THE LIMITING DISTRIBUTION OF THE COEFFICIENTS OF THE q-CATALAN NUMBERS

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

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.

.

1. 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

$$C_n(q) = \frac{1}{[n+1]}\begin{bmatrix}2n\\n\end{bmatrix}, \tag{1.1}$$

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

$$\begin{bmatrix}n\\k\end{bmatrix} = \frac{[n]!}{[k]![n-k]!}.$$

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.

Date: August 17, 2007; in revised form, December 17, 2007.

2000 *Mathematics Subject Classification.* Primary 05A16, 60C05.

Key words and phrases. Bernoulli number, q-Catalan number, unimodality, log-concavity, moment generating function.

This work was supported by the 973 Project, the PCSIRT Project of the Ministry of Education, the Ministry of Science and Technology, and the National Science Foundation of China.

2. 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 ξ_n corresponding to the probability generating function

$$\phi_n(q) = C_n(q)/C_n.$$

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 ξ_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 ξ_n.

Lemma 2.1. *We have*

$$E(\xi_n) = \frac{n(n-1)}{2} \quad \textit{and} \quad \mathrm{Var}(\xi_n) = \frac{n(n-1)(n+1)}{6}. \tag{2.1}$$

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

$$\begin{aligned} C_n(q)''|_{q=1} &= \left. \left(\frac{F''}{G} - \frac{FG''}{G^2} - \frac{2G'F'}{G^2} + \frac{2G'^2F}{G^3} \right) \right|_{q=1} \\ &= \frac{1}{12} n(n-1)(3n^2 - n - 4) C_n, \end{aligned}$$

we obtain

$$\mathrm{Var}(\xi_n) = \frac{C_n(q)''|_{q=1}}{C_n} + E(\xi_n) - E(\xi_n)^2 = \frac{1}{6} n(n-1)(n+1).$$

This completes the proof. ∎

Lemma 2.2. *When $n \to \infty$, we have*

$$\sum_{k=2}^{\infty} B_{2k} \frac{t^{2k}}{2k(2k)!\sigma^{2k}} \sum_{i=2}^{n} \left((n+i)^{2k} - i^{2k} \right) \to 0$$

uniformly for t from any bounded set, where B_j's are the Bernoulli numbers and σ^2 is the variance of ξ_n as given in (2.1).

Proof. The second summation can be expanded as follows:

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

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

$$\sum_{i=2}^{n}n^j i^{2k-j}<n^j\int_1^{n+1}t^{2k-j}dt=n^j\cdot\frac{(n+1)^{2k-j+1}-1}{2k-j+1}.$$

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

$$\begin{aligned}&\left|\sum_{2\nmid k,k\geq 3}B_{2k}\frac{t^{2k}}{2k(2k)!\sigma^{2k}}\sum_{i=2}^{n}\left((n+i)^{2k}-i^{2k}\right)\right|\\ &<\ n^{-1/3}\sum_{2\nmid k,k\geq 3}|B_{2k}|\frac{t^{2k}}{2k(2k)!}64^{2k}n^{-k/3}\\ &=\ n^{-1/3}\sum_{2\nmid k,k\geq 3}|B_{2k}|\frac{(64tn^{-\frac{1}{6}})^{2k}}{2k(2k)!}.\end{aligned}$$

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 $\sum_{2\nmid k,k\geq 3}|B_{2k}|\frac{t^{2k}}{2k(2k)!}$ equals 2π. Since t is from a bounded set, when n is large enough, the series

$$\sum_{2\nmid k,k\geq 3}|B_{2k}|\frac{(64tn^{-\frac{1}{6}})^{2k}}{2k(2k)!}$$

converges. Moreover, it is evident that $64tn^{-\frac{1}{6}}<1$, we can bound the above summation by the constant

$$M_1=\sum_{2\nmid k,k\geq 3}|B_{2k}|\frac{1}{2k(2k)!}.$$

Similarly, it can be deduced that

$$\sum_{2|k,k\geq 2}B_{2k}\frac{t^{2k}}{2k(2k)!\sigma^{2k}}\sum_{i=2}^{n}\left((n+i)^{2k}-i^{2k}\right)<\frac{M_2}{n^{\frac{1}{3}}},$$

