Maximally Nontransitive Dice

Joe Buhler, Ron Graham, and Al Hales

Abstract. We construct arbitrarily large sets of dice with some remarkable nontransitivity properties. In a sense made precise later, each set exhibits all possible pairwise win/loss relationships by summing different numbers of rolls. The proof of this fact relies on an asymptotic formula for the difference between the median and mean of sums of multiple rolls of dice. This formula is a consequence of a suitable Edgeworth series, an asymptotic refinement of the central limit theorem, but we also give a detailed sketch of a proof in the final section.

1. INTRODUCTION. Nontransitive triples of dice have been known since (at least) 1959 and are, along with many generalizations, a perpetually fascinating topic; see [**5**], [**8**], [**10**], [**12**] [**14**], and the many references therein. Indeed, during the course of writing this paper there were a number interesting new developments, including the Polymath project [**12**]. One of the simplest examples of a nontransitive triple of dice is due to Moser [**11**], and can be written

$$A = [2, 6, 7], \qquad B = [1, 5, 9], \qquad C = [3, 4, 8], \tag{1}$$

where $[x, y, z]$ denotes the die with equally likely values x, y, and z. These dice will be called the "magic-square" dice since their values, suitably ordered, are the rows of a magic square. Note that a physical realization of these dice can be constructed by putting each of the three values on two faces of a cubical die.

The nontransitivity of A, B, and C could be written

$$A > B,\ B > C,\ \text{and } C > A,$$

where $A > B$ means that the probability that a roll of A exceeds a roll of B is strictly larger than the reverse, i.e., $\mathbf{P}(A > B) > \mathbf{P}(B > A)$. Curiously, it turns out that the direction of nontransitivity is reversed if the dice are rolled twice so that

$$A[2] < B[2],\ B[2] < C[2],\ \text{and } C[2] < A[2],$$

where $A[2]$ denotes the sum of two different rolls of A, etc.

It is convenient to make some definitions. For us a *die* X is a bounded integer-valued random variable; this is equivalent to a probability distribution on a finite set of integers. The *dominance indicator* $W(X, Y)$ between dice X and Y is the sign of the difference $\mathbf{P}(X > Y) - \mathbf{P}(Y > X)$ between their respective winning probabilities, i.e.,

$$W(X, Y) := \operatorname{sgn}\left(\mathbf{P}(X > Y) - \mathbf{P}(Y > X)\right) = \operatorname{sgn}\left(\mathbf{P}(X \geq Y) - \mathbf{P}(Y \geq X)\right),$$

where

$$\operatorname{sgn}(x) := \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{if } x = 0 \\ -1 & \text{if } x < 0. \end{cases}$$

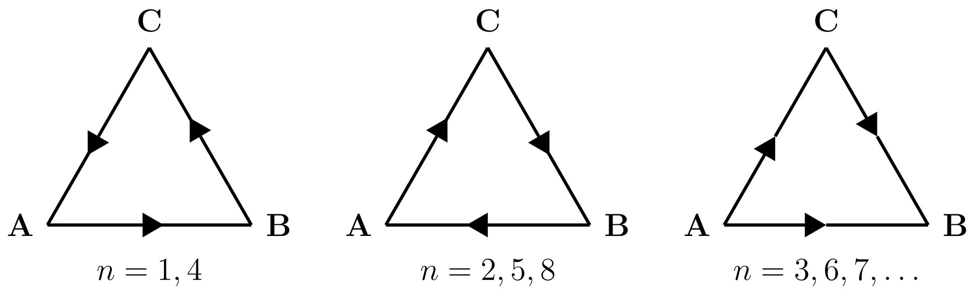

Figure 1. T_n, for all n

Roughly, if $W(X,Y) = 1$ (or -1) then X (or Y) would be the winner in a long run of head-to-head rolls of the dice. If $X_1, \ldots, X_k$ is a set of k dice then the antisymmetric k by k matrix

$$T = [W(X_i, X_j)]_{k\times k}\,,$$

with entries in $\{0, \pm 1\}$, is a *tournament* if all entries off the diagonal are nonzero, i.e., there are no ties between any two of the dice. A tournament gives a choice of a "winner" between each pair of dice, and is sometimes visualized as an orientation of (i.e., a choice of direction on each edge of) the complete graph on the dice. The usual convention is that if X_i dominates X_j then the arrow between i and j points towards j.

If n is a positive integer then $X[n]$ denotes the sum of n different rolls of X. A probabilist would call this a sum of IID (independent and identically distributed) copies of X. If a set of k dice X_i is fixed, then

$$T_n = [W(X_i[n], X_j[n])]_{k\times k}$$

is a tournament if there are no ties between the $X_i[n]$. It might seem that if m and n were close (and large) then T_m and T_n would be very similar tournaments, and even that T_n might be constant for large enough n.

The magic-square dice in (1) give an example since, as it happens, there are no ties between $A[n], B[n], C[n]$, for any $n > 0$. The resulting tournaments T_n are depicted in Figure 1.

The sequence of tournaments has a limiting value, in the sense that $T_n = T_3$ for all $n \geq 9$. However, there are $8 = 2^3$ possible tournaments on a set with three elements, and it is natural to ask whether there are dice that realize all of those tournaments. In fact such sets exist, though it is not entirely trivial to find them; explicit examples will be given later.

