Asymptotic normality determined by high moments, and submap counts of random maps

Zhicheng Gao*
Faculty of Business Administration
University of Macau
Macau
China zcgao@umac.mo

Nicholas C. Wormald†
Department of Combinatorics and Optimization
University of Waterloo
Waterloo ON
Canada N2L 3G1 nwormald@uwaterloo.ca

Abstract

We give a general result showing that the asymptotic behaviour of high moments determines the shape of distributions which are asymptotically normal. Both the factorial and non-factorial (non-central) moments are treated. This differs from the usual moment method in combinatorics, as the expected value may tend to infinity quite rapidly. Applications are given to submap counts in random planar triangulations, where we use a simple argument to asymptotically determine high moments for the number of copies of a given subtriangulation in a random 3-connected planar triangulation. Similar results are also obtained for 2-connected triangulations and quadrangulations with no multiple edges.

1 Introduction

It is well known that the moments, or factorial moments, determine many distributions (such as Poisson or normal). In combinatorial situations, another version of this effect is often used: if for a sequence $\{X_n\}$ of nonnegative integer random variables, the factorial moments $\{\mathbf{E}[X_n]_k\}$ for each fixed integer k tend towards those of a Poisson random variable X, then X_n tends towards X in distribution. Here $[x]_k$ denotes $x(x-1)\cdots(x-k+1)$.

*Research supported by NSERCC and University of Macau

†Research supported by the Australian Research Council and the Canada Research Chairs program. Research carried mainly while the author was at the Department of Mathematics and Statistics, University of Melbourne

The condition that X is fixed is not necessary: if the factorial moment of X_n is asymptotic to that of Y_n, where the variables Y_n are Poisson with bounded expectation, then the total variation distance between the distributions of X_n and Y_n tends to 0. The boundedness condition is needed if there is no knowledge of the rate of convergence of the moments. However, since the moments of a Poisson or normal random variable determine its distribution, perfect knowledge of the moments would imply a convergence result regardless of the behaviour of $\mathbf{E}X_n$. In many combinatorial applications, the moments can only be determined approximately (asymptotically). In that case one method which incorporates the factorial moments, at least for an asymptotically Poisson random variable, is to use Bonferroni's inequalities. These quantify the determination of the distribution from the moments, but with an alternating sum which contains potentially damaging cancellation if $\mathbf{E}X_n$ grows with n. For instance, Erdős and Rényi [5] showed asymptotic normality of the number of trees in a certain sort of random graph, where $\mathbf{E}X_n$ grows logarithmically and their asymptotic evaluation of the factorial moments determines the asymptotic behaviour of the centralised, standardised moments. However, their method required them to bound the convergence rate of the asymptotics for the factorial moments. Indeed, without such bounds, one cannot deduce asymptotic normality from the raw asymptotic behaviour of a fixed set of moments when $\mathbf{E}X_n \to \infty$, a point overlooked in the discussion in Ruciński [13].

A number of other ways have been used to prove asymptotic normality in combinatorics (see Janson et al. [7, Chapter 6] for example) but they tend to rely on the random variable in question being a sum of nearly independent indicator variables, such that the expectation of each of them can be estimated. However, for our main application, submaps of planar maps as in Section 4, the variables in question have never been successfully put into such a form.

However, a given factorial moment is commonly easy to determine asymptotically, and sometimes this applies to the higher factorial moments as well, by which we mean the kth factorial moment where $k \to \infty$. We show in Section 2 that the asymptotic behaviour of a certain set of higher factorial moments suffices to determine the shape of distributions which are asymptotically normal, provided the variance is neither too small nor too large compared with the expectation. Part of the difficulty here is that we are dealing with non-central moments, which are the simplest to compute in combinatorial situations. Two short applications are then given in Section 3 to show how to obtain an existing result on the number of components in a certain random forest, and the distribution of nonoverlapping subwords of a word. In Section 4, applications are given to submap counts in random planar triangulations and quadrangulations.

Although the ordinary (non-factorial) moments tend to be less natural to compute in combinatorial applications, we also provide the corresponding result for these in Section 2. Note that approximation by Poisson is doomed to failure when the expectation and variance are not asymptotically equal, which is the general case for both of these results.

We use $\mathbf{P}$, $\mathbf{E}$, and $\mathbf{V}$ to denote the probability, expectation, and variance of a random variable, respectively.

2 Basic moment results

All asymptotics refer to $n \to \infty$. It is easy to verify that if $(X_n - \mu_n)/\sigma_n$, $\sigma_n > 0$, has standard normal distribution, then

$$\mathbf{E}[X_n]_k \sim \mu_n^k \exp\left(\frac{k^2(\sigma_n^2 - \mu_n)}{2\mu_n^2}\right)$$

provided $k = O(\mu_n/\sigma_n)$ and μ_n grows significantly faster than σ_n. The first result given here shows that this statement is to a large extent reversible.