where $M_2 = \sum_{2|k, k\geq 2} B_{2k} \frac{1}{2k(2k)!}$ is a constant. Hence

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

which tends to zero as $n \to \infty$. This completes the proof. ∎

In [7], Margolius applied 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*

$$\eta_n = \frac{\xi_n - E(\xi_n)}{\mathrm{Var}(\xi_n)^{\frac{1}{2}}}$$

has the standard normal distribution.

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

$$\begin{aligned}
M_n(q) &= \frac{n+1}{\binom{2n}{n}} \frac{1-e^q}{1-e^{(n+1)q}} \cdot \prod_{i=1}^{n} \frac{1-e^{(n+i)q}}{1-e^{iq}} \\
&= \prod_{i=2}^{n} \frac{i}{n+i} \cdot \prod_{i=2}^{n} \frac{1-e^{(n+i)q}}{1-e^{iq}} \\
&= \prod_{i=2}^{n} \frac{(1-e^{(n+i)q})/(n+i)}{(1-e^{iq})/i} \\
&= \exp\left\{\frac{1}{2}\sum_{i=2}^{n} ((n+i)q - iq)\right\} \prod_{i=2}^{n} \frac{(e^{(n+i)q/2} - e^{-(n+i)q/2})/\frac{n+i}{2}}{(e^{iq/2} - e^{-iq/2})/\frac{i}{2}} \\
&= \exp\left\{\frac{n(n-1)q}{2}\right\} \prod_{i=2}^{n} \frac{\sinh\left((n+i)q/2\right)/\frac{n+i}{2}}{\sinh\left(iq/2\right)/\frac{i}{2}}.
\end{aligned}$$

Recalling the following relation on the Bernoulli numbers [7]

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

we find that

$$\begin{aligned}
\ln M_n(q) &= \frac{n(n-1)}{2} q + \sum_{i=2}^{n} \left(\ln\left(\frac{\sinh((n+i)q/2)}{(n+i)/2}\right) - \ln\left(\frac{\sinh(iq/2)}{i/2}\right)\right) \\
&= \frac{n(n-1)}{2} q + \sum_{k=1}^{\infty} B_{2k} \frac{q^{2k}}{2k(2k)!} \sum_{i=2}^{n} \left((n+i)^{2k} - i^{2k}\right).
\end{aligned}$$

Setting $q = t/\sigma$, where σ is the standard deviation of ξ_n as given in Theorem 2.1, we are led to the expansion

$$\ln M_n(t/\sigma) = \frac{n(n-1)t}{2\sigma} + \sum_{k=1}^{\infty} B_{2k} \frac{t^{2k}}{2k(2k)!\sigma^{2k}} \sum_{i=2}^{n} \left((n+i)^{2k} - i^{2k}\right).$$

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,

$$\begin{aligned}
&\lim_{n\to\infty} M_n(t/\sigma) \exp\left\{-\frac{n(n-1)t}{2\sigma}\right\} \\
&= \lim_{n\to\infty} \exp\left\{\sum_{k=1}^{\infty} B_{2k} \frac{t^{2k}}{2k(2k)!\sigma^{2k}} \sum_{i=2}^{n} \left((n+i)^{2k} - i^{2k}\right)\right\} \\
&= \lim_{n\to\infty} \exp\left\{B_2 \frac{t^2}{2(2)!\sigma^2} \sum_{i=2}^{n} \left((n+i)^2 - i^2\right)\right\} \\
&= e^{t^2/2},
\end{aligned}$$

which coincides with the moment generating function of the standard normal distribution. Employing Curtiss's theorem [8], we reach the conclusion that η_n has the standard normal distribution when n approaches infinity. ∎

