{(j)}$ in a collection $\mathbf{D} = (D^{(1)}, \dots, D^{(\ell)})$ are always prescribed deterministic numbers rather than nontrivial random variables.

As explained after Theorem 1, when investigating existence of *no-tie* intransitive dice, only the relative ordering of the faces of the dice matters, but not their particular value. Thus, in this section we also restrict to dice whose faces' entries are positive integer numbers, and always pairwise distinct.

4.1. **A bijection between dice and words.** We look at the set of dice labels $\mathcal{A} = \{D^{(1)}, \dots, D^{(\ell)}\}$ as an alphabet, and now explain how to map dice to words. Let $\mathcal{W}(n_1, \dots, n_\ell)$ be the set of strings (or words) with $n_1 + \cdots + n_\ell$ letters in the alphabet $\mathcal{A}$, such that each letter $D^{(k)}$ appears exactly n_k times. There is a natural bijection[1] $\mathcal{D}(n_1, \dots, n_\ell) \overset{\pi}{\mapsto} \mathcal{W}(n_1, \dots, n_\ell)$: a collection of dice $\mathbf{D}$ is mapped to the word $\mathbf{W} = W_1 \cdots W_n$ determined uniquely by the rule that the letter W_i is equal to $D^{(k)}$ if the number $n - i + 1$ appears in a face of the die $D^{(k)}$. This bijection for $\mathcal{D}(4, 4, 4)$ is represented in Figure 1.



A	12				8				4	3		
B		11				7	6				2	
C			10	9				5				1
$\mathbf{W}$	A	B	C	C	A	B	B	C	A	A	B	C

Figure 1. An example of a triple of dice $\mathbf{D} = (D^{(1)}, D^{(2)}, D^{(3)}) = (A, B, C) \in \mathcal{D}(4, 4, 4)$ and its representation as a 12-letter word $\pi(\mathbf{D}) = \mathbf{W} = ABCCABBCAABC \in \mathcal{W}(4, 4, 4)$ in the alphabet $\mathcal{A} = \{A, B, C\}$.

Recall that $\mathcal{D}_\triangleright(n_1, \dots, n_\ell)$ is the subset of $\mathcal{D}(n_1, \dots, n_\ell)$ that consists of intransitive words, and denote by $\mathcal{W}_\triangleright(n_1, \dots, n_\ell)$ the corresponding image of $\mathcal{D}_\triangleright(n_1, \dots, n_\ell)$ by the bijection π. Given the word representation $\mathbf{W}$ of a collection of dice $\mathbf{D}$, it is also possible to compare which one of two dice $D^{(i)}$ and $D^{(j)}$ in $\mathbf{D}$ is stronger: one sums how many letters $D^{(i)}$ are to the right of every letter $D^{(j)}$. The result is how many possible victories $D^{(j)}$ has over $D^{(i)}$, and if this

[1] The idea of translating the dice as strings were inspired by a video on the YouTube channel Polylog: "We designed special dice using math, but there's a catch", available at https://youtu.be/-64UT8yikng.

result is larger than half the total number of combinations $n_j n_i$, then $D^{(j)} \triangleright D^{(i)}$. In particular, repeating this process over consecutive letters in a given word $\mathbf{W}$, it is possible to determine whether it belongs to $\mathcal{W}_{\triangleright}(n_1, \ldots, n_\ell)$.

To illustrate this process in the dice from Figure 1, introduce some auxiliary labels in the letters $\boldsymbol{B}$ from $\mathbf{W}$ as

$$\mathbf{W} = ABCCABBCAABC = A_1BCCA_2BBCA_3A_4BC. \tag{4.1}$$

There are 4 B's to the right of A_1, 3 B's to the right of A_2, and 1 B to the right of each of A_3 and A_4. Thus, the number of victories of the die A over B is $4+3+1+1=9$. By symmetry, the number of victories of B over A is $16-9=7$, and in this case $A \triangleright B$.

Also, to compare which of two given dice $D^{(i)}$ and $D^{(j)}$ of a collection $\mathbf{D}$ is better, it suffices to know the sub-word in $\pi(\mathbf{D})$ obtained when we remove all letters different from $D^{(i)}$ and $D^{(j)}$. For instance, in the example just explained we could have compared the dice A and B by looking solely at the sub-word $ABABBAAB$ obtained when we remove the C's from $\mathbf{W}$ in (4.1).

It is convenient to introduce the quantities

$$N_{i,j}(\mathbf{D}) \coloneqq \sum_{\substack{k_1,k_2 \\ D^{(i)}_{k_1} > D^{(j)}_{k_2}}} 1, \quad \text{and its induced version on } \mathbf{W}, \text{ namely } N_{i,j}(\mathbf{W}) \coloneqq N_{i,j}(\pi^{-1}(\mathbf{D})). \tag{4.2}$$

In general, for $\mathbf{W} \in \mathcal{W}(n_1, \ldots, n_\ell)$ the numbers $N_{i,j}(\mathbf{W})$ satisfy

$$N_{i,j}(\mathbf{W}) + N_{j,i}(\mathbf{W}) = n_i n_j, \tag{4.3}$$

and the statement $\mathbf{D}^{(i)} \triangleright \mathbf{D}^{(j)}$ is equivalent to saying that

$$N_{i,j}(\mathbf{W}) \;>\; \frac{n_i n_j}{2}. \tag{4.4}$$

Furthermore, if $\mathbf{W}$ is any sub-word of $\widetilde{\mathbf{W}}$ obtained without removing two given letters $D^{(j)}$ and $D^{(k)}$, then

$$N_{i,j}(\mathbf{W}) \;=\; N_{i,j}(\widetilde{\mathbf{W}}), \tag{4.5}$$

which follows from the interpretation of $N_{i,j}(\cdot)$ as the number of $D^{(i)}$'s to the left of $D^{(j)}$'s in the given letter.

4.2. **Proof of Theorem 1.**

We now focus on dice with the same number of faces n, that is, we fix ℓ and look at $\mathcal{D}_\ell(n)$ and $\mathcal{D}_{\triangleright,\ell}(n)$ from (2.2), and their corresponding images

$$\mathcal{W}_\ell(n) \coloneqq \pi(\mathcal{D}_\ell(n)) \quad \text{and} \quad \mathcal{W}_{\triangleright,\ell}(n) \coloneqq \pi(\mathcal{D}_{\triangleright,\ell}(n)).$$

The proof of Theorem 1 is based on the next result.

Proposition 8. *The following properties holds.*

(i) For $\ell \geq 3$, the sets $\mathcal{W}_{\triangleright,\ell}(2)$ is empty.
(ii) The sets $\mathcal{W}_{\triangleright,3}(3)$ and $\mathcal{W}_{\triangleright,3}(4)$ are both non-empty.
(iii) If the set $\mathcal{W}_{\triangleright,\ell}(n)$ is non-empty, then both sets $\mathcal{W}_{\triangleright,\ell}(n+2)$ and $\mathcal{W}_{\triangleright,\ell+1}(n)$ are also non-empty.

Proof. To prove (i), let $\mathbf{W} \in \mathcal{W}_\ell(2)$ for which $D^{(1)} \triangleright D^{(2)} \triangleright \ldots \triangleright D^{(\ell)}$. In this case, we have that

$$N_{j,k}(\mathbf{W}) + N_{k,j}(\mathbf{W}) = 4 \quad \text{for any } j \neq k,$$

so in this case $N_{j,k} \geq 3$ whenever $D^{(j)} \triangleright D^{(k)}$. We learned the following: any sub-word of $\mathbf{W}$ in two different letters $D^{(j)}$ and $D^{(k)}$ for which $D^{(j)} \triangleright D^{(k)}$ has to be either one of the following two words

$$D^{(j)}D^{(j)}D^{(k)}D^{(k)} \quad \text{or} \quad D^{(j)}D^{(k)}D^{(j)}D^{(k)}. \tag{4.6}$$

Thus, in $\mathbf{W}$ there is always a $D^{(1)}$ to the left of the two occurrences of $D^{(2)}$, there is always a $D^{(2)}$ to the left of the two ocurrences of $D^{(3)}$ etc. Consequently, there is always a $D^{(1)}$ to the left of the two occurrences of $D^{(k)}$, for any $k \geq 1$. Hence, the sub-word of $\mathbf{W}$ in $D^{(1)}$ and $D^{(\ell)}$ cannot be of the form (4.6) with $j = \ell$ and $k = 1$, so the relation $D^{(\ell)} \triangleright D^{(1)}$ is not verified in $\mathbf{W}$.

For (ii), examples of words in $\mathcal{W}_{\triangleright,3}(3)$ and $\mathcal{W}_{\triangleright,3}(4)$, and their corresponding dice, are displayed in Figure 2. The proof of part (iii) is postponed to Section 4.3. □

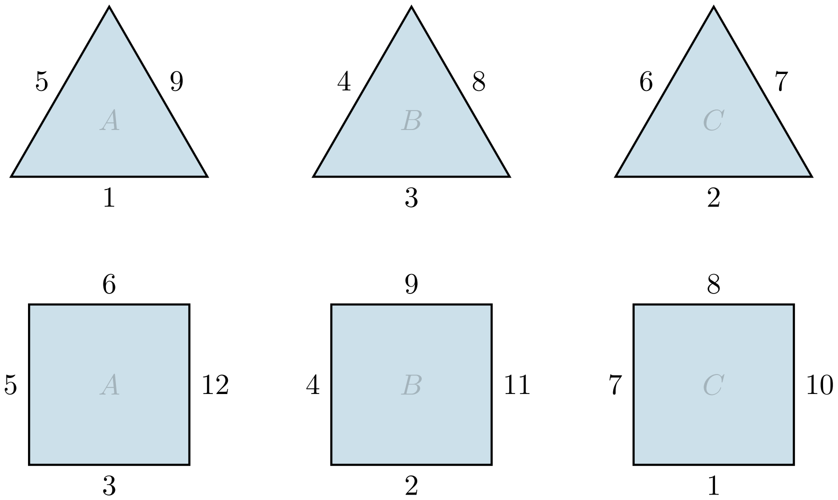


FIGURE 2. On top, a collection of three 3-sided dice corresponding to the word $\mathbf{W} = ABCCABBCA \in \mathcal{W}_{\triangleright,3}(3)$. On bottom, a collection of three 4-sided dice corresponding to the word $\mathbf{W} = ABCBCCAABABC \in \mathcal{W}_{\triangleright,3}(4)$

With Proposition 8 at hand, we are ready to prove Theorem 1.

Proof of Theorem 1. From the definition of $\mathcal{D}_{\triangleright,\ell}(n)$ it immediately follows that Theorem 1 is equivalent to the following two claims:

(1) For any $\ell \geq 3$, the set $\mathcal{D}_{\triangleright,\ell}(2)$ is empty.
(2) For any $n \geq 3$ and $\ell \geq 3$, the set $\mathcal{D}_{\triangleright,\ell}(n)$ is non-empty.

The claim (1) above follows from the definition of the bijection π and Proposition 8–(i). In turn, claim (2) follows applying Proposition 8–(iii) recursively, having in mind that $\mathcal{D}_{\triangleright,3}(3)$ and $\mathcal{D}_{\triangleright,4}$ are both non-empty by Proposition 8–(ii). □

4.3. **Proof of Proposition 8–(iii).**

To prove Proposition 8 we need some additional notions and lemmas. For what comes next, we recall that the numbers $N_{i,j}(\mathbf{W})$ were introduced in (4.2).

The **dual word** $\mathbf{W}^* \in \mathcal{W}(n_1, \dots, n_\ell)$ is obtaining reversing the ordering of the letters in a given word $\mathbf{W} \in \mathcal{W}(n_1, \dots, n_\ell)$; in the Example (4.1) the result is

$$\mathbf{W}^* = CBAACBBACCBA.$$

Lemma 9. *Let $\mathbf{W} \in \mathcal{W}_\ell(n)$ and $\mathbf{W}^*$ its dual word. Then*

$$N_{i,j}(\mathbf{W}) = N_{j,i}(\mathbf{W}^*).$$

Proof. The number $N_{i,j}(\mathbf{W})$ counts the number of times the letter $D^{(i)}$ appears to the left of each $D^{(j)}$ in $\mathbf{W}$. The dual letter $\mathbf{W}^*$ is obtained from $\mathbf{W}$ by reading the letters from $\mathbf{W}$ in the backwards manner, interpretation from which the lemma follows. □

We say a collection of dice $\mathbf{D} \in \mathcal{D}(n_1, \dots, n_\ell)$, or its corresponding letter $\mathbf{W} = \pi(\mathbf{D})$, is **neutral** if any given die in $\mathbf{D}$ beats any other given die in $\mathbf{D}$ the same amount of times. In terms of $N_{i,j}(\mathbf{W})$ this is equivalent to verifying that

$$N_{i,j}(\mathbf{W}) = N_{j,i}(\mathbf{W}). \tag{4.7}$$

From this relation and (4.3) it follows that if $\mathbf{W} \in \mathcal{W}_\ell(n)$ is neutral, then n must be even.

