THE LIMITING DISTRIBUTION OF THE COEFFICIENTS OF THE \( q \)-CATALAN NUMBERS

WILLIAM Y. C. CHEN, CAROL J. WANG, AND LARRY X. W. WANG

(Communicated by Jim Haglund)

Abstract. We show that the limiting distributions of the coefficients of the \( q \)-Catalan numbers and the generalized \( q \)-Catalan numbers are normal. Despite the fact that these coefficients are not unimodal for small \( n \), we conjecture that for sufficiently large \( n \), the coefficients are unimodal and even log-concave except for a few terms of the head and tail.

Introduction

The main objective of this paper is to show that the limiting distribution of the coefficients of the \( q \)-Catalan numbers is normal. The Catalan numbers

\[ C_n = \frac{1}{n+1} \binom{2n}{n} \]

have many combinatorial interpretations; see Stanley [10]. The usual \( q \)-analog of the Catalan numbers is given by

\[ C_n(q) = \frac{1}{[n+1]} \binom{2n}{n}, \]

where \( [n] = 1 + q + q^2 + \cdots + q^{n-1} \), and

\[ \binom{n}{k} = \frac{[n]!}{[k]![n-k]!}. \]

There are also other types of \( q \)-analogs of the Catalan numbers; see, for example, Andrews [2], Gessel and Stanton [4], Krattenthaler [5].

We further consider the limiting distribution of the coefficients of the quotient of two products, which includes the result for the \( q \)-Catalan numbers as a special case. We conclude this paper with two conjectures on the unimodality and log-concavity for almost all the coefficients of the \( q \)-Catalan numbers and the generalized \( q \)-Catalan numbers provided that \( n \) is sufficiently large.
The limiting distribution

In this section, we use the moment generating function technique to obtain the limiting distribution of the coefficients of the \(q\)-Catalan numbers. We introduce the random variable \(\xi_n\) corresponding to the probability generating function
\[
\phi_n(q) = \frac{C_n(q)}{C_n}.
\]
As far as the computations are concerned, we will not need the following combinatorial interpretation of \(C_n(q)\). However, for the sake of completeness, we recall that \(\xi_n\) reflects the distribution of the major indices of Catalan words of length \(2n\); see, for example, [3]. Moreover, we write
\[
C_n(q) = \sum_{m_n(k)} q^k,
\]
where \(m_n(k)\) stands for the number of Catalan words of length \(2n\) with major index \(k\). The following lemma gives the expectation and variance of \(\xi_n\).

**Lemma 2.1.** We have
\[
E(\xi_n) = \frac{n(n-1)}{2}, \quad \text{and} \quad \text{Var}(\xi_n) = \frac{n(n-1)(n+1)}{6}.
\]

**Proof.** By the definition of \(C_n(q)\), it is easy to check the following symmetry property of \(m_n(k)\):
\[
m_n(k) = m_n(n(n-1)-k).
\]
Hence
\[
E(\xi_n) = \frac{n(n-1)}{2}.
\]
Let
\[
F = F(q) = \prod_{i=1}^{n-1} (1 + q + \cdots + q^{n+i}) \quad \text{and} \quad G = G(q) = \prod_{i=1}^{n-1} (1 + q + \cdots + q^i).
\]
It is easily verified that \(C_n(q) = F/G\). Since
\[
C_n(q)''|_{q=1} = \left. \left(\frac{F''}{G} - \frac{F'G''}{G^2} - \frac{2G'F'}{G^2} + \frac{2G''F}{G^3} \right) \right|_{q=1}
\]
we obtain
\[
\text{Var}(\xi_n) = \frac{C_n(q)''|_{q=1} + E(\xi_n) - E(\xi_n)^2}{C_n} = \frac{1}{6} n(n-1)(n+1).
\]
This completes the proof. \(\Box\)

