CONFIRMING KLEITMAN–WINSTON CONJECTURE ON THE LARGEST COEFFICIENT IN A q-CATALAN NUMBER.

Jeong Han Kim and Boris Pittel*

Abstract. In 1983 Kleitman and Winston conjectured that the largest coefficient in an n-th q-Catalan number is of order $O(4^n/n^{3/2})$. Assuming its truth, they proved that the total number of n-tournament score sequences is $O(4^n/n^{5/2})$, thus matching their own lower bound. Our purpose is to confirm the conjecture.

1. Introduction. An n-tournament score sequence is a nondecreasing sequence of nonnegative integers $\mathbf{s} = (s_1, \dots, s_n)$ such that

$$\sum_{i=1}^{m} s_i \begin{cases} \geq \binom{m}{2}, & \text{if } 1 \leq m \leq n-1, \\ = \binom{n}{2}, & \text{if } m = n. \end{cases} \tag{1.1}$$

It is a classical result of Landau [5], (see Moon [6], Ford and Fulkerson [3]) that the above conditions are both necessary and sufficient for existence of a complete digraph on $[n]$ whose (size-ordered) out-degree sequence is $\mathbf{s}$. Such a digraph is interpreted as an outcome of a round-robin tournament, with $[n]$ as the set of players, and $\mathbf{s}$ comprised by scores put in order of increase. Let S_n denote the total number of all such $\mathbf{s}$. In a seminal paper [4] Kleitman and Winston were able to show that for some absolute (positive) constants c_1, c_2

$$c_1 \frac{4^n}{n^{5/2}} \leq S_n \leq c_2 \frac{4^n}{n^2}, \tag{1.2}$$

thus improving the earlier bounds due to Erdös and Moser, see [6], by factors n^2 and $n^{-1/2}$ respectively. Kleitman and Winston demonstrated that an upper bound in (1.2) could be upgraded to a best possible bound

$$S_n \leq c_3 \frac{4^n}{n^{5/2}}, \tag{1.3}$$

* Research supported by the NSF grant # DMS-9803410 and by Microsoft Research during the second author visit in Spring, 1999.

(hence confirming the conjecture made by Moser [7] in 1968), provided that the following is true. Let $c_n(q)$ be a sequence of polynomials, with nonnegative coefficients, defined by an initial condition $c_1(q) = 1$ and a recurrence

$$c_n(q) = \sum_{i=1}^{n-1} q^{i(n-i-1)} c_i(q) c_{n-i}(q), \quad n \geq 2. \tag{1.4}$$

$c_n(q)$ is the generating function of the subdiagonal "up-or-right" paths from $(0,0)$ to (n,n) in $\mathbf{Z}^2$, classified according to value of the area below. Then $c_n(1)$ was identified as the n-th Catalan number $C_n = n^{-1}\binom{2(n-1)}{n-1}$, whence the name "$q$-Catalan number" for $c_n(q)$. (Its degree equals $\binom{n-1}{2}$.) The conjecture was that

$$\max_m \, [q^m] c_n(q) = O\left(\frac{4^n}{n^3}\right). \tag{1.5}$$

(For a polynomial $a(q)$, $[q^j]\, a(q)$ denotes the coefficient by q^j.) Kleitman and Winston showed that the variance of the probability distribution $p_{nm} := [q^m]c_n(q)/C_n, (0 \leq m \leq \binom{n-1}{2})$, was asymptotic to n^3, so that one should expect (1.5) be true, considering that $C_n \sim c4^n n^{-3/2}$ and guessing that

$$p_{nm} = O(n^{-3/2}). \tag{1.6}$$

The latter was indeed plausible, in the light of the local limit theorems for convergence to normal distribution, but no technique available (Canfield [1]) worked. Implicit in [4] was a doubt about an even weaker property of asymptotic normality, in part probably because the mean of the distribution was found to be $n^2/2 - cn^{3/2}$ roughly, thus close to $\binom{n-1}{2} - cn^{3/2}$. Notice that $\phi_n(q) := q^{\binom{n-1}{2}} c_n(q^{-1})$ is also a polynomial, and (1.4) transforms then into a simpler looking

$$\phi_n(q) = \sum_{i=1}^{n-1} q^{i-1} \phi_i(q) \phi_{n-i}(q), \quad \phi_1(q) = 1. \tag{1.7}$$