However, much more is true: for any positive integer k there are k dice such that the corresponding T_n include all of the $2^{k(k-1)/2}$ possible tournaments, and each tournament occurs for infinitely many n. In particular, for these sets of dice there is no limiting tournament.

Theorem 1. *For all $k > 0$ there is a set of k dice X_i such that for every one of the $2^{k(k-1)/2}$ tournaments T on the X_i there are infinitely many n such*

$$T = [W(X_i[n], X_j[n])]_{k\times k}\,.$$

This wildly oscillating nontransitivity may seem surprising or even shocking, but the variety of examples of nontransitive dice in the literature, as well as [**5**] and [**12**], indicate that nontransitivity might not be all that rare. Perhaps the real surprise is not so much that such a thing could be true, but that it is not particularly hard to prove, as will be seen below.

Various ideas about tournaments have been motivated by dice. It is known that any tournament can be realized by appropriate dice, and "efficient" examples are known for small k (see [**1**], [**2**], [**4**]). The existence of sets of dice for any tournament is of course implied by the above theorem, though a construction of a set that realizes *all* possible T as n varies is not likely to give a particularly efficient construction for a specific T.

2. FACTS ABOUT DICE. It is convenient to recall notation about dice and then state a very special case of the asymptotic refinement of the central limit theorem. This reduces the proof of Theorem 1 to a geometric problem involving hypercubes in $k(k-1)/2$-space. Two further preliminary results follow: a description of how properties of dice behave when taking sums, and an explicit construction of dice with given "shift and span." If X is a die, as above, then there is a finite set of integers x, called the "values" of X, and a probability $p_x > 0$ associated to each x, such that the p_x sum to 1. We say that the probability of a roll of X taking the value x is p_x. The *mean* or *expected value* of X is

$$\mu_1 = \mu_1(X) = \mathbf{E}(X) = \sum_x p_x\, x,$$

where the sum is over all values x of X. The sum $X + Y$ denotes the resulting of rolling X, rolling Y and adding; expectation is additive in the sense that $\mathbf{E}(X+Y) = \mathbf{E}(X) + \mathbf{E}(Y)$. The means of the magic-square dice given earlier are all equal to 5, so in some sense the head-to-head competition between them appears to be fair. On the other hand, if X and Y have distinct means then it can be shown, e.g., using the central limit theorem, that the dominance indicator $W(X[n], Y[n])$ has a fixed value from some n_0 onwards.

If X has mean 0, then we simplify the notation by introducing a single-argument dominance indicator

$$W(X) := \operatorname{sgn}(\mathbf{P}(X > 0) - \mathbf{P}(X < 0)).$$

If A and B are the magic-square dice above, let $X = A - B$, so that the dominance indicator $W(A, B)$ is equal to $W(X)$. The positive values of X are $1, 2, 5, 6$, and the negative values are $-2, -3, -7$, where all values have probability $1/9$ except that 1 and -3 have probability $2/9$. The mean of X is 0, and

$$W(X) = \operatorname{sgn}(5/9 - 4/9) = \operatorname{sgn}(1/9) = 1.$$

This confirms $W(A, B) = W(X)$ as in Figure 1.

The k^{th} moment of X (sometimes called a *central moment* if the mean is 0) is

$$\mu_k = \mu_k(X) = \mathbf{E}(X^k) = \sum p_x\, x^k.$$

The *variance* of a mean-zero random variable X is μ_2, and its *standard deviation* is $\sigma(X) = \sqrt{\mu_2(X)}$.

To avoid trivialities, dice here will always be assumed to have at least two values. Then the *span* of a die X is defined to be the largest integer b such that the values of X are contained in a coset of the set $b\mathbf{Z}$ of integer multiples of b. This means that b is the largest integer such that there exists an a such that all values x are in $a + b\mathbf{Z}$ (or, equivalently, $x \equiv a \bmod b$). The spans of the magic-square dice A and C above are equal to 1, but the span of B is 4 because all three values are congruent to 1 modulo 4, i.e., 1,5, and 9 all lie in $1 + 4\mathbf{Z}$. The integer a is called the *shift* of X, and is only well defined modulo b.

We remind the reader that the gcd g of a set of integers x_i (not all zero) can be defined in several equivalent ways, including the largest integer g such that all x_i lie in the set of integer multiples $g\mathbf{Z}$ of g, or the largest positive integer that divides all of the x_i, or unique positive integer g such that the set of all integer linear combinations of the x_i is equal to $g\mathbf{Z}$. In particular the span b of X as above is the gcd of the set of all $x - a$, where a is the shift and x ranges over all values.

The first theorem in this section is the key ingredient of the proof of Theorem 1. It reduces the proof of whether a set of k dice is maximally nontransitive to a question in $k(k-1)/2$-dimensional geometry.

Theorem 2. *Let X be a die with shift a, span b, and first and third moments equal to 0. Then there is an n_0 such that if $n \geq n_0$ and na is not congruent to 0 or $b/2$ modulo b, then $W(X[n]) := \mathbf{P}(X[n] > 0) - \mathbf{P}(X[n] < 0)$ is nonzero and has the same sign as $1/2 - \{na/b\}$, i.e.,*

$$W(X[n]) = \operatorname{sgn}(1/2 - \{na/b\})$$

where $\{na/b\}$ denotes the fractional part of na/b.

