Counting tournament score sequences

(Extended abstract)

Anders Claesson[*]   Mark Dukes[†]   Atli Fannar Franklín[‡]

Sigurður Örn Stefánsson[§]

Abstract

The score sequence of a tournament is the sequence of the out-degrees of its vertices arranged in nondecreasing order. The problem of counting score sequences of a tournament with n vertices is more than 100 years old (MacMahon 1920). In 2013 Hanna conjectured a surprising and elegant recursion for these numbers. We settle this conjecture in the affirmative by showing that it is a corollary to our main theorem, which is a factorization of the generating function for score sequences with a distinguished index. We also derive a closed formula and a quadratic time algorithm for counting score sequences.

1   Introduction

This extended abstract summarises the results of our paper [4]. In 1953 Landau [9] used oriented complete graphs—also called *tournaments*—to model pecking orders. If the vertices of the complete graph represent players (rather than chickens), then the initial vertex of a directed edge signifies the winner of a game between the two end-point players. The number of wins of a player is equal to the number of outgoing edges from that vertex. A *score sequence* is a sequence of these number of wins given in a nondecreasing order. For instance, with 3 players there are two possible score sequences, namely $(0, 1, 2)$ and

---

[*]Science Institute, University of Iceland, Iceland. Email: `akc@hi.is`.

[†]School of Mathematics & Statistics, University College Dublin, Ireland. E-mail: `mark.dukes@ucd.ie`.

[‡]Science Institute, University of Iceland, Iceland. Email: `aff6@hi.is`.

[§]Science Institute, University of Iceland, Iceland. Email: `sigurdur@hi.is`.

$(1, 1, 1)$. Note that non-isomorphic tournaments may give rise to the same score sequence. With 5 players there are, up to isomorphism, 12 tournaments but only 9 score sequences. To be even more specific, here are two non-isomorphic tournaments:
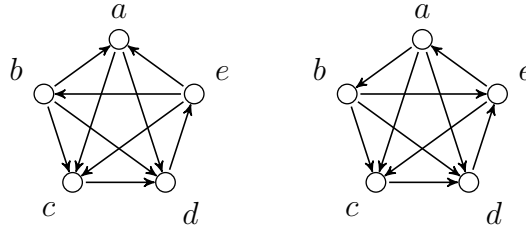
The score sequence associated with both is $(1, 1, 2, 3, 3)$. The following characterization of score sequences is known as Landau's theorem.

Theorem 1 (Landau [9]). *A sequence of integers $s = (s_0, \ldots, s_{n-1})$ is a score sequence if and only if*

(1) $0 \le s_0 \le s_1 \le \cdots \le s_{n-1} \le n - 1$,

(2) $s_0 + \cdots + s_{k-1} \ge \binom{k}{2}$ *for* $1 \le k < n$, *and*

(3) $s_0 + \cdots + s_{n-1} = \binom{n}{2}$.

Let S_n be the set of score sequences of length n. There is no known closed formula for the associated cardinalities (A000571 in the OEIS [7])

$$\bigl(|S_n|\bigr)_{n \ge 0} = (1, 1, 1, 2, 4, 9, 22, 59, 167, 490, 1486, 4639, 14805, \ldots)$$

or their generating function.

It should be noted that Landau was not the first person to study score sequences, or attempt to count them. MacMahon [10] used symmetric functions and hand calculations to determine $|S_n|$ for $n \le 9$ in 1920. Building on Landau's work, Narayana and Bent [11], in 1964, derived a multivariate recursive formula for determining $|S_n|$. They used it to give a table for $n \le 36$. In 1968 Riordan [12] gave a simpler and more efficient recursion, but unfortunately it turned out to be incorrect [13].

Let $[a, b]$ denote the interval of integers $\{a, a+1, \ldots, b\}$. We may view a score sequence $s \in S_n$ as an endofunction $s : [0, n-1] \to [0, n-1]$. We now introduce the notion of a *pointed score sequence*. Define $S_n^\bullet$ as the Cartesian product $S_n^\bullet = S_n \times [0, n-1]$. We call the members of $S_n^\bullet$ *pointed score sequences*; e.g. there are 6 pointed score sequences in $S_3^\bullet$:

$$((0, 1, 2), 0), ((0, 1, 2), 1), ((0, 1, 2), 2),$$
$$((1, 1, 1), 0), ((1, 1, 1), 1), ((1, 1, 1), 2).$$