(This equation is missing in [4].) Obviously, $\phi_n(q)$ enumerates the paths by the difference between maximum possible area $\binom{n-1}{2}$ and an actual area below a path. We will denote by X_n a random variable whose distribution is given by

$$\mathbf{Pr}(X_n = m) = \frac{[q^m]\phi_n(q)}{C_n}, \quad 0 \leq m \leq \binom{n-1}{2}. \tag{1.8}$$

Then clearly,

$$\mathbf{Pr}(X_n = m) = p_{n, \binom{n-1}{2} - m}.$$

In a wide-ranging study Takács [8] showed that several interesting random variables, including the "area" under a Bernoulli excursion of duration $2(n-1)$, have the same distribution as $n-1+2X_n$. Using the moments method, based on (1.7), Takács found that, in terms of X_n, the distribution of $n^{-3/2}X_n$ converges to a certain classical distribution (with a density) for the Brownian excursion process. This result ruled out asymptotic normality of X_n, but the scaling factor $n^{-3/2}$ was still in place! This led Takács to the conclusion that Kleitman-Winston's conjecture (1.5) would be settled, if the integral limit theorem for X_n were strengthened to a corresponding local limit theorem for individual probabilities $\mathbf{Pr}(X_n = k)$.

In this paper we upperbound $\mathbf{Pr}(X_n = k)$ by reducing the problem to estimating $\mathbf{Pr}(\mathcal{S}_n = k)$ for $\mathcal{S}_n$ being a sum of independent random variables. A key point in our argument is the fact that, according to [8], we may just as well consider the distribution of the "area" under the Bernoulli excursion. This is an excursion on the nonnegative integers of the symmetric random walk, conditioned on return to 0 after $2n$ steps. And it opens the door for a healthy dose of independence!

2. Proofs. We begin with the precise definition of the Bernoulli excursion and its "area" considered in [8]. Introduce the set of all $2n$-long sequences $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_{2n})$ such that $\varepsilon_i \in \{-1, 1\}$ and

$$\sum_{i=1}^{m} \varepsilon_i \begin{cases} \geq 0, & \text{if } 1 \leq m \leq 2n-1, \\ = 0, & \text{if } m = 2n. \end{cases} \tag{2.1}$$

The total number of such sequences is the $(n+1)$-th Catalan number $C_n = (n+1)^{-1}\binom{2n}{n}$. The Bernoulli excursion is the sequence $\boldsymbol{\varepsilon}$ chosen uniformly at random among all such sequences, and the "area" associated with $\boldsymbol{\varepsilon}$ is defined by

$$Y_n = Y_n(\boldsymbol{\varepsilon}) = \sum_{m=1}^{2n} \sum_{i=1}^{m} \varepsilon_i = \sum_{i=1}^{2n} (2n-i+1)\varepsilon_i = -\sum_{i=1}^{2n} i\varepsilon_i, \tag{2.2}$$

as $\sum_{i=1}^{2n} \varepsilon_i = 0$. Then $0 \leq Y_n \leq n^2$, with the upper bound attained at the sequence $\varepsilon_1 = \cdots = \varepsilon_n = 1$, $\varepsilon_{n+1} = \cdots = \varepsilon_{2n} = -1$. (The reader has certainly noticed that we are dealing with excursions of duration $2n$, rather than $2(n-1)$ mentioned in introduction. The difference is immaterial though.)

We need to prove

Theorem. *Uniformly for all* $0 \leq k \leq n^2$,

$$\mathbf{Pr}(Y_n = k) = O(n^{-3/2}).$$

The proof consists of three lemmas, first two basically probabilistic, and third analytic.