Theorem 1 *Let $s_n > -\mu_n^{-1}$ and*

$$\sigma_n = \sqrt{\mu_n + \mu_n^2 s_n}, \tag{2.1}$$

where $0 < \mu_n \to \infty$. Suppose that

$$\mu_n = o(\sigma_n^3), \tag{2.2}$$

and a sequence $\{X_n\}$ of nonnegative random variables satisfies

$$\mathbf{E}[X_n]_k \sim \mu_n^k \exp\left(\frac{k^2 s_n}{2}\right) \tag{2.3}$$

uniformly for all integers k in the range $c\mu_n/\sigma_n \le k \le c'\mu_n/\sigma_n$ for some constants $c' > c > 0$. Then $(X_n - \mu_n)/\sigma_n$ tends in distribution to the standard normal as $n \to \infty$.

Before proving this, we deal with a corresponding result for to the central moments $\mathbf{E}X^k$, as it is slightly simpler and proved more easily. For $(X_n - \mu_n)/\sigma_n$ with the standard normal distribution,

$$\mathbf{E}X_n^k \sim \mu_n^k \exp\left(\frac{k^2\sigma_n^2}{2\mu_n^2}\right)$$

provided $k = O(\mu_n/\sigma_n)$ and μ_n grows significantly faster than σ_n.

Theorem 2 *Suppose that $\mu_n/\sigma_n \to \infty$ and that a sequence $\{X_n\}$ of nonnegative random variables satisfies*

$$\mathbf{E}X_n^k \sim \mu_n^k \exp\left(\frac{k^2\sigma_n^2}{2\mu_n^2}\right) \tag{2.4}$$

uniformly for all integers k in the range $c\mu_n/\sigma_n \le k \le c'\mu_n/\sigma_n$ for some constants $c' > c > 0$. Then $(X_n - \mu_n)/\sigma_n$ tends in distribution to the standard normal as $n \to \infty$.

Proof: Let

$$\zeta_n = \frac{\mu_n}{\sigma_n} \ln\left(\frac{X_n}{\mu_n}\right). \tag{2.5}$$

Then for $t = k\sigma_n/\mu_n$, (2.4) gives

$$\mathbf{E}e^{t\zeta_n} \to e^{t^2/2} \quad (n \to \infty) \tag{2.6}$$

uniformly for t taking a certain set of values which are in an asymptotic sense dense in (c, c'). Since the function $e^{t^2/2}$ is monotonic for $t > 0$, the convergence in (2.6) is equivalent to convergence for all $t \in (c, c')$. Hence ζ_n converges weakly to the standard normal variable η (see for example [3, Problem 30.4, p. 397]). Then, since $\sigma_n/\mu_n \to 0$ and $X_n = \mu_n e^{\zeta_n \sigma_n/\mu_n}$, the Taylor expansion of the exponential function yields $(X_n - \mu_n)/\sigma_n \xrightarrow{\text{d}} \eta$ as required. ∎

Proof of Theorem 1: From (2.2) it follows that $[\mu_n]_k \sim \mu_n^k \exp(-k^2/2\mu_n)$ for $k = O(\mu_n/\sigma_n)$. Thus, as with (2.6), if we set

$$Q_n = \frac{[X_n]_k}{[\mu_n]_k}, \quad \text{and} \quad t = k\frac{\sigma_n}{\mu_n}$$

the assumption (2.3) implies the convergence

$$\mathbf{E}Q_n \to e^{t^2/2} \tag{2.7}$$

uniformly for $t \in (c, c')$. We will use the fact that for $a \geq b > k$,

$$k\log(a/b) \leq \log\frac{[a]_k}{[b]_k} \leq k\log\left(\frac{a-k}{b-k}\right) \leq \frac{k\log(a/b)}{1-k/b} \tag{2.8}$$

where the last step is easily verified by differentiation with respect to a. Define ζ_n as in (2.5), put $\epsilon = k/(r\mu_n)$ and fix $0 < r < 1$. Since $\epsilon \to 0$, applying (2.8) with $a = X_n$ and $b = \mu_n$ when $X_n \geq \mu_n$, and with $a = \mu_n$ and $b = X_n$ when $r\mu_n \leq X_n \leq \mu_n$, we obtain

$$T_n := |e^{t\zeta_n} - Q_n| \leq |Q_n - Q_n^{1-\epsilon}| \leq \begin{cases} \min(Q_n, \epsilon Q_n \log Q_n) & X_n \geq \mu_n \\ \epsilon Q_n^{1-\epsilon}\log(1/Q_n) & r\mu_n \leq X_n < \mu_n. \end{cases}$$