We also talk about concatenation of two words $\mathbf{W}_1$ and $\mathbf{W}_2$ into a new word $\mathbf{W} = \mathbf{W}_1\mathbf{W}_2$; in the example (4.1), for instance, we can write $\mathbf{W} = \mathbf{W}_1\mathbf{W}_2$ with $\mathbf{W}_1 = ABCCA$ and $\mathbf{W}_2 = BBCAABC$.

When dealing with concatenations, the involved words need not be in the same letters, neither need they have the same size. However, when $\mathbf{W}_1 \in \mathcal{W}_\ell(n_1)$ and $\mathbf{W}_2 \in \mathcal{W}_\ell(n_2)$, then obviously $\mathbf{W}_1\mathbf{W}_2 \in \mathcal{W}_\ell(n_1 + n_2)$ and the identity

$$N_{i,j}(\mathbf{W}_1\mathbf{W}_2) = N_{i,j}(\mathbf{W}_1) + n_1 n_2 + N_{i,j}(\mathbf{W}_2) \tag{4.8}$$

holds true.

Lemma 10. *Given any word $\mathbf{W} \in \mathcal{W}_\ell(n)$, the concatenation $\widetilde{\mathbf{W}} = \mathbf{W}\mathbf{W}^* \in \mathcal{W}_\ell(2n)$ is neutral.*

Proof. Applying Lemma 9 to (4.8), we obtain that $N_{i,j}(\mathbf{W}) = N_{j,i}(\mathbf{W}^*)$ for any $i \neq j$. The proof is then completed using (4.7). □

We are ready to start applying recursive arguments that preserve intransitive words, starting with adding a new letter to a word known to be transitive.

Lemma 11. *If $\mathcal{W}_{\triangleright,\ell}(n)$ is non-empty, then $\mathcal{W}_{\triangleright,\ell+1}(n)$ is non-empty.*

Proof. From a given $\mathbf{W} \in \mathcal{W}_\ell(n)$, create a new word $\widetilde{\mathbf{W}} \in \mathcal{W}_{\ell+1}(n)$ obtained by replacing every occurrence of $D^{(\ell)}$ by $D^{(\ell)}D^{(\ell+1)}$. In the example $\mathbf{W} \in \mathcal{W}_3(4)$ from (4.1), the new word $\widetilde{\mathbf{W}} \in \mathcal{W}_4(4)$ is

$$\widetilde{\mathbf{W}} = ABCDCDABBCDAABCD.$$

In virtue of (4.5), we see that the relation $D^{(k)} \triangleright D^{(k+1)}$ for $k = 1, \dots, n-1$ is preserved when going from a letter $\mathbf{W} \in \mathcal{W}_{\triangleright,\ell}(n)$ to the corresponding letter $\widetilde{\mathbf{W}} \in \mathcal{W}_{\ell+1}(n)$. Still by construction, the relations

$$N_{\ell,\ell+1}(\widetilde{\mathbf{W}}) > N_{\ell+1,\ell}(\widetilde{\mathbf{W}}) N_{\ell,1}(\mathbf{W}) = N_{\ell+1,1}(\widetilde{\mathbf{W}})$$

are of straightforward verification, since each letter $\mathbf{D}^{(\ell)}$ appears immediately to the right of a letter $D^{(\ell)}$ in $\widetilde{\mathbf{W}}$. When $\mathbf{W} \in \mathcal{W}_{\triangleright,\ell}(n)$, the inequality above shows that $D^{(\ell)} \triangleright D^{(\ell+1)}$ in $\widetilde{\mathbf{W}}$, and the equality above shows that the relation $D^{(\ell)} \triangleright D^{(1)}$ in $\mathbf{W}$ transfers to the relation $D^{(\ell+1)} \triangleright D^{(1)}$ in $\widetilde{\mathbf{W}}$. $\square$

Adding new faces whilst preserving intransitivity is a little bit more involved, and will be based on the next two lemmas.

Lemma 12. *Fix $k \geq 1$, and suppose that $\mathbf{I} \in \mathcal{W}_\ell(2k)$ is a neutral word. Let $\mathbf{W} \in \mathcal{W}_\ell(n)$. Then $\mathbf{W} \in \mathcal{W}_{\triangleright,\ell}(n)$ if, and only if, $\mathbf{IW} \in \mathcal{W}_{\triangleright,\ell}(n+2k)$.*

Proof. From (4.8) we learn that for any $i \neq j$,

$$N_{i,j}(\mathbf{IW}) \;=\; N_{i,j}(\mathbf{I}) + 2kn + N_{i,j}(\mathbf{W})\,.$$

Using the symmetry (4.7) for the neutral word $\mathbf{I}$, we obtain

$$N_{k,k+1}(\mathbf{IW}) - N_{k+1,k}(\mathbf{IW}) \;=\; N_{k,k+1}(\mathbf{W}) - N_{k+1,k}(\mathbf{W})\,, \quad k = 1, \ldots, \ell - 1.$$

The result then follows from (4.4). $\square$

Lemma 13. *If $\mathcal{W}_\ell(n)$ is non-empty, then $\mathcal{W}_\ell(n+2)$ is non-empty.*

Proof. By Lemma 10, the word $\mathbf{I} = \mathbf{SS}^* \in \mathcal{W}_\ell(2)$ constructed from the choice

$$\mathbf{S} \;=\; D^{(1)} D^{(2)} \cdots D^{(\ell)}$$

is neutral. The result now follows from Lemma 12. $\square$

We finally complete the proof of Proposition 8.

Proof of Proposition 8–(iii). Proposition 8–(iii) is now simply a combination of Lemmas 11 and 13. $\square$

4.4. **On the number of intransitive words.**
The proof of Theorem 2 is now a consequence of some of the results already established.

Proof of Theorem 2. For given integers n_1, n_2, let $\mathbf{W}_j \in \mathcal{W}_{\triangleright,\ell}(n_j)$, $j = 1, 2$, so that by (4.4),

$$N_{k,k+1}(\mathbf{W}_j) > \frac{(n_j)^2}{2}\,, \quad k = 1, \ldots, \ell - 1, \quad \text{and} \quad N_{\ell,1}(\mathbf{W}_j) > \frac{(n_j)^2}{2}, \quad j = 1, 2.$$

Using these inequalities and (4.8), it follows that $\mathbf{W}_1\mathbf{W}_2 \in \mathcal{W}_\ell(n_1 + n_2)$ satisfies

$$N_{k,k+1}(\mathbf{W}_1\mathbf{W}_2) > \frac{(n_1+n_2)^2}{2},\ k = 1, \ldots, \ell - 1, \quad \text{and} \quad N_{\ell,1}(\mathbf{W}_1\mathbf{W}_2) > \frac{(n_1+n_2)^2}{2}.$$

Using again (4.8) we conclude that $\mathbf{W}_1\mathbf{W}_2 \in \mathcal{W}_{\triangleright,\ell}(n_1 + n_2)$. Hence,

$$|\mathcal{W}_{\triangleright,\ell}(n_1 + n_2)| \;\geq\; |\mathcal{W}_{\triangleright,\ell}(n_1)|\,|\mathcal{W}_{\triangleright,\ell}(n_2)|\,,$$

so the sequence $(\log|\mathcal{W}_{\triangleright,\ell}(n)|)_n$ is superadditive, and thus by Fekete's Lemma,

$$\lim_{n\to\infty} \frac{\log|\mathcal{W}_{\triangleright,\ell}(n)|}{n} \;=\; \sup_n \frac{\log|\mathcal{W}_{\triangleright,\ell}(n)|}{n} \;=\; L(\ell)$$

for some constant $L(\ell) > 0$. Since $\mathcal{W}_{\triangleright,\ell}(n) = \pi(\mathcal{D}_{\triangleright,\ell}(n))$ and π is a bijection, the result follows. $\square$

4.5. **Some numerical aspects on the number of intransitive words.**

Through a numerical study, we are able to estimate the number $L(3)$ from Theorem 2 as follows.

A simple algorithm computes $|\mathcal{D}_{\triangleright,3}(n)|$ in a straightforward way: we iterate through every word in the set $\mathcal{W}_3(n)$ and check whether each word is intransitive, a task that can be accomplished in $\Theta(n)$ operations. A drawback of this approach is that the number of words that need checking grows exponentially with respect to n. In fact, by Stirling's approximation, we have that $|\mathcal{D}_3(n)| = \Theta(27^n/n)$, resulting in a total time complexity of $\Theta(27^n)$. We optimize this algorithm by partitioning the set $\mathcal{W}_3(n)$ into words with the same "prefixes" (that is, the same sequence of letters for the first n positions in the word). We then avoid prefixes which are already known to not yield intransitive words, thus performing early exits while checking for intransitivity. The algorithm was implemented in C++ and executed on the Euler cluster maintained by the Center for Mathematical Sciences Applied to Industry (CeMEAI). Using this method, we were able to compute $|\mathcal{D}_{\triangleright,3}(n)|$ for $n \leq 11$ (see Table 1).

Using Table 1, we can estimate $L(3)$. By Fekete's lemma, we have that

$$L(3) = \sup_n \frac{\log|\mathcal{D}_3(n)|}{n} \geq \frac{\log|\mathcal{D}_3(11)|}{11} > 2.445.$$

On the other hand, as $\mathcal{W}_{\triangleright,\ell}(n) \subset \mathcal{W}_\ell(n)$, we obtain

$$L(\ell) \leq \lim_n \frac{\log|\mathcal{W}_\ell(n)|}{n} = \ell \log \ell.$$

In particular, the estimate $2.445 < L(3) \leq 3\log 3 \approx 3.296$ is valid.

n	$\lvert\mathcal{D}_{\triangleright,3}(n)\rvert$	$\lvert\mathcal{D}_3(n)\rvert$	$\lvert\mathcal{D}_{\triangleright,3}(n)\rvert/\lvert\mathcal{D}_3(n)\rvert$
3	15	1 680	0.008929
4	39	34 650	0.001126
5	5 196	756 756	0.006866
6	32 115	17 153 136	0.001872
7	2 093 199	399 072 960	0.005245
8	19 618 353	9 465 511 770	0.002073
9	960 165 789	227 873 431 500	0.004214
10	11 272 949 151	5 550 996 791 340	0.002031
11	479 538 890 271	136 526 995 463 040	0.003512

TABLE 1. Values of $|\mathcal{D}_{\triangleright,3}(n)|$, $|\mathcal{D}_3(n)|$ and $|\mathcal{D}_{\triangleright,3}(n)|/|\mathcal{D}_3(n)|$ for $3 \leq n \leq 11$. In the values in the table, the ratios $|\mathcal{D}_{\triangleright,3}(n)|/|\mathcal{D}_3(n)|$ are lower for even n. The decrease in ratios for even n may be due to neutral strings only being possible for even n, leading to fewer intransitive strings proportionally.