Introduce the sequence of i.i.d. random variables δ_i, $(i \ge 1)$, such that

$$\mathbf{Pr}(\delta_i = 1) = \mathbf{Pr}(\delta_i = -1) = \frac{1}{2}.$$

Let $D_m = \sum_{i=1}^m \delta_i$. Then the Bernoulli excursion $\boldsymbol{\varepsilon}$ has the same distribution as the sequence $\boldsymbol{\delta} = (\delta_1, \dots, \delta_{2n})$ conditioned on the event A_n given by the conditions $D_1, \dots, D_{2n-1} \ge 0$, $D_{2n} = 0$. So, introducing also $I_m = \sum_{i=1}^m i\delta_i$ and using (2.2), we see that

$$\mathbf{Pr}(Y_n = k) = \mathbf{Pr}(I_{2n} = -k | A_n). \tag{2.3}$$

Notice that

$$\mathbf{Pr}(A_n) = \frac{C_{n+1}}{2^{2n}} = \frac{1}{2^{2n}(n+1)}\binom{2n}{n} \sim \frac{1}{\sqrt{\pi}} n^{-3/2}.$$

So, by (2.3), it suffices to show that

$$P_n := \max_k \mathbf{Pr}\big(\{I_{2n} = k\} \cap A_n\big) = O(n^{-3}). \tag{2.4}$$

It is enough to consider the case n is even. Set $\nu = n/2$.

Lemma 1. *Uniformly for all integers a, b,*

$$P_n = O\left(n^{-1} \max_{a,b} \mathbf{Pr}(D_n = a, I_n = b)\right). \tag{2.4}$$

Proof of Lemma 1. Introduce

$$D'_m = \sum_{i=m}^{2n} \delta_i, \quad I'_m = \sum_{i=m}^{2n} i\delta_i, \quad (m \le 2n).$$

The conditions of the event A_n imply that

(2.5)

$$D_m \ge 0 \ \forall m \le \nu, \quad D'_m \le 0 \ \forall m > 3\nu; \quad \sum_{i=\nu+1}^{3\nu} \delta_i = -D_\nu - D'_{3\nu+1}, \sum_{i=\nu+1}^{3\nu} i\delta_i = k - I_\nu - I'_{3\nu+1}.$$

Denote by $\mathcal{A}_n$ the event obtained by dropping the third and the forth conditions in (2.5), so that $\mathcal{A}_n$ depend only on δ_i $(i \le \nu,$ or $i > 3\nu)$. Introduce

$$F(a,b) = \mathbf{Pr}\left(\sum_{i=\nu+1}^{3\nu} \delta_i = a, \ \sum_{i=\nu+1}^{3\nu} i\delta_i = b\right), \quad a, b \in \mathbf{Z};$$

obviously $F(a,b) = \mathbf{Pr}(D_n = a, I_n = b - \nu a)$. Since the middle δ_i $(\nu < i \le 3\nu)$ are independent of the left- and right-wing δ_i $(i \le \nu$ or $i > 3\nu)$, we have then

$$\begin{aligned}
&\mathbf{Pr}(\{I_{2n} = k\} \cap A_n) \\
&\le \mathbf{E}\left(\mathbf{1}_{\mathcal{A}_n}\mathbf{Pr}\left(\sum_{i=\nu+1}^{3\nu}\delta_i = -D_\nu - D'_{3\nu+1},\ \sum_{i=\nu+1}^{3\nu} i\delta_i = k - I_\nu - I'_{3\nu+1} \,\middle|\, \delta_i\,(i \le \nu \text{ or } i > 3\nu\right)\right) \\
&= \mathbf{E}\left(\mathbf{1}_{\mathcal{A}_n}F\big(-D_\nu - D'_{3\nu+1},\, k - I_\nu - I'_{3\nu+1}\big)\right) \\
&\le \mathbf{Pr}(\mathcal{A}_n)\cdot\max_{a,b} F(a,b).
\end{aligned}$$

It remains to notice that

$$\mathbf{Pr}(\mathcal{A}_n) = \mathbf{Pr}^2(D_m \ge 0\,\forall m \le \nu) = O(n^{-1}),$$

since (see e.g. Durrett [2], Ch. 3)

$$\mathbf{Pr}(D_m \ge 0\,\forall m \le \nu) = O\big(\nu^{-1/2}\big) = O\big(n^{-1/2}\big).$$

□

Next

Lemma 2. *Let $Z_1, \ldots, Z_n$ be i.i.d. random variables such that*

$$\mathbf{Pr}(Z_1 = 1) = \mathbf{Pr}(Z_1 = -1) = 1/4, \quad \mathbf{Pr}(Z_1 = 0) = 1/2.$$