Let $(s, i) \in S_n^\bullet$. Depending on the context, the element i will be interpreted as a position (element in the domain) or a value (element in the codomain) of s. If i is a value, then

the cardinality of the fiber $s^{-1}(i)$ is the number of times i occurs in s; this number may be zero. Let

$$S_n^\bullet(t) = \sum_{(s,i) \in S_n^\bullet} t^{|s^{-1}(i)|}$$

be the polynomial recording the distribution of the statistic $(s, i) \mapsto |s^{-1}(i)|$ on $S_n^\bullet$. As an example, $S_3^\bullet(t) = 2 + 3t + t^3$. Let

$$S^\bullet(x, t) = \sum_{n \geq 1} S_n^\bullet(t) x^n.$$

To present the bijection that is the main result of this paper, we will first introduce a particular type of multiset that is an essential ingredient in our deconstruction of a pointed score sequence. At first glance it is not obvious what the relevance of these multisets to score sequences is.

We define EGZ_n as the set of multisets of size n with elements in the cyclic group $\mathbb{Z}_n$ whose sum is $\binom{n}{2}$ modulo n. To understand what the elements of EGZ_n look like it may be helpful to note that $\binom{n}{2}$, as an element of $\mathbb{Z}_n$, is 0 if n is odd and $n/2$ if n is even. For instance, EGZ_3 consists of the 4 multisets $\{0, 0, 0\}$, $\{0, 1, 2\}$, $\{1, 1, 1\}$, and $\{2, 2, 2\}$.

The notation EGZ_n refers to the Erdős-Ginzburg-Ziv Theorem [5], which is stated below. Following it we give a proposition motivating this terminology; its proof gives a simple one-to-one correspondence between EGZ_n and the sets considered by Erdős, Ginzburg, and Ziv.

Theorem 2 (Erdős, Ginzburg, and Ziv [5]). *Each set of $2n - 1$ integers contains some subset of n elements the sum of which is a multiple of n.*

Proposition 3. *There is a one-to-one correspondence between EGZ_n and n-element subsets of $[1, 2n - 1]$ whose sum is a multiple of n.*

Proof. Let $A = \{a_1, \ldots, a_n\}$ be a subset of $[1, 2n - 1]$ such that $a_1 + \cdots + a_n$ is divisible by n. Without loss of generality we can further assume that $a_1 < a_2 < \cdots < a_n$. Let $b_i = a_i - i$. The mapping $A \mapsto \{b_1, \ldots, b_n\}$ is a bijection onto EGZ_n. Further proof details are omitted but can be found in [4], as with other results presented in this abstract. $\square$

The sequence of cardinalities

$$\left(|\mathrm{EGZ}_n|\right)_{n \geq 1} = (1, 1, 4, 9, 26, 76, 246, 809, 2704, 9226, 32066, \ldots)$$

is entry A145855 in the OEIS [7]. As recorded in that OEIS entry, Jovović conjectured and Alekseyev [1] proved in 2008 that

$$|\mathrm{EGZ}_n| = \frac{1}{2n} \sum_{d|n} (-1)^{n-d} \varphi(n/d) \binom{2d}{d}, \tag{4}$$

where the sum runs over all positive divisors of n and φ is Euler's totient function. A generalization of this result was given by Chern [3] in 2019.

The zeros in a multiset $M \in \mathrm{EGZ}_n$ play a prominent role in our construction. We now introduce a generating function to record their number. For a multiset $M \in \mathrm{EGZ}_n$ let $|M|_i$ be the number of occurrences of i in M. Furthermore, let

$$\mathrm{EGZ}_n(t) = \sum_{M \in \mathrm{EGZ}_n} t^{|M|_0}$$

be the polynomial recording the distribution of zeros in multisets belonging to EGZ_n. For instance, $\mathrm{EGZ}_3(t) = 2 + t + t^3$ (looking at the distribution of 1s or 2s in EGZ_3 would result in the same polynomial). Define the generating functions

$$\mathrm{EGZ}(x,t) = \sum_{n \geq 1} \mathrm{EGZ}_n(t)x^n \quad \text{and} \quad S(x) = \sum_{n \geq 0} |S_n|x^n.$$