The algorithm just described yields exact values of $|\mathcal{D}_{\triangleright,3}(n)|$, producing the values shown in Table 1. However, its performance is very slow in n, and even computing $|\mathcal{D}_{\triangleright,3}(n)|$ for, say, $n = 12$ already, becomes out of reach. With this issue in mind, we also performed an stochastic simulation to estimate $|\mathcal{D}_{\triangleright,3}(n)|/|\mathcal{D}_3(n)|$ for a sample of values of n up to 500, an arbitrary cut

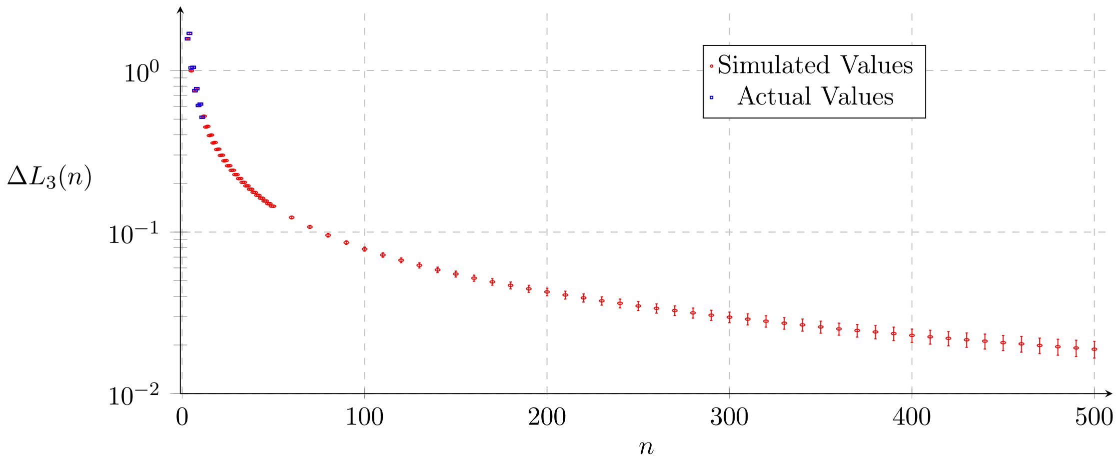


FIGURE 3. $\Delta L_3(n)$ for various values of n. The blue data points were calculated using Table 1, and the red data points were generated through a stochastic simulation. The vertical axis is represented in a logarithmic scale

where fluctuations in the estimation should still be controlled. We sample uniformly from $D_3(n)$. We can then estimate $\Delta L_3(n) = -\log(|\mathcal{D}_{\triangleright,3}(n)|/|\mathcal{D}_3(n))|/n$. The results are displayed in Figure 3, showing good agreement with the values computed in Table 1. Observe how the values seem to tend to 0. Since $\lim_n \Delta L_3(n) = 3\log 3 - L(3)$, we conjecture that $L(3) = 3\log 3$, so that the fraction $|\mathcal{D}_{\triangleright,3}(n)|/|\mathcal{D}_3(n)|$ should decay sub-exponentially with n. Compare it with (2.3). All the algorithms and data presented here are publicly available in our repository in GitHub[2].

5. Some generalities on the counting functions and Gaussian vectors

The goal here is to prove Theorem 4. In this section $\mathbf{D}_m = (D^{(1)}, \dots, D^{(\ell)})$ will always denote a collection of $\ell \geq 3$ independent random dice, where each face of $D^{(i)} = D^{(i)}(m)$ has law $\mathcal{L}^{(i)} = \mathcal{L}^{(i)}_m$, and such that the sequence $\{\mathbf{D}_m\}_m$ satisfies Assumption 3.

5.1. Properties of the counting functions.

Recall the counting variables N_j, their normalized versions $\widetilde{N}_j$, and E_j, which were defined in (2.6), (2.13) and (2.7), respectively, and the quantities $\mathbf{p}_k, \mathbf{q}_k, \mathbf{r}_k$ and $\mathbf{s}_k$, which depend only on the collection of laws $\mathcal{L}^{(1)}, \dots, \mathcal{L}^{(\ell)}$, were introduced in (2.9)–(2.12). Note that $\mathbf{q}_k$ and $\mathbf{r}_k$ are not necessarily equal because each die has its law. As said before, these quantities also depend on m but we will not write down this dependence explicitly.

The next lemma establishes (2.16).

Lemma 14. *We have*

$$\begin{aligned}
\mathbb{E}N_k &= n_k n_{k+1} \mathbf{p}_k\,, \\
\operatorname{Var}(N_k) &= n_k n_{k+1}\big[n_k(\mathbf{q}_k - \mathbf{p}_k^2) + n_{k+1}(\mathbf{r}_k - \mathbf{p}_k^2) + \mathbf{p}_k^2 + \mathbf{p}_k - \mathbf{q}_k - \mathbf{r}_k\big]\,, \qquad \textit{and} \\
\operatorname{Cov}(N_{k-1}, N_k) &= n_{k-1} n_k n_{k+1}\left(\mathbf{s}_k - \mathbf{p}_{k-1}\mathbf{p}_k\right).
\end{aligned}$$

Consequently, under Assumption 3 we have that as $m \to \infty$,

$$\mathbb{E}N_k = f_k(\infty) f_{k+1}(\infty) \mathbf{p}_k(\infty) m^2 + o(m^2)\,,$$

[2]https://github.com/NonTransitiveDices/NonTransitiveDices.git

$$\begin{aligned}\mathrm{Var}\,(N_k) &= \sigma_k(\infty)^2 m^3 + O(m^2)\,,\\ \mathrm{Corr}\,(N_{k-1}, N_k) &= \gamma_k(\infty) + O(1/m).\end{aligned}$$

Proof. The calculation of $\mathbb{E}N_k$ is immediate from the definition. For the variance, we begin noticing that

$$\mathbb{E}\left[N_k^2\right] = \sum_{i_1=1}^{n_k}\sum_{j_1=1}^{n_{k+1}}\sum_{i_2=1}^{n_k}\sum_{j_2=1}^{n_{k+1}} \mathbb{P}\left(\{D_{i_1}^{(k)} > D_{j_1}^{(k+1)}\} \cap \{D_{i_2}^{(k)} > D_{j_2}^{(k+1)}\}\right).$$

The probability of such intersections is always in $\{\mathbf{p}_k^2, \mathbf{p}_k, \mathbf{r}_k, \mathbf{q}_k\}$, depending on whether indices i_1 and i_2 or j_1 and j_2 coincide. Decomposing into all possibilities, we have

$$\begin{aligned}\mathbb{E}\left[N_k^2\right] &= \sum_{\substack{i_2\neq i_1\\ j_2\neq j_1}} \mathbf{p}_k^2 + \sum_{\substack{i_2=i_1\\ j_2\neq j_1}} \mathbf{r}_k + \sum_{\substack{j_2=j_1\\ i_2\neq i_1}} \mathbf{q}_k + \sum_{\substack{i_2=i_1\\ j_2=j_1}} \mathbf{p}_k\\ &= n_k(n_k-1)n_{k+1}(n_{k+1}-1)\mathbf{p}_k^2 + n_k n_{k+1}(n_{k+1}-1)\mathbf{r}_k\\ &\quad + n_k(n_k-1)n_{k+1}\mathbf{q}_k + n_k n_{k+1}\mathbf{p}_k.\end{aligned}$$

Hence, the variance of N_k is given by

$$\begin{aligned}\mathrm{Var}\,(N_k) &= \mathbb{E}\left[N_k^2\right] - \mathbb{E}\left[N_k\right]^2 = \mathbb{E}\left[N_k^2\right] - (n_k n_{k+1}\mathbf{p}_k)^2\\ &= n_k n_{k+1}\big[n_k(\mathbf{q}_k - \mathbf{p}_k^2) + n_{k+1}(\mathbf{r}_k - \mathbf{p}_k^2) + \mathbf{p}_k^2 + \mathbf{p}_k - \mathbf{q}_k - \mathbf{r}_k\big].\end{aligned}$$

The calculation of $\mathrm{Cov}\,(N_{k-1}, N_k)$ is similar, and the asymptotic expressions follow immediately. □

For the similar result for the E_k's, we need to introduce certain quantities analogous to $\mathbf{p}_k, \mathbf{q}_r, \mathbf{s}_k$. For E_k as in (2.7), let us introduce

$$\mathbf{p}_k^{=} := \mathbb{P}\left(D_1^{(k)} = D_1^{(k+1)}\right) = \mathbb{E}\left(\mathbb{1}_{D_1^{(k)}=D_1^{(k+1)}}\right) \tag{5.1}$$

which is the that probability a given face of the k-th die coincide with a given face of the $(k+1)$-th die;

$$\mathbf{q}_k^{=} := \mathbb{P}\left(D_1^{(k)} = D_1^{(k+1)}, D_2^{(k)} = D_1^{(k+1)}\right) \tag{5.2}$$

which is the probability that two given faces of the k-th die coincide with a given face of the $(k+1)$-th die; and

$$\mathbf{s}_k^{=} := \mathbb{P}\left(D_1^{(k-1)} = D_1^{(k)} = D_1^{(k+1)}\right), \tag{5.3}$$

which is the probability that three given faces, one from each of the dice $D^{(k-1)}$, $D^{(k)}$ and $D^{(k+1)}$, coincide.

The next result is the analogue of Lemma 14 for the variables E_k's.

Lemma 15. *The random variables E_k, $k = 1, \dots, \ell$, satisfy*

$$\mathbb{E}\,(E_k) = n_k n_{k+1}\mathbf{p}_k^{=} \tag{5.4}$$

$$\mathrm{Var}\,(E_k) = n_k n_{k+1}\big[(n_k + n_{k+1})(\mathbf{q}_k^{=} - (\mathbf{p}_k^{=})^2) + (\mathbf{p}_k^{=})^2 + \mathbf{p}_k^{=} - 2\mathbf{q}_k^{=}\big]. \tag{5.5}$$

In particular, the estimates (2.23) *hold true.*

Proof. The proof of (5.4)–(5.5) is done following the same steps used in the proof of Lemma 14, we skip the details. The estimates (2.23) then follow, having in mind (2.4) and the fact that each $\mathbf{p}_k^{=}, \mathbf{q}_k^{=}, \boldsymbol{s}_k^{=}$ is a probability, and thus bounded as functions of m. □

Using the previous result, we are able to compare (2.21) with (2.23).

Lemma 16. *If condition (2.21) holds, then for every $k = 1, \ldots, \ell$, it is valid*

$$\mathbb{E}(E_k) = o(m^2) \quad \textit{and} \quad \mathrm{Var}\,(E_k) = o(m^3) \quad \textit{as} \quad m \to \infty.$$

Proof. Condition (2.21) is the same as saying that $\boldsymbol{p}_k^{=} \to 0$ for every k. Thus, the claim on $\mathbb{E}(E_k)$ is immediate from (5.4). From (5.5) and the fact that $\mathbf{p}_k^{=}$ and $\mathbf{q}_k^{=}$ both remain bounded as $m \to \infty$, we see that

$$\mathrm{Var}\,(E_k) \;=\; m^3 f_k(\infty) f_{k+1}(\infty)(f_k(\infty) + f_{k+1}(\infty))(\mathbf{q}_k^{=} - (\mathbf{p}_k^{=})^2) + O(m^2)\,.$$

A comparison of (5.1) and (5.2) shows that $0 \leq q_k^{=} \leq p_k^{=}$, so that (2.21) implies also that $\boldsymbol{q}_k^{=} \to 0$ and the claim on $\mathrm{Var}\,(E_k)$ follows. ☐

Next, we turn our attention to the structure of the covariance matrix (2.17). It turns out that in the case of particular interest to our problem, the coefficients $\gamma_j(\infty)$ admit a particularly interesting structure, as we now compute.

Proposition 17. *Let $\{\mathbf{D}_m\}_m$ be a sequence satisfying the conditions of Theorem 5. Then the coefficients $(\gamma_k(\infty))$ and $(f_k(\infty))$ from Assumption 3 are related by*

$$\gamma_k(\infty) \;=\; -\frac{f_{k-1}(\infty) f_k(\infty) f_{k+1}(\infty)}{\sqrt{f_{k-1}(\infty) f_k(\infty)(f_{k-1}(\infty) + f_k(\infty))}\sqrt{f_k(\infty) f_{k+1}(\infty)(f_k(\infty) + f_{k+1}(\infty))}}$$

for $k = 1, \ldots, \ell$.

Proof. From (5.1),(5.2),(5.3), it follows that $0 \leq \mathbf{s}_k^{=}, \mathbf{q}_k^{=} \leq \mathbf{p}_k^{=}$. Condition (16) says that $\boldsymbol{p}_k^{=} \to 0$, and in this case we therefore have $\mathbf{s}_k^{=}, \mathbf{q}_k^{=} \to 0$ as well.

