THE PROBABILITY THAT A RANDOM TRIPLE OF DICE IS TRANSITIVE

D. H. J. POLYMATH

Abstract. An *n-sided die* is an n-tuple of positive integers. We say that a die $(a_1, \dots, a_n)$ *beats* a die $(b_1, \dots, b_n)$ if the number of pairs (i, j) such that $a_i > b_j$ is greater than the number of pairs (i, j) such that $a_i < b_j$. We show that for a natural model of random n-sided dice, if A, B and C are three random dice then the probability that A beats C given that A beats B and B beats C is approximately 1/2. In other words, the information that A beats B and B beats C has almost no effect on the probability that A beats C. This proves a statement that was conjectured by Conrey, Gabbard, Grant, Liu and Morrison for a different model.

1. Introduction

It is an amusing fact, first observed by Bradley Efron in the 1960s, that there can be four dice A_0, A_1, A_2, A_3, each with six sides but with non-standard numberings such that if they are all rolled, then each of the four events "A_i shows a higher number than A_{i+1}" occurs with probability 2/3 (where $i + 1$ is interpreted mod 4). Thus, if we say that one die *beats* another if it has a better-than-50% chance of showing a higher number, then the relation "beats" is not transitive.

Much more recently, Conrey, Gabbard, Grant, Liu and Morrison decided to investigate how common a phenomenon intransitivity is. They defined a notion of random n-sided dice, defined a suitable relation "beats" for such dice, and did some computer experiments that indicated, to their surprise, that if A, B and C are three random dice, then the probability that A beats C given that A beats B and B beats C is, for large n, approximately equal to 1/2. That is, the information that A beats B and B beats C gives almost no clue about whether A beats C.

The definition they gave of a random n-sided die, which we shall refer to as the *multiset model*, is as follows: they define an *n-sided die* to be a multiset with n elements that add up to $n(n+1)/2$ (or equivalently average $(n + 1)/2$), and a random n-sided die is simply an n-sided die chosen uniformly at random. An equivalent definition is that a random n-sided die is a random non-decreasing sequence $(a_1, \dots, a_n)$ of positive integers between 1 and n that add up to $n(n + 1)/2$. For example, the 4-sided dice are $(1, 1, 4, 4), (1, 2, 3, 4), (1, 3, 3, 3), (2, 2, 2, 4)$, and $(2, 2, 3, 3)$.

Given two random n-sided dice $A = (a_1, \dots, a_n)$ and $B = (b_1, \dots, b_n)$, we say that A *beats* B if the number of pairs (i, j) such that $a_i > b_j$ is greater than the number of pairs (i, j) such that $a_i < b_j$. For example, the die $A = (1, 1, 4, 4)$ beats the die $B = (1, 3, 3, 3)$ because there are eight

pairs (i, j) with $a_i > b_j$ and only six with $a_i < b_j$. If the two numbers are equal we say that *A ties with B*.

Conrey, Gabbard, Grant, Liu and Morrison made the following two conjectures.

{ties}

Conjecture 1.1. *Let n be a positive integer and let A and B be independent random n-sided dice in the multiset model. Then the probability that A ties with B is* $o(1)$.

transitive}

Conjecture 1.2. *Let n be a positive integer and let A, B and C be independent random n-sided dice in the multiset model. Then the probability that A beats C given that A beats B and B beats C is* $\frac{1}{2} + o(1)$.

They also conjectured a strengthening of Conjecture 1.2, which is the following. Recall that a *tournament* is a complete graph for which every edge is given a direction. We shall regard it as a set T of ordered pairs of distinct elements of a set V such that for any two distinct elements v, w of V exactly one of (v, w) or (w, v) belongs to T.

uasirandom}

Conjecture 1.3. *Let T be a tournament with vertices* $1, 2, \dots, k$. *Then if* $A_1, \dots, A_k$ *are independent random n-sided dice in the multiset model, then the probability that for each* $1 \leq i < j \leq k$ *we have that* A_i *beats* A_j *if and only if* $(i, j) \in T$ *is* $2^{-\binom{k}{2}} + o(1)$.

The conclusion about the tournament T is stating that it is *quasirandom* in a sense introduced by Chung and Graham [**?**]. It turns out to be equivalent to the statement that for all but a fraction $o(1)$ of the pairs of vertices x, y, the fraction of vertices z such that either (x, z) and (y, z) belong to T or (z, x) and (z, y) belong to T is $\frac{1}{2} + o(1)$. There are many different equivalent conditions for quasirandomness: the one conjectured to hold by Conrey, Gabbard, Grant, Liu and Morrison states that all small tournaments occur in T with approximately the frequency one would expect in a random tournament.

Conrey, Gabbard, Grant, Liu and Morrison also looked at other models, and the experimental evidence was surprisingly sensitive to the model chosen, with Conjecture 1.2 (and hence also Conjecture 1.3) appearing to be false for most of them. However, there was one other model for which it seemed to be true, which we shall refer to as the *balanced sequences model*. Here an n-sided die is simply a sequence $(a_1, \dots, a_n)$ of elements of $\{1, 2, \dots, n\}$ that adds up to $n(n+1)/2$ and a random n-sided die is an n-sided die chosen uniformly at random. Note that permuting a sequence does not affect which other sequences it beats. If we say that two dice that are permutations of one another are *equivalent*, then the difference between the balanced sequences model and the multiset model is that the multiset model gives the same weight to each equivalence class, while the balanced sequences model gives the same weight to each individual sequence.

The main results of this paper are that Conjectures 1.1 and 1.2 are true for the balanced sequences model. We also report on experimental evidence that suggests that the stronger conjecture, Conjecture 1.3 is false for both models.

The method of proof can be summarized as follows. We begin by showing that Conjecture 1.2 is equivalent to the statement that almost every random die beats approximately half the other dice and is beaten by approximately half the other dice (and therefore ties with almost no dice). We then argue that unless a die A has a very "atypical" distribution, then it is indeed the case that it beats approximately half the other dice and is beaten by approximately half the other dice. We can regard this last statement as the claim that if $(b_1, \dots, b_n)$ is a random sequence of elements of $\{1, 2, \dots, n\}$, then the probability that it is beaten by A given that it sums to $n(n+1)/2$ is approximately 1/2, as is the probability that it beats A given the same condition. It turns out that this is true provided that a certain sum of independent random variables with values in $\mathbb{Z}$ is sufficiently close to a discrete Gaussian distribution. In order to prove this, we need a rather explicit quantitative local central limit theorem, which we prove by standard Fourier-analytic means, using probabilistic arguments to prove that the behaviour we need the Fourier transform (or characteristic function) to satisfy holds for almost all dice A.

This paper is the result of an open online collaboration between several authors. A complete record of the discussion that led to its existence can be found in a series of five consecutive blog posts and comments on them, of which the first is `https://gowers.wordpress.com/2017/04/28/a-potential-new-polymath-project-intransitive-dice/`. The posts belong to a category entitled polymath13.

2. The preliminary reduction

{reductior

We begin with a lemma about tournaments, or rather about near tournaments, by which we mean directed graphs with n vertices and $(1-o(1))\binom{n}{2}$ edges. Given a triple of vertices (x, y, z), we shall call it *intransitive* if the subgraph induced by the three vertices is a directed cycle of length 3, and *transitive* if it is a triangle but not a directed 3-cycle. The *out-degree* $d_+(x)$ of a vertex x is the number of vertices y such that (x, y) is an edge, and the *in-degree* $d_-(x)$ is the number of vertices y such that (y, x) is an edge.

{equivaler

Lemma 2.1. *Let T be a directed graph with n vertices and $(1-o(1))\binom{n}{2}$ edges. Then the following two statements are equivalent.*

(1) *The probability that a random triple of vertices is intransitive is $\frac{1}{4} + o(1)$.*
(2) *If x is a random vertex, then with probability $1 - o(1)$ $d_+(x) = (\frac{1}{2} + o(1))n$.*

Proof. Write $x \to y$ if (x, y) is an edge of T. First let us count the number of triples (x, y, z) such that $x \to y \to z$. A directed triangle xyz in T gives rise to three such triples, namely $(x, y, z), (y, z, x)$ and (z, x, y). Any other triangle gives rise to just one: for example, if $x \to y$, $x \to z$ and $y \to z$, then the only triple we obtain is (x, y, z). Since the number of triangles is $(1 - o(1))\binom{n}{3}$, we find that the number of triples (x, y, z) such that $x \to y \to z$ is $(1 - o(1))\binom{n}{3}$ plus twice the number of directed triangles. Note that $\binom{n}{3} = (1 + o(1))n^3/6$.

