Ann. Inst. Statist. Math.
28 (1976), Part A, 499-506

THE ASYMPTOTIC NORMALITY OF CERTAIN COMBINATORIAL DISTRIBUTIONS

CH. A. CHARALAMBIDES

(Received Mar. 29, 1973)

Summary

The numbers $C(m, n, s)$ and $|C(m, n, -s)|$, $s>0$, appearing in the n-fold convolution of truncated binomial and negative binomial distributions, respectively, are shown to be asymptotically normal. Moreover a concavity property for these numbers is concluded.

1. Introduction

A combinatorial distribution, as suggested by Harper [7], is defined to be any "generalized Pascal's triangle" on the lattice points of the positive quadrant, defined by a difference equation of the form:

$$A(m+1, n)=g(m, n)A(m, n)+h(m, n)A(m, n-1) \tag{1.1}$$

with boundary conditions:

$$A(0, 0)=1 \quad \text{and} \quad A(0, n)=0 \qquad \text{if } n\neq 0$$

and with $g(m, n)$, $h(m, n)$ positive for $n=0, 1,\cdots, m$, $m=0, 1,\cdots$.

The asymptotic behavior of certain combinatorial distributions has been studied by several authors. Among them Feller [3] has shown that the distributions given by the number of permutations of m elements with n inversions and the number of permutations of m elements with n cycles, are asymptotically normal. Gončarov [4] has shown asymptotic normality for the above and other combinatorial distributions. Harper [7] has shown that the Stirling numbers of the second kind are asymptotically normal.

In the present paper we consider two other special cases of (1.1); the numbers $C(m, n, s)$, appearing in the n-fold convolution of a truncated binomial distribution and satisfying the difference equation (see [1] and [2]):

$$C(m+1, n, s)=(sn-m)C(m, n, s)+sC(m, n-1, s) , \qquad s>0 \tag{1.2}$$

with boundary conditions:

$$C(0,0,s)=1 \quad \text{and} \quad C(0,n,s)=0 \qquad \text{if } n\neq 0,\ s>0\,,$$

and the numbers $|C(m,n,-s)|$, appearing in the n-fold convolution of a truncated negative binomial distribution and satisfying the difference equation (see [1] and [2]):

$$(1.3)\qquad |C(m+1,n,-s)|=(sn+m)|C(m,n,-s)| \\ +s|C(m,n-1,-s)|\,, \qquad s>0$$

with boundary conditions:

$$|C(0,0,-s)|=1 \quad \text{and} \quad |C(0,n,-s)|=0 \qquad \text{if } n\neq 0,\ s>0\,.$$

These two combinatorial distributions are shown to be asymptotically normal. The results for the numbers $|C(m,n,-s)|$ are given in Section 4 without proof since this is similar to that for $C(m,n,s)$.

2. Preparatory results

Let

$$(2.1)\qquad f_m(t)=\sum_{n=0}^{m} C(m,n,s)t^n$$

denote the generating function of $C(m,n,s)$ with respect to the index n. Then we have the following

LEMMA 2.1. *The generating function $f_m(t)$ of $C(m,n,s)$ has m distinct real non-positive roots for all $m=1,2,\cdots,s$, $s>0$.*

PROOF. From (1.2) and (2.1) we obtain for the generating function f_m the following difference-differential equation

$$(2.2)\qquad f_{m+1}(t)=st\frac{d}{dt}f_m(t)+(st-m)f_m(t)\,.$$

By the definition (2.1) and since $C(1,1,s)=s$ we have

$$f_1(t)=st$$

and from (2.2)

$$f_2(t)=s^2t^2+s(s-1)t$$

that is the statement holds for $m=1,2$.

Now suppose $m>2$. By the induction hypothesis f_m has m distinct real non-positive roots. If we define the function

$$h_m(t)=f_m(t)t^{-m/s}e^t$$

then h_m has exactly the same finite roots that f_m does and moreover h_m has a zero at $-\infty$; the identity (2.2) becomes

$$h_{m+1}(t)=st^{(s-1)/s}\frac{d}{dt}h_m(t) .$$

By Rolle's theorem between any two zeroes of h_m the derivative $(d/dt)h_m$ will have a zero. Therefore f_{m+1} has m distinct real negative roots; obviously $t=0$ is another one and since f_{m+1} is of degree $m+1$, by induction, we have found all the roots and the proof is completed.