Next, the events defining $\mathbf{p}_k, \mathbf{r}_k, \mathbf{q}_k, \mathbf{s}_k$ amount to observing specific orderings of the faces involved. Since each face has the same distribution, any fixed ordering has the same probability. For instance, we have

$$1 \;=\; \mathbb{P}(D_1^{(k)} > D_1^{(k+1)}) + \mathbb{P}(D_1^{(k)} < D_1^{(k+1)}) + \mathbb{P}(D_1^{(k)} = D_1^{(k+1)}) \;=\; 2\mathbf{p}_k + \mathbf{p}_k^{=}\,,$$

implying that $\mathbf{p}_k \to \frac{1}{2}$ as $m \to \infty$. In an analogous way we obtain that $\mathbf{q}_k \to \frac{1}{3}$, $\mathbf{r}_k \to \frac{1}{3}$ and $\mathbf{s}_k \to \frac{1}{6}$. The formula for $\gamma_k(\infty)$ then follows plugging these limits into (2.14) and (2.15). ☐

5.2. Gaussian vectors associated to the structured covariance matrix.

Under the conditions of Theorem 5, Proposition 17 ensures that the nontrivial entries $\gamma_k(\infty)$ of the covariance matrix (2.17) have a particular structure, which ultimately yield that the probability in the right-hand side of (2.22) vanishes, and proving this last claim is the main goal of this subsection.

To avoid cumbersome notation, for the calculations that come next we denote

$$\mathfrak{f}_k \;\coloneqq\; f_k(\infty), \quad k = 1, \ldots, \ell, \quad \mathfrak{f}_{\ell+1} \coloneqq f_1(\infty)\,,$$

so that

$$\gamma_k(\infty) \;=\; -\frac{\mathfrak{f}_{k-1}\mathfrak{f}_k\mathfrak{f}_{k+1}}{\sqrt{\mathfrak{f}_{k-1}\mathfrak{f}_k(\mathfrak{f}_{k-1} + \mathfrak{f}_k)}\sqrt{\mathfrak{f}_k\mathfrak{f}_{k+1}(\mathfrak{f}_k + \mathfrak{f}_{k+1})}}\,. \tag{5.6}$$

To study the covariance matrix Σ from (2.17) with coefficients (5.6), we start by collecting some properties of these $\gamma_k(\infty)$'s.

Proposition 18. *The coefficients $\gamma_k = \gamma_k(\infty)$, $k = 1, \ldots, \ell$, in (5.6) satisfy the following properties.*

(i) $\gamma_k^2 = \dfrac{\mathfrak{f}_{k-1}}{\mathfrak{f}_{k-1}+\mathfrak{f}_k} \cdot \dfrac{\mathfrak{f}_{k+1}}{\mathfrak{f}_k+\mathfrak{f}_{k+1}}$.

(ii) $\gamma_k \in (-1,0)$ *for every* k.

(iii) As functions of the $\mathfrak{f}_j$*'s, the coefficients* $\gamma_k = \gamma_k(\mathfrak{f}_1,\dots,\mathfrak{f}_\ell)$ *are scale-invariant: for every* $k \in [\ell]$ *and* $r > 0$ *we have*

$$\gamma_k(r\mathfrak{f}_1,\dots,r\mathfrak{f}_\ell) = \gamma_k(\mathfrak{f}_1,\dots,\mathfrak{f}_\ell).$$

(iv) $\prod_k \gamma_k = (-1)^\ell \prod_k \frac{\mathfrak{f}_k}{\mathfrak{f}_k+\mathfrak{f}_{k+1}}$.

(v) $|\prod_k \gamma_k| \le 2^{-\ell}$*, with equality being valid if, and only if,* $\mathfrak{f}_1 = \cdots = \mathfrak{f}_\ell$.

Proof. Items *(i)*, *(iii)* and *(iv)* are immediate from (5.6). It is obvious that $\gamma_k < 0$, so to prove *(ii)* it suffices to show that $\gamma_k^2 < 1$ which, in turn, by *(i)* is equivalent to the inequality

$$\mathfrak{f}_{k-1}\mathfrak{f}_{k+1} < (\mathfrak{f}_{k-1}+\mathfrak{f}_k)(\mathfrak{f}_k+\mathfrak{f}_{k+1}), \quad \text{that is,} \quad 0 < \mathfrak{f}_{k-1}\mathfrak{f}_k + \mathfrak{f}_k^2 + \mathfrak{f}_k\mathfrak{f}_{k+1}.$$

Since $\mathfrak{f}_k > 0$ for every k, part *(ii)* follows. Finally, for item *(v)* we apply the inequality between arithmetic and geometric means to obtain

$$\frac{f_k + f_{k+1}}{2} \ge \sqrt{f_k f_{k+1}}, \text{ for every } k \in [\ell].$$

Multiplying all inequalities above, the result follows using item *(iv)*. □

With the aforementioned properties of $\gamma_k = \gamma_k(\infty)$ from (5.6) at hand, we now need to collect some important information on the associated covariance matrix Σ from (2.17). From a linear algebra perspective, this is an example of a periodic Jacobi matrix (see for instance [12, 8, 4]). However, we could not explore these interpretations for the results needed later. Instead, in our case, we use the additional structure (5.6) in a fundamental way to obtain the next results.

Lemma 19. *Let* Σ *be as in* (2.17) *with coefficients* $\gamma_k = \gamma_k(\infty)$ *as in* (5.6)*. Then*

$$\det\Sigma \;=\; 1 + 2(-1)^{\ell-1}\gamma_1\dots\gamma_\ell + \sum_{m=1}^{\ell}(-1)^m \sum_{\substack{i_1<i_2<\dots<i_m:\\ i_j \text{ non-consecutive}}} \gamma_{i_1}^2\dots\gamma_{i_m}^2 \,. \tag{5.7}$$

Proof. We start from the expression of $\det\Sigma$ as a sum over all permutations $\sigma \in S_\ell$, the symmetric group of degree ℓ,

$$\det\Sigma \;=\; \sum_{\sigma\in S_\ell} \operatorname{sgn}\sigma \cdot \prod_{i\in[\ell]} \Sigma_{i\sigma(i)}\,.$$

To prove (5.7) we will now show that many terms do not contribute to the sum.

From the explicit form of Σ, we can see that the permutations σ such that $\sigma(i) \notin \{i-1, i, i+1\}$ for some i have $\Sigma_{i\sigma(i)} = 0$, recalling that we consider indices modulo ℓ.

For the remaining terms, consider the cyclic decomposition of $\sigma = \tau_1 \cdots \tau_m$ where τ_j are disjoint cycles. Using the disjointness, we can compute the product of $\Sigma_{i\sigma(i)}$ by evaluating it for each cycle.

The contribution of cycles of order 1 is always 1, since when $\tau = (i)$ we have that $\Sigma_{i\tau(i)} = \Sigma_{ii} = 1$.

The contribution of a cycle of order 2, say $\tau = (i\ j)$, is non-zero if, and only if, i and j are consecutive. If $i = j-1$ then $\Sigma_{j-1,j}\Sigma_{j,j-1} = \gamma_j^2$.

By a similar reasoning, for a cycle $\tau = (i_1\ i_2\ \cdots\ i_o)$ of order $o \ge 3$ to have a non-zero contribution one must have $o = \ell$. Indeed, we have $i_2 \in \{i_1-1, i_1+1\}$. If $i_2 = i_1+1$, since all indices must be consecutive we have $i_j = i_{j-1} + 1$ for every j. After i_o the cycle returns to i_1, implying that it has length ℓ. The case $i_2 = i_1 - 1$ is analogous.

Hence, apart from cycles of order 1 and cycles of the form $(i\ i+1)$, the only cycles whose product is non-zero are $(1\ 2\ \cdots\ \ell)$ and $(1\ \ell\ \ell-1\ \cdots\ 2)$. Both have the same product:

$$\prod_{i\in[\ell]} \Sigma_{i,i+1} \;=\; \prod_{i\in[\ell]} \gamma_i\,.$$

Finally, if σ is a permutation different from the identity and the two cycles of order ℓ, in order for its product to be non-zero one must have a cyclic decomposition $\sigma = \tau_1 \dots \tau_t$ with every τ_i being a cycle of order 1 or 2. As cycles of order 1 contribute with 1 to the product, we can focus on the cycles of order 2. Suppose there are m cycles of order 2 and reorder if necessary so that they are given by $\tau_1, \dots, \tau_m$ with $\tau_j = (i_j - 1\ i_j)$ and $i_1 < i_2 < \cdots < i_m$. The formula in (5.7) follows, since the i_j are non-consecutive by construction. □

Using Lemma 19 we are able to verify that $\det \Sigma$ is always zero.

Lemma 20. *Let Σ be as in (2.17) with coefficients $\gamma_k = \gamma_k(\infty)$ as in (5.6). Then* $\det \Sigma = 0$.

Proof. We have to prove that the right-hand side of (5.7) is zero. We will replace the expressions for $\gamma_k = \gamma_k(\infty)$ given by Proposition 18–*(i)*, *(iv)* and verify that the right-hand side of (5.7) vanishes. In order to make the computation more streamlined, it is convenient to reinterpret it as an estimate of probabilities, as we describe below.

Let us define $a_k := \frac{\mathfrak{f}_{k-1}}{\mathfrak{f}_{k-1}+\mathfrak{f}_k} \in (0,1)$, which satisfies $1 - a_{k+1} = \frac{\mathfrak{f}_{k+1}}{\mathfrak{f}_k+\mathfrak{f}_{k+1}}$. Consider a collection $(U_j : j \in [\ell])$ of i.i.d. uniform random variables in $(0,1)$ and set

$$A_k \;:=\; \{U_k \le a_k\}\,. \tag{5.8}$$

The collection $(A_k : k \in [\ell])$ consists of mutually independent events such that $\mathbb{P}(A_k) = a_k$. Defining $B_k := A_k \cap A_{k+1}^c$, it holds that

$$\mathbb{P}(B_k) \;=\; a_k(1 - a_{k+1}) \;=\; \gamma_k^2\,. \tag{5.9}$$

Let us compute the probability of the event $\cup_k B_k$ in two different ways. By the inclusion-exclusion principle, we have

$$\begin{aligned}\mathbb{P}(\cup B_k) \;&=\; \sum_{m=1}^{\ell} (-1)^{m-1} \sum_{i_1<i_2<\cdots<i_m} \mathbb{P}(B_{i_1}\cap\cdots\cap B_{i_m}) \;=\; \sum_{m=1}^{\ell} (-1)^{m-1} \sum_{\substack{i_1<i_2<\cdots<i_m\\ i_j \text{ non-consecutive}}} \mathbb{P}(B_{i_1}\cap\ldots\cap B_{i_m})\\ &=\; \sum_{m=1}^{\ell} (-1)^{m-1} \sum_{\substack{i_1<i_2<\ldots<i_m\\ i_j \text{ non-consecutive}}} \gamma_{i_1}^2 \ldots \gamma_{i_m}^2\,,\end{aligned} \tag{5.10}$$

where the second equality is due to $B_k \cap B_{k+1} = (A_k \cap A_{k+1}^c) \cap (A_{k+1} \cap A_{k+2}^c) = \varnothing$, and the last equality follows from (5.9) and independence. On the other hand, from the identity

$$\cap B_k^c \;=\; (\cap_{k\in[\ell]} A_k) \cup (\cap_{k\in[\ell]} A_k^c)$$

we compute

$$\begin{aligned}\mathbb{P}(\cup B_k) \;&=\; 1 - \mathbb{P}(\cap B_k^c) \;=\; 1 - \mathbb{P}(\cap A_k^c) - \mathbb{P}(\cap A_k)\\ &=\; 1 - a_1 \cdots a_\ell - (1-a_1)\cdots(1-a_\ell)\end{aligned}$$

$$= \ 1 - 2 \prod \frac{\mathfrak{f}_k}{\mathfrak{f}_k + \mathfrak{f}_{k+1}} = 1 - 2(-1)^{\ell} \prod \gamma_k \,. \tag{5.11}$$

Equating (5.10) and (5.11) the result follows. □

Lemma 20 ensures that zero is an eigenvalue of Σ, and we now collect some info about the associated eigenspace.