But the number of such triples is also $\sum_y d_+(y)d_-(y)$. Since the number of edges is $(1-o(1))\binom{n}{2}$, this is equal to $(1 + o(1)) \sum_y d_+(y)(n - d_+y)$. Also, $\sum_y d_+(y) = (1 - o(1))\binom{n}{2} = (1 - o(1))n^2/2$, so this is $(1 + o(1))(n^3/2 - \sum_y d_+(y)^2)$. Therefore, twice the number of directed triangles is $(1 + o(1))(n^3/3 - \sum_y d_+(y)^2)$.

If a random triple of vertices has a probability $\frac{1}{4} + o(1)$ of being intransitive, then twice the number of directed triangles is also $(\frac{1}{2} + o(1))\binom{n}{3} = (\frac{1}{12} + o(1))n^3$. It follows that $\sum_y d_+(y)^2 = (\frac{1}{4}+o(1))n^3$, and therefore that $\mathbb{E}_y d_+(y)^2 = (\frac{1}{4}+o(1))n^2$. But $\mathbb{E}_y d_+(y) = (\frac{1}{2}+o(1))n$, so $\mathrm{var}(d_+(y)) = o(n^2)$, which implies that $d_+(y) = (\frac{1}{2} + o(1))n$ with probability $1 - o(1)$.

The steps in the previous paragraph can also be reversed, so the lemma is proved. □

3. A random variable related to an n-sided die and a second reduction

{rvs}

Write $[n]$ for the set $\{1, 2, \dots, n\}$. Given an n-sided die $A = (a_1, \dots, a_n)$ (in fact the definition we are about to give applies to any sequence in $[n]^n$) we define a cumulative distribution function f_A by

$$f_A(j) = |\{i \in [n] : a_i < j\}| + \frac{1}{2}|\{i \in [n] : a_i = j\}|.$$

We typically expect $f_A(j)$ to be around $j - \frac{1}{2}$, so it is convenient also to define a function g_A by $g_A(j) = f_A(j) - j + \frac{1}{2}$. For a fixed A, we shall be interested in the random variable $(g_A(j), j - (n + 1)/2)$, which is defined on $[n]$. More precisely, we choose j uniformly from $[n]$ and evaluate the pair $(g_A(j), j - (n + 1)/2)$.

To see why this is useful to look at, let us do a few simple calculations.

First of all,

$$\begin{aligned}\sum_j f_A(j) &= \sum_j \sum_i \Big(\mathbb{1}_{[a_i<j]} + \frac{1}{2}\mathbb{1}_{[a_i=j]}\Big) \\ &= \sum_i \sum_j \Big(\mathbb{1}_{[a_i<j]} + \frac{1}{2}\mathbb{1}_{[a_i=j]}\Big) \\ &= \sum_i (n - a_i + 1/2) \\ &= n^2/2,\end{aligned}$$

where the last equality follows from the fact that A is an n-sided die and therefore $\sum_i a_i = n(n+1)/2$. This gives us that

$$\sum_j g_A(j) = n^2/2 - \sum_j (j - 1/2) = n^2/2 - n(n+1)/2 + n/2 = 0,$$

and therefore that the mean of the random variable $(g_A(j), j - (n+1)/2)$ is $(0, 0)$.

Next, let $B = (b_1, \dots, b_n)$ be another n-sided die. Then

$$\sum_j f_A(b_j) = \sum_j \sum_i \Big(\mathbb{1}_{[a_i<b_j]} + \frac{1}{2}\mathbb{1}_{[a_i=b_j]}\Big) = \Big|\{(i,j) : a_i < b_j\}\Big| + \frac{1}{2}\Big|\{(i,j) : a_i = b_j\}\Big|.$$

But

$$\sum_j g_A(b_j) = \sum_j (f_A(b_j) - b_j + 1/2) = \sum_j f_A(b_j) - n^2/2,$$

where the last inequality follows from the fact that $\sum_j b_j = n(n+1)/2$. It follows that

$$\sum_j g_A(b_j) = \Big|\{(i,j) : a_i < b_j\}\Big| + \frac{1}{2}\Big|\{(i,j) : a_i = b_j\}\Big| - n^2/2.$$

Since there are n^2 pairs (i, j), this tells us that $\sum_j g_A(b_j) > 0$ if and only if

$$\Big|\{(i,j) : a_i < b_j\}\Big| + \frac{1}{2}\Big|\{(i,j) : a_i = b_j\}\Big| > \Big|\{(i,j) : a_i > b_j\}\Big| + \frac{1}{2}\Big|\{(i,j) : a_i = b_j\}\Big|,$$

which is true if and only if B beats A. Similarly A beats B if and only if $\sum_j g_A(b_j) < 0$.

We will therefore be done if we can prove the following claim.

{main}

Claim 3.1. *If A is a random n-sided die, then with probability $1-o(1)$ we have that the proportion of n-sided dice $B = (b_1, \dots, b_n)$ with $\sum_j g_A(b_j) > 0$ is $\frac{1}{2} + o(1)$.*

The proof that this claim is sufficient requires one small observation. Given a die $A = (a_1, \dots, a_n)$, define the *complementary die* $\overline{A}$ to be the sequence $(n + 1 - a_1, \dots, n + 1 - a_n)$. Then A beats B if and only if $\overline{B}$ beats $\overline{A}$. So if the claim is true, then with probability $1 - o(1)$, the proportion of B such that A beats B is $\frac{1}{2} + o(1)$ and the proportion of B such that $\overline{A}$ beats $\overline{B}$ is also $\frac{1}{2} + o(1)$, which implies that the proportion of B such that A ties with B is $o(1)$.

4. A heuristic argument for Claim 3.1

We begin by explaining why one would expect Claim 3.1 to be true. Once we have done that, we shall turn our heuristic argument into a rigorous one. It is at that point that we shall need to prove a local central limit theorem with sufficiently explicit bounds.

Let $(b_1, \dots, b_n)$ be a purely random sequence belonging to $[n]^n$ – that is, one where the b_i are chosen uniformly and independently from $[n]$ and there is no restriction on the sum. Then to prove Claim 3.1 for a fixed A we need to show that

$$\mathbb{P}\Big[\sum_j g_A(b_j) > 0 \;\Big|\; \sum_j b_j = n(n+1)/2\Big] = \frac{1}{2} + o(1),$$

which is equivalent to the assertion that

$$\mathbb{P}\Big[\sum_j g_A(b_j) > 0 \;\Big|\; \sum_j (b_j - (n+1)/2) = 0\Big] = \frac{1}{2} + o(1).$$

But $b_1, \dots, b_n$ are uniformly and independently chosen from $[n]$. Thus, if we write (X_j, Y_j) for the random variable $(g_A(b_j), b_j - (n+1)/2)$, then $(X_1, Y_1), \dots, (X_n, Y_n)$ are n independent copies of the random variable $(g_A(j), j - (n+1)/2)$ mentioned earlier, and we are concerned with the sum $\sum_{j=1}^n (X_j, Y_j)$, which we shall write as (X, Y).

The central limit theorem suggests that the distribution of this sum will be approximately Gaussian, and since each (X_i, Y_i) has mean $(0, 0)$ we would in particular expect that the distribution would be approximately symmetric about the origin. Also, we would expect a typical value of $g_A(j)$ to have magnitude around $\sqrt{n}$, so the standard deviation of X ought to be around n. Also Y has standard deviation of order $n^{3/2}$ and the two random variables, though correlated, will probably not be too heavily correlated.

If all these heuristics are correct, then the probability that $X = 0$ given that $Y = 0$ should be of order n^{-1}, and certainly $o(1)$. The symmetry should imply that $\mathbb{P}[X > 0 | Y = 0] \approx \mathbb{P}[X < 0 | Y = 0]$, and these statements taken together would give us that $\mathbb{P}[X > 0 | Y = 0] = \frac{1}{2} + o(1)$ and $\mathbb{P}[X < 0 | Y = 0] = \frac{1}{2} + o(1)$, which is equivalent, as we have seen, to the statement that the proportion of dice that beat A is $\frac{1}{2} + o(1)$ and the proportion of dice that A beats is $\frac{1}{2} + o(1)$.