**Lemma 2.2.** When \(n \to \infty\), we have
\[
\sum_{k=2}^{\infty} B_{2k} \frac{t^{2k}}{2k(2k)!} \sigma^{2k} \sum_{i=2}^{n} ((n+i)^{2k} - i^{2k}) \to 0
\]
uniformly for \(t\) from any bounded set, where the \(B_j\)’s are the Bernoulli numbers and \(\sigma^2\) is the variance of \(\xi_n\) as given in (2.1).
Proof: The second summation can be expanded as follows:

\[
\sum_{i=2}^{n} \left( (n+i)^{2k} - i^{2k} \right) = \sum_{i=2}^{n} \sum_{j=1}^{2k} \left( \binom{2k}{j} \right) n^{j-1} (2k-j) \left( \sum_{i=2}^{n} i^{j-1} \right).
\]

For \( k > 1 \), the second factor in the preceding summation is bounded by the following integral:

\[
\sum_{i=2}^{n} n^{j-1} \left( 2k-j \right) < n^j \int_{1}^{n+1} t^{2k-j} \, dt = n^j \cdot \frac{(n+1)^{2k-j+1} - 1}{2k-j+1}.
\]

Consequently,

\[
\sum_{i=2}^{n} \left( (n+i)^{2k} - i^{2k} \right) < 2^{2k} (n+1)^{2k+1} - 8^{2k} n^{2k+1}.
\]

Since \( \sigma^2 = \frac{n^3-n}{6} > \frac{n^3}{8} \) when \( n \) is sufficiently large, we have

\[
\sigma^{-2k} \sum_{i=2}^{n} \left( (n+i)^{2k} - i^{2k} \right) < 64^{2k} n^{1-k} \leq n^{-1/3} 64^{2k} n^{-k/3},
\]

for large \( n \) and \( k > 1 \). Thus

\[
\left| \sum_{2|k,k\geq3} B_{2k} \frac{t^{2k}}{2k(2k)!} \sum_{i=2}^{n} (n+i)^{2k} - i^{2k} \right| < n^{-1/3} \sum_{2|k,k\geq3} \frac{|B_{2k}| t^{2k}}{2k(2k)!} \frac{64^{2k} n^{-k/3}}{2k(2k)!}
\]

\[
= n^{-1/3} \sum_{2|k,k\geq3} |B_{2k}| \frac{(64tn^{-\frac{1}{2}})^{2k}}{2k(2k)!}.
\]

In view of the following asymptotic expansion of the Bernoulli numbers [1],

\[
|B_{2n}| \sim \frac{2(2n)!}{(2\pi)^{2n}},
\]

the convergent radius \( R \) of the series \( \sum_{2|k,k\geq3} |B_{2k}| \frac{t^{2k}}{2k(2k)!} \) equals \( 2\pi \). Since \( t \) is from a bounded set, when \( n \) is large enough, the series

\[
\sum_{2|k,k\geq3} |B_{2k}| \frac{(64tn^{-\frac{1}{2}})^{2k}}{2k(2k)!}
\]

converges. Moreover, it is evident that \( 64tn^{-\frac{1}{2}} < 1 \); we can bound the above summation by the constant

\[
M_1 = \sum_{2|k,k\geq3} |B_{2k}| \frac{1}{2k(2k)!}.
\]

Similarly, it can be deduced that

\[
\sum_{2|k,k\geq3} B_{2k} \frac{t^{2k}}{2k(2k)!} \sigma^{2k} \sum_{i=2}^{n} (n+i)^{2k} - i^{2k} < \frac{M_2}{n^{\frac{k}{2}}}
\]
Proof. Let \( M_2 = \sum_{2^{k}, k \geq 2} B_{2k} \frac{1}{2k(2k)!} \) be a constant. Hence

\[
\sum_{k=2}^{\infty} B_{2k} \frac{\epsilon^{2k}}{2k(2k)!} \sigma^{2k} \sum_{i=2}^{n} \left( (n+i)^{2k} - i^{2k} \right) < \frac{M_1 + M_2}{n^{1/3}},
\]

which tends to zero as \( n \to \infty \). This completes the proof. \( \square \)

In [7], Margolius used Bernoulli numbers to show that the distribution of the number of inversions in a random permutation is asymptotically normal. In [6], Louchard and Prodinger used the saddle point method to derive some stronger results. Based on Lemma 2.2, we obtain the following theorem.