Denote $\mathcal{S}_m = \sum_{i=1}^m (2i-1)Z_i$. Then there exists $c > 0$ such that

$$\max_{a,b}\mathbf{Pr}(D_n = a, I_n = b) = O\Big(e^{-cn} + n^{-1/2}\max_b \mathbf{Pr}(\mathcal{S}_\nu = b)\Big). \tag{2.6}$$

Proof of Lemma 2. Clearly $\boldsymbol{\delta}$ has the same distribution as $(\delta_1, \ldots, \delta_\nu; \delta'_\nu\delta_\nu, \ldots, \delta'_1\delta_1)$, where $(\delta'_1, \ldots \delta'_\nu)$ is an independent copy of $(\delta_1, \ldots, \delta_\nu)$. The corresponding sums are

$$\begin{aligned}
D_n &= \sum_{i=1}^{\nu}(1+\delta'_i)\delta_i; \\
I_n &= \sum_{i=1}^{\nu}\big[i + (n-i+1)\delta'_i\big]\delta_i \\
&= \sum_{i=1}^{\nu}(2i-n-1)\frac{(1-\delta'_i)\delta_i}{2} + (n+1)\sum_{i=1}^{\nu}\frac{(1+\delta'_i)\delta_i}{2}.
\end{aligned}$$

Consequently the conditions $D_n = a$, $I_n = b$ are equivalent to

$$U_n := \sum_{i=1}^{\nu} \frac{(1+\delta_i')\delta_i}{2} = a', \quad V_n := \sum_{i=1}^{\nu} (2i-n-1)\frac{(1-\delta_i')\delta_i}{2} = b', \tag{2.7}$$

where $a' = a/2$, $b' = b-(n+1)a'$. A key observation here is that, conditioned on $\boldsymbol{\delta}' = (\delta_1', \dots, \delta_\nu')$, U_n and V_n are independent, since U_n depends on $\{\delta_i : \delta_i' = 1\}$, and V_n on $\{\delta_i : \delta_i' = -1\}$, and the two sets are disjoint! Furthermore, given $\boldsymbol{\delta}'$, U_n has the distribution of the number of heads minus number of tails in $\nu(\boldsymbol{\delta}') = \sum_{i=1}^{\nu}(1+\delta_i')/2$ coin tosses, whence

$$\max_a \mathbf{Pr}(U_n = a' | \boldsymbol{\delta}') = O(1+\nu(\boldsymbol{\delta}'))^{-1/2}, \tag{2.8}$$

uniformly for all a and $\boldsymbol{\delta}'$. In its turn, $\nu(\boldsymbol{\delta}')$ is binomial, with parameters ν and $1/2$. So the term on the right side of (2.8) is at most $c'/n^{1/2}$, with probability $1-e^{-cn}$ at least; here c and c' are two positive constants. We conclude then that

$$\begin{aligned} \mathbf{Pr}(U_n = a', V_n = b') =& O\left(e^{-cn} + n^{-1/2}\mathbf{E}\big(\mathbf{Pr}(V_n = b'|\boldsymbol{\delta}')\big)\right) \\ =& O\left(e^{-cn} + n^{-1/2}\mathbf{Pr}\big(V_n = b'\big)\right). \end{aligned} \tag{2.9}$$

Now the random variables $(1-\delta_i')\delta_i/2$ in the formula (2.7) for V_n are independent, and distributed as Z_1, whence as $-Z_1$. Thus V_n has the same distribution as

$$-\sum_{i=1}^{\nu}(2i-n-1)Z_{\nu-(i-1)} = \sum_{j=1}^{\nu}(2j-1)Z_j,$$

and (2.9) implies (2.6). □

From Lemma 1 and Lemma 2 we get immediately

Corollary. *There exists* $c' > 0$ *such that*

$$P_n = O\big(e^{-c'n} + n^{-3/2}\max_b \mathbf{Pr}(\mathcal{S}_\nu = b)\big).$$

Thus the bound $P_n = O(n^{-3})$ will be proven once we establish the last

Lemma 3.