The theorem is equivalent to saying, for $n \geq n_0$, that $W_n := W(X[n])$ has the same sign as $b/2 - (na \bmod b)$ if $na \bmod b$ is neither 0 nor $b/2$. Here $na \bmod b$ denotes the unique integer congruent to na modulo b that lies in the interval $[0, b)$, so that $na \bmod b = b\{na/b\}$.

Note that the sign of W_n for large n, which might be called the "asymptotic dominance indicator," depends only on $na \bmod b$ in the sense that for large enough n, W_n is constant on congruence classes modulo b.

Theorem 2 is an asymptotic refinement of the central limit theorem. It is nontrivial to extract a proof of this from the literature on asymptotic statistics, so in the final section an overview of a proof will be given.

The next preliminary result is a straightforward statement of how moments, shifts, and spans behave when taking the sum of dice.

Theorem 3. **(a)** *If two independent dice have their first and third moments equal to zero, then so does their sum.*
(b) *The span of the sum of two dice is the gcd (greatest common divisor) of their spans, and the shift of a sum is the sum of the shifts.*

Proof. The proof of (a) is an enjoyable calculation that will be left to the reader; note that independence is necessary.

To prove the statement about the span of a sum $X + X'$, where X and X' have shifts a and a' and spans b and b', observe that by the earlier remarks

$$b\mathbf{Z} = \sum_x (x - a)\mathbf{Z}, \qquad b'\mathbf{Z} = \sum_{x'} (x' - a')\mathbf{Z},$$

where x and x' range over all values of X and X' respectively. The span b'' of $X + X'$ is defined by

$$b''\mathbf{Z} = \sum_{x,x'}(x + x' - a - a')\mathbf{Z} = \sum_{x}(x - a)\mathbf{Z} + \sum_{x'}(x' - a')\mathbf{Z} = b\mathbf{Z} + b'\mathbf{Z}.$$

The last sum is $g\mathbf{Z}$, where g is the gcd of b and b', finishing the proof that the span of the sum is the gcd of the spans. The sum $a + a'$ is obviously a shift of $X + X'$. ■

Note that part (b) applies even if the dice are equal, and by induction it follows that $X[n]$ has span b and shift $na \bmod b$.

The final preliminary result gives an explicit construction of the dice that will be useful in the sequel. Curiously, it suffices to consider dice that only have three values.

Theorem 4. *Given integers a and b, with b positive, there is a die X with shift a, span b, and $\mu_1(X) = \mu_3(X) = 0$.*

Proof. Assume, without loss of generality, that $0 \leq a < b$. Set

$$x_1 = a, \quad x_2 = a - 2b, \quad x_3 = a + 3b.$$

These will be the three values of an X with the desired properties.

The vector cross-product $[1,1,1] \times [x_1^2, x_2^2, x_3^2] = [x_3^2 - x_2^2, x_1^2 - x_3^2, x_2^2 - x_1^2]$ is orthogonal to $[1,1,1]$ and $[x_1^2, x_2^2, x_3^2]$. The vector

$$v = \left[\frac{x_3^2 - x_2^2}{x_1}, \frac{x_1^2 - x_3^2}{x_2}, \frac{x_2^2 - x_1^2}{x_3}\right]$$

is therefore orthogonal to both $[x_1, x_2, x_3]$ and $[x_1^3, x_2^3, x_3^3]$, and it is easy to use $x_1^2 < x_2^2 < x_3^2$ to show that all coordinates of v are positive. Normalizing, by dividing by the sum of the coordinates, gives a probability distribution on the three integers x_i which gives an X with the desired properties. Indeed, a is obviously a shift, and the span is the gcd of $-2b = x_2 - a$ and $3b = x_3 - a$. ■

3. CONSTRUCTING DICE.

Theorem 2 suggests that we may be very interested in whether $\{na/b\}$ is above or below $1/2$, where a and b are the shift and span of a die. Moreover, the dominance indicator is unknown for the boundary cases where $\{na/b\}$ is 0 or $1/2$, so we are uninterested in those n.

It is convenient to introduce notation to capture this information by defining

$$S(x) := \begin{cases} 1 & \text{if } 0 < \{x\} < 1/2 \\ 0 & \text{if } \{x\} = 0 \text{ or } 1/2 \\ -1 & \text{if } 1/2 < \{x\} < 1. \end{cases}$$

Thus $S(x)$ encodes which half of the unit interval $[0, 1]$ the fractional part $\{x\}$ lies in. The boundary cases 0 and $1/2$ are outliers that we hope to avoid. If $v \in \mathbf{R}^k$ is a real vector then we let $S_k(v)$ denote the result of applying S to each coordinate, i.e.,

$$S_k((v_1, \ldots, v_k)) := (S(v_1), \ldots, S(v_k)) \in \{-1, 0, 1\}^k \; .$$

Vectors whose coordinates lie in $\{\pm 1\}^k$ are said to be *sign vectors*; there are 2^k such vectors.