**Theorem 2.3.** When \( n \to \infty \), the random variable

\[
\eta_n = \frac{\xi_n - E(\xi_n)}{\sqrt{\text{Var}(\xi_n)}}
\]

has the standard normal distribution.

**Proof.** Let \( M_n(q) \) denote the moment generating function of \( \xi_n \). Then we have \( M_n(q) = \phi_n(e^q) \); see Sachkov [8]. Hence

\[
M_n(q) = \frac{n+1}{(2n)} \frac{1 - e^q}{1 - e^{(n+1)q}} \prod_{i=1}^{n} \frac{1 - e^{(n+i)q}}{1 - e^{iq}}
\]

\[
= \prod_{i=2}^{n} \frac{i}{n+i} \cdot \frac{n}{\prod_{i=2}^{n} 1 - e^{(n+i)q}}
\]

\[
= \prod_{i=2}^{n} \frac{(1 - e^{(n+i)q})/(n+i)}{(1 - e^{iq})/i}
\]

\[
= \exp \left\{ \frac{1}{2} \sum_{i=2}^{n} ((n+i)q - iq) \right\} \prod_{i=2}^{n} \left( \frac{e^{(n+i)q/2} - e^{-(n+i)q/2}/i}{e^{iq/2} - e^{-iq/2}/i} \right)^{1/2}
\]

\[
= \exp \left\{ \frac{n(n-1)q}{2} \right\} \prod_{i=2}^{n} \frac{\sinh ((n+i)q/2)/i}{\sinh (iq/2)/i}.
\]

Recalling the following relation on the Bernoulli numbers [7]

\[
(2.2) \quad \ln \left( \frac{\sinh(x/2)}{x/2} \right) = \sum_{k=1}^{\infty} B_{2k} \frac{x^{2k}}{2k(2k)!},
\]

we find that

\[
\ln M_n(q) = \frac{n(n-1)}{2} q + \sum_{i=2}^{n} \left( \ln \left( \frac{\sinh((n+i)q/2)}{(n+i)/2} \right) - \ln \left( \frac{\sinh(iq/2)}{i/2} \right) \right)
\]

\[
= \frac{n(n-1)}{2} q + \sum_{k=1}^{\infty} B_{2k} \frac{q^{2k}}{2k(2k)!} \sum_{i=2}^{n} ((n+i)^{2k} - i^{2k}).
\]

Setting \( q = t/\sigma \), where \( \sigma \) is the standard deviation of \( \xi_n \) as given in Lemma 2.1, we are led to the expansion

\[
\ln M_n(t/\sigma) = \frac{n(n-1)t}{2\sigma} + \sum_{k=1}^{\infty} B_{2k} \frac{t^{2k}}{2k(2k)!} \sigma^{2k} \sum_{i=2}^{n} ((n+i)^{2k} - i^{2k}).
\]
Applying Lemma 2.2, we have, when \( n \to \infty \),
\[
\sum_{k=2}^{\infty} B_{2k} \frac{t^{2k}}{2k(2k)! \sigma^{2k}} \sum_{i=2}^{n} ((n + i)^{2k} - i^{2k}) \to 0
\]
uniformly for \( t \) from any bounded set. Finally,
\[
\lim_{n \to \infty} M_n(t/\sigma) \exp \left\{ -\frac{n(n-1)t}{2\sigma} \right\}
\]
\[
= \lim_{n \to \infty} \exp \left\{ \sum_{k=1}^{\infty} B_{2k} \frac{t^{2k}}{2k(2k)! \sigma^{2k}} \sum_{i=2}^{n} ((n + i)^{2k} - i^{2k}) \right\}
\]
\[
= \lim_{n \to \infty} \exp \left\{ B_2 \frac{t^2}{2(2)! \sigma^2} \sum_{i=2}^{n} ((n + i)^2 - i^2) \right\}
\]
\[
= e^{t^2/2},
\]
which coincides with the moment generating function of the standard normal distribution. Employing Curtiss’s theorem [8], we reach the conclusion that \( \eta_n \) has the standard normal distribution when \( n \) approaches infinity. \( \square \)