Proposition 21. *Let Σ be as in (2.17) with coefficients $\gamma_k = \gamma_k(\infty)$ as in (5.6). Then 0 is an eigenvalue of Σ, its eigenspace has dimension 1 and is generated by a vector $x \in (0,\infty)^{\ell}$.*

Proof. Let $x = (x_1, \dots, x_\ell)$ be a non-zero vector satisfying $\Sigma x = 0$. Then, for every $k \in [\ell]$ we have

$$\gamma_k x_{k-1} + x_k + \gamma_{k+1} x_{k+1} \ = \ 0\,. \tag{L_k}$$

It is possible to solve the system of equations above explicitly, but the formulas obtained this way are cumbersome. Instead, we show that if some coordinate x_j is positive then x_{j+1} is positive as well. Since we can always choose the sign of one entry of x, by the cyclic symmetry of the problem we then conclude that there is $x \in (0,\infty)^{\ell}$ with $\Sigma x = 0$, as wanted. Therefore, assume without loss of generality that $x_{\ell-1} \geq 0$. From $(L_1) - \gamma_1(L_\ell)$, we obtain

$$-\gamma_1\gamma_\ell x_{\ell-1} + (1-\gamma_1^2)x_1 + \gamma_2 x_2 \ = \ 0\,.$$

Defining $P_0 := 1$ and $P_1 := 1 - \gamma_1^2$, the equation above becomes

$$-\gamma_1\gamma_\ell x_{\ell-1} + P_1 x_1 + \gamma_2 P_0 x_2 \ = \ 0\,. \tag{L_1'}$$

Equation (L_1') relates $x_{\ell-1}$ to x_1 and x_2. By successive applications of the same reasoning we can relate $x_{\ell-1}$ to x_k and x_{k+1} for any k. Indeed, suppose that it holds

$$(-1)^k \gamma_1 \dots \gamma_k \gamma_\ell x_{\ell-1} + P_k x_k + \gamma_{k+1} P_{k-1} x_{k+1} = 0, \tag{L_k'}$$

for some already defined P_{k-1} and P_k. Then, from $P_k(L_{k+1}) - \gamma_{k+1}(L_k')$, we obtain

$$\begin{aligned}
&\big(P_k\gamma_{k+1}x_k + P_k x_{k+1} + P_k \gamma_{k+2} x_{k+2}\big) \\
&\qquad - \big(\gamma_{k+1}(-1)^k\gamma_1 \dots \gamma_k\gamma_\ell x_{\ell-1} + \gamma_{k+1}P_k x_k + \gamma_{k+1}^2 P_{k-1} x_{k+1}\big) \\
&= (-1)^{k+1}\gamma_1 \dots \gamma_{k+1}\gamma_\ell x_{\ell-1} + (P_k - \gamma_{k+1}^2 P_{k-1})x_{k+1} + \gamma_{k+2}P_k x_{k+2} \ = \ 0\,.
\end{aligned}$$

Defining $P_{k+1} \coloneqq P_k - \gamma_{k+1}^2 P_{k-1}$ for $k \leq \ell - 2$, we conclude that (L_{k+1}') also holds. Since we know that (L_1') holds, it follows by induction that $(L_{\ell-2}')$ holds, implying that

$$0 \ = \ (-1)^{\ell-1}\gamma_1 \dots \gamma_\ell x_{\ell-1} + P_{\ell-1}x_{\ell-1} + \gamma_\ell P_{\ell-2} x_\ell\,,$$

that is,

$$\gamma_\ell P_{\ell-2} x_\ell \ = \ \big((-1)^{\ell}\gamma_1 \dots \gamma_\ell - P_{\ell-1}\big)x_{\ell-1}\,.$$

To finish the proof, we need to control the sign of the coefficients appearing above. Once again, the strategy of expressing relevant quantities using the independent events A_k plays a role.

Lemma 22. *Define*

$$P_0 \coloneqq 1, \quad P_1 \coloneqq 1 - \gamma_1^2, \quad P_k \coloneqq P_{k-1} - \gamma_k^2 P_{k-2},\ 2 \leq k \leq \ell - 1,$$

and

$$P_\ell \coloneqq 2(-1)^{\ell}\gamma_1 \cdots \gamma_\ell.$$

Then

$$1 \ = \ P_0 \ > \ P_1 \ > \ P_2 \ > \ \cdots \ > \ P_{\ell-1} > P_\ell \ = \ 2(-1)^{\ell}\gamma_1 \cdots \gamma_\ell \ > \ 0\,. \tag{5.12}$$

Proof. Recall A_k as defined in (5.8) and $B_k = A_k \cap A_{k+1}^c$. We simply notice that the sequence P_k in the statement can be alternatively described by the equation

$$P_k \;=\; 1 - \mathbb{P}\Big(\bigcup_{j=1}^{k} B_j\Big), \quad k \leq \ell - 1. \tag{5.13}$$

Indeed, it is straightforward to check (5.13) for $k = 0, 1$. Now, suppose that (5.13) holds for $k-1$. Then

$$1 - \mathbb{P}\Big(\bigcup_{j=1}^{k} B_j\Big) \;=\; 1 - \mathbb{P}\Big(\bigcup_{j=1}^{k-1} B_j\Big) - \mathbb{P}(B_k) + \mathbb{P}\Big(B_k \cap \bigcup_{j=1}^{k-1} B_j\Big).$$

Since $B_k \cap B_{k-1} = \varnothing$, the intersection above is given by

$$\mathbb{P}\Big(B_k \cap \bigcup_{j=1}^{k-1} B_j\Big) \;=\; \mathbb{P}\Big(B_k \cap \bigcup_{j=1}^{k-2} B_j\Big) \;=\; \mathbb{P}(B_k)\mathbb{P}\Big(\bigcup_{j=1}^{k-2} B_j\Big) \;=\; \gamma_k^2(1 - P_{k-2}),$$

where in the second identity we used that $B_k = A_k \cap A_{k+1}^c$ and $\cup_{j\leq k-1} B_j = \cup_{j \leq k-2}(A_j \cap A_{j+1}^c)$ are mutually independent, because the A_j's are. Putting together the equations above, we obtain

$$1 - \mathbb{P}\Big(\bigcup_{j=1}^{k} B_j\Big) \;=\; P_{k-1} - \gamma_k^2 + \gamma_k^2(1 - P_{k-2}) \;=\; P_{k-1} - \gamma_k^2 P_{k-2} \;=\; P_k\,,$$

completing the induction step. The inequalities in (5.12) are now evident from the fact that the sequence of events $\cup_{j=1}^{k} B_j$ is increasing in k and that we already know $P_\ell = \mathbb{P}(\cap B_j^c) = 2(-1)^\ell \gamma_1 \dots \gamma_\ell > 0$, see (5.11). □

With Lemma 22, we can finish the proof by noticing that

$$\gamma_\ell P_{\ell-2} \;<\; 0 \quad \text{and} \quad (-1)^\ell \gamma_1 \dots \gamma_\ell - P_{\ell-1} \;=\; \frac{1}{2} P_\ell - P_{\ell-1} \;<\; 0\,,$$

which imply that $x_{\ell-1}$ and x_ℓ have the same sign. The reasoning above actually shows that any eigenvector of zero with some positive entry is in $(0,\infty)^\ell$. Finally, we argue that the eigenspace of zero has dimension 1. The Spectral Theorem ensures that Σ has an orthonormal basis of eigenvectors. Now, suppose that v_1, v_2 are two orthogonal eigenvectors of zero. Replacing v_j by $-v_j$ if needed, we can assume $v_j \in (0,\infty)^\ell$ for $j = 1, 2$. But then their inner product is positive, leading to a contradiction. □

The final result of this section is a consequence of the previous proposition, and it is the essential outcome of this section which will be used later.

Theorem 23. *For $\ell \geq 3$, suppose that $X = (X_1, \dots, X_\ell)$ is a centered Gaussian vector with covariance matrix Σ as in (2.17), whose coefficients $\gamma_k = \gamma_k(\infty)$ are of the form (5.6). Then $\mathbb{P}(X_j \geq 0,\ j = 1, \dots, \ell) = 0$.*

Proof. Recall that the support of a Gaussian vector Z is given by $\mathbb{E}(Z) + \mathrm{Ker}(\mathrm{Cov}\,(Z))^\perp$. Thus, in our case the support of X is $\mathrm{Ker}(\Sigma)^\perp$, and by Proposition 21, $\mathrm{Ker}(\Sigma)$ is spanned by a vector $v = (v_1, \dots, v_\ell)$ with $v_j > 0$ for every j. If $y = (y_1, \dots, y_\ell)$ is such that $y_j > 0$ for every j, then we must have $\langle y, v\rangle > 0$, so $y \notin \mathrm{Ker}(\Sigma)^\perp$. Thus,

$$\mathrm{supp}(X) \cap \{y \in \mathbb{R}^d : y_j > 0, j = 1, \dots, \ell\} \;=\; \{0\}\,,$$

and the result follows. □

6. Proofs of Theorems 5 and 6

Theorem 4 will be proved in the next section. In this section we assume its validity in order to prove Theorems 5 and 6.

Proof of Theorem 5. Recall that the mean and variance of the N_k's were computed in Lemma 14, the quantities $f_k = f_k(m)$ and $\mathbf{p}_k = \mathbf{p}_k(m)$ are as in (2.4) and (2.9), and for $k = 1, \dots, \ell$ denote

$$v_k \;=\; v_k(m) \;:=\; \frac{1}{m^{3/2}} \operatorname{Var}(N_k)^{1/2} \;=\; \sigma_k(\infty)(1+o(1)), \quad m \to \infty,$$

so that $\widetilde{N}_k$ from (2.13) reads as

$$\widetilde{N}_k \;=\; \frac{N_k - m^2 f_k f_{k+1} \mathbf{p}_k}{m^{3/2} v_k}, \quad k = 1, \dots, \ell.$$

In an analogous way, and with Lemma 15 in mind, introduce the normalized version $\widetilde{E}_k$ of E_k from (2.7), namely

$$\widetilde{E}_k \;:=\; \frac{E_k - \mathbb{E}(E_k)}{m^{3/2} v_k^{=}}, \quad k = 1, \dots, \ell,$$

with

$$v_k^{=} \;=\; v_k^{=}(m) \;:=\; \frac{1}{m^{3/2}} \operatorname{Var}(E_k)^{1/2} \;=\; o(1), \tag{6.1}$$

where the last identity is valid thanks to Lemma 16. Finally, introduce the events

$$\begin{aligned} A_k \;&:=\; \left\{ \widetilde{N}_k > \frac{f_k f_{k+1} m^2}{m^{3/2} v_k} \left(\frac{1}{2} - \mathbf{p}_k \right) - \frac{1}{2 m^{3/2} v_k} E_k \right\} \\ &=\; \left\{ \widetilde{N}_k > \frac{m^{1/2} f_k f_{k+1}}{v_k} \left(\frac{1}{2} - \mathbf{p}_k - \frac{1}{2} \mathbf{p}_k^{=} \right) - \frac{v_k^{=}}{2 v_k} \widetilde{E}_k \right\}. \end{aligned}$$

These notations were introduced so that the identity (2.8) writes simply as

$$\mathbb{P}(D^{(1)} \triangleright \cdots \triangleright D^{(\ell)} \triangleright D^{(1)}) \;=\; \mathbb{P}(A), \quad \text{where} \quad A \;:=\; \bigcap_{k=1}^{\ell} A_k.$$

If we were to set $\widetilde{E}_k = 0$, then the probability $\mathbb{P}(A)$ would be already suited for a direct application of the Theorem 4. However, in the general case that we are considering here, we need to estimate the possible contributions from the E_k's in a more careful manner.