The reason this heuristic argument cannot immediately be turned into a proof is that the central limit theorem is too blunt a tool. There are two reasons for this. The first is that although it tells us that a sum of i.i.d. random variables will converge to a Gaussian, it does not tell us how fast that convergence will occur, and we need it to have occurred (to within a small error) when we take a sum of n copies of $(g_A(j), j)$. And we cannot just let n tend to infinity because the random variables themselves depend on n. This second problem applies not just to the central limit theorem but also to the Berry-Esseen theorem, which gives a rate of convergence in the central limit theorem, but with a constant that (necessarily) depends on the random variable.

A second problem is that the notion of convergence in the central limit theorem and the Berry-Esseen theorem is not suitable for our purposes. We need to be able to estimate the probability that (X, Y) belongs to the positive x-axis, which is a "probability zero event" from the point of view of the central limit theorem and Berry-Esseen theorem. Instead, we need a *local central limit theorem*, the name given to versions of the central limit theorem that can give us estimates for the density function at individual values. Unfortunately, the local central limit theorems that appear in the literature tend still to involve inexplicit constants that depend on the random variable, again necessarily. (We did find an exception to this, but it proved a one-dimensional theorem where we need a two-dimensional one [**?**].)

In the end, we have proved for ourselves a local central limit theorem that is tailored to our application. It is not hard to prove using Fourier analysis, which is one of the standard methods for proving such results, but it requires the random variable to have certain properties, as we shall explain later, in order for us to be able to make the implied constant explicit. So the rest of the proof splits into two parts: first we shall prove that the random variable $(U, V) = (g_A(j), j - (n + 1)/2)$ has certain properties with high probability (when A is a random n-sided die). Then we shall use those properties to establish a suitable local central limit theorem, after which the argument will essentially be finished.

5. Properties of the random variable (U, V)

In this section we shall obtain an upper bound for $\|U\|_\infty$, a lower bound for $\|U\|_2$, and an upper bound on the size of the characteristic function of (U, V). All these bounds will hold with probability $1 - o(1)$ when A is a random n-sided die in the balanced sequences model.

5.1. **An upper bound for $\|U\|_\infty$.** We begin with an almost standard fact (Lemma 5.2 below), but for convenience we provide a complete proof. (The fact and its proof could be thought of as a weakening of a very special case of a one-dimensional local central limit theorem.) First we prove an even more basic lemma.

Lemma 5.1. *Let I_n be the set $\{-(n-1)/2, -(n-3)/2, \dots, (n-3)/2, (n-1)/2\}$ and let f be defined on $\frac{1}{2}\mathbb{Z}$ by taking $f(x) = n^{-1}$ if $x \in I_n$ and $f(x) = 0$ otherwise. (Thus, $f(x) = \mathbb{P}[V = x]$.) Then the k-fold convolution f^{*k} of f is supported on $\mathbb{Z}$ except if k is odd and n is even, in which case it is supported on $\mathbb{Z} + \frac{1}{2}$. In all cases, f^{*k} is an even function, and its non-zero values increase when $x < 0$ and decrease when $x > 0$.*

Proof. The statements about the support and the symmetry are trivial. To prove the increasing and decreasing properties, we note that they follow easily by induction. Indeed, let g be any even function supported on an arithmetic progression of common difference 1 that increases towards the middle, and let $x \geq 0$. Then the inner product of g with $I_n + x$ is greater than or equal to the inner product of g with $I_n + x + 1$, since $g(x - (n-1)/2) \geq g(x + 1 + (n-1)/2)$. Therefore, $g * f$ decreases when x is positive, and by symmetry it increases when x is negative (when we restrict to appropriate supports). □

What we care about here is that when $k = n$, the maximum of f^{*n} is attained at zero. And all we really need from the next lemma is that the probability that $Y = 0$ is not tiny.

{probzero}

Lemma 5.2. *Let $(a_1, \dots, a_n)$ be an element of $[n]^n$ chosen uniformly at random. Then the probability that $\sum_i a_i = n(n+1)/2$ is at least $n^{-3/2}/4$.*

Proof. The probability that $\sum_i a_i = n(n+1)/2$ is $f^{*n}(0)$. Equivalently, it is the probability that $Y = 0$, where Y is the sum of n independent copies of V. The variance of V is at most $n^2/4$, so the variance of Y is at most $n^3/4$. Therefore, by Chebyshev's inequality, the probability that $|Y| \geq n^{3/2}$ is at most $1/4$, which implies that the probability that $|Y| \leq n^{3/2}$ is at least $3/4$. Since 0 is the most likely value of Y, it follows that $Y = 0$ with probability at least $3/8(n^{3/2} + 1) \geq n^{-3/2}/4$. □

We shall now obtain an upper bound for $\|U\|_\infty$. Our method is to obtain an upper bound that holds with such high probability for a purely random element of $[n]^n$ that it continues to hold with high probability even when we condition on the sum being $n(n+1)/2$.

{maxnorm}

Lemma 5.3. *Let A be a random n-sided die. Then with probability $1 - 8n^{-9/2}$ we have that $\max_j |g_A(j)| \leq 6\sqrt{n \log n}$.*

Proof. Let $(a_1, \dots, a_n)$ be a purely random sequence – that is, an element of $[n]^n$ chosen uniformly at random. For each j, let $n_A(j)$ be the number of i such that $a_i \leq j$. Note that $f_A(j)$ is the average of $n_A(j-1)$ and $n_A(j)$.

Now $n_A(j)$ is a sum of n independent Bernoulli random variables of mean j/n. By Chernoff's bounds, the probability that $|n_A(j) - j| \geq m$ is at most $2\exp(-m^2/6n)$. Therefore, the probability that there exists j such that $|n_A(j) - j| \geq m$ is at most $2n\exp(-m^2/6n)$. Setting $m = 6\sqrt{n \log n}$,

this is at most $2\exp(-6\log n) = 2n^{-6}$. By Lemma 5.2, if we now condition on the event that $\sum_i a_i = n(n+1)/2$, then this probability rises to at most $8n^{-9/2}$.

If no such j exists, then for every j we have that

$$|(n_A(j-1) + n_A(j))/2 - (j-1+j)/2| \leq 6\sqrt{n\log n},$$

by the triangle inequality. The left-hand side of this inequality is $|g_A(j)|$. □

5.2. **A lower bound for $\|U\|_2$.** We begin with a lemma that we have not found in the literature precisely as stated. However, it is similar to results such as Hoeffding's inequality and can be proved in essentially the same way, so it has probably been formulated before.

{deviation

Lemma 5.4. *Let $c > 0$, let $X_1, \dots, X_k$ be 01-valued random variables, and suppose that for every i we have that*

$$\mathbb{P}[X_i = 1 | X_1, \dots, X_{i-1}] \leq c.$$

Then $\mathbb{P}[\sum_i X_i \geq (c+\epsilon)k] \leq e^{-\epsilon^2 k/4c}$.

Proof. Let $\lambda > 0$. Then

$$\begin{aligned}
\mathbb{E}(e^{\lambda(X_1+\cdots+X_k)}) &= \mathbb{E}(e^{\lambda(X_1+\cdots+X_{k-1})}\mathbb{E}^{\lambda X_k}) \\
&= \sum_t e^{\lambda t}\mathbb{P}[X_1+\cdots+X_{k-1}=t]\,\mathbb{E}[e^{\lambda X_k}|X_1+\cdots+X_{k-1}=t] \\
&\leq \sum_t e^{\lambda t}\mathbb{P}[X_1+\cdots+X_{k-1}=t](1-c+ce^{\lambda}) \\
&= (1+(e^{\lambda}-1)c)\mathbb{E}e^{\lambda(X_1+\cdots+X_{k-1})}.
\end{aligned}$$

Therefore, by induction,

$$\mathbb{E}(e^{\lambda(X_1+\cdots+X_k)}) \leq (1+(e^{\lambda}-1)c)^k \leq e^{ck(e^{\lambda}-1)}.$$

By Markov's inequality it follows that

$$\mathbb{P}[X_1+\cdots+X_k \geq (c+\epsilon)k] \leq e^{ck(e^{\lambda}-1)}e^{-\lambda(c+\epsilon)k} = e^{-\epsilon\lambda k}e^{ck(e^{\lambda}-\lambda-1)} \leq e^{(-\epsilon\lambda+c\lambda^2)k},$$

where the last inequality uses the fact that $e^{\lambda} \leq 1+\lambda+\lambda^2$, which is valid if $\lambda \leq 1$. Choosing $\lambda = \epsilon/2c$, we obtain an upper bound of $e^{-\epsilon^2 k/4c}$. □

We shall also need to show that certain sums of independent random variables lie in certain ranges with probability bounded away from zero. The Berry-Esseen theorem is sufficient for

this. (More elementary approaches are possible too, but a bit more cumbersome.) The version of the Berry-Esseen theorem we shall use is the following.