$$\max_k \mathbf{Pr}(\mathcal{S}_n = k) = O(n^{-3/2}). \tag{2.10}$$

Proof of Lemma 3. By the definition of $\mathcal{S}_n$, the characteristic function $f_n(\theta) = \mathbf{E}\exp(i\theta\mathcal{S}_n)$ is given by

$$\begin{aligned} f_n(\theta) &= \prod_{j=1}^{n} \mathbf{E}\left[\exp\bigl(i\theta(2j-1)Z_j\bigr)\right] \\ &= \prod_{j=1}^{n}\left(\frac{1}{4}\exp(-i(2j-1)\theta) + \frac{1}{2} + \frac{1}{4}\exp(i(2j-1)\theta\right) \\ &= \prod_{j=1}^{n} \cos^2\bigl((j-1/2)\theta\bigr). \end{aligned}$$

By the Fourier inversion formula,

$$\mathbf{Pr}(\mathcal{S}_n = k) = \frac{1}{2\pi i}\int_{-\pi}^{\pi} e^{-ik\theta} f_n(\theta)\, d\theta \le \frac{1}{2\pi}\int_{-\pi}^{\pi} f_n(\theta)\, d\theta. \tag{2.11}$$

Let us bound $f_n(\theta)$ from above. We have

$$\begin{aligned} f_n(\theta) &= \prod_{j=1}^{n}\bigl(1 - \sin^2((j-1/2)\theta\bigr) \le \exp(-g_n(\theta)); \\ g_n(\theta) &:= \sum_{j=1}^{n} \sin^2((j-1/2)\theta). \end{aligned}$$

Here

$$\begin{aligned} g_n(\theta) =& \frac{1}{2}\sum_{j=1}^{n}\bigl(1 - \cos((2j-1)\theta\bigr) \\ =& \frac{1}{2}\left[n - \Re\sum_{j=1}^{n}\exp\bigl(i(2j-1)\theta\bigr)\right] \\ =& \frac{1}{2}\left[n - \Re\frac{e^{i\theta}(1-e^{i2n\theta})}{1-e^{i2\theta}}\right] \\ =& \frac{1}{2}\left[n - \frac{\sin 2n\theta}{2\sin\theta}\right]. \end{aligned}$$

We have assumed that $\theta \ne 0, \pm\pi$. We see directly that $g_n(\theta) = 0$ for $\theta = 0$, and $g_n(\pm\pi) = n$. Suppose first that $|\sin\theta| \ge \frac{2}{3n}$. Then

$$g_n(\theta) \ge \frac{1}{2}\left(n - \frac{|\sin 2n\theta|}{2|\sin\theta|}\right) \ge \frac{1}{2}(n - 3n/4) = n/8,$$

and the contribution of those θ's to the integral in (2.11) is $O\big(e^{-n/8}\big)$.
Consider now the case $|\sin\theta| \le \frac{2}{3n}$. For n sufficiently large, this condition implies that either (1) $|\theta| \le n^{-1}$, or (2) $|\pm\pi - \theta| \le n^{-1}$. In the case (2), denoting $\varepsilon_n = \pm\pi - \theta$, so that $|\varepsilon_n| \le n^{-1}$, we have: for $\varepsilon_n \ne 0$,

$$\frac{\sin(2n\theta)}{2\sin 2\theta} = \frac{\sin(\pm 2n\pi - 2n\varepsilon_n)}{2\sin(\pm\pi - \varepsilon_n)} = -\frac{\sin(2n|\varepsilon_n|)}{2\sin|\varepsilon_n|} < 0,$$

since $|\varepsilon_n|$, $2n|\varepsilon_n| \in (0,\pi)$. So here $g_n(\theta) \ge n/2$, hence those θ's contribute only $O(e^{-n/2})$ to the integral. Consider the case (1). Introduce a new variable $\eta = n\theta$, so that $|\eta| \le 1$. Then, setting $G_n(\eta) = g_n(\theta)$,

$$\begin{aligned} G_n(\eta) =& \frac{1}{2}\left[n - \frac{\sin 2\eta}{2\sin(\eta/n)}\right] \\ =& \frac{n}{2}\left(1 - \frac{\sin 2\eta}{2\eta}\right) + O(n^{-1}). \end{aligned}$$