To that end, let us fix $\varepsilon > 0$ and consider the events

$$B_k(\varepsilon) := \left\{ \frac{v_k^{=}}{2 v_k} |\widetilde{E}_k| > \varepsilon \right\}, \quad k = 1, \dots, \ell, \quad B(\varepsilon) := \bigcup_{k=1}^{\ell} B_k(\varepsilon),$$

$$C_k(\varepsilon) := B_k(\varepsilon)^c = \left\{ \frac{v_k^{=}}{2 v_k} |\widetilde{E}_k| \le \varepsilon \right\}, \quad k = 1, \dots, \ell, \quad C(\varepsilon) := \bigcap_{k=1}^{\ell} C_k(\varepsilon) = B(\varepsilon)^c,$$

and write

$$\mathbb{P}(A) \;=\; \mathbb{P}(A \cap B(\varepsilon)) + \mathbb{P}(A \cap C(\varepsilon)). \tag{6.2}$$

Given any $\varepsilon > 0$, a simple union bound combined with Chebyshev's inequality gives

$$\mathbb{P}(A \cap B(\varepsilon)) \;\le\; \mathbb{P}(B(\varepsilon)) \;\le\; \frac{1}{4\varepsilon^2} \sum_{k=1}^{\ell} \left(\frac{v_k^{=}}{v_k} \right)^2. \tag{6.3}$$

Thanks to (6.1), we thus conclude that

$$\mathbb{P}(A \cap B(\varepsilon)) \overset{m\to\infty}{\longrightarrow} 0\,, \quad \text{for any } \varepsilon > 0 \text{ fixed.} \tag{6.4}$$

To handle the second term in the right-hand side of (6.2), for $t \in \mathbb{R}$ we introduce yet another event $D_k(t)$, namely

$$D_k(t) \;:=\; \left\{ \widetilde{N}_k > \frac{f_k f_{k+1} m^{1/2}}{v_k}\left(\frac{1}{2} - \mathbf{p}_k - \frac{1}{2}\mathbf{p}_{\overline{k}}^{=}\right) - t \right\},\ k = 1,\dots,\ell, \quad D(t) \;:=\; \bigcap_{k=1}^{\ell} D_k(t)\,.$$

From the definition of $A_k, D_k(\varepsilon)$ and $C_k(\varepsilon)$, we obtain that

$$A_k \cap C_k(\varepsilon) \;\subset\; D_k(\varepsilon) \cap C_k(\varepsilon), \quad k = 1, \dots, \ell. \tag{6.5}$$

We now estimate the probability of the events on the right-hand side above. Conditioning, we compute

$$\mathbb{P}(D(\varepsilon)\cap C(\varepsilon)) \;=\; \mathbb{P}(D(\varepsilon)\mid C(\varepsilon))\mathbb{P}(C(\varepsilon)) \;=\; \mathbb{P}(D(\varepsilon)) - \mathbb{P}(D(\varepsilon)\mid C(\varepsilon)^c)\mathbb{P}(C(\varepsilon)^c)\,,$$

and using that $C(\varepsilon)^c = B(\varepsilon)$ and (6.3), we obtain

$$\mathbb{P}(D(\varepsilon)\cap C(\varepsilon)) \;=\; \mathbb{P}(D(\varepsilon)) + o(1)\,, \quad \text{as } m \to \infty, \text{ for any } \varepsilon > 0 \text{ fixed.}$$

Finally, a combination of (6.2), (6.4), the inclusion (6.5) and this last estimate, we obtain that for any $\varepsilon > 0$ fixed,

$$\mathbb{P}(A) \;\leq\; \mathbb{P}(D(\varepsilon)) + o(1), \quad \text{as } m \to \infty.$$

Thus,

$$\limsup_{m\to\infty} \mathbb{P}(A) \;\leq\; \limsup_{m\to\infty} \mathbb{P}(D(\varepsilon)), \quad \text{for any } \varepsilon > 0.$$

But from condition (2.20) and Theorem 4, for any $\varepsilon > 0$, the inequality

$$\begin{aligned}\limsup_{m\to\infty} \mathbb{P}(D(\varepsilon)) \;&\leq\; \limsup_{m\to\infty} \mathbb{P}\left(\widetilde{N}_k \geq -\frac{\delta f_k f_{k+1}}{v_k r(m)} - \varepsilon,\ k = 1,\dots,\ell\right)\\ &\leq\; \mathbb{P}(X_k \geq -\varepsilon,\ k = 1,\dots,\ell)\end{aligned}$$

holds true, and the result follows. □

The proof of Theorem 6 is now a simple consequence of a combination of our results.

Proof of Theorem 6. Under the conditions of Theorem 6, we apply Theorem 23 to conclude that the right-hand side of (2.22) vanishes, and the proof is complete. □

7. Proof of Theorem 4

We now move to the proof of the last standing Theorem 4. So during this section, $\{\mathbf{D}_m\}_m$ is a collection of ℓ independent random dice, each with number of faces $n_k = f_k m$ satisfying Assumption 3. Recall also that the random variables $\widetilde{N}_1(m), \dots, \widetilde{N}_\ell(m)$ were introduced in (2.6) and (2.13), they depend on the index m of the sequence but we keep omitting this dependence and write $\widetilde{N}_k = \widetilde{N}_k(m)$. Likewise, the associated quantities $\mathbf{p}_k = \mathbf{p}_k(m), \mathbf{q}_k = \mathbf{q}_k(m), \mathbf{r}_k = \mathbf{r}_k(m), \mathbf{s}_k = \mathbf{s}_k(m), \sigma_k = \sigma_k(m), \gamma_k = \gamma_k(m)$, $k = 1, \dots, \ell$, were all defined by (2.9)–(2.15); we also omit their dependence on m, and recall that they are instrumental in computing the leading terms in $\mathbb{E}(N_k), \mathrm{Var}\,(N_k)$ and $\mathrm{Corr}\,(N_{k-1}, N_k)$ as in (2.16).

Thanks to Assumption 3 and Lemma 14, we see that

$$\widetilde{N}_k = \frac{N_k - n_k n_{k+1}\mathbf{p}_k}{m^{3/2} v_k} = \frac{N_k - m^2 f_k f_{k+1}\mathbf{p}_k}{m^{3/2} v_k}, \quad k = 1, \dots, \ell, \tag{7.1}$$

with

$$\mathbf{p}_k = \mathbf{p}_k(\infty) + o(1), \quad v_k := \frac{1}{m^{3/2}} \operatorname{Var}(N_k)^{1/2} = \sigma_k(\infty) + o(1), \quad m \to \infty, \quad \sigma_k(\infty) > 0.$$

Our proof of the Central Limit Theorem will be based on the moment method, so for completeness we record here the moments of a general Gaussian random vector. For its statement, recall that

$$n!! = \prod_{k=0}^{\lceil \frac{n}{2}-1 \rceil} (n-2k)$$

is the double factorial of a positive integer n, which is given by the product of all the positive integers up to n that have the same parity as n.

Proposition 24. *Let $X = (X_1, \cdots, X_\ell)^T \sim \mathcal{N}_\ell(0, \mathbf{\Sigma})$ be a centered Gaussian vector of size ℓ and covariance matrix $\mathbf{\Sigma}$ with rank $r \geq 1$. Fix a column vector $\alpha = (\alpha_1, \cdots, \alpha_\ell)^T \in \mathbb{R}^\ell$ for which $\alpha^T \mathbf{\Sigma} \alpha \neq 0$. Then*

$$\mathbb{E}\Big[\Big(\sum_{j=1}^{\ell+1} \alpha_j X_j\Big)^s\Big] = \begin{cases} 0, & \text{if } s \text{ is odd,} \\ (\alpha^T \mathbf{\Sigma} \alpha)^{s/2}(s-1)!!, & \text{if } s \text{ is even.} \end{cases} \tag{7.2}$$

Proof. The proof follows standard textbook arguments, we include it here for sake of completeness. The matrix $\mathbf{\Sigma}$ is positive semi-definite, so it admits a Cholesky decomposition of the form

$$\mathbf{\Sigma} = \boldsymbol{L}\boldsymbol{L}^T,$$

where $\boldsymbol{L}$ is a real matrix of size $(\ell+1) \times r$ and r is the rank of $\mathbf{\Sigma}$. At the level of the random variable X, it induces the identity

$$X = \boldsymbol{L} Z,$$

where $Z \sim \mathcal{N}_r(0, \boldsymbol{I}_r)$ is a normalized Gaussian vector of size r. Now, set

$$G = \frac{1}{(\alpha^T \mathbf{\Sigma} \alpha)^{1/2}} \alpha^T \boldsymbol{L} Z.$$

Observe that $\alpha^T \Sigma \alpha > 0$ so G as above is well defined. In fact, G is a linear combination of independent centered scalar Gaussian random variables, so G is a centered Gaussian itself. Its variance is

$$\mathbb{E}(G^2) = \mathbb{E}(GG^T) = \frac{1}{\alpha^T \mathbf{\Sigma} \alpha} \alpha^T \boldsymbol{L} \mathbb{E}[ZZ^T] \boldsymbol{L}^T \alpha = 1.$$

Hence G is actually a standard Gaussian, so

$$\mathbb{E}(G^s) = \begin{cases} 0, & \text{if } s \text{ is odd,} \\ (s-1)!!, & \text{if } s \text{ is even.} \end{cases}$$

The proof is now completed observing that the term inside the expectation on the left-hand side of (7.2) is $\big(\alpha X \alpha^T\big)^s = (\alpha^T \mathbf{\Sigma} \alpha)^{s/2} G^s$. □

From the Cramér–Wold Criteria, in order to prove Theorem 4 it suffices to show that for any $\alpha = (\alpha_1, \ldots, \alpha_\ell) \in \mathbb{R}^\ell$ we have

$$\sum_{k=1}^{\ell} \alpha_k \widetilde{N}_k \xrightarrow{d} \sum_{k=1}^{\ell} \alpha_k X_k \quad \text{as } m \to \infty,$$

where $X = (X_1, \ldots, X_\ell)^T \sim \mathcal{N}(0, \mathbf{\Sigma})$ with $\mathbf{\Sigma}$ as in Theorem 4.

To prove this, the method of moments will be used (see [3, Theorem 3.12, page 109]), as the normal is a random variable uniquely determined by its moments. Thus, by Proposition 24, we need to show that for each $s \in \mathbb{N}$,

$$\mathbb{E}\Big[\Big(\sum_{k=1}^{\ell} \alpha_k \widetilde{N}_k\Big)^s\Big] \longrightarrow \begin{cases} 0, & \text{if } s \text{ is odd,} \\ (\alpha^T \mathbf{\Sigma} \alpha)^{s/2}(s-1)!!, & \text{if } s \text{ is even,} \end{cases} \tag{7.3}$$

as $m \to \infty$.

The overall strategy we take is the following. The sum inside the expectation can be seen as a weighted sum over all pairs of dice faces that are being compared. We identify each term in this weighted sum with a sum over graphs with appropriate properties. This is done in Section 7.1 below.

Depending on certain properties of these graphs, they can either give an asymptotic neglegible contribution, or contribute to the leading order. In fact, we will show that at the end only graphs with a very particular structure contribute to the leading order of the sum. The second step of the proof consists in pinpointing the neglegible contributions, and also identifying the structure of the graphs that give the leading contribution. This part is done in Section 7.2

The last part of the proof then consists in counting exactly the graphs that give the leading contributions, and this will be done in Section 7.3, which completes the proof of Theorem 4.

7.1. **From moments to combinatorics of graphs.**

We now show how to identify the terms in the sum on the left-hand side of (7.3) with a graph representation.

Using Lemma 14 and the definition of N_k in (2.6), we write

$$\alpha_k \widetilde{N}_k \;=\; \frac{\alpha_k}{\sigma_k m^{3/2}}(1+O(m^{1/2})) \cdot \sum_{i=1}^{n_k} \sum_{j=1}^{n_{k+1}} (\mathbb{1}_{D_i^{(k)} > D_j^{(k+1)}} - \mathbf{p}_k),$$

and therefore

$$\sum_{k=1}^{\ell} \alpha_k \widetilde{N}_k \;=\; m^{-3/2}(1+O(m^{-1/2})) \sum_{k=1}^{\ell} \sum_{i=1}^{n_k} \sum_{j=1}^{n_{k+1}} \frac{\alpha_k}{\sigma_k} (\mathbb{1}_{D_i^{(k)} > D_j^{(k+1)}} - \mathbf{p}_k)\,. \tag{7.4}$$

Raising equation (7.4) to the power s can be seen combinatorially as choosing s indexes (k, i, j) from the triple sum above, multiplying their terms together and finally summing over all possible choices. We now introduce a graph representation of this procedure.

Define

$$\begin{aligned} V &:= \{(k,i); k \in [\ell], i \in [n_k]\}, \\ E &:= \{e = \big((k,i),\, (k+1,j)\big) : k \in [\ell], i \in [n_k], j \in [n_{k+1}]\}\,. \end{aligned} \tag{7.5}$$

The graph $\mathcal{G} = (V, E)$ has vertices representing all faces of all dice and edges e that represent the triples (k, i, j) that appear in equation (7.4). Graph $\mathcal{G}$ already has some structure inherited from the situation it encodes: it is clearly ℓ-partite, with parts $V_k := \{(k, i) : i \in [n_k]\}$ and edges exist only between V_k and V_{k+1}.