{erryesseen}

Theorem 5.5. *Let $X_1, \dots, X_n$ be independent and identically distributed random variables with $\mathbb{E}X_i = 0$, $\mathbb{E}X_i^2 = \sigma^2$ and $\mathbb{E}|X_i|^3 = \rho$ for each i. Let $X = X_1 + \cdots + X_n$, let x be a real number, and let Y be a random variable with the standard normal distribution. Then*

$$\Big|\mathbb{P}[X \le x\sigma\sqrt{n}] - \mathbb{P}[Y \le x]\Big| \le \frac{\rho}{2\sigma^2\sqrt{n}}.$$

Berry and Esseen obtained the same theorem but with a larger absolute constant on the right-hand side. The constant of 1/2 (in fact, slightly better) was obtained by Shevtsova [**?**].

{inrange}

Corollary 5.6. *Let $X_1, \dots, X_n$ and X be as in the previous lemma. Then*

$$\mathbb{P}[\mu n + \sigma\sqrt{n} \le X \le \mu n + 2\sigma\sqrt{n}] \ge \frac{1}{8} - \frac{\rho}{\sigma^2\sqrt{n}}$$

and

$$\mathbb{P}[\mu n - 2\sigma\sqrt{n} \le X \le \mu - \sigma\sqrt{n}] \ge \frac{1}{8} - \frac{\rho}{\sigma^2\sqrt{n}}.$$

Proof. By Theorem 5.5, the first probability differs from $\mathbb{P}[1 \le Y \le 2]$ by at most $\rho/\sigma^2\sqrt{n}$. But a standard normal lies in the interval $[1, 2]$ with probability greater than 1/8, which gives the first estimate. The second is proved in the same way. □

{ell2bound}

Lemma 5.7. *Let A be an element of $[n]^n$ chosen uniformly at random, and let k, m be positive integers with $2km \le n/3$. Then with probability at least $e^{-k/12800}$ we have that $\sum_j g_A(j)^2 \ge km^2/1600$.*

Proof. For $i = 1, 2, \dots, k$ let S_i be the interval $[2(i-1)m + 1, \dots, 2im]$. That is, the S_i are k consecutive intervals, each of length $2m$.

Let E_i be the event that $\sum_{j \in S_i} g_A(j)^2 \ge m^2/40$. The plan is to apply Lemma 5.4, and for this purpose we need to obtain a lower bound for $\mathbb{P}[E_i|E_1, \dots, E_{i-1}]$. Now $E_1, \dots, E_{i-1}$ depend only on the restriction of A to $S_1 \cup \cdots \cup S_{i-1}$ (that is, on the subsequence of A that consists of all values that belong to $S_1 \cup \cdots \cup S_{i-1}$, so it will be enough to condition on that subsequence.

Let $s = 2m(i-1)$ and let r be the number of h with $a_h \le s$. Then for each $j \in [n-s]$ we have that $f_A(s+j) = r + |\{h : s < a_h < s + j\}| + \frac{1}{2}|\{h : a_h = s + j\}|$. Also, once we know r, the rest of A can be thought of as an element of $[s+1, n]^{n-r}$ chosen uniformly at random, which tells us that $|\{h : s < a_h < s + j\}| + \frac{1}{2}|\{h : a_h = s + j\}|$ is a sum of $n - r$ independent random variables $X_1, \dots, X_{n-r}$, each of which takes value 1 with probability $(j-1)/(n-s)$, 1/2 with probability $1/(n-s)$, and 0 with probability $(n-s-j)/(n-s)$. Let $X = X_1 + \cdots + X_{n-r}$.

By hypothesis, $s \le n/3$. Suppose also that $m < j \le 2m$. If $r \ge s + 3m$, then $f_A(s + j) \ge s + 3m$, so

$$g_A(s + j) \ge s + 3m - (s + j) + 1/2 \ge m.$$

That is way better than we need and gives us that the event E_i holds with probability 1. So now let us assume that $r \le n/2$.

We shall now obtain a lower bound for $\mathbb{P}[|g_A(s + j)| \ge \sqrt{m/2}]$ under the same assumption about j.

Let $\mu = (j - 1/2)/(n - s)$, the mean of each X_i. Note first that since $g_A(s + j)$ differs from $f_A(s + j)$ by a constant, the difference between $g_A(s + j)$ and its expectation is the same as the difference between $f_A(s + j)$ and its expectation. Next, note that since $\mu \le 1/2$, the variance σ^2 of each X_i is at least $(j - 1)/4(n - s) \ge m/4n$. Also, $\mathbb{E}|X_i - \mu|^3$ is at most $j/(n - s) + \mu^3 \le 2\mu$.

Suppose now that $\mathbb{E}X \ge s + j - 1/2$. Then by Corollary 5.6 it follows that

$$\mathbb{P}[X \ge s + j - 1/2 + \sigma\sqrt{n - r}] \ge \frac{1}{8} - \frac{\rho}{\sigma^2\sqrt{n - r}} \ge \frac{1}{8} - \frac{8\mu n}{m\sqrt{n - r}} \ge \frac{1}{8} - \frac{50}{\sqrt{n}},$$

where we have used the bounds $\mu \le 3m/n$ and $r \le n/2$.

For sufficiently large n, this implies that $g_A(s + j) \ge \sigma\sqrt{n - r} \ge \sqrt{m/2}$ with probability at least $1/10$. This is true for each j between $m + 1$ and $2m$, so the expected number of j with $g_A(s + j) \ge \sqrt{m/2}$ is at least $m/10$. Since the total number cannot be greater than m, it follows that the probability that there are at least $m/20$ values of j with $g_A(s+j) \ge \sqrt{m/2}$ is at least $1/20$. Therefore with probability at least $1/20$ we find that $\sum_{j=m+1}^{2m} g_A(s + j) \ge m^2/40$.

If $\mathbb{E}X \le s + j - 1/2$, then we can run the same argument but this time using a lower bound for the probability that $X \le s + j - 1/2 - \sigma\sqrt{n - r}$, and we obtain the same conclusion.

We now apply Lemma 5.4 to the random variables Y_i, where $Y_i = 1$ if E_i does not hold and 0 otherwise. Our argument so far proves that $\mathbb{P}[Y_i = 1|Y_1, \dots, Y_{i-1}] \ge 1/20$. If we take $c = 19/20$ and $\epsilon = 1/40$, then we find that the probability that E_i holds for at least $39k/40$ values of i is at most $e^{-k/12800}$. Therefore, with probability at least $1 - e^{-k/12800}$ we have that E_i holds for at least $k/40$ values of i, and when that is the case, we have that $\sum_j g_A(j)^2 \ge km^2/1600$. □

The larger the value of k, the weaker the lower bound we obtain for $\sum_A g_A(j)^2$ but the more likely it is to hold. When we apply the above lemma, we shall want the probability to be bounded above by a negative power of n, so k will be logarithmic and the lower bound will be of order $n/\log n$.

5.3. **Bounding the magnitude of the characteristic function away from 1.** As with many proofs of central-limit-type theorems, we shall use characteristic functions, or equivalently Fourier

analysis. Recall that the characteristic function of (U, V) is the function

$$\hat{f}(\alpha,\beta) = \mathbb{E}e(\alpha U + \beta V),$$

where $e(x)$ is shorthand for $\exp(2\pi i x)$. We shall say more about the importance of the characteristic function later, but for now we simply ask the reader to take it on trust that it will be useful to us to bound $|\hat{f}(\alpha,\beta)|$ away from 1, except when α and β are very small. (Note that when $\alpha = \beta = 0$ then $\hat{f}(\alpha,\beta) = 1$, so some dependence is necessary.)