Our main result, Theorem 4, is a factorization of the generating function for pointed score sequences:

$$S^\bullet(x,t) = \mathrm{EGZ}(x,t)S(x). \tag{5}$$

Let $(s,i) \in S_n^\bullet$. Viewing i is an element of the codomain of s we find that $S^\bullet(x,0)$ consists of terms stemming from pairs (s,i) such that $s^{-1}(i)$ is empty; i.e. i is outside the image of s. Thus, $S^\bullet(x,1) - S^\bullet(x,0)$ counts pairs (s,i) for which i is in the image of s. Let

$$S_n^\circ = \{(s,i) \in S_n^\bullet : i \in \mathrm{Im}(s)\} = \{(s,i) \in S_n^\bullet : i = s_j \text{ for some } j \in [n]\}$$

and let $S^\circ(x) = S^\bullet(x,1) - S^\bullet(x,0)$ be the corresponding generating function. For instance, S_3° consists of the 4 elements $((0,1,2),0)$, $((0,1,2),1)$, $((0,1,2),2)$, and $((1,1,1),1)$. We will show (in Corollary 6) that $S^\circ(x) = xC(x)S(x)$, where $C(x) = (1-\sqrt{1-4x})/(2x)$ is the generating function for the Catalan numbers $C_n = \binom{2n}{n}/(1+n)$. This striking occurrence of the Catalan numbers was in fact the original inspiration for our work. It was in the summer of 2019 that we experimented with score sequences and conjectured the identity. Despite ample attempts we were for the longest time unable to prove it.

By setting $t = 1$ in Equation 5 and noting that $S^\bullet(x,1) = xS'(x)$ it follows that

$$xS'(x) = \mathrm{EGZ}(x,1)S(x), \tag{6}$$

a fact conjectured by Paul D. Hanna as recorded in the OEIS entry A000571 in 2013. Equation 6 may alternatively be written $\big(\log S(x)\big)' = \mathrm{EGZ}(x,1)/x$ and so

$$S(x) = \exp\left(\sum_{n \geq 1} \frac{|\mathrm{EGZ}_n|}{n} x^n \right),$$

which arguably is the most elegant way of expressing the relation between $|S_n|$ and $|\mathrm{EGZ}_n|$. The most efficient way of computing the numbers $|S_n|$ is, however, to use the recursion underlying Equation 6. Namely, $|S_0| = 1$ and, for $n \geq 1$,

$$|S_n| = \frac{1}{n} \sum_{k=1}^{n} |S_{n-k}||\mathrm{EGZ}_k|.$$

See Corollary 8 and the discussion following it.

2 The main theorem and its bijection

Let the generating functions $S^\bullet(x,t)$, $\mathrm{EGZ}(x,t)$ and $S(x)$ be defined as in Section 1.

Theorem 4. *We have $S^\bullet(x,t) = \mathrm{EGZ}(x,t)S(x)$.*

The proof of Theorem 4 is combinatorial and is achieved by creating a bijection

$$\Phi : S_n^\bullet \to \bigcup_{k=1}^n \mathrm{EGZ}_k \times S_{n-k}$$

that maps a pointed score sequence to a pair consisting of a multiset and a score sequence. A property of this bijection is that, for $(M,v) = \Phi(s,i)$, the number of occurrences of i in s is equal to the multiplicity of zero in M. Before defining Φ we need to introduce several necessary concepts.

A nonempty directed graph is said to be *strongly connected* if there is a directed path between each pair of vertices of the graph. Note that we do not consider the empty graph to be strongly connected. A *strong score sequence* is one which stems from a strongly connected tournament. Equivalently (see Harary and Moser [6, Theorem 9]), $s = (s_0, \ldots, s_{n-1})$, with $n \geq 1$, is a strong score sequence if the inequality (2) of Theorem 1 is always strict; that is, $s_0 + \cdots + s_{k-1} > \binom{k}{2}$ for $1 \leq k < n$. Let us define the *direct sum* of two score sequences $u \in S_k$ and $v \in S_\ell$ by $u \oplus v = uv'$, where v' is obtained from v by adding k to each of its letters and juxtaposition indicates concatenation. For instance, $(0) \oplus (0) \oplus (1,1,1) = (0,1,3,3,3)$. If U and V are tournaments having score sequences u and v, one may view the direct sum $u \oplus v$ as the score sequence of the tournament where arrows are placed between the vertices of U and V such that they all point towards U:
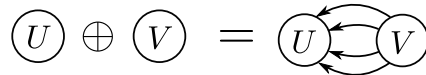
This may easily be seen to be independent of the choice of tournaments.