Any choice $H = \{(k_t, i_t, j_t) : t \in [s]\}$ of s indices can be seen as an ordered collection of s (possibly repeated) edges of $\mathcal{G}$, and we refer to the set of all possible H as $\mathcal{G}_s$. Any fixed $H \in \mathcal{G}_s$ can be interpreted as a weighted subgraph of $\mathcal{G}$: for each edge $e \in \mathcal{G}$, we assign the weight $w(e) = \#\{t \in [s] : (k_t, i_t, j_t) = e\}$, i.e., its multiplicity. For a graph $H \in \mathcal{G}_s$ introduce $\varphi(H)$ by

$$\varphi(H) \;=\; \prod_{t \in [s]} \frac{\alpha_{k_t}}{\sigma_{k_t}} \big(\mathbb{1}_{D_{i_t}^{(k_t)} > D_{j_t}^{(k_t+1)}} - \mathbf{p}_{k_t}\big). \tag{7.6}$$

When we raise (7.4) to the power s, we re-index the resulting sum on the right-hand side by $H \in \mathcal{G}_s$, and the factor $\varphi(H)$ is precisely the term in this sum that corresponds to a given graph $H \in \mathcal{G}_s$. Taking expectation, we thus obtain

$$\mathbb{E}\Big[\Big(\sum_{k=1}^{\ell} \alpha_k \tilde{N}_k\Big)^s\Big] \;=\; m^{-\frac{3s}{2}}\big(1+O(m^{-\frac{1}{2}})\big) \sum_{H\in\mathcal{G}_s} \mathbb{E}[\varphi(H)]\,. \tag{7.7}$$

Equation (7.7) expresses the expectation we want to compute in terms of a weighted sum over graphs, and the next step is to identify which structure on these graphs lead to leading and negligible asymptotic contributions as $m \to \infty$.

7.2. **Estimating the contributions from each class of graphs.**
The next step is to estimate the terms inside the sum in (7.7). The following claims emphasize some of the main properties that will play a role in our computations.

Claim 25. *Quantity $\varphi(H)$ is uniformly bounded for all $H \in \mathcal{G}_s$.*

Proof. Since σ_k is bounded away from zero as m tends to infinity (see Assumption 3-(ii)) we have

$$|\varphi(H)| \;\le\; \Big(2 \max_{k\in[\ell]} \frac{\alpha_k}{\sigma_k}\Big)^s\,. \qquad \square$$

Let $H \in \mathcal{G}_s$. We say that edges e_0 and $\tilde{e}$ in H are in the same connected component if there is a sequence of edges $(e_j \in H; j \in [t])$ such that e_{j-1} and e_j have a vertex in common for every $j \in [t]$ and $e_t = \tilde{e}$. This forms an equivalence relation and we partition the edges of H into connected components. This is helpful to take advantage of independence when evaluating expected values.

Claim 26. *Suppose $H \in \mathcal{G}_s$ has t connected components $H_1, \ldots, H_t$. Then*

$$\mathbb{E}[\varphi(H)] = \prod_{i\in[t]} \mathbb{E}[\varphi(H_i)].$$

Proof. It is immediate from the definitions, since for $i \neq j$ the random variables $\varphi(H_i)$ and $\varphi(H_j)$ depend on disjoint sets of dice faces. $\square$

Claims 25 and 26 allow us to disregard the contribution of some classes graphs. In the next two claims, we take advantage of the factor $m^{-\frac{3s}{2}}$ to conclude that the contribution of graphs with too few or too many connected components is negligible.

Claim 27. *There are at most $K_1 m^{(3s-1)/2}$ graphs in $\mathcal{G}_s$ with less than $s/2$ connected components, where K_1 does not depend on m.*

Proof. We give an upper bound on the number of graphs in $\mathcal{G}_s$ with t connected components. Define $f_{\max} := \max_{k\in[\ell]} f_k$. The total number of edges in $\mathcal{G}$ is

$$|E| \;=\; \sum_{k\in[\ell]} (f_k m)(f_{k+1} m) \;=\; m^2 \sum_{k\in[\ell]} f_k f_{k+1} \;\le\; (\ell f_{\max}^2) m^2\,. \tag{7.8}$$

To count the number of graphs in $\mathcal{G}_s$, we begin by building such graphs $H \in \mathcal{G}_s$ in a specific ordering. Let H_j with $j \in [t]$ denote the t connected components of a given H. First, we choose one edge e_j from E for each H_j, without any restriction. For these initial choices, we have at most $((\ell f_{\max}^2) m^2)^t$ possibilities. Since H has s edges, we still have to choose $s - t$ edges. For the remaining choices e_j with $j \in [s] \setminus [t]$ we will always choose e_j so that it has some vertex in common with some previously chosen e_i with $i \in [j-1]$, to ensure that we do not create any new connected components. Hence, on the second round of choices, for choosing e_j we have at most $(2(j-1))$ options for the common vertex and at most $2f_{\max}m$ options for the other vertex. Hence, we have at most

$$((\ell f_{\max}^2)m^2)^t\,(2(s-1)2f_{\max}m)^{s-t} \;=\; Km^{t+s}$$

possibilities, where $K = K(\ell, s, t, f_{\max})$ is a positive constant. Finally, observe that any graph $H \in \mathcal{G}_s$ with exactly t connected components can have its edges reordered to a graph $H' \in \mathcal{G}_s$ so that the edges of H' were chosen according to the procedure above. It follows that the number of graphs in $\mathcal{G}_s$ with t connected components is at most $s!Km^{t+s}$. Therefore, there are at most

$$s!K(m^{s+1} + m^{s+2} + \cdots + m^{s+t}) \;\leq\; s!Ktm^{s+t} \;\leq\; \tilde{K}m^{s+\frac{s-1}{2}}$$

graphs in $\mathcal{G}_s$ with less than $s/2$ connected components (as $t < s/2$, then $t \leq (s-1)/2$, because t and s are integer numbers). The positive constant $\tilde{K}$ does not depend on m, and the claim is proved. ☐

Claim 28. *If $H \in \mathcal{G}_s$ has more than $s/2$ connected components, then $\mathbb{E}\left[\varphi(H)\right] = 0$.*

Proof. As there are more than $s/2$ connected components and only s edges, at least one of the components must be an isolated edge, say H_1 is just the edge $(k,i)(k+1,j)$. Then, we have

$$\mathbb{E}[\varphi(H_1)] \;=\; \mathbb{E}\Big[\frac{\alpha_k}{\sigma_k}(\mathbb{1}_{D_i^{(k)} > D_j^{(k+1)}} - \mathbf{p}_k)\Big] \;=\; \frac{\alpha_k}{\sigma_k}\mathbb{E}\big[\mathbb{1}_{D_i^{(k)} > D_j^{(k+1)}} - \mathbf{p}_k\big] \;=\; 0,$$

where the expectation vanishes because of the definition of $\mathbf{p}_k$ in (2.9), and the result follows by Claim 26. ☐

As a consequence of the claims above, we are able to pinpoint the leading order of the s moment in (7.7) by focusing on a very specific class of graphs in $\mathcal{G}_s$. We say that a connected component H_j of a graph $H \in \mathcal{G}_s$ is a **cherry** if it is composed by two distinct edges, and we say that a graph $H \in \mathcal{G}_s$ is a **cherry graph** if all its connected components are cherries. In particular, if $H \in \mathcal{G}_s$ is a cherry graph then s must be even and H must have exactly $s/2$ components.

The vertex of degree 2 in a cherry will be called joint and the other two will be called tips. Let us denote by $\mathcal{C}_s$ the set of all graphs $H \in \mathcal{G}_{2s}$ that are cherry graphs. In words, if $H \in \mathcal{C}_s$ then it has s connected components of size two with no repeating edges.

It turns out that the leading contribution to the right-hand side of (7.7) comes precisely from cherry graphs, as claimed by our next result.

Proposition 29. *For any positive integer s, the estimates*

$$\mathbb{E}\Big[\Big(\sum_{k=1}^{\ell}\alpha_k\widetilde{N}_k\Big)^{2s+1}\Big] \;=\; O(m^{-1/2})\,, \tag{7.9}$$

$$\mathbb{E}\Big[\Big(\sum_{k=1}^{\ell}\alpha_k\widetilde{N}_k\Big)^{2s}\Big] \;=\; m^{-3s}\sum_{H\in\mathcal{C}_s}\mathbb{E}[\varphi(H)] + O(m^{-1})\,. \tag{7.10}$$

hold true as $m \to \infty$.

Proof. Let us estimate the s-moment via equation (7.7). From Claims 25 and 27 we conclude that when estimating the sum in equation (7.7) the contribution of graphs in $\mathcal{G}_s$ with less than $s/2$ connected components is too small when compared to $m^{3s/2}$. By Claim 28, the contribution of graphs with more than $s/2$ is precisely zero. Hence, equation (7.9) is immediate: for the $(2s+1)$-moment we can write

$$\begin{aligned} m^{-\frac{3}{2}(2s+1)} \sum_{H \in \mathcal{G}_{2s+1}} \mathbb{E}[\varphi(H)] &= m^{-\frac{3}{2}(2s+1)} \sum_{1 \le t \le s} \sum_{\substack{H \text{ with } t \text{ connected} \\ \text{components}}} \mathbb{E}[\varphi(H)] \\ &\le m^{-3s-3/2} \cdot K m^{3s+1} = K m^{-1/2} . \end{aligned}$$

When estimating the $2s$-moment, the same argument shows that we only have to worry with the contribution of graphs with exactly s connected components. By Claim 26 if H has some component with only one edge then $\mathbb{E}[\varphi(H)] = 0$. Consequently, for a non-zero contribution, each of the s components must have at least 2 edges. But then we already have $2s$ edges in total, and we conclude that each component has exactly 2 edges. In principle, we can have multiple edges in such graphs. However, another simple counting argument shows that the number of such graphs containing at least one multiple edge is at most Km^{3s-1}. This implies we can focus on the sum over $H \in \mathcal{C}_s$, as claimed in (7.10). □

With Proposition (29) at hand, the remaining step is to estimate the sum over cherry graphs in the right-hand side of (7.10), we will see that this remaining sum is in fact $\Theta(m^{3s})$, so the leading order is indeed given by it, and we will in fact be able to compute its contribution precisely.

7.3. Computing the leading contribution, and the conclusion of the proof of Theorem 4.

What remains is to count all the cherry graphs $H \in \mathcal{C}_s$ and compute $\mathbb{E}[\varphi(H)]$. Since cherries are disjoint, we break any $H \in \mathcal{C}_s$ into s cherries, which we denote $H_1, \dots, H_s$. For a cherry H_j the value of $\mathbb{E}[\varphi(H_j)]$ will depend on which dice are used for its joints and tips. We say that cherry H_j has:

Type $(k,1)$**:** If its joint is on die $D^{(k)}$, one tip is on $D^{(k-1)}$ and the other is on $D^{(k+1)}$.
Type $(k,2)$**:** If its joint is on die $D^{(k)}$, and both tips are on $D^{(k+1)}$.
Type $(k,3)$**:** If its joint is on die $D^{(k)}$, and both tips are on $D^{(k-1)}$.

By the construction of the graph $\mathcal{G}$ in (7.5), these are the only cherries that can occur as components of a graph $H \in \mathcal{C}_s$. It is straightforward to compute $\mathbb{E}[\varphi(H_j)]$.

Proposition 30. *If cherry H_j has type (k,t) then $\mathbb{E}[\varphi(H_j)]$ depends only on (k,t). Denoting its value by $\varphi_{k,t}$, we have*

$$\varphi_{k,t} := \mathbb{E}[\varphi(H_j)] = \begin{cases} \dfrac{\alpha_{k-1}\alpha_k}{\sigma_{k-1}\sigma_k}(\mathbf{s}_k - \mathbf{p}_{k-1}\mathbf{p}_k) & \text{if } t = 1; \\ \left(\dfrac{\alpha_k}{\sigma_k}\right)^2 (\mathbf{r}_k - \mathbf{p}_k^2) & \text{if } t = 2; \\ \left(\dfrac{\alpha_{k-1}}{\sigma_{k-1}}\right)^2 (\mathbf{q}_k - \mathbf{p}_{k-1}^2) & \text{if } t = 3. \end{cases} \tag{7.11}$$

Proof. It is straightforward from the definition of $\varphi(H_j)$ given in (7.6) and the definitions of $\mathbf{p}_k, \mathbf{q}_k, \mathbf{r}_k, \mathbf{s}_k$ in (2.9)–(2.12). □

Since $k \in [\ell]$, we have in total 3ℓ different types of cherries. Recall that the total number of cherries is s. It is useful to classify $H \in \mathcal{C}_s$ with respect to the number of occurrences of each type. We define $M_{k,t} = M_{k,t}(H)$ as the number of cherries of type (k,t) on the cherry graph H, we encode these numbers in the matrix $M = (M_{k,t})_{\ell\times 3}$, and say that H has type M. Observe that $M_{k,t} \in \mathbb{N}$ and $\sum_{k,t} M_{k,t} = s$.