Our argument will be fairly similar to the the proof in the previous section, but we need a small modification of Lemma 5.4. Roughly speaking, Lemma 5.4 works fine except if certain events of very low probability occur, so we now prove a variant of the lemma that shows that low-probability events do not mess up the conclusion by too much.

deviation2}

Lemma 5.8. *Let $0 < \epsilon < c$ let $0 < \delta \leq \epsilon^2 k^{-1} e^{-\lambda k}/16c$, let $X_1, \dots, X_k$ be 01-valued random variables, let $D_1, \dots, D_{k-1}$ be events of probability at least $1 - \delta$, and suppose that for every i we have that*

$$\mathbb{P}[X_i = 1 | X_1, \dots, X_{i-1}, D_{i-1}] \leq c.$$

Then $\mathbb{P}[\sum_i X_i \geq (c+\epsilon)k] \leq e^{-\epsilon^2 k/8c}$.

Proof. Let $\lambda > 0$. Then as in the proof of Lemma 5.4, we have that

$$\begin{aligned}\mathbb{E}(e^{\lambda(X_1+\cdots+X_k)}) &= \mathbb{E}(e^{\lambda(X_1+\cdots+X_{k-1})}\mathbb{E}^{\lambda X_k}) \\ &\leq \sum_t e^{\lambda t}\mathbb{P}[X_1 + \cdots + X_{k-1} = t]\,\mathbb{E}[e^{\lambda X_k} | X_1 + \cdots + X_{k-1} = t].\end{aligned}$$

Now

$$\begin{aligned}\mathbb{E}[e^{\lambda X_k} | X_1 + \cdots + X_{k-1} = t] &\leq \mathbb{E}[e^{\lambda X_k} | X_1 + \cdots + X_{k-1} = t, D_{k-1}] + e^{\lambda}\mathbb{P}[\neg D_{k-1} | X_1 + \cdots + X_{k-1} = t] \\ &\leq \mathbb{E}[e^{\lambda X_k} | X_1 + \cdots + X_{k-1} = t, D_{k-1}] + \frac{e^{\lambda}\mathbb{P}[\neg D_{k-1}]}{\mathbb{P}[X_1 + \cdots + X_{k-1} = t]},\end{aligned}$$

so the expression we are trying to bound is at most the sum of

$$\sum_t e^{\lambda t}\mathbb{P}[X_1 + \cdots + X_{k-1} = t]\,\mathbb{E}[e^{\lambda X_k} | X_1 + \cdots + X_{k-1} = t, D_{k-1}]$$

and

$$e^{\lambda}\mathbb{P}[\neg D_{k-1}] \sum_t e^{\lambda t}.$$

The second term we can bound above crudely by $\delta \sum_{t=0}^{k} e^{\lambda(t+1)} \leq 2\delta k e^{\lambda k}$, while the first is at most $(1+(e^\lambda-1)c)\mathbb{E}e^{\lambda(X_1+\cdots+X_{k-1})}$ as in the previous argument. Writing θ for $2\delta k e^{\lambda k}$, we therefore obtain an upper bound of

$$(1+(e^\lambda-1)c)\mathbb{E}e^{\lambda(X_1+\cdots+X_{k-1})}+\theta \leq (1+(e^\lambda-1)c+\theta)\mathbb{E}e^{\lambda(X_1+\cdots+X_{k-1})}.$$

Since θ increases with k (when $\lambda > 0$, which it will be), this gives us an upper bound for $\mathbb{E}e^{\lambda(X_1+\cdots+X_k)}$ of $(1+(e^\lambda-1)c+\theta)^k \leq e^{k((e^\lambda-1)c+\theta)}$. By Markov's inequality and the inequality $e^\lambda \leq 1+\lambda+\lambda^2$ again, the probability that $X_1+\cdots+X_k \geq c+\epsilon$ is at most $e^{k(\lambda^2 c-\lambda\epsilon+\theta)}$. Setting $\lambda = \epsilon/2c$ and noting that $\theta \leq \epsilon^2/8c$, we obtain the result. □

We shall also make use of the following small technicality.

{quadruple

Lemma 5.9. *Let $\theta_1,\theta_2,\theta_3$ and θ_4 and ϵ be real numbers such that the distance from $\theta_1-\theta_2-\theta_3+\theta_4$ to the nearest integer is at least ϵ. Then $|e(\theta_1)+e(\theta_2)+e(\theta_3)+e(\theta_4)| \leq 4-\epsilon^2$.*

Proof. First note that for any θ and ϕ with $|\theta-\phi| \leq 1/2$ we have the inequality

$$|e(\theta)+e(\phi)|^2 = 2+2\cos(2\pi(\theta-\phi)) \leq 4-(2\pi)^2(\theta-\phi)^2+(2\pi)^4(\theta-\phi)^4/12 \leq 4(1-(\theta-\phi)^2)^2,$$

and therefore

$$|e(\theta)+e(\phi)| \leq 2(1-(\theta-\phi)^2).$$

Since adding an integer makes no difference, we may assume that $|\theta_1-\theta_2|$ and $|\theta_3-\theta_4|$ are both at most $1/2$. It follows that

$$|e(\theta_1)+e(\theta_2)+e(\theta_3)+e(\theta_4)| \leq 4-2(\theta_1-\theta_2)^2-2(\theta_3-\theta_4)^2 \leq 4-(\theta_1-\theta_2-\theta_3+\theta_4)^2,$$

which proves the result, since $|\theta_1-\theta_2-\theta_3+\theta_4| \geq \epsilon$. □

The main result of this subsection is the following.

{fcbound}

Lemma 5.10. *There is an absolute constant $c>0$ with the following property. Let $m \leq cn/\log n$. Then with probability at least $1-n^{-10}$, $|\hat{f}(\alpha,\beta)| \leq 1-\alpha^2 m/960000$ for every α such that $\alpha^2 m \leq 1/100$.*

Proof. Since $(U,V) = (g_A(j), j)$ for a randomly chosen $j \in [n]$, it follows that

$$\hat{f}(\alpha,\beta) = n^{-1}\sum_{j=1}^{n} e(\alpha g_A(j)+\beta j).$$

Note that if $j_1 - j_2 = j_3 - j_4$ and $\epsilon \leq |\alpha(g_A(j_1) - g_A(j_2) - g_A(j_3) + g_A(j_4))| \leq 1/2$, then for every β we have that

$$|e(\alpha g_A(j_1) + \beta j_1) + e(\alpha g_A(j_2) + \beta j_2) + e(\alpha g_A(j_3) + \beta j_3) + e(\alpha g_A(j_4) + \beta j_4)| \leq 1 - \epsilon^2/4.$$

This follows by setting $\theta_i = \alpha g_A(j_i) + \beta j_i$ and observing that since $\beta(j_1 - j_2 - j_3 + j_4) = 0$, the distance from $\theta_1 - \theta_2 - \theta_3 + \theta_4$ to the nearest integer is the same as the distance from $\alpha(g_A(j_1) - g_A(j_2) - g_A(j_3) + g_A(j_4))$ to the nearest integer, which by hypothesis is at least ϵ. This allows us to apply Lemma 5.9.

Our strategy now will be to prove that with high probability there are many non-overlapping quadruples (j_1, j_2, j_3, j_4) with $j_1 - j_2 = j_3 - j_4$ and $c\alpha\sqrt{m} \leq |\alpha(g_A(j_1) - g_A(j_2) - g_A(j_3) + g_A(j_4))| \leq 1/2$.

To do this, we first pick k maximal such that $4km \leq n/2$ and we let $S_1, \dots, S_k$ be consecutive intervals of length $4m$: that is, $S_i = [4(i-1)m + 1, \dots, 4im]$. For each i, let E_i be the event that $(4m)^{-1}|\sum_{j \in S_i} e(\alpha g_A(j) + \beta j)| \leq 1 - \alpha^2 m/4000$. We shall apply Lemma 5.8 to these events (that is, with $X_i = 1$ if E_i holds and 0 otherwise), and shall take D_i being the event that the number of h such that $a_h \in S_1 \cup \dots \cup S_i$ is at most $3n/4$. Since $|S_1 \cup \dots \cup S_i| \leq n/2$, D_i has probability at least $1 - 2e^{-n/8}$, by Hoeffding's inequality.

It remains to obtain a lower bound for the probability $\mathbb{P}[E_i | E_1, \dots, E_{i-1}, D_{i-1}]$. We shall do this as follows. Let $s = 4m(i-1)$ and let r be the number of h with $a_h \leq s$. Since D_{i-1} holds, $r \leq 3n/4$.

Let t be the number of h such that $a_h \leq s + j + 3m$. If we condition on r, then the expected value of t is $r + (j + 3m)(n - r)/(n - s) \leq r + 8m$. Since $t \geq r$, it follows from Markov's inequality that $t \leq r + 16m \leq 4n/5$ with probability at least $1/2$.