**A general setting**

In this section, we will determine the limiting distribution of the coefficients of a quotient of products and will give two special cases.

**Theorem 3.1.** Let \( a_1, a_2, a_3, \ldots \) and \( b_1, b_2, b_3, \ldots \) be two sequences of positive numbers, and let
\[
\phi_n(x) = \sum_{k} p_n(k) x^k = \frac{(1 - q^{a_1})(1 - q^{a_2}) \cdots (1 - q^{a_n})}{(1 - q^{b_1})(1 - q^{b_2}) \cdots (1 - q^{b_n})}.
\]
Suppose that \( \xi_n \) is the random variable corresponding to the generating function \( \phi_n(x) \), that is,
\[
P(\xi_n = k) = \frac{p_n(k)}{\sum_k p_n(k)}.
\]
Then \( \xi_n \) is normally distributed as \( n \to \infty \) if and only if
\[
\sum_{k=1}^{\infty} B_{2k} \frac{t^{2k}}{2k(2k)!} \left( \sum_{i=1}^{n} (a_i^{2k} - b_i^{2k}) \right) \frac{1}{\left( \sum_{i=1}^{n} (a_i^2 - b_i^2) \right)^k} \to 0 \text{ as } n \to \infty.
\]

**Proof.** The expectation of \( \xi_n \) is easy to compute, as given below:
\[
E(\xi_n) = \phi_n(x)^{q=1} = \frac{1}{2} \sum_{i=1}^{n} (a_i - b_i).
\]
Proceeding analogously as in the proof of Lemma 2.1, we find
\[
\sigma^2 = \text{Var}(\xi_n) = \frac{1}{12} \sum_{i=1}^{n} (a_i^2 - b_i^2).
\]
Hence,
\[
B_2 \frac{t^2}{2(2)!\sigma^2} \left( \sum_{i=1}^{n} (a_i^2 - b_i^2) \right) = \frac{1}{6} \cdot \frac{t^2}{4} \cdot \frac{1}{\sigma^2} \left( \sum_{i=1}^{n} (a_i^2 - b_i^2) \right) \cdot \left( \sum_{i=1}^{n} (a_i^2 - b_i^2) \right) = \frac{t^2}{2}.
\]

By the same procedure as in the proof of Theorem 2.3, we obtain
\[
\lim_{n \to \infty} M_n(t/\sigma) \exp \left\{ \frac{1}{2} \sum_{i=1}^{n} (a_i^2 - b_i^{2k}) \right\} = e^{t^2/2} \lim_{n \to \infty} \exp \left\{ \sum_{k=2}^{\infty} B_{2k} \frac{t^{2k}}{2k(2k)!\sigma^{2k}} \left( \sum_{i=1}^{n} (a_i^{2k} - b_i^{2k}) \right) \right\}.
\]
It follows that the limiting distribution of \( p_n(k) \) is normal if and only if
\[
\sum_{k=2}^{\infty} B_{2k} \frac{t^{2k}}{2k(2k)!\sigma^{2k}} \left( \sum_{i=1}^{n} (a_i^{2k} - b_i^{2k}) \right) \to 0 \quad \text{as} \quad n \to \infty,
\]
for \( t \) from any bounded set. By virtue of the variance formula (3.1), the condition (3.2) is equivalent to
\[
\sum_{k=1}^{\infty} B_{2k} \frac{t^{2k}}{2k(2k)!} \left( \frac{\sum_{i=1}^{n} (a_i^{2k} - b_i^{2k})}{\sum_{i=1}^{n} (a_i^2 - b_i^2)} \right)^k \to 0 \quad \text{as} \quad n \to \infty
\]
for \( t \) from any bounded set. Thus (3.2) is verified. This completes the proof. □