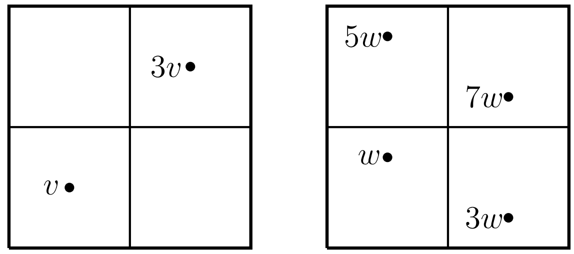

Figure 2. Non-saturation vs. saturation in the plane

Definition. If $v \in \mathbf{R}^k$ is a k-vector than we say that v *saturates* if the set

$$\{S_k(nv) : n \in \mathbf{Z}\}$$

contains all sign vectors.

Consider the plane, i.e., $k = 2$. If v is a nonzero vector in $\mathbf{R}^2$ then the points nv, for integers n, lie on a line through the origin. Define the "color" of a vector v, none of whose coordinates are integers or half-integers, to be $S_2(v)$. Thus the four sign-vectors are the four possible colors for such v. The plane can be tiled by squares of size $1/2$ that are (the closures of) the open squares on which $S_2(v)$ is defined and constant. A vector v in the plane saturates if all four colors can be realized as $S_2(nv)$ as n ranges over all integers. For instance, if $v = (1/4, 1/4)$ then the color is defined for multiples nv of v by odd integers n. (All multiples nv for even n hit the boundaries, so we ignore them.) Odd multiples have the shape $(m + 1/4, m + 1/4)$ or $(m + 3/4, m + 3/4)$ for some integer m. Thus v does *not* saturate in the plane since the colors $S_2(1/4, 3/4)$ and $S_2(3/4, 1/4)$ are not hit. On the other hand if $w = (1/4, 3/8)$ then it is easy to check that, for odd n, $S(nw)$ is one of the four points in Figure 2 and therefore w saturates in $\mathbf{R}^2$.

In k-space, a vector saturates if its multiples nv hit each one of the 2^k "colors" of open half-hypercubes whose closures tile k-space.

Remark. It is a standard fact that the multiples of an irrational number v are dense modulo 1 in the sense that if I is any open subinterval of $[0, 1)$ then there is some integer n such that $\{nv\} = nv \bmod 1$ lies in I. An n-dimensional generalization says that if a vector v in $\mathbf{R}^k$ has irrational entries that are linearly independent over $\mathbf{Q}$, then the images of nv in $\mathbf{R}^k/\mathbf{Z}^k$ are dense in the sense that they intersect every open subset of $[0, 1)^k$. In our case, the entries are rational and the sequence nv is periodic with period (dividing) the lcm (least common multiple) of the denominators of the coordinates. So the sequence of nv can not be dense. However, we are not interested in denseness, but rather merely that the nv hit translates of all of the open half-size hypercubes that tile the unit square.

Before applying these ideas to dice it is convenient to prove a simple criterion for saturation.

Lemma 1. *If a vector v in $\mathbf{R}^k$ has rational coordinates whose denominators (expressed in least terms) are distinct powers of 2, each at least 4, then v saturates.*

Proof. The proof is by induction; the case $k = 1$ is clear.

Assume that vectors in $\mathbf{Q}^k$ satisfying the condition saturate, and let v be an element of $\mathbf{Q}^{k+1}$ where $v_i = a_i/b_i$, in least terms, for $1 \le i \le k+1$, The a_i are odd, and the b_1 are distinct powers of 2, each at least 4. For notational ease we will assume that the v_i are ordered so that the b_i are decreasing; in particular, b_1 is the largest power of 2, and $b_1 \ge 2b_2$.

It suffices to consider the case $a_1 = 1$. Indeed, choose an odd integer r such that $ra_1 \equiv 1 \bmod b_1$, so that $ra_1 = 1 + sb_1$ for some integer s. Since the truth of the lemma for rv implies the truth for v, and $S(nra_1/b_1) = S(n/b_1 + s) = S(n/b_1)$ it suffices to assume that $a_1 = 1$.

The last k coordinates of $v' = (v_2, \ldots v_k)$ satisfy the hypotheses of the lemma and by induction v' saturates in $\mathbf{R}^k$. Therefore, for $0 < n < b_2 \le b_1/2$, the vectors

$$\begin{aligned} S_{k+1}(nv) &= (S(n/b_1), S(nv_2), \ldots, S(nv_{k+1})) \\ &= ((1, S(nv_2), \ldots, S(nv_{k+1})) \end{aligned}$$

hit all sign vectors whose first component is 1. (Note that $S(nv')$ is periodic of period b_2 since all of the denominators of v_i, for $i \ge 2$, are powers of 2 that are at most b_2.) Similarly, for $b_1/2 < n < b_1/2 + b_2$,

$$\begin{aligned} S_{k+1}(nv) &= (S(n/b_1), S(nv_2), \ldots, S(nv_{k+1})) \\ &= (-1, S(nv_2), \ldots, S(nv_{k+1})) \end{aligned}$$

hits all sign vectors whose first component is -1. This finishes the proof of the lemma. ■