Let us condition on the value of t and suppose that this inequality holds. Then

$$f_A(s + j + 4m) = t + |\{h : s + j + 3m < a_h < s + j + 4m\}| + \frac{1}{2}|\{h : a_h = s + j + 4m\}|,$$

which is a sum of $n - t$ independent random variables $X_1, \dots, X_{n-t}$, each of which takes the value 1 with probability $(m-1)/(n - s - j - 3m)$, $1/2$ with probability $1/(n - s - j - 3m)$, and 0 with probability $(n - s - j - 4m)/(n - s - j - 3m)$.

As in the proof of Lemma 5.7, we shall now apply Corollary 5.6 to these random variables. Back-of-envelope calculations similar to those of Lemma 5.7 give the bounds $\rho \leq 4m/n$ and $\sigma^2 \geq m/4n$. Therefore, writing X for $X_1 + \dots + X_{n-t}$, we have that

$$\mathbb{P}[X \geq M + \sigma\sqrt{n - t}] \geq \frac{1}{8} - \frac{\rho}{\sigma^2\sqrt{n - t}}$$

and

$$\mathbb{P}[X \le M - \sigma\sqrt{n-t}] \ge \frac{1}{8} - \frac{\rho}{\sigma^2\sqrt{n-t}}.$$

Note that $\rho/(\sigma^2\sqrt{n-t}) \le 50/\sqrt{n}$ (since $t \le 4n/5$), so for n sufficiently large, both these probabilities are at least $1/10$.

Since $f_A(s+j) - g_A(s+j) = s + j - 1/2$ for every $j \in [4m]$, we have

$$g_A(s+j)-g_A(s+j+m)-g_A(s+j+2m)+g_A(s+j+3m) = f_A(s+j)-f_A(s+j+m)-f_A(s+j+2m)+f_A(s+j+3m)$$

for every $j \in [m]$.

Therefore, if $\alpha(f_A(s+j)-f_A(s+j+m)-f_A(s+j+2m)) = \theta$, then there is a probability of at least $1/10$ that $\alpha(g_A(s+j)-g_A(s+j+m)-g_A(s+j+2m)+g_A(s+j+3m)) \in [\theta+\alpha\sigma\sqrt{n-t}, \theta+2\alpha\sigma\sqrt{n-t}]$ and also a probability of at least $1/10$ that $\alpha(g_A(s+j)-g_A(s+j+m)-g_A(s+j+2m)+g_A(s+j+3m)) \in [\theta - 2\alpha\sigma\sqrt{n-t}, \theta - \alpha\sigma\sqrt{n-t}]$.

As we have shown, the probability that $t \le 4n/5$ is at least $1/2$. Assuming that this is the case, we have that $\sqrt{m}/6 \le \sigma\sqrt{n-t} \le 2\sqrt{m}$. Because $\alpha\sqrt{m} \le 1/10$, it follows in particular that $\alpha\sigma\sqrt{n-t} \le 1/5$.

From this it follows that with probability at least $1/20$, the distance from $\alpha(g_A(s+j) - g_A(s+j+m) - g_A(s+j+2m) + g_A(s+j+3m))$ to the nearest integer equals its modulus, which is at least $\alpha\sqrt{m/6}$. And if that holds, then by Lemma 5.9 it follows that

$$(1/4)|e(\alpha g_A(j_1)+\beta j_1) + e(\alpha g_A(j_2)+\beta j_2) + e(\alpha g_A(j_3)+\beta j_3) + e(\alpha g_A(j_4)+\beta j_4)| \le 1 - \alpha^2 m/24$$

when we take $j_1 = s + j$, $j_2 = s + j + m$, $j_3 = s + j + 2m$ and $j_4 = s + j + 3m$.

Therefore, the expected number of $j \in [m]$ for which the above upper bound holds is at least $m/20$. It follows that with probability at least $1/40$ it holds for at least $m/40$ values of j. And in that case the event E_i holds, by the triangle inequality.

Thus, the probability that E_i holds, given $E_1, \dots, E_{i-1}$ and D_{i-1} is at least $1/40$. Lemma 5.8 now implies that the probability that E_i holds for fewer than $k/80$ values of i is at most $e^{-k/80^2.8} = e^{-k/51200}$. If E_i holds for at least $k/80$ values of i, then since $4km \ge n/3$, we have that $n^{-1}|\sum_{j=1}^{n} e(\alpha g_A(j) + \beta j)| \le 1 - \alpha^2 m/960000$, as claimed. □

{smallouts

Corollary 5.11. *Let A be a random element of $[n]^n$ and let $\hat{f}$ be the characteristic function of the random variable (U, V), which is uniformly distributed on the set $\{(g_A(j), j-(n+1)/2) : j \in [n]\}$. Then with probability at least $1 - 3n^{-6}$ we have the bound*

$$|\hat{f}(\alpha,\beta)|^n \le n^{-10}$$

for every $(\alpha,\beta) \in [-1/2, 1/2]^2$ such that either $|\alpha| \ge 10^5\sqrt{\log n}/n$ or $|\beta| \ge 2^{23}\sqrt{\log n}/n^{3/2}$.

Proof. By Lemma 5.10 we have with probability at last $1 - n^{-9}$ that $|\hat{f}(\alpha,\beta)| \leq 1 - \alpha^2 m/960000$ for every α,β and every $m \leq n/20\log n$ such that $\alpha^2 m \leq 1/100$. If $|\alpha| \geq 100000\sqrt{\log n}/n$, then we can pick $m \leq n/20\log n$ such that $\alpha^2 m/960000 \geq 20\log n/n$, giving us that $|\hat{f}(\alpha,\beta)| \leq 1 - 20\log n/n$. But then $|\hat{f}(\alpha,\beta)|^n \leq (1 - 20\log n/n) \leq n^{-10}$, as claimed.

Now suppose that $|\alpha| \leq 100000\sqrt{\log n}/n$. By the proof of Lemma 5.3 we have that every $g_A(j)$ has absolute value at most $6\sqrt{n\log n}$.

Suppose first that $|\beta| \leq n^{-1}/2$ and let $m = \lfloor n/2 \rfloor$. For each $j \leq m$ let us obtain an upper bound for $|e(g_A(j)\alpha + j\beta) + e(g_A(j+m)\alpha + (j+m)\beta)|$. By our bound on $|\alpha|$, we have that $|g_A(j)\alpha|$ and $|g_A(j+m)\alpha|$ are both at most $600000\log n/\sqrt{n}$. It follows that

$$|g_A(j)\alpha + j\beta - (g_A(j+m)\alpha + (j+m)\beta)| \geq n|\beta|/3 - 1200000\log n/\sqrt{n}.$$

Our lower bound for $|\beta|$ implies that this is at least $n|\beta|/6$. We also have that

$$|g_A(j)\alpha + j\beta - (g_A(j+m)\alpha + (j+m)\beta)| \leq n|\beta|/2 + 120000\log n/\sqrt{n} \leq n|\beta| \leq 1/2.$$

The proof of Lemma 5.9 included the inequality that $|e(\theta) + e(\phi)| \leq 2(1 - (\theta-\phi)^2)$ when $|\theta - \phi| \leq 1/2$, so using this and the two bounds just noted, we find that

$$|e(g_A(j)\alpha + j\beta) + e(g_A(j+m)\alpha + (j+m)\beta)| \leq 2(1 - n^2\beta^2/36),$$

from which it follows that $|\hat{f}(\alpha,\beta)| \leq 1 - n^2\beta^2/40$ when n is sufficiently large. (The slight worsening of the absolute constant is to allow for the fact that n may be odd, in which case we do not get an exact partition of $[n]$ into pairs $\{j, j+m\}$.)

Using our lower bound on $|\beta|$ again, we find that $n^2\beta^2/40 \geq 2^{40}\log n/n$, from which it follows readily that $|\hat{f}(\alpha,\beta)|^n \leq n^{-10}$.