**Corollary 3.2.** Let \( p_n(k) \) be given as in the above theorem. Suppose that for \( k \geq 2 \), there exist constants \( \alpha > 0 \), \( \beta < 0 \) and \( \gamma < 0 \) such that
\[
\frac{\sum_{i=1}^{n} (a_i^{2k} - b_i^{2k})}{\left( \sum_{i=1}^{n} (a_i^2 - b_i^2) \right)^k} < n^{\gamma} (\alpha n^\beta)^{2k},
\]
for \( t \) from any bounded set. Then the limiting distribution of \( p_n(k) \) is normal.

**Proof.** Note that the convergent radius \( R \) of the series
\[
\sum_{2k,k \geq 3} |B_{2k}| \frac{x^{2k}}{2k(2k)!}
\]
is \( 2\pi \). If (3.4) holds for \( k > 1 \), then for \( t \) from any bounded set, and for sufficiently large \( n \), we have
\[
\left| t^{2k} \frac{\sum_{i=1}^{n} (a_i^{2k} - b_i^{2k})}{\sigma^{2k}} \right| \leq n^{\gamma} (\tan^\beta)^{2k},
\]
where \( \tan^\beta < 2\pi \). It is clear that \( n^{\gamma} \to 0 \) since \( \gamma < 0 \). □

If we choose \( \alpha = 32\sqrt{3}/3 \), \( 2\beta = \gamma = -\frac{4}{3} \), Corollary 3.2 contains Theorem 2.3 as a special case. We now give two more examples. One is the following \( q \)-analog of the Catalan numbers:
\[
c_n(q) = \frac{[2]}{[2n]} \frac{2n}{n-1},
\]
which are symmetric and unimodal; see Stanley [9].

Using Theorem 3.1, we reach the following assertion.

**Corollary 3.3.** The distribution of the coefficients in \( c_n(q) \) is asymptotically normal.

**Proof.** First, we write \( c_n(q) \) in the following form:

\[
\frac{\prod_{i=1}^{n} (1 - q^{n+i-1})}{(1 - q) \prod_{i=3}^{n-1} (1 - q^i)}
\]

Set \( a_1 = a_2 = 1, a_i = n + i - 1, 3 \leq i \leq n, \) and \( b_1 = b_2 = 1, b_i = i - 1, 4 \leq i \leq n. \) Then we have

\[
\sum_{i=1}^{n} (a_i^{2k} - b_i^{2k}) = (a_3^{2k} - b_3^{2k}) + \sum_{i=4}^{n} (a_i^{2k} - b_i^{2k})
\]

\[
= (n + 2)^{2k} - 1 + \sum_{i=3}^{n-1} ((n + i)^{2k} - i^{2k})
\]

and

\[
\left( \sum_{i=1}^{n} (a_i^2 - b_i^2) \right)^k = \left( (n + 2)^2 - 1 + \sum_{i=3}^{n-1} ((n + i)^2 - i^2) \right)^k
\]

\[
= (n - 1)^k (n + 1)^k (2n - 3)^k.
\]

By the same arguments as in the proof of Lemma 2.2, we may set \( \alpha = 32\sqrt{3}/3 \) and \( 2\beta = \gamma = -\frac{4}{3} \) such that the condition (3.4) is satisfied. Therefore, Theorem 3.1 implies the limiting distribution of the coefficients of \( c_n(q) \). \( \square \)

The \( m \)-Catalan numbers are defined by

\[
C_{n,m} = \frac{1}{(m - 1)n + 1} \binom{mn}{n},
\]

for \( n \geq 1. \) Accordingly, the generalized \( q \)-Catalan numbers are given by

\[
C_{n,m}(q) = \frac{1}{[(m - 1)n + 1]} \left\lfloor \binom{mn}{n} \right\rfloor.
\]

Theorem 3.1 has the following consequence.