Finally, we are ready to apply these ideas to our dice questions. Let $X_1, \ldots, X_k$ be dice with shifts a_i and spans b_i, and with first and third moments equal to 0. The tournament that they determine, when rolled n times, could be written

$$T_n = [W(X_i[n], X_j[n])]_{k\times k} = [W(X_{ij}[n])]_{k\times k}\,,$$

where $X_{ij} = X_j - X_i$. By Theorem 3, each $X_{ij}[n]$ has first and third moments equal to 0, and has span $b_{ij} := \gcd(b_i, b_j)$, and shift $na_{ij} := n(a_i - a_j)$. The statement that $X_i[n]$ dominates $X_j[n]$ is equivalent to $W(X_{ij}[n]) = 1$ which is in turn equivalent, by Theorem 2, to

$$0 < \left\{ \frac{na_{ij}}{b_{ij}} \right\} < \frac{1}{2}$$

for large enough n.

The assertion that all tournaments can be realized as T_n for some n can be converted from realizing all tournaments into turned into a question about saturation. There are $K := k(k-1)/2$ pairs of dice. Define a K-tuple v, indexed by pairs i, j with $1 \le i < j \le k$ by

$$v = (v_{ij}), \qquad v_{ij} = \frac{a_i - a_j}{b_j}, \qquad 1 \le i < j \le K,$$

The dice are maximally nontransitive exactly when this vector saturates in $\mathbf{R}^K$, since this says implies that all tournaments are realized as some T_n.

Thus the proof of Theorem 1 follows from the specification of a suitable set of a_i, b_i. The idea of the proof can perhaps best be visualized by an example for small (but not too small) k.

Let $k = 5$, and $K = 10$, and choose a_i and b_i as in the following table.

$$\begin{array}{llllll} a_i: & 2^0 & 2^1 & 2^2 & 2^3 & 2^4 \\ b_i: & 2^{5+4+3+2+1} & 2^{5+4+3+2} & 2^{5+4+3} & 2^{5+4} & 2^5. \end{array}$$

The 10-tuple $v = [(a_i - a_j)/b_i]$ has components

$$\frac{2^1-2^0}{2^{5+4+3+2}}, \frac{2^2-2^0}{2^{5+4+3}}, \frac{2^2-2^1}{2^{5+4+3}}, \frac{2^3-2^0}{2^{5+4}},$$
$$\frac{2^3-2^1}{2^{5+4}}, \frac{2^3-2^2}{2^{5+4}}, \frac{2^4-2^0}{2^5}, \frac{2^4-2^1}{2^5}, \frac{2^4-2^2}{2^5}, \frac{2^4-2^3}{2^5}$$
$$= \frac{1}{2^{14}}, \frac{3}{2^{12}}, \frac{2}{2^{11}}, \frac{7}{2^9}, \frac{3}{2^8}, \frac{1}{2^7}, \frac{15}{2^5}, \frac{7}{2^4}, \frac{3}{2^3}, \frac{1}{2^2}.$$

The expressions in least terms show that the hypothesis of lemma 1 hold, and therefore that these dice are maximally nontransitive.

Proof of Theorem 1. Fix k and, for notational convenience, index vectors from 0 to $k-1$ rather than 1 to k. Let $t(i) = 1 + 2 + \cdots + i = i(i+1)/2$. Set $a_i = 2^i$, and $b_i = 2^{t(k)-t(i)}$ for $0 \le i < k$. We claim that the K-tuple with coordinates

$$v_{ij} = \frac{2^j - 2^i}{\gcd(2^{t(k)-t(j)}, 2^{t(k)-t(i)})} = \frac{2^{j-i}-1}{2^{t(k)-t(j)-i}}, \qquad 0 \le i < j < k,$$

satisfies the hypotheses of lemma 1.

The smallest denominator is $2^{t(k)-t(k-1)-(k-2)} = 4$. The proof of distinctness is by contradiction. Suppose that two of the denominators are equal, i.e., $t(k) - t(j) - i = t(k) - t(j') - i'$. Then $t(j') - t(j) = i - i'$; if (without loss of generality) $j' > j$, then the left hand side is at least j' and the right hand side is at most i, but this is a contradiction since $j' > j > i$.

By lemma 1 this proves saturation, which shows that $T_n = [W(X_i[n]) - W(X_j[n])]$ realizes all possible tournaments. Since the spans are rational numbers, the tournaments are periodic in n, which proves that each tournament occurs infinitely often.

This finishes the proof of Theorem 1, i.e., the existence of maximally nontransitive dice! ■

Remarks. 1. It would be interesting to allow ties, i.e., to show that all "partial tournaments" could be realized by a suitable set of dice.

2. One measure of the size of a set of k maximally nontransitive dice is the size of the least common multiple N of the spans b_i of the k dice. Not surprisingly, our construction is not optimal. For instance, for $k = 3$, the largest denominator in the above construction is $2^{t(3)-1} = 32$, whereas a computer search shows that there are three maximally nontransitive dice with span 10, and that $N = 10$ is in fact the best possible size for $k = 3$. Similarly, for four dice the smallest N satisfies $64 < N \le 512$ (the upper bound comes from using the construction in the proof.). A computer search shows $n = 68$ is possible, and that this is best

possible. Roughly speaking, these examples suggest that nontransitive sets of dice exist almost as soon as there is room enough for them to exist.