As a consequence of Lemma 2.1 we obtain by using Newton's inequality (c.f. [6]) the following

COROLLARY 2.1. *The number $C(m, n, s)$ for $s>0$, is a strong logarithmic concave function of n, that is*

$$[C(m, n, s)]^2>C(m, n+1, s)C(m, n-1, s) .$$

Let us now define the numbers C_{ms}, $m=1, 2, \cdots$ by

$$C_{ms}=\sum_{n=0}^{m} C(m, n, s) .$$

Since (see [2])

$$\lim_{s\to\pm\infty} s^{-m}C(m, n, s)=S(m, n)$$

where $S(m, n)$ denotes the Stirling numbers of the second kind, we obtain the following limiting property of the numbers C_{ms}:

$$\lim_{s\to\pm\infty} s^{-m}C_{ms}=B_m$$

where

$$B_m=\sum_{n=0}^{m} S(m, n)$$

denotes the Bell numbers.

The numbers C_{ms} may be considered as generated by the function $\exp[(1+z)^s-1]$, that is

$$\sum_{m=0}^{\infty} C_{ms}\frac{z^m}{m!}=\exp[(1+z)^s-1] . \qquad (2.3)$$

This can be easily verified by expanding the right-hand side of (2.3) into powers of z and using the relation (see [2])

$$\sum_{m=n}^{\infty} C(m, n, s)\frac{z^m}{m!}=\frac{1}{n!}[(1+z)^s-1]^n .$$

An asymptotic formula for the numbers C_{ms} is given in the following

LEMMA 2.2. *Let R be a real solution of the equation: $sR(1+R)^{s-1}=m$, $m=1, 2, \cdots, s$, $s=1, 2, \cdots$.* Then*

$$C_{ms}=s^m(1+R)^{m(s-1)+1/2}(1+sR)^{-1/2}\exp\{m[(1+R)(sR)^{-1}-1]-1\} \times\left\{1-\frac{R^2[2s^3(s-1)R^2+(7s^2-4s-3)sR+10s^2-12s+2]}{24m(1+R)(1+sR)^3}+o\left(\frac{R^2}{m^2}\right)\right\} . \tag{2.4}$$

SKETCH OF PROOF. Since the method of proof is analogous to that given by Moser and Wyman [9] and Szekeres and Binet [10] for the Bell numbers, details are omitted. From (2.3) and Cauchy's theorem, we obtain

$$C_{ms}=\frac{m!}{2\pi i}\oint_c z^{-(m+1)}\exp[(1+z)^s-1]dz$$

where c is the circle $z=Re^{i\theta}$. Hence

$$C_{ms}=A\int_{-\pi}^{\pi}\exp[F(\theta)]d\theta \tag{2.5}$$

where

$$A=(2\pi)^{-1}R^{-m}m!\exp[(1+R)^s-1] \tag{2.6}$$

and

$$F(\theta)=(1+Re^{i\theta})^s-im\theta-(1+R)^s .$$

Expanding $F(\theta)$ in a Maclaurin series about $\theta=0$ and choosing R to be a real solution of the equation

$$sR(1+R)^{s-1}=m \tag{2.7}$$

we get

$$F(\theta)=-\frac{1}{2}(1+sR)(sR)(1+R)^{s-2}\theta^2+\sum_{k=1}^{\infty}[\Theta^{k+2}(1+R)^s]\frac{(i\theta)^{k+2}}{(k+2)!} \tag{2.8}$$

where Θ denotes the operator

$$\Theta=R\frac{d}{dR} .$$

* Such a solution always exists since the function $\varphi(R)=sR(1+R)^{s-1}-m$ has a root $R\in(0, a)$ where $a>\max\{(m/s)^{1/2}, (m/s)^{1/2(s-1)}-1\}$, because of $\varphi(0)=-m<0$, $\varphi(a)>0$ and the continuity of φ.

Introducing the following notation:

$$\phi=\left[\frac{1}{2}sR(1+sR)(1+R)^{s-2}\right]^{1/2}\theta$$

$$B=A\left[\frac{1}{2}sR(1+sR)(1+R)^{s-2}\right]^{-1/2}$$

(2.9) $$a_k=(1+R)^{-(s-2)}\Theta^{k+2}(1+R)^s(i\phi)^{k+2}/(k+2)!\left[\frac{1}{2}sR(1+sR)\right]^{(k+2)/2}$$

$$z=(1+R)^{-(s-2)/2}$$

$$f(z)=\sum_{k=1}^{\infty}a_k z^k$$

and making the substitution (2.8) in (2.5) we finally arrive at the asymptotic formula