**Corollary.** The coefficients of the generalized \( q \)-Catalan numbers \( C_{n,m}(q) \) are normally distributed when \( n \to \infty. \)

**Proof.** First, express \( C_{n,m}(q) \) as follows:

\[
\prod_{i=2}^{n} \frac{1 - q^{(m-1)n+i}}{1 - q^i}.
\]

Set \( a_1 = 1, a_i = (m - 1)n + i, 2 \leq i \leq n, \) and \( b_1 = 1, b_i = i, 2 \leq i \leq n. \) Then we have

\[
\sum_{i=1}^{n} (a_i^{2k} - b_i^{2k}) = \sum_{i=2}^{n} (a_i^{2k} - b_i^{2k}) = \sum_{i=2}^{n} \sum_{j=1}^{2k} \binom{2k}{j} ((m - 1)n)^{2k-j} i^j.
\]
The same argument as in the proof of Lemma 2.2 yields the following bound:

$$\sum_{i=1}^{n} (a_i^{2k} - b_i^{2k}) < 8^{2k} ((m - 1)n)^{2k+1}.$$  

Now,

$$\left(\sum_{i=1}^{n} (a_i^2 - b_i^2)\right)^k = \left(\sum_{i=2}^{n} ((m - 1)n + i)^2 - i^2\right)^k > (m - 1)^{2k} n^{2k} (n - 1)^k > (m - 1)^{2k+1} n^{3k} / (2m)^k.$$  

It follows that

$$\frac{\sum_{i=1}^{n} (a_i^{2k} - b_i^{2k})}{(\sum_{i=1}^{n} (a_i^2 - b_i^2))^k} < (8\sqrt{2m})^{2k} n^{1-k}.$$  

Again, by the same arguments as in the proof of Lemma 2.2, we may set $\alpha = 8\sqrt{2m}$ and $2\beta = \gamma = -\frac{1}{3}$ such that the condition (3.4) holds. Finally, we may use Theorem 3.1 to get the desired distribution. \[\Box\]

**Open problems**

While the $q$-Catalan numbers are not unimodal for small $n$, see Stanley [9], the limiting distribution suggests that the coefficients are almost unimodal in a certain sense for sufficiently large $n$. Obviously, the first and the last term should not be taken into account; otherwise one can never expect to have unimodality. In fact, an easy computation indicates that $C_n(q)$ are unimodal for $n \geq 16$.

**Conjecture 4.1.** The sequence $\{m_n(1), \ldots, m_n(n(n - 1) - 1)\}$ is unimodal when $n$ is sufficiently large.

When $n > 70$, numerical evidence is suggestive of a stronger conjecture:

**Conjecture 4.2.** There exists an integer $t$ such that when $n$ is sufficiently large, the sequence $\{m_n(t), \ldots, m_n(n(n - 1) - t)\}$ is log-concave, namely,

$$(m_n(k))^2 \geq m_n(k+1)m_n(k-1)$$

for $t + 1 \leq k \leq n(n - 1) - t - 1$. Moreover, the minimum value of $t$ seems to be 75.

We also conjecture that similar properties hold for the generalized $q$-Catalan numbers.

**References**


Center for Combinatorics, LPMC-TJKLC, Nankai University, Tianjin 300071, People’s Republic of China

E-mail address: chen@nankai.edu.cn

E-mail address: wangjian@cfc.nankai.edu.cn

E-mail address: wxw@cfc.nankai.edu.cn

License or copyright restrictions may apply to redistribution; see http://www.ams.org/journal-terms-of-use