4. A FIRST-ORDER LATTICE EDGEWORTH EXPANSION. The only truly non-elementary fact used in the proof of Theorem 1 is the formula in Theorem 2 giving the sign of the dominance indicator for large n. This formula follows from a very special case of the so-called "Edgeworth expansion at the mean for bounded lattice random variables." It is not easy to extract a proof of this from the literature, and in fact only slightly less nontrivial for a novice to extract a succinct statement of the theorem. We first found a statement implying the formula in [**9**], and there are many sources for general Edgeworth expansions such as [**7**] or [**6**]. The purpose of this final section is to sketch a proof of the theorem, giving all of the steps, leaving out any of the proofs that the supposedly negligible terms are in fact negligible for sufficiently large n.

It is convenient to restrict to random variables with mean 0, and to scale so that the span is 1. From now on, X is a random variable with finitely many values all of the form $x = a + j$, where the shift a is fixed and j ranges over a finite set of integers. The goal is to approximate the cumulative distribution function of an IID sum of n copies of X by an expression of the form

$$\mathbf{P}(X[n] < 0) = \frac{1}{2} + \frac{1}{\sqrt{2\pi n}} \left(\frac{\{na\} - 1/2}{\sigma} - \frac{\mu_3}{6\sigma^3} \right) + o(1/\sqrt{n}). \tag{2}$$

Applying this to $-X$ and subtracting will allow us to determine the sign of

$$\Pr(X[n] > 0) - \Pr(X[n] < 0) = \Pr((-X)[n] < 0) - \Pr(X[n] < 0)$$

for n large enough so that the $o(1/\sqrt{n})$ terms can be ignored.

The probability generating function (PGF) of X is a function of a complex variable z defined by

$$F(z) = \mathbf{E}(z^X) = \sum_x p_x z^x = z^a \sum_j p_{a+j} z^j,$$

where $j = \lfloor x \rfloor$. Note that $z^{-a}F(z)$ is a finite Laurent series:

$$z^{-a}F(z) = \sum_x p_x z^{x-a} = \sum_j p_{a+j} z^j.$$

Applying Cauchy's famous theorem from complex analysis gives

$$p_{a+j} = \mathbf{P}(X = a + j) = [z^j]\, z^{-a}F(z) = \frac{1}{2\pi i} \oint_{C(r)} \frac{z^{-a}F(z)}{z^j} \frac{dz}{z},$$

where $[z^j]z^{-a}F(z)$ denotes the coefficient of z^j in the polynomial $z^{-a}F(z)$, and the contour γ can be chosen to be a counterclockwise circle $C(r)$ of radius r around the origin.

The set of negative values x is the set of $\{a\} + j$ where j ranges over negative integers. Therefore,

$$\mathbf{P}(X < 0) = \sum_{j<0} p_{\{a\}+j} = \frac{1}{2\pi i} \oint_{C(r)} \left(z + z^2 + z^3 \ldots\right) z^{-\{a\}} F(z) \frac{dz}{z}$$

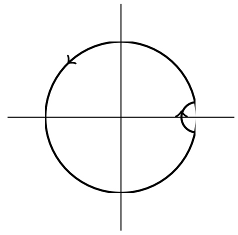

Figure 3. A notched contour

$$= \frac{1}{2\pi i} \oint_{C(r)} \frac{z^{1-\{a\}}\, F(z)}{1-z}\, \frac{dz}{z},$$

where the radius r is taken to be less than 1 to insure that the geometric series converges.

The independence of the summands in $X[n]$ can be used to verify that the PGF of $X[n]$ is $F(z)^n$. Applying the preceding formula to $X[n]$ gives

$$\mathbf{P}(X[n] < 0) = \frac{1}{2\pi i} \oint_{C(r)} \frac{z^{1-\{na\}}\, F(z)^n}{1-z}\, \frac{dz}{z}.$$

Expand the contour outward to the unit circle except for a small semicircular divot centered at, and to the left of, 1, as in Figure 3.

This contour follows the unit circle counterclockwise from $z = e^{i\varepsilon}$ to $z = e^{-i\varepsilon}$ followed by a clockwise small circular arc back to $e^{i\varepsilon}$. For very small ε, the integrand around the notch is close to $-1/(z-1)$, and the contour is basically a small clockwise semicircle; Cauchy's Theorem implies that the value of the integral over the divot is very close to $1/2$. Taking the limit as ε goes to zero gives

$$\mathbf{P}(X[n] < 0) = \frac{1}{2} + \frac{1}{2\pi i} \oint \frac{z^{1-\{na\}} F(z)^n}{1-z}\, \frac{dz}{z},$$

where the contour is the unit circle punctured at $z = 1$, with the "principal value interpretation" at the puncture, which means that the value is the limit, as ε goes to 0, of the counterclockwise integral along the unit circle from $e^{i\varepsilon}$ to $e^{-i\varepsilon}$.