(2.10) $$C_{ms}\sim B\sum_{k=0}^{\infty}\left[\int_{-\infty}^{\infty}e^{-\phi^2}b_{2k}d\phi\right](1+R)^{-k(s-2)}$$

where b_k is the coefficient of z^k in the expansion of $e^{f(z)}$ in a convergent Maclaurin series, that is

$$e^{f(z)}=\sum_{k=0}^{\infty}b_k z^k\,,\qquad b_0=1\,.$$

Consider now only the first two terms in the expansion (2.10). By calculation we obtain

$$b_0=1\,,$$

$$b_2=a_2+\frac{1}{2}a_1^2=\frac{s^3R^3(6s^2-4s+1)R^2+(7s-4)R+1}{6sR(1+R)^2(1+sR)^2}\phi^4$$

$$-\frac{[s^2R^2+(3s-1)R+1]^2}{9sR(1+R)^2(1+sR)^3}\phi^6\,.$$

Hence from (2.10) and using (2.7) we find:

$$C_{ms}\sim\pi^{1/2}B\left[1-\frac{2s^4R^4+(9s^3+2s^2-3s)R^3+(16s^2-6s+2)R^2+(6s+2)R+2}{24m(1+R)(1+sR)^3}\right]$$

which by (2.6), (2.9) and using Stirling's expansion of $m!$ implies the required formula (2.4).

LEMMA 2.3.

(2.14) $$\frac{C_{m+k,s}}{s^kC_{m,s}}=\left(\frac{m}{sR}\right)^k\left[1+\frac{1}{m}Q(k,R,s)+o\left(\frac{R^2}{m^2}\right)\right]$$

where $Q(k, R, s)$ is a polynomial in k of degree 2.

SKETCH OF PROOF. Expand $C_{m+k,s}$ and $C_{m,s}$ as in Lemma 2.2 and consider the ratio $C_{m+k,s}/C_{m,s}$ as a product of three terms each of which can be expressed as a sum of reciprocal powers of m with coefficients which are functions of k, R and s (c.f. Lemma 1.4 in [5]).

3. The asymptotic normality of the numbers $C(m, n, s)$

THEOREM 3.1. *The numbers $C(m, n, s)$ for positive s are asymptotically normal in the sense that*

$$C_{ms}^{-1}\sum_{k=0}^{[x_m]} C(m, n, s) \to (2\pi)^{-1/2}\int_{-\infty}^{x} e^{-t^2/2}dt \qquad as\ m\to\infty$$

where

$$x_m=\left[\frac{C_{m+2,s}+C_{m+1,s}}{s^2C_{ms}}-\left(\frac{C_{m+1,s}}{sC_{ms}}\right)^2-1\right]^{1/2}\cdot x+\left[\frac{C_{m+1,s}}{sC_{ms}}+\frac{m-s}{s}\right].$$

PROOF. Let $-x_{mk}$, $k=1, 2, \cdots, m$ be the roots of the generating function (2.1) of the numbers $C(m, n, s)$ and define the sequence of independent random variables X'_{mk}, $k=1, 2, \cdots, m$ by

$$\Pr[X'_{mk}=y]=\begin{cases}\dfrac{x_{mk}}{1+x_{mk}} & \text{if } y=0\\[2ex] \dfrac{1}{1+x_{mk}} & \text{if } y=1,\end{cases}$$

then the random variable

$$S'_m=\sum_{k=0}^{m} X'_{mk}$$

has probability function:

$$\Pr[S'_m=n]=C_{ms}^{-1}C(m, n, s), \qquad n=0, 1, \cdots, m$$

and

$$\mathrm{E}(S'_m)=\sum_{n=0}^{m}\frac{nC(m, n, s)}{C_{ms}}=\frac{C_{m+1,s}}{sC_{ms}}+\frac{m-s}{s}$$

$$\begin{aligned}\mathrm{Var}(S'_m)&=\sum_{n=0}^{m}\frac{n^2C(m, n, s)}{C_{ms}}-\left[\frac{C_{m+1,s}}{sC_{ms}}+\frac{m-s}{s}\right]^2\\ &=\frac{C_{m+2,s}+C_{m+1,s}}{s^2C_{ms}}-\left[\frac{C_{m+1,s}}{sC_{ms}}\right]^2-1.\end{aligned}$$