3. 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 ξ_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 ξ_n is normally distributed as $n \to \infty$, if and only if

$$\sum_{k=1}^{\infty} B_{2k} \frac{t^{2k}}{2k(2k)!} \left(\sum_{i=1}^{n} (a_i^{2k} - b_i^{2k})\right) \frac{1}{\left(\sum_{i=1}^{n} (a_i^2 - b_i^2)\right)^k} \to 0 \quad \text{as} \quad n \to \infty.$$

Proof. The expectation of ξ_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 Theorem 2.1, we find

$$\sigma^2 = \mathrm{Var}(\xi_n) = \frac{1}{12}\sum_{i=1}^{n}\left(a_i^2 - b_i^2\right). \tag{3.1}$$

Hence,

$$B_2\frac{t^2}{2(2)!\sigma^2}\left(\sum_{i=1}^{n}(a_i^2-b_i^2)\right) = \frac{1}{6}\cdot\frac{t^2}{4\cdot\frac{1}{12}\left(\sum\limits_{i=1}^{n}(a_i^2-b_i^2)\right)}\cdot\left(\sum_{i=1}^{n}(a_i^2-b_i^2)\right) = \frac{t^2}{2}.$$

By the same procedure as in the proof of Theorem 2.3, we obtain

$$\begin{aligned}
&\lim_{n\to\infty} M_n(t/\sigma)\exp\left\{\frac{1}{2}\sum_{i=1}^{n}\left(a_i^{2k}-b_i^{2k}\right)\right\}\\
&= \quad e^{t^2/2}\lim_{n\to\infty}\exp\left\{\sum_{k=2}^{\infty}B_{2k}\frac{t^{2k}}{2k(2k)!\sigma^{2k}}\left(\sum_{i=1}^{n}\left(a_i^{2k}-b_i^{2k}\right)\right)\right\}.
\end{aligned}$$

It follows that the limiting distribution of $p_n(k)$ is normal if and only if

$$\sum_{k=2}^{\infty}B_{2k}\frac{t^{2k}}{2k(2k)!\sigma^{2k}}\left(\sum_{i=1}^{n}\left(a_i^{2k}-b_i^{2k}\right)\right)\to 0 \quad \text{as} \quad n\to\infty, \tag{3.2}$$

for t from any bounded set. By virtue of the variance formula (3.1), the condition (3.2) is equivalent to

$$\sum_{k=1}^{\infty}B_{2k}\frac{t^{2k}}{2k(2k)!}\frac{\sum\limits_{i=1}^{n}\left(a_i^{2k}-b_i^{2k}\right)}{\left(\sum\limits_{i=1}^{n}\left(a_i^2-b_i^2\right)\right)^k}\to 0 \quad \text{as} \quad n\to\infty \tag{3.3}$$

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*

$$\frac{\sum\limits_{i=1}^{n}\left(a_i^{2k}-b_i^{2k}\right)}{\left(\sum\limits_{i=1}^{n}(a_i^2-b_i^2)\right)^k} < n^{\gamma}(\alpha n^{\beta})^{2k}, \tag{3.4}$$

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

$$\sum_{2\nmid k, k\geq 3}|B_{2k}|\frac{x^{2k}}{2k(2k)!}$$

is 2π. If (3.4) holds for $k > 1$, then for t from any bounded set, and for sufficiently large n, we have

$$\left|t^{2k}\sum_{i=1}^{n}\left(a_i^{2k}-b_i^{2k}\right)/\sigma^{2k}\right| \leq n^{\gamma}(t\alpha n^{\beta})^{2k},$$

where $t\alpha n^{\beta} < 2\pi$. It is clear that $n^{\gamma}\to 0$ since $\gamma < 0$. ∎

If we choose $\alpha = 32\sqrt{3}/3$, $2\beta = \gamma = -\frac{1}{3}$, Theorem 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

$$c_n(q) = \frac{[2]}{[2n]} \begin{bmatrix} 2n \\ n-1 \end{bmatrix},$$

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:

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

Set $a_1 = a_2 = 1,\ a_i = n+i-1,\ 3 \le i \le n$, and $b_1 = b_2 = 1,\ b_3 = 1,\ b_i = i-1,\ 4 \le i \le n$. Then we have

$$\begin{aligned}\sum_{i=1}^{n}\left(a_i^{2k} - b_i^{2k}\right) &= \left(a_3^{2k} - b_3^{2k}\right) + \sum_{i=4}^{n}\left(a_i^{2k} - b_i^{2k}\right) \\ &= (n+2)^{2k} - 1 + \sum_{i=3}^{n-1}\left((n+i)^{2k} - i^{2k}\right)\end{aligned}$$

and

$$\begin{aligned}\left(\sum_{i=1}^{n}(a_i^2 - b_i^2)\right)^k &= \left((n+2)^2 - 1 + \sum_{i=3}^{n-1}\left((n+i)^2 - i^2\right)\right)^k \\ &= (n-1)^k(n+1)^k(2n-3)^k.\end{aligned}$$

By the same arguments as in the proof of Lemma 2.2, we may set $\alpha = 32\sqrt{3}/3$ and $2\beta = \gamma = -\frac{1}{3}$ such that the condition (3.4) is satisfied. Therefore, Theorem 3.1 implies the limiting distribution of the coefficients of $c_n(q)$. ∎

The *m-Catalan numbers* are defined by

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

for $n \ge 1$. Accordingly, the generalized q-Catalan numbers are given by

$$C_{n,m}(q) = \frac{1}{[(m-1)n+1]}\begin{bmatrix} mn \\ n \end{bmatrix}.$$

Theorem 3.1 has the following consequence.

Corollary 3.4. *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 $a_1 = 1,\ a_i = (m-1)n + i,\ 2 \le i \le n$, and $b_1 = 1, \quad b_i = i,\ 2 \le i \le n$. Then we have

$$\sum_{i=1}^{n} \left(a_i^{2k} - b_i^{2k}\right) = \sum_{i=2}^{n} \left(a_i^{2k} - b_i^{2k}\right) = \sum_{i=2}^{n} \sum_{j=1}^{2k} \binom{2k}{j} ((m-1)n)^{2k-j}\, i^j.$$

The same argument as in the proof of Lemma 2.2 yields the following bound

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

Now,

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

It follows that

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

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. ∎

4. 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 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 $n \ge 16$.

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,*

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

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

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

Acknowledgments. The authors are grateful to the referee for valuable suggestions. Thanks are also due to Barbara Margolius and Helmut Prodinger for helpful comments.

References

1. H. Alzer, Sharp bounds for the Bernoulli numbers, *Arch. Math.* **74** (2000), 207–211.
2. G.E. Andrews, Catalan numbers, q-Catalan numbers and hypergeometric series, *J. Combin. Theory, Ser. A* **44** (1987), 267–273.
3. J. Fürlinger and J. Hofbauer, q-Catalan Numbers, *J. Combin. Theory, Ser. A* **40** (1985), 248–264.
4. I. Gessel and D. Stanton, Applications of q-Lagrange inversion to basic hypergeometric series, *Trans. Amer. Math. Soc.* **177** (1983), 173–201.
5. C. Krattenthaler, A new q-Lagrange formula and some applications, *Proc. Amer. Math. Soc.* **90** (1984), 338–344.
6. G. Louchard, H. Prodinger, The number of inversions in permutations: a saddle point approach, *J. Integer Seq.* **6** (2003), A03.2.8.
7. B.H. Margolius, Permutations with inversions, *J. Integer Seq.* **4** (2001), A01.2.4.
8. V.N. Sachkov, *Probabilisitic Methods in Combinatorial Analysis*, Cambridge University Press, New York, 1997.
9. R.P. Stanley, Log-concave and unimodal sequences in algebra, combinatorics, and geometry, *Graph Theory and Its Applications: East and West*, Ann. New York Acad. Sci., Vol. 576, 1989, pp. 500–535.
10. R.P. Stanley, *Enumerative Combinatorics*, Vol. 2, Cambridge University Press, Cambridge, 1999.

Center for Combinatorics, LPMC-TJKLC, Nankai University, Tianjin 300071, P. R. China

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

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

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