Anticipating a change of variables $z = e^{it}$, let $f(t)$ be the *characteristic function* (CF) of X:

$$f(t) = F(e^{it}) = \mathbf{E}(e^{itX}) = \sum_x p_x\, e^{itx}. \tag{3}$$

Making that change of variables, and doing some algebraic juggling, leads to

$$\mathbf{P}(X[n] < 0) = \frac{1}{2} - \frac{1}{2\pi i} \int_{-\pi}^{\pi} e^{i(1/2-\{an\})t}\, f(t)^n\, \frac{t/2}{\sin(t/2)}\, \frac{dt}{t},$$

where the principal value interpretation is used at $t = 0$, i.e.,

$$\int_{-\pi}^{\pi} := \lim_{\varepsilon \to 0} \left(\int_{-\pi}^{-\varepsilon} + \int_{\varepsilon}^{\pi} \right).$$

Expanding each of the exponential functions in (3) gives

$$f(t) = 1 + \sum_{k=1}^{\infty} \frac{\mu_k (it)^k}{k!} = 1 - \mu_2 t^2 - i\mu_3 t^3/6 + \cdots .$$

This suggests that the largest contribution to the integral will come from a small neighborhood of the origin of size $O(1/\sqrt{n})$, where

$$\begin{aligned} f(t/(\sigma\sqrt{n}))^n &= \exp(n \log(1 - t^2/n - i\mu_3 t^3/(6\sigma^3 n^{3/2} + \cdots) \\ &\simeq \exp(-t^2 - i\mu_3 t^3/(6\sigma^3\sqrt{n}) \cdots). \end{aligned}$$

Transform this integral as follows:

- Make a substitution $t \to t/(\sigma\sqrt{n})$ in the integral, and restrict the integral to an interval of size $O(1/\sqrt{n})$ in which the power series for $\exp(n\log(f(n)))$ converges.
- Ignore all power series terms which are $O(1/n)$.
- Extend the interval of integration to the real line.

Note, for instance, that the power series

$$\frac{t/(2\sigma\sqrt{n})}{\sin(t/(2\sigma)\sqrt{n})} = 1 - \frac{t^2}{48\mu_2 n} + \cdots \simeq 1$$

disappears. The end result of these steps is

$$\mathbf{P}(X[n] < 0) = \frac{1}{2} - \frac{1}{2\pi i} \int_{-\infty}^{\infty} \exp\left(\frac{i(Ct - Dt^3)}{\sqrt{n}}\right) e^{-t^2/2}\,\frac{dt}{t},$$

where

$$C = \frac{1/2 - \{an\}}{\sigma}, \qquad D = \frac{\mu_3}{6\sigma^3}. \tag{4}$$

Expand the imaginary exponential, ignore $O(1/n)$ terms, and use the fact that an odd function integrates to 0 under the principal value interpretation. The result is

$$\begin{aligned} \mathbf{P}(X[n] < 0) &= \frac{1}{2} - \frac{1}{2\pi i\sqrt{n}} \int_{-\infty}^{\infty} \left(1 + iCt - iDt^3\right) e^{-t^2/2}\,\frac{dt}{t} \\ &= \frac{1}{2} + \frac{1}{2\pi\sqrt{n}} \int_{-\infty}^{\infty} (-C + Dt^2)\, e^{-t^2/2}\, dt. \end{aligned}$$

The identities

$$\int_{-\infty}^{\infty} e^{-t^2/2} dt = \int_{-\infty}^{\infty} t^2 e^{-t^2/2} dt = \sqrt{2\pi},$$

are well known; the first is equivalent to the fact that the normal density $\varphi(t) = e^{-t^2/2}/\sqrt{2\pi}$ is a probability density, and the second follows from the first by an easy integration by parts. Applying these gives

$$\mathbf{P}(X[n] < 0) = \frac{1}{2} + \frac{1}{\sqrt{2\pi n}}\,(-C + D) + o(1/\sqrt{n}). \tag{5}$$

Heuristic proof of Theorem 2. Let Y be a die with shift a, span b, and first and third moments equal to 0. Then (5) applies to $X = Y/b$. Note that $W(X[n]) = W(Y[n])$. The only change in the equation in Theorem 2 when X is replaced by $-X$ is that a and μ_3 change signs in the definitions of C and D. Subtract these two expression to get

$$\mathbf{P}((-X)[n] < 0) - \mathbf{P}(X[n] < 0) = \frac{\{-na/b\} - \{na/b\}}{\sigma} + o(1/\sqrt{n}).$$

In Theorem 1 it is assumed that a is not 0 or $b/2$ modulo b, which implies that $\{na/b\}$ is neither 0 nor $1/2$ and that the first term on the right hand side of the above equation is nonzero. If n is large enough so that the $o(1/\sqrt{n})$ term is less than the first term then the sign of the left hand side is the same as the sign of the right hand side. Noting that

$$\{na/b\} - \{-na/b\} = 1 - 2\{na/b\} = 2\ (1/2 - \{na/b\})$$

the statement of Theorem 2 follows immediately. ∎

Remarks. 1. It is not hard to show that if all of the lattice corrections are 0, e.g., if all dice have span 1, then there is an asymptotic dominance relation that is constant from some n_0 onwards. In other words, the "infinitely-often" nontransitivity in the theorem can occur only for dice with nontrivial lattice correction factors, and hence spans $b > 1$. In particular, the differences of the magic-square dice all have span 1, and without further fuss we know that they have a limiting tournament.