Here $1 - \sin 2\eta/(2\eta)$ attains its zero minimum value at $\eta = 0$, and

$$1 - \frac{\sin 2\eta}{2\eta} \ge \frac{2\eta^2}{3}.$$

Therefore we bound the contribution of $|\theta| \le n^{-1}$ to the integral in (2.11) as follows:

$$\begin{aligned} \int\limits_{|\theta|\le n^{-1}} f_n(\theta)\,d\theta \le& \frac{1}{n}\int\limits_{|\eta|\le 1} \exp(-G_n(\eta))\,d\eta \\ \le& \frac{2}{n}\int\limits_{|\eta|\le 1} \exp\big(-n\eta^2/3\big)\,d\eta \\ =& O(n^{-3/2}). \end{aligned}$$

Thus, uniformly for all k, $\mathbf{Pr}(\mathcal{S}_n = k) = O(n^{-3/2})$. □

Note. The inequality

$$\mathbf{E}\exp(i\theta\mathcal{S}_n) \le \exp\left(-\frac{1}{2}\big(n - \sin 2n\theta/(2\sin\theta)\big)\right),$$

crucial in our argument, could have been used to prove a local limit theorem for $\mathcal{S}_n$ that is stronger than (2.10):

$$\limsup_{n\to\infty}\max_{|k|\le n^2}\left|\sigma_n\mathbf{Pr}(\mathcal{S}_n = k) - \frac{\exp(-k^2/(2\sigma_n^2))}{\sqrt{2\pi}}\right| = 0, \quad \sigma_n^2 = \mathbf{Var}\mathcal{S}_n.$$

As an afterthought, we could have obtained in exactly the same way an analogous bound for the *joint* characteristic function

$$\mathbf{E}\exp\big(i(\theta I_n+\psi D_n)\big)=\prod_{j=1}^{n}\cos(j\theta+\psi), \tag{2.12}$$

namely

$$\left|\prod_{j=1}^{n}\cos(j\theta+\psi)\right|\le\exp\left(-\frac{1}{4}\left(n-\frac{\sin n\theta\cos(n\theta+2\psi)}{\sin\theta}\right)\right).$$

And then we could have proved directly that

$$\max_{a,b}\mathbf{Pr}(D_n=a,I_n=b)=O\big(n^{-2}\big),$$

thus not using Lemma 2. We have decided to keep the initial argument, as more probabilistically revealing. Besides, had it not been for the *squared* cosines in the formula for $\mathbf{E}\exp(i\theta\mathcal{S}_n)$, we might have overlooked the surprisingly simple, yet efficient, bound whose derivation started with using $\cos^2\alpha+\sin^2\alpha=1$! Our inspiration came from studying Section 6 of [4]; in fact, the bound (2.12) might have simplified considerably the difficult line of argument pursued there.

Acknowledgement. The second author is grateful to Eric Schmutz for drawing his attention to the Kleitman-Winston conjecture. The research was done during the second author visit to Microsoft, and the generous hospitality of Jennifer Chayes, Christian Borgs, and the coauthor is greatly appreciated. We thank a referee for carefully reading the paper and making useful comments.

References

1. R. Canfield, *Central and local limit theorems for the coefficients of polynomials of binomial type*, J. Comb. Theory, A **23** (1977,), 275–290.
2. R. Durrett, *Probability: Theory and Examples*, Wadsworth& Brooks/Cole Advanced Books & Software, Pacific Grove, California, 1991.
3. L. R. Ford and D. R. Fulkerson, *Flows in Networks*, Princeton University Press, Princeton, 1962.
4. D. J. Kleitman and K. J. Winston, *On the asymptotic number of tournament score sequences*, J. Comb. Theory, A **35** (1983), 208–230.
5. H. G. Landau, *The condition for a score structure*, Bull. M. Biophys. **15** (1953), 143–148.
6. J. W. Moon, *Topics on Tournaments*, Holt, Rinehart & Winston, New York, 1968.
7. L. Moser, *Asymptotics of tournament scores*, Combinatorics (Proc. Sympos. Pure Math. vol. XIV), Univ. California, Los Angeles, 1968), Amer. Math. Soc. Providence, R. I., 1971, pp. 165–166.
8. L. Takács, *A Bernoulli excursion and its various applications*, Adv. Appl. Prob. **23** (1991), 557–585.

Microsoft, One Microsoft Way, Redmond, Washington 98052-6399

Department of Mathematics, Ohio State University, Columbus, Ohio, 43210-1174