Lemma 5. *Let $s \in S_n$. If $s_0 + \cdots + s_{k-1} = \binom{k}{2}$ for some $k < n$, then $u = (s_0, \ldots, s_{k-1})$ and $v = (s_k - k, \ldots, s_{n-1} - k)$ are both score sequences, and $s = u \oplus v$.*

A direct consequence of Lemma 5 is that every score sequence s can be uniquely written as a direct sum $s = t_1 \oplus t_2 \oplus \cdots \oplus t_k$ of nonempty strong score sequences; in this context, the t_i will be called the *strong summands* of s. In terms of underlying tournaments we have the picture:
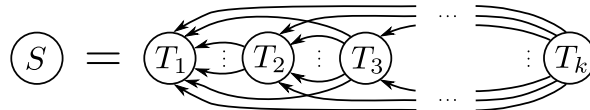
We are now almost in a position to define the promised map Φ, but first a couple of definitions. Assume that we are given a score sequence $s = (s_0, s_1, \ldots, s_{n-1}) \in S_n$.

- For any integer j, let $s+j$ denote the sequence obtained by adding j to each element of s, reducing modulo n, and sorting the outcome in nondecreasing order. Note that $s+j$ need not be a score sequence even though s is. E.g. $s = (1,1,1)$ is a score sequence, but $s+1 = (2,2,2)$ is not. On the other hand, if $s = (0,1,2)$ then $s+1 = s$ is a score sequence. A characterization of when $s+j$ is a score sequence is given in [4, Lemma 7].

- Let $\mu(s+j)$ denote the multiset $\{s_0 + j, s_1 + j, \ldots, s_{n-1} + j\}$ with elements in the cyclic group $\mathbb{Z}_n$.

Given a pointed score sequence $(s,i) \in S_n^{\bullet}$, write $s = t_1 \oplus t_2 \oplus \cdots \oplus t_k$ and let j be the smallest index such that $i < |t_1 \oplus \cdots \oplus t_j|$. Another way to define j is as the smallest prefix $t_1 \oplus \cdots \oplus t_j$ of strong summands of s that begins $s_0, s_1, \ldots, s_i$. Define the two score sequences u and v by

$$u = t_1 \oplus \cdots \oplus t_j \quad \text{and} \quad v = t_{j+1} \oplus \cdots \oplus t_k.$$

Finally, we let

$$\Phi(s,i) := \big(\mu(u-i), v\big).$$

As an example, consider the score sequence $s = (0,2,2,3,3,5,7,7,7)$; its decomposition into strong summands is $s = (0) \oplus (1,1,2,2) \oplus (0) \oplus (1,1,1)$. With $i = 3$ we get $u = (0) \oplus (1,1,2,2) = (0,2,2,3,3)$, $v = (0) \oplus (1,1,1) = (0,2,2,2)$, $u - 3 = (0,0,2,4,4)$ and so $\Phi(s,3) = \big(\{0,0,2,4,4\}, (0,2,2,2)\big)$.

Corollary 6. *We have $S^{\circ}(x) = xC(x)S(x)$, where $C(x)$ is the generating function for the Catalan numbers.*

Corollary 7. *We have $S(x) = \exp\bigg(\sum_{n \geq 1} |\mathrm{EGZ}_n| x^n / n\bigg)$.*

We end by comparing our result (Corollary 7) with earlier results on the enumeration of the *ordered* score sequences $(s_0, s_1, \ldots, s_{n-1})$, also called *score vectors*. That is, if G is a tournament on the vertex set $\{v_0, v_1, \ldots, v_{n-1}\}$, then s_i is the out-degree of v_i in G. For instance, while there are only two score sequences of length 3, namely $(1,1,1)$ and $(1,2,3)$, there are 7 score vectors of length 3: the vector $(1,1,1)$ together with the 6 permutations of $(1,2,3)$.

Stanley and Zaslavsky [14] have shown that the number of score vectors of length n equals the number of (labeled) forests on n nodes. A combinatorial proof was subsequently given by Kleitman and Winston [8]. Cayley [2] famously gave the formula n^{n-2} for the number of trees on n nodes. From the theory of exponential generating functions it immediately follows that $\exp\big(\sum_{n \geq 1} n^{n-2} x^n / n!\big)$ is the exponential generating function of forests, and thus also of score vectors.