If $1/4 \geq |\beta| \geq n^{-1}/2$, then the proof is similar, but this time we choose m maximal such that $m|\beta| \leq 1/4$. Note that $m \leq n/2$ if we do this. It follows that we can partition $[n]$ into at least $n/3$ disjoint pairs $\{j, j+m\}$. For each such pair we have that

$$|g_A(j)\alpha + j\beta - (g_A(j+m)\alpha + (j+m)\beta)| \geq m|\beta| - 120000\log n/\sqrt{n} \geq 1/5$$

and also

$$|g_A(j)\alpha + j\beta - (g_A(j+m)\alpha + (j+m)\beta)| \leq m|\beta| + 120000\log n/\sqrt{n} \leq 1/3,$$

when n is sufficiently large. It follows that

$$|e(g_A(j)\alpha + j\beta) + e(g_A(j+m)\alpha + (j+m)\beta)| \leq 2(1 - 1/25),$$

and hence that $|\hat{f}(\alpha,\beta)| \leq 74/75$, which for sufficiently large n gives us that $|\hat{f}(\alpha,\beta)|^n \leq n^{-10}$.

If $1/4 \leq |\beta| \leq 1/2$ then we can argue very similarly but taking $m = 1$. We can bound $|\hat{f}(\alpha,\beta)|$ away from 1 by an even better absolute constant and the required estimate holds with a lot of room to spare. □

6. A local central limit theorem for (U, V)

At this point, we note that if we choose a random n-sided die A, then with probability $1 - o(1)$ (where the $o(1)$ term is a power of n) we have the conclusions of the main results of the previous section: that is, Lemma 5.3 and Corollary 5.11. The first of these gives an upper bound for $\|U\|_\infty$ and the second shows that $|\hat{f}(\alpha,\beta)|^n$ is small when (α,β) lie outside a small box about the origin.

Given these properties, it is now reasonably easy to follow the standard Fourier method to prove a local central limit theorem for the sum of n independent copies of (U, V) that will be strong enough and explicit enough to enable us to prove our main result. So let us fix an n-sided die A that has the properties, and let (U, V) be the random variable defined earlier that we associate with A.

Let us briefly recall a few standard facts about characteristic functions. One is that if $\hat{f}$ is the characteristic function of (U, V), then the characteristic function of the sum of n independent copies of (U, V) is $\hat{f}^n$. This follows from the convolution law in Fourier analysis (since if we regard (U, V) as a function f from $\mathbb{Z}^2$ to $\mathbb{R}$, then $\hat{f}$ is its Fourier transform, and the function corresponding to the distribution of the sum of n independent copies of (U, V) is the n-fold convolution of f). For similar reasons we have the inversion formula

$$\mathbb{P}[(U, V) = (x, y)] = \int_{\mathbb{T}^2} \hat{f}(\alpha,\beta)^n e(-\alpha x - \beta y)\, d\alpha\, \delta\beta.$$

We shall also need to know that the characteristic function of (U, V) relates in a simple way to its moments. We have that

$$\frac{\partial^{r+s}}{\partial^r \alpha\, \partial^s \beta} \hat{f}(\alpha,\beta) = (2\pi i)^{r+s} \mathbb{E}(U^r V^s e(\alpha U + \beta V)),$$

and evaluating this at zero we get $(2\pi i)^{r+s}\mathbb{E}(U^r V^s)$.

Writing ∂_1 and ∂_2 for the operators of partially differentiating with respect to the first and second variables, respectively, we shall use the following estimate, which follows from Taylor's theorem and the observation about the partial derivatives. (We also use the fact that (U, V) has mean $(0, 0)$.)

{taylor}

Lemma 6.1. *Let f be as above. Then*

$$\hat{f}(\alpha,\beta) = 1 - 2\pi^2(\alpha^2\mathbb{E}U^2 + 2\alpha\beta\mathbb{E}UV + \beta^2\mathbb{E}V^2) + R(\alpha,\beta),$$

where $|R(\alpha,\beta)| \le \frac{4\pi^3}{3}(|\alpha|\|U\|_\infty + |\beta|\|V\|_\infty)^3$. □

From Lemma 6.1, we see that for small α,β, $\hat{f}(\alpha,\beta)^n$ is approximately equal to $\exp(-q(\alpha,\beta))$, where $q(\alpha,\beta) = \alpha^2\mathbb{E}U^2 + 2\alpha\beta\mathbb{E}UV + \beta^2\mathbb{E}V^2$, which is a positive semidefinite quadratic form in α and β. Next, we prove a small technical lemma in order to help us determine sufficient conditions for this approximation to be a good one.

xpestimate}

Lemma 6.2. *For every positive integer n and every pair of real numbers x, y such that $x^2 \le 1/4n$ and $|y| \le 1/4n$, we have the inequality*

$$\exp(-nx)\exp(-n(|y| + 2x^2 + 2y^2)) \le (1 - x + y)^n \le \exp(-nx)\exp(n(|y| + 2x^2 + 2y^2)).$$

It follows that the ratio of $(1-x+y)^n$ to $\exp(-nx)$ lies between $1-4n(|y|+x^2)$ and $1+4n(|y|+x^2)$.

Proof. For every u with $|u| \le 1/2$, we have the inequality $-u - u^2 \le \log(1-u) \le -u + u^2$, which implies that $\exp(-n(u+u^2)) \le (1-u)^n \le \exp(-n(u-u^2))$. Applying this with $u = x - y$ and noting that $u^2 \le 2(x^2+y^2)$, we obtain the first inequality.

Also, since nx^2 and $n|y|$ are both at most $1/4$, $n(|y| + 2x^2 + 2y^2) \le 1/2$ and $2y^2 \le |y|$. But when $|w| \le 1/2$ we have $1 - 2|w| \le e^{-w} \le 1 + 2|w|$, and the second inequality follows. □

Combining the above lemmas with Corollary 5.11, we obtain the following result, which tells us that in a suitable sense $\hat{f}^n$ is approximated by a Gaussian.

roximation}

Lemma 6.3. *Let $\hat{f}$ be the characteristic function of (U,V). Define a function $h : \mathbb{R}^2 \to \mathbb{R}$ by setting $h(\alpha,\beta) = \hat{f}(\alpha,\beta)^n$ when $|\alpha| \le 1/2$ and $|\beta| \le 1/2$ and $h(\alpha,\beta) = 0$ otherwise. And let g be the two-dimensional Gaussian*

$$g(\alpha,\beta) = \exp(-2\pi^2 n(\alpha^2\mathbb{E}U^2 + 2\alpha\beta\mathbb{E}UV + \beta^2\mathbb{E}V^2)).$$

Then $\|h - g\|_1 \le 2^{123}(\log n/n)^4$.

Proof. Since the conclusion of Lemma 5.3, is satisfied, we have $\|U\|_\infty \le 6\sqrt{n\log n}$. We also have that $\|V\|_\infty \le n$ for trivial reasons.

It follows that for any α and β we have

$$2\pi^2(\alpha^2\mathbb{E}U^2 + 2\alpha\beta\mathbb{E}UV + \beta^2\mathbb{E}V^2) \le 2\pi^2(6\alpha\sqrt{n\log n} + \beta n)^2 \le 1440\alpha^2 n\log n + 40\beta^2 n^2.$$

Also, in Lemma 6.1 we have that

$$R(\alpha,\beta) \le 50(6|\alpha|\sqrt{n\log n} + |\beta|n)^3 \le 200(216|\alpha|^3(n\log n)^{3/2} + |\beta|^3 n^3) \le 2^{16}|\alpha|^3(n\log n)^{3/2} + 2^8|\beta|^3 n^3.$$

Setting $Q(\alpha,\beta) = 2\pi^2(\alpha^2\mathbb{E}U^2 + 2\alpha\beta\mathbb{E}UV + \beta^2\mathbb{E}V^2)$, so $g(\alpha,\beta) = \exp(-nQ(\alpha,\beta))$, we have that

$$\hat{f}(\alpha,\beta) = 1 - Q(\alpha,\beta) + R(\alpha,\beta),$$

and therefore by Lemma 6.2 that

$$g(\alpha,\beta)(1 - 4n(|R(\alpha,\beta)| + Q(\alpha,\beta)^2)) \leq \hat{f}(\alpha,\beta)^n \leq g(\alpha,\beta)(1 + 4n(|R(\alpha,\beta)| + Q(\alpha,\beta)^2)),$$

provided that $Q(\alpha,\beta) \leq 1/2\sqrt{n}$ and $|R(\alpha,\beta)| \leq 1/4n$.