From (2.11) we have:

$$\mathrm{Var}\,(S'_m)=\frac{m}{(sR)^2}[Q(2,\,R,\,s)-2Q(1,\,R,\,s)+R]+o(1)\;.$$

Hence

$$\mathrm{Var}\,(S'_m)\to\infty \qquad \text{as } m\to\infty\;.$$

Now let

$$S_m=[\mathrm{Var}\,(S'_m)]^{-1/2}[S'_m-\mathrm{E}\,(S'_m)]=\sum_{k=1}^{m}X_{mk}$$

where

$$X_{mk}=[\mathrm{Var}\,(S'_m)]^{-1/2}[X'_{mk}-\mathrm{E}\,(X'_{mk})]\;.$$

Since for a given $\varepsilon>0$ there exists M such that

$$|X_{mk}|\leqq[\mathrm{Var}\,(S_m)]^{-1/2} \qquad \text{for all } m>M$$

we obtain:

$$\lim_{m\to\infty}\sum_{k=1}^{m}\int_{|x|>\varepsilon}x^2dF_{mk}=0 \qquad \text{for all } m>M$$

with F_{mk} the distribution function of X_{mk}.

Therefore all the conditions of the "bounded variance normal convergence criterion" (see [8], p. 295) are fulfilled. Hence the statement of the theorem holds.

4. The asymptotic normality of the numbers $|C(m,\,n,\,-s)|$

The results for the numbers $|C(m,\,n,\,-s)|$, $s>0$, may be summarized in the following

PROPOSITION 4.1. *The generating function* $g_m(t)=\sum_{n=1}^{m}|C(m,\,n,\,-s)|t^n$ *of the numbers* $|C(m,\,n,\,-s)|$ *has m distinct real non-positive roots for all* $m=1,2,\cdots,s>0$.

PROPOSITION 4.2. *The number* $|C(m,\,n,\,-s)|$ *for* $s>0$ *is a strong logarithmic concave function n.*

THEOREM 4.1. *The numbers* $|C(m,\,n,\,-s)|$ *for positive s are asymptotically normal in the sence that:*

$$|C_{m,-s}|^{-1}\sum_{n=0}^{[x_m]}|C(m,\,n,\,-s)|\to(2\pi)^{-1/2}\int_{-\infty}^{x}e^{-t^2/2}dt \qquad as\ m\to\infty$$

where

$$x_m=\left[\frac{|C_{m+2,-s}|-|C_{m+1,-s}|}{s^2|C_{m,-s}|}-\left(\frac{|C_{m+1,-s}|}{s|C_{m,-s}|}\right)^2-1\right]^{1/2}\cdot x+\left[\frac{|C_{m+1,-s}|}{s|C_{m,-s}|}-\frac{m+s}{s}\right]$$

and

$$|C_{m,-s}|=(-1)^m C_{m,-s} .$$

Acknowledgement

The author is indebted to Professor T. Cacoullos for many helpful suggestions.

UNIVERSITY OF ATHENS

REFERENCES

[1] Cacoullos, T. and Charalambides, Ch. (1975). On minimum variance unbiased estimation for truncated binomial and negative binomial distributions, *Ann. Inst. Statist. Math.*, **27**, 235-244.

[2] Charalambides, Ch. (1972). A new kind of numbers appearing in the n-fold convolution of truncated binomial and negative binomial distributions, submitted for publication to the *J. Comb. Theory*.

[3] Feller, W. (1968). *An Introduction to Probability Theory and Its Applications*, Vol. 1 (3rd ed.) Wiley, New York.

[4] Gončarov, V. (1944). Du domaine d'analyse combinatoroire, *Bull. Acad. Sci. U.S.S.R., Ser. Math.*, 8, 3-48 (in Russian with French summary).

[5] Haigh, J. (1972). Random equivalence relations, *J. Comb. Theory*, Ser. A, **13**, 287-295.

[6] Hardy, Littlewood and Polya (1959). *Inequalities*, 2nd ed., Cambridge Univ. Press, Cambridge, England, pp. 104-105 and pp. 51-54.

[7] Harper, L. H. (1967). Stirling behavior is asymptotically normal, *Ann. Math. Statist.*, **38**, 410-414.

[8] Loève, M. (1960). *Probability Theory*, 2nd ed., Van Nostrand, Princeton.

[9] Moser, L. and Wyman, M. (1955). An asymptotic formula for the Bell numbers, *Trans. Roy. Soc. Canada*, **49**, 49-53.

[10] Szekeres, G. and Binet, F. E. (1957). On Borel fields over finite sets, *Ann. Math. Statist.*, **28**, 494-498.