3 The number of score sequences

If two power series $A(x) = 1 + \sum_{n \geq 1} a_n x^n$ and $B(x) = \sum_{n \geq 1} b_n x^n$ satisfy $xA'(x)/A(x) = B(x)$ and hence $\log A(x) = \sum_{n \geq 1} b_n x^n/n$, then one readily obtains a closed formula for a_n by expanding and identifying coefficients in $A(x) = \exp(b_1 x^1/1) \exp(b_2 x^2/2) \cdots$. Applying this to the equation in Corollary 7 we arrive at

$$|S_n| = \frac{1}{n!} \sum_{\pi \in \text{Sym}(n)} \prod_{\ell \in C(\pi)} |\text{EGZ}_\ell|, \tag{7}$$

where $\text{Sym}(n)$ is the symmetric group of degree n and $C(\pi)$ encodes the cycle type of π; i.e. there is an $\ell \in C(\pi)$ for each ℓ-cycle of π. While having the virtue of being closed, this formula does not lend itself to quickly calculating $|S_n|$. For that purpose the following recursion is better suited.

Corollary 8. *For $n \geq 1$,* $|S_n| = \sum_{k=1}^{n} |S_{n-k}||\text{EGZ}_k| = \frac{1}{n} \sum_{k=1}^{n} \frac{|S_{n-k}|}{2nk} \sum_{d|k} (-1)^{k-d} \varphi(k/d) \binom{2d}{d}.$

This allows us to calculate all values of $|S_k|$ for $k \leq n$ in $\Theta(n^2)$ time, assuming constant time integer operations. This is an improvement on earlier results by Narayana and Bent [11]. Their recursive formula can be implemented to find $|S_n|$ in $\Theta(n^3)$ time, but no faster since their recursive function must always visit $\Theta(n^3)$ states to do so; to get all $|S_k|$ for $k \leq n$ takes $\Theta(n^4)$ time due to lack of overlap in the states recursively visited for different k.

Since $S(x) = (1 - T(x))^{-1}$, where $T(x)$ is the generating function for the number of strong score sequences $|T_k|$ having length k, this recursive computation method can be extended to $|T_k|$. We first calculate the values $|S_k|$ and use this recursion to calculate all the values $|T_k|$ for $k \leq n$ in $\Theta(n^2)$ time. This is the same method as used by Stockmeyer [15], just calculating the underlying $|S_k|$ faster which brings the total time complexity down from $\Theta(n^4)$ to $\Theta(n^2)$.

References

[1] Max Alekseyev. https://oeis.org/A145855/a145855.txt, 2008.

[2] Arthur Cayley. *A theorem on trees*, volume 13 of *Cambridge Library Collection - Mathematics*, pages 26—-28. Cambridge University Press, 2009.

[3] Shane Chern. An extension of a formula of Jovovic. *Integers*, 19:Paper No. A47, 2019.

[4] Anders Claesson, Mark Dukes, Atli Fannar Franklín, and Sigurdur Örn Stefánsson. Counting tournament score sequences. *Proceedings of the American Mathematical Society*, to appear. doi:10.1090/proc/16425. arXiv:2209.03925.

[5] P. Erdős, A. Ginzburg, and A. Ziv. Theorem in the additive number theory. *Bull. Res. Council Israel Sect. F*, 10F(1):41–43, 1961.

[6] Frank Harary and Leo Moser. The theory of round robin tournaments. *Amer. Math. Monthly*, 73:231–246, 1966.

[7] OEIS Foundation Inc. The Online Encyclopedia of Integer Sequences. `https://oeis.org`, 2021.

[8] Daniel J Kleitman and Kenneth J Winston. Forests and score vectors. *Combinatorica*, 1(1):49–54, 1981.

[9] H. G. Landau. On dominance relations and the structure of animal societies: III. The condition for a score structure. *Bull. Math. Biophys.*, 15:143–148, 1953.

[10] P. A. MacMahon. An American tournament treated by the calculus of symmetric functions. *Quart. J. Math.*, 49:1–36, 1920. Reprinted in *Percy Alexander MacMahon Collected Papers, Volume I*, George E. Andrews, ed., MIT Press, 1978.

[11] T. V. Narayana and D. H. Bent. Computation of the number of score sequences in round-robin tournaments. *Can. Math. Bull*, 7(1):133–136, 1964.

[12] John Riordan. The number of score sequences in tournaments. *J. Combinatorial Theory*, 5:87–89, 1968.

[13] John Riordan. Erratum: "The number of score sequences in tournaments". *J. Combinatorial Theory*, 6:226, 1969.

[14] Richard P. Stanley. Decompositions of rational convex polytopes. In J. Srivastava, editor, *Combinatorial Mathematics, Optimal Designs and Their Applications*, volume 6 of *Annals of Discrete Mathematics*, pages 333–342. Elsevier, 1980.

[15] Paul K. Stockmeyer. Counting various classes of tournament score sequences. *arXiv:2202.05238v1*, 2022.