For $Q(\alpha,\beta)$ to be at most $1/2\sqrt{n}$ it is enough if $1440\alpha^2 n\log n \leq 1/4\sqrt{n}$ and $40\beta^2n^2 \leq 1/4\sqrt{n}$, and for that it is enough if $|\alpha| \leq 1/80n^{3/4}\sqrt{\log n}$ and $|\beta| \leq 1/16n^{5/4}$. For $R(\alpha,\beta)$ to be at most $1/4n$ it is enough if $2^{16}|\alpha|^3(n\log n)^{3/2} \leq 1/8n$ and $2^8|\beta|^3n^3 \leq 1/8n$, and for that it is enough if $|\alpha| \leq 1/100n^{5/6}\sqrt{\log n}$ and $|\beta| \leq 1/16n^{4/3}$. The second pair of conditions is more stringent, so provided that they hold, we have the required bounds on $Q(\alpha,\beta)$ and $R(\alpha,\beta)$.

To bound $\|h - g\|_1$, we shall look at the contribution within a small box and outside it. Corollary 5.11 tells us that $|\hat{f}(\alpha,\beta)|^n \leq n^{-10}$ when $|\alpha| \geq 10^5\sqrt{\log n}/n$ or $|\beta| \geq 2^{23}\sqrt{\log n}/n^{3/2}$. For sufficiently large n, $10^5\sqrt{\log n}/n \leq 1/100n^{5/6}$ and $2^{23}\sqrt{\log n}/n^{3/2} \leq 1/16n^{4/3}$. It follows that on the boundary of the box $B = [-10^5\sqrt{\log n}/n, 10^5\sqrt{\log n}/n] \times [-2^{23}\sqrt{\log n}/n^{3/2}, 2^{23}\sqrt{\log n}/n^{3/2}]$ we have that $g(\alpha,\beta) \leq (1 - 4n(|R(\alpha,\beta)| + Q(\alpha,\beta)^2))^{-1}n^{-10}$. For sufficiently large n this implies that $g(\alpha,\beta) \leq 2n^{-10}$ everywhere on the boundary.

Let $G(t)$ equal the integral of $g(\alpha,\beta)$ round the boundary of the box tB. Note that for each (α,β) on the boundary of B, we have that $g(t\alpha,t\beta)$ is a Gaussian in t that takes the value at most n^{-10} when $t = 1$. Therefore, it is bounded above by $\exp(-10t^2\log n)$. The perimeter of B is at most $10^6\sqrt{\log n}/n$ (again assuming that n is sufficiently large), so

$$\int_{(\alpha,\beta)\notin B} g(\alpha,\beta)\,d\alpha\,d\beta \leq 10^6\frac{\sqrt{\log n}}{n}\int_1^\infty t\exp(-10t^2\log n)\,dt \leq 50000n^{-11}.$$

Since $h = 0$ outside $[-1/2,1/2]^2$ and $|h| \leq n^{-10}$ outside B, we also have that

$$\int_{(\alpha,\beta)\notin B} |h(\alpha,\beta)|\,d\alpha\,d\beta \leq n^{-10}.$$

For all (α,β) we have that $|f(\alpha,\beta)^n| \leq 1$. Therefore, our estimate for the ratio of g to f implies that for all $(\alpha,\beta) \in B$ we have that

$$|g(\alpha,\beta) - \hat{f}(\alpha,\beta)^n| \leq 4n(|R(\alpha,\beta)| + Q(\alpha,\beta)^2).$$

Inside B, one can check that $Q(\alpha,\beta) \leq 10^{12}(\log n)^2/n$ for n sufficiently large, and hence that $Q(\alpha,\beta)^2 \leq 10^{24}(\log n)^4/n^2$. Also, $|R(\alpha,\beta)| \leq 2^{77}(\log n)^3/n^{3/2}$. So, for sufficiently large n, we

have the bound

$$|g(\alpha,\beta) - \hat{f}(\alpha,\beta)^n| \leq 2^{80}(\log n)^3/n^{3/2}$$

everywhere in B. Since B has area at most $2^{42}\log n/n^{5/2}$, it follows that the contribution to $\|g-h\|_1$ from inside B is at most $2^{122}(\log n)^4/n^4$.

Combining these estimates gives us the bound stated (as always, assuming that n is sufficiently large). □

In the statement of the next result, we refer to a "discrete Gaussian". By this we mean a function defined on $\mathbb{Z}^2$ with a formula of the form $f(x,y) = c\exp(-\lambda q(x,y))$ for some positive semidefinite quadratic form q.

tegaussian}

Corollary 6.4. *There is a discrete Gaussian G such that*

$$\mathbb{P}[(U,V) = (x,y)] - G(x,y)| \leq 2^{123}(\log n/n)^4$$

for every $(x,y) \in \mathbb{Z}^2$. Furthermore, $\mathbb{P}[(U,V) = (0,0)] \leq 2^{43}\log n/n^{5/2}$.

Proof. We have

$$\mathbb{P}[(U,V) = (x,y)] = \int_{\mathbb{T}^2} \hat{f}(\alpha,\beta)^n e(-\alpha x - \beta y)\,d\alpha\,d\beta = \int_{\mathbb{R}^2} h(\alpha,\beta)e(-\alpha x - \beta y)\,d\alpha\,d\beta\,.$$

By Lemma 6.3 and the fact that $e(-\alpha x - \beta y)$ has modulus 1 for every x, y, if we replace h by g on the right-hand side, the difference to the integral is at most $2^{123}(\log n/n)^4$. But $\int_{\mathbb{R}^2} g(\alpha,\beta)e(-\alpha x - \beta y)\,d\alpha\,d\beta$ has a Gaussian dependence on (x,y), since it is the Fourier transform of a Gaussian (the Gaussian being defined on $\mathbb{R}^2$ even if we are evaluating its Fourier transform at points of $\mathbb{Z}^2$). This proves the first part.

For the second part, recall from the proof of Lemma 6.3 that the integral of h outside the box B is at most n^{-10} and that B has area at most $2^{42}\log n/n^{5/2}$. It follows that

$$\mathbb{P}[(U,V) = (0,0)] = \int_{\mathbb{R}^2} h(\alpha,\beta)\,d\alpha\,d\beta \leq 2^{43}\log n/n^{5/2}.$$

when n is sufficiently large. (Indeed, this is a bound for $\mathbb{P}[(U,V) = (x,y)]$ for all (x,y).) □

7. The main theorem

We are almost ready to prove the main theorem. However, a uniform bound on the probabilities is not quite enough for our purposes. As is customary, we need to combine it with tail estimates. However, this is straightforward.

{tail}

Lemma 7.1. $\mathbb{P}[|U^{*n}| \geq 6Cn\sqrt{\log n}] \leq 2\exp(-2C^2)$.

Proof. The distribution of U^{*n} is given by the sum of n independent copies of U, which has mean zero. Since $\|U\|_\infty \leq 6\sqrt{n \log n}$, Hoeffding's inequality gives us the required estimate. □

Theorem 7.2. *The probability that A beats another random die is* $\frac{1}{2} + o(1)$.

Proof. As remarked earlier, we will be done if we can prove that

$$\mathbb{P}[U^{*n} > 0 | V^{*n} = 0] = \frac{1}{2} + o(1).$$

Recall first that $\mathbb{P}[V^{*n} = 0] \geq n^{-3/2}/4$, by Lemma 5.2.

Next, note that by Corollary 6.4 we have for every x that

$$|\mathbb{P}[(U, V)^{*n} = (x, 0)] - \mathbb{P}[(U, V)^{*n} = (-x, 0)]| \leq 2^{124}(\log n/n)^4,$$

since G is an even function. We also have that $\mathbb{P}[(U, V) = (0, 0)] \leq 2^{43} \log n/n^{5/2}$.

Thirdly, Lemma 7.1 gives us that $\mathbb{P}[|U^{*n}| > 12n \log n] \leq 2n^{-8}$.

Putting these estimates together, we find that

$$|\mathbb{P}[U^{*n} > 0 \wedge V^{*n} = 0] - \mathbb{P}[U^{*n} < 0 \wedge V^{*n} = 0]| \leq (12n \log n)(2^{124}(\log n/n)^4 \leq 2^{128}(\log n)^5/n^3,$$

and therefore that

$$|\mathbb{P}[U^{*n} > 0 | V^{*n} = 0] - \mathbb{P}[U^{*n} < 0 | V^{*n} = 0]| \leq 2^{130}(\log n)^5/n^{3/2}.$$

We also have that

$$\mathbb{P}[U^{*n} = 0 | V^{*n} = 0] \leq 2^{45} \log n/n.$$

The result follows. □

As observed in Section 2, this statement is equivalent to Conjecture 1.2 for the balanced sequences model, and the proof also yields Conjecture 1.1 for this model.