2. The only details needed to turn the above argument into a proof are several, perhaps slightly tedious, estimates. In fact, the energetic reader could use the above template to prove a general Edgeworth expansion for lattice variables. This proof is "constructive" in the sense that explicit bounds on the errors can be given, with one exception. The compactness of $[\varepsilon, \pi]$ implies that there is some bound $B = B(\varepsilon)$ such that $|f(t)| \leq B < 1$ on that interval, which implies that high powers of the characteristic function are small. However, making this explicit requires more work, and it seems that this is best done by finding an explicit c such that $|f(t)|^2 < ct^2$ for $|t| < \pi$.

3. The paper [**3**] works through all of these details and gives explicit and (reasonably) effective bounds on the errors in theorem 2. That paper was partially motivated by a question from this paper: when does asymptopia actually arrive for some explicit set of dice?

4. The dice X, Y, and Z with respective PGFs

$$\frac{16t^2 + 7t^{-8} + 2t^{12}}{25}, \quad \frac{11t^{-9} + 33t + 6t^{11}}{50}, \quad \frac{6t^{-11} + 33t^{-1} + 11t^9}{50}$$

have span 10, and are maximally nontransitive. In fact all 8 possible tournaments on $\{X[n], Y[n], Z[n]\}$ occur for $n = 1, 2, 3, 4, 6, 7, 8, 9$ and it is possible to use the results in [**3**] to show that the n_0 in Theorem (2) is $n_0 = 1$, so "asymptopia" arrives instantly.

ACKNOWLEDGMENT. We thank Richard Arratia, Persi Diaconis, Anthony Gamst, Larry Goldstein, Fred Kochman, and Sandy Zabell for interesting and useful conversations.

1. Noga Alon, Graham Brightwell, H.A. Kierstead, A.V. Kostochka, and Peter Winkler, Dominating sets in k-majority tournaments, *J. Combin. Theory Ser B*, **96** (2006) 374–387.
2. Levi Angel, and Matt Davis, A Direct Construction of Non-Transitive Dice Sets, `arXiv:1610.08595`, (2016).
3. J.P. Buhler, A.C. Gamst, A.W. Hales, and Graham, R.L. Graham, Explicit error bounds for lattice Edgeworth expansions, To appear in *Connections in Discrete Mathematics*, Cambridge University Press, 2018.
4. S. Bozóki, S., Nontransitive dice sets realizing the Paley tournaments for solving Schütte's tournament problem, *Miskolc Math. Notes* **15** (2014) 39–50.
5. Brian Conrey, James Gabbard, Katie Grant, Andrew Liu, and Kent E. Morrison, Intransitive Dice, *Mathematics Magazine* **89** (2016) 133–143.
6. Anirban Das Gupta, *Asymptotic Theory of Statistics and Probability*, Springer, New York, 2008.
7. William Feller, *An introduction to probability theory and its applications ,vol. 2*, John Wiley & Sons, Inc., New York-London-Sydney, 1971.
8. Martin Gardner, *Wheels, life and other mathematical amusements*, W. H. Freeman and Co., San Francisco, CA, 1983.
9. B.B. Gnedenko, and A.N. Kolmogorov, B.V. Gnedenko and A.N. Kolmogorov, *Limit distributions for sums of independent random variables*, Revised edition, Addison-Wesley Publishing Co., Reading, Mass. 1968.
10. James Grime, The bizarre world of nontransitive dice: games for two or more players, *College Math. J.* **48** (2017), 2–9.
11. J.W. Moon and L. Moser, Generating oriented graphs by means of team comparisons, *Pacific J. Math.* **21** (1967) 531–535.
12. Polymath 13: Non-transitive Dice, `https://polymathprojects.org` (2017).
13. Alex Schaefer and Jay Schweig, Balanced Non-Transitive Dice, `arXiv.1602.00969` (2016)
14. Allen Schwenk, Beware of Geeks Bearing Grifts, *Math Horizons*, 2000, 10–13.

JOE BUHLER received his PhD. from Harvard University, and taught at Reed College before becoming the Deputy Director of MSRI in 1999, and then Director of the Center for Communications Research in San Diego in 2004, succeeding Al Hales. One of his avocations is juggling, including both the practice thereof, as well as the writing of mathematical papers thereon with Ron Graham.
Center of Communications Research, San Diego, CA 92122
buhler@ccrwest.org

RON GRAHAM received his PhD from UC Berkeley, and worked at AT&T Bell Labs, where he was Director of Information Sciences and then Chief Scientist, until he joined the faculty at UC San Diego in 1999, where he is currently Distinguished Research Professor. He served as MAA President 2003–2004.
University of California, San Diego, CA
graham@ucsd.edu

AL HALES Al Hales received his PhD from Caltech and was a Postdoc at Cambridge University and Harvard University before he joined the UCLA Mathematics Department in 1966. He served as Department Chair at UCLA during 1988-1991, and became the Director of the Center for Communications Research in San Diego in 1992.
Center of Communications Research, San Diego, CA 92122
hales@ccrwest.org