With this codification in mind, for a cherry graph H of type M we have

$$\mathbb{E}[\varphi(H)] \;=\; \prod_{j=1}^{s} \mathbb{E}[\varphi(H_j)] \;=\; \prod_{k,t} \varphi_{k,t}^{M_{k,t}} \,. \tag{7.12}$$

Finally, to estimate the sum over all $H \in \mathcal{C}_s$ we partition the cherry graphs according to the possible types. Let $\mathcal{C}_s(M)$ denote the set of all cherry graphs $H \in \mathcal{C}_s$ of type M. We need estimates on the number of elements of $\mathcal{C}_s(M)$ for each M.

Lemma 31. *For each cherry type (k,t), define*

$$c_{k,t} \;=\; \begin{cases} f_{k-1}f_k f_{k+1} & \text{if } t=1;\\ \frac{1}{2}f_k f_{k+1}^2 & \text{if } t=2;\\ \frac{1}{2}f_k f_{k-1}^2 & \text{if } t=3. \end{cases} \tag{7.13}$$

As n tends to infinity, the size $|\mathcal{C}_s(M)|$ of $\mathcal{C}_s(M)$ satisfies

$$|\mathcal{C}_s(M)| \;=\; m^{3s}\Big[\prod_{k,t}\frac{c_{k,t}^{M_{k,t}}}{M_{k,t}!}\Big]\,((2s)!)\,(1+O(1/m)) \tag{7.14}$$

Since $f_k = f_k(m)$ depends on m, we also have that $c_{k,t} = c_{k,t}(m)$ depends on m, but in virtue of Assumption 3 each $c_{k,t}$ has a nonzero limit as $m \to \infty$.

Proof. We begin by counting the number of cherries of a given specific type, considering that its edges are *not ordered*. Recall that $n_k = f_k m$ denotes the number of faces in the die $D^{(k)}$, and let us define

$$C_{k,t}(n_1,\dots,n_\ell) \;\coloneqq\; \frac{1}{2}|\{H \in \mathcal{G}_2 : H \text{ is a cherry of type } (k,t)\}|\,, \tag{7.15}$$

where the factor $\frac{1}{2}$ is precisely to disregard the order of edges in a cherry. By a simple counting argument, we have that

$$C_{k,t}(n_1,\dots,n_\ell) \;=\; \begin{cases} n_{k-1}n_k n_{k+1} & \text{if } t=1;\\ n_k\dbinom{n_{k+1}}{2} & \text{if } t=2;\\ n_k\dbinom{n_{k-1}}{2} & \text{if } t=3. \end{cases} \tag{7.16}$$

Therefore, the estimate

$$C_{k,t}(n_1,\dots,n_\ell) = c_{k,t}m^3 + O(m^2), \quad \text{as } m \to \infty,$$

is valid, where $c_{k,t}$ are the values in (7.13).

Now, let us fix a type $M = (M_{k,t})$. First, we compute in how many ways we can choose an unordered collection of s cherries with exactly $M_{k,t}$ occurrences of each type (k,t). Given M, we will choose its cherries one by one following the sequence of types $\{(k_j,t_j) : j \in [s]\}$ in lexicographic order.

The first cherry, with type (k_1,t_1), is chosen from all possible edges of $\mathcal{G}$. When choosing the following cherries, we have to successively remove the vertices that appear in the previous cherries, to ensure disjointness. Hence, when choosing the vertices of cherry (k_j,t_j) we have

$C_{k_j,t_j}(n_1^{(j)},\dots,n_\ell^{(j)})$ options, where $n_i^{(j)}$ is the number of faces of die $D^{(i)}$ that do not appear in the $j-1$ previously chosen cherries. It is clear that $(n_i^{(j)})$ will depend on the sequence $\{(k_j,t_j)\}$. However, for our estimates it is enough to notice that since we only choose s cherries, we have $n_i^{(j)} = f_i m + O(1)$. Finally, the above procedure chooses the s cherries following the ordering $\{(k_j,t_j)\}$. Hence, the number of choices of an unordered collection of s cherries is given by

$$\frac{\prod_{j\in[s]}|C_{k_j,t_j}(n_1^{(j)},\dots,n_\ell^{(j)})|}{\prod_{k,t}M_{k,t}!} = \frac{\prod_{j\in[s]}\left(c_{k_j,t_j}m^3+O(m^2)\right)}{\prod_{k,t}M_{k,t}!} = \Big[\prod_{k,t}\frac{c_{k,t}^{M_{k,t}}}{M_{k,t}!}\Big]m^{3s}(1+O(1/m))\,.$$

To conclude the argument, just notice that when summing over $H\in\mathcal{C}_s$ we are actually summing over all fixed unordered collection and considering all possible permutations of the $2s$ edges that compose H. The estimate in equation (7.14) follows. □

Now, we proceed with the estimate in equation (7.10). Breaking the sum on the right-hand side with respect to the type M of the cherry graphs $H\in\mathcal{C}_s$ and using equation (7.12), we have that

$$\begin{aligned}\mathbb{E}\Big[\Big(\sum_{k=1}^{\ell}\alpha_k\widetilde{N}_k\Big)^{2s}\Big] &= m^{-3s}\sum_{H\in\mathcal{C}_s}\mathbb{E}[\varphi(H)]+O(m^{-1})\\ &= m^{-3s}\sum_{M}\sum_{H\in\mathcal{C}_s(M)}\prod_{k,t}\varphi_{k,t}^{M_{k,t}}+O(m^{-1})\\ &= \sum_{M}(2s)!\cdot\Big[\prod_{k,t}\frac{(c_{k,t}\varphi_{k,t})^{M_{k,t}}}{M_{k,t}!}\Big]+O(m^{-1})\,. \end{aligned}\tag{7.17}$$

To obtain a more meaningful expression, we recognize the sum over M as a sum to the s power. Indeed, recall that $M=(M_{k,t})$ is such that $M_{k,t}\in\mathbb{N}$ must sum to s. Hence, we can write

$$\begin{aligned}\sum_{M}(2s)!\cdot\Big[\prod_{k,t}\frac{(c_{k,t}\varphi_{k,t})^{M_{k,t}}}{M_{k,t}!}\Big] &= \frac{(2s)!}{s!}\sum_{(M_{k,t});\sum M_{k,t}=s}\frac{s!}{\prod_{k,t}M_{k,t}!}(c_{k,t}\varphi_{k,t})^{M_{k,t}}\\ &= \frac{(2s)!}{s!}\Big(\sum_{k,t}c_{k,t}\varphi_{k,t}\Big)^s\\ &= (2s-1)!!\Big(\sum_{k,t}2c_{k,t}\varphi_{k,t}\Big)^s, \end{aligned}\tag{7.18}$$

where we used that $(2s-1)!!=\frac{(2s)!}{s!2^s}$.

Using equation (7.17) the next step is to identify the limit of this $2s$-moment. From equations (7.13) and (7.11) we have that

$$2c_{k,t}\varphi_{k,t} = \begin{cases}2f_{k-1}f_kf_{k+1}\frac{\alpha_{k-1}\alpha_k}{\sigma_{k-1}\sigma_k}\cdot(\mathbf{s}_k-\mathbf{p}_{k-1}\mathbf{p}_k) & \text{if } t=1;\\ f_kf_{k+1}^2\big(\frac{\alpha_k}{\sigma_k}\big)^2\cdot(\mathbf{r}_k-\mathbf{p}_k^2) & \text{if } t=2;\\ f_kf_{k-1}^2\big(\frac{\alpha_{k-1}}{\sigma_{k-1}}\big)^2\cdot(\mathbf{q}_k-\mathbf{p}_{k-1}^2) & \text{if } t=3,\end{cases}\tag{7.19}$$

and we can recognize the sum over $(k,t) \in [\ell] \times [3]$ as a quadratic form in the vector $\alpha = (\alpha_1, \dots, \alpha_\ell)^T$. The coefficient of α_k^2 is given by

$$\frac{1}{\sigma_k^2}\Big[f_k f_{k+1}^2(\mathbf{r}_k - \mathbf{p}_k^2) + f_k^2 f_{k+1}(\mathbf{q}_{k+1} - \mathbf{p}_k^2)\Big] \;=\; 1\,,$$

recalling the definition of σ_k in (2.14). The coefficient of $\alpha_{k-1}\alpha_k$ is precisely the value $\gamma_k = \gamma_k(m)$ given by (2.15)

$$\gamma_k \;=\; \frac{1}{\sigma_{k-1}\sigma_k} f_{k-1} f_k f_{k+1}(\mathbf{s}_k - \mathbf{p}_{k-1}\mathbf{p}_k)\,.$$

Writing $\alpha = (\alpha_1, \dots, \alpha_\ell)^T$, and defining

$$\Sigma(m) \;\coloneqq\; \begin{pmatrix} 1 & \gamma_2(m) & 0 & \cdots & 0 & \gamma_1(m) \\ \gamma_2(m) & 1 & \gamma_3(m) & \cdots & 0 & 0 \\ 0 & \gamma_3(m) & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & \gamma_\ell(m) \\ \gamma_1(m) & 0 & 0 & \cdots & \gamma_\ell(m) & 1 \end{pmatrix},$$

with $\gamma_k(m)$ as in (2.15), we just unraveled the identity

$$\sum_{k,t} 2c_{k,t}\varphi_{k,t} = \alpha^T \Sigma(m) \alpha.$$

By Assumption 3, we learn that $\Sigma(m)$ converges to the matrix Σ in (2.17). This last convergence thus shows (7.3), and concludes the proof of Theorem 4.

References

[1] B. M. Brown and T. P. Hettmansperger. Kruskal-Wallis, multiple comparisons and Efron dice. *Aust. N. Z. J. Stat.*, 44(4):427–438, 2002.

[2] B. Conrey, J. Gabbard, K. Grant, A. Liu, and K. E. Morrison. Intransitive dice. *Math. Mag.*, 89(2):133–143, 2016.

[3] R. Durrett. *Probability: Theory and Examples*, volume 31 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 3ª edition, 2010.

[4] W. E. Ferguson, Jr. The construction of Jacobi and periodic Jacobi matrices with prescribed spectra. *Math. Comp.*, 35(152):1203–1220, 1980.

[5] M. Gardner. The paradox of the nontransitive dice. *Sci. Amer.*, 223:110–111, 1970.

[6] J. Hązła, E. Mossel, N. Ross, and G. Zheng. The probability of intransitivity in dice and close elections. *Probab. Theory Related Fields*, 178(3-4):951–1009, 2020.

[7] A. V. Lebedev. The nontransitivity problem for three continuous random variables. *Avtomat. i Telemekh.*, (6):91–103, 2019.

[8] L. G. Molinari. Determinants of block tridiagonal matrices. *Linear algebra and its applications*, 429(8-9):2221–2226, 2008.

[9] K. E. Morrison. Exact probabilities for three and four dice in the balanced uniform model of 3-sided dice. *ArXiv*, 2023. https://arxiv.org/abs/2304.07427.

[10] D. H. J. Polymath. The probability that a random triple of dice is transitive. *ArXiv*, 2022. https://arxiv.org/abs/2211.16156v1.

[11] H. Steinhaus and S. Trybuła. On a paradox in applied probabilities. *Bull. Acad. Polon. Sci.*, 7:67–69, 1959.

[12] P. van Moerbeke. The spectrum of Jacobi matrices. *Invent. Math.*, 37(1):45–81, 1976.

(LC) Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, Brazil
Email address: luisguilhermecoelho@usp.br

(TF) UFBA, Instituto de Matemática, Campus de Ondina, Av. Adhemar de Barros, S/N. CEP 40170-110, Salvador, Brazil
Email address: tertu@ufba.br

(LL) Instituto de Matemática, Estatística e Computação Científica, Universidade Estadual de Campinas, Brazil
Email address: l176809@dac.unicamp.br

(JPa) Instituto de Matemática, Estatística e Computação Científica, Universidade Estadual de Campinas, Brazil
Email address: j246527@dac.unicamp.br

(JP) Instituto de Física de São Carlos, Universidade de São Paulo, Brazil
Email address: joaovictorpimenta@usp.br

(GS) Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, Brazil.
Email address: silvag@usp.br

(DU) Instituto de Matemática, Universidade Federal do Rio de Janeiro, Brazil
Email address: daniel@im.ufrj.br