Finally, for $X_n < r\mu_n$, note that $Q_n \leq (X_n/\mu_n)^k = e^{t\zeta_n} \to 0$. Noting that $X_n < 1$ iff $Q_n < 1$, we may split the expectation of T_n into four re ions: $X_n < r\mu_n$, $r\mu_n \leq X_n < \mu_n$, $1 \leq Q_n \leq \epsilon^{-1/2}$ and $Q_n > \epsilon^{-1/2}$, and obtain

$$ET_n \leq o(1) + O(\epsilon) + \epsilon(\epsilon^{-1/2}\log\epsilon^{-1/2}) + E[Q_n I_{(Q_n \geq \epsilon^{-1/2})}] \to 0$$

where the last term is estimated by applying (2.7). From this, (2.7) implies (2.6) which then implies the theorem by the argument in the proof of Theorem 1. ∎

3 Two quick applications

Pittel and Weishaar [8] obtained the exact factorial moments of a random variable $T = T(n)$ which counted the trees in a certain random forest: $\mathbf{E}[T_n]_k = ([n]_k)^2/[2n-1]_k$. Applying Theorem 1 with $\mu_n = n/2$, $\sigma_n = \sqrt{n/8}$, and $s_n = -3/(2n)$, we deduce the convergence of $(T_n - \mu_n)/\sigma_n$ to standard normal, a result which they obtained using an entirely different argument.

It is well-known that the distribution of subword occurrences in a random word is asymptotically normal. (See [9] for references on this subject.) Here we give a quick application of Theorem 1 to this problem. Let W be a word of length r in an alphabet of a letters, and let X_n denote the number of occurrences of W in a random word of length n. Assume that W has the property that separate occurrences of W cannot have nonzero overlap. Then

$$\mathbf{E}[X_n]_k = [n - k(r-1)]_k/a^{rk}$$

since k occurrences of W can be determined by selecting the k positions of its first letter in the sequence of length $n-k(r-1)$ obtained by deleting all other letters of the k occurrences.

Set $\mu_n = n/a^r$ and suppose $r \geq 1$ and $a \geq 2$ are any functions of n such that $\mu_n \to \infty$. Then $r = O(\log n)$, and (2.3) holds with $s_n = (1-2r)/n$ for all $k = O(\sqrt{n})$, so that $\sigma_n^2 = \mu_n(1-(2r-1)/a^r)$. It is easy to see that $c\sqrt{\mu_n} \leq \sigma_n < \sqrt{\mu_n}$ for some positive constant c . So X_n is asymptotically normal with mean μ_n and variance σ_n^2.

4 Submaps of random maps

Throughout this section, a *map* is a connected graph G embedded in the plane with no edge crossings. Loops and multiple edges are allowed in G. A map is *rooted* if an edge is distinguished together with a vertex on the edge and a side of the edge. The distinguished vertex and edge are called the root vertex and the root edge of the map. The face on the distinguished side of the root edge is called the root face. Two rooted maps are considered the same if there is a homeomorphism from the plane to itself which transforms one rooted map to the other and preserves the rooting. A *triangulation (quadrangulation)* is a map such that all faces are triangles (quadrangles). It is well-known that a triangulation is 2-connected if and only if it contains no loops, and it is 3-connected if and only if it contains no loops or multiple edges. Let $T_n(\overline{T}_n)$ be the number of rooted 3-connected (2-connected) triangulations with $n+2$ vertices, and let Q_n be the number of rooted quadrangulations with $n+2$ vertices, and no multiple edges. It is known by [4, 14] that

$$T_n = \frac{\sqrt{6}}{32\sqrt{\pi}} n^{-5/2}(256/27)^n(1+c_1/n+O(1/n^2)), \tag{4.1}$$

$$\overline{T}_n = \frac{\sqrt{3}}{4\sqrt{\pi}} n^{-5/2}(27/2)^n(1+c_2/n+O(1/n^2)), \tag{4.2}$$

$$Q_n = \frac{8\sqrt{3}}{27\sqrt{\pi}} n^{-5/2}(27/4)^n(1+c_3/n+O(1/n^2)), \tag{4.3}$$

where c_1, c_2 and c_3 are constants. We will see that the actual values of c_1, c_2 and c_3 do not contribute to our results. Throughout this section, all probability distributions are uniform over a given family of rooted maps. We consider rooted maps for accessability by generating function techniques. By the results in [12], any almost sure property of one of the classes of rooted maps in this paper is also an almost sure property of the corresponding unrooted versions.

The theory of submaps of a random map was begun in [10] and [11] and extended in a general way in [2], where it is shown that a random rooted map with n edges almost surely contains at least cn copies of any given planar submap for some positive constant c. In [6] it was shown that the number X_M of copies of a given map in a random 3-connected triangulation with $n+2$ vertices is sharply concentrated around $2rn(27/256)^j$, where $j+3$ is the number of vertices in M and there are r ways to root M. This result, which also applies to near-triangulations M (i.e., maps with all internal faces triangles), was obtained by deriving asymptotic expressions for the first two moments using somewhat complicated multivariate asymptotic analysis of the generating functions. With Theorem 1 available, we can derive a stronger result—the asymptotic distribution of X_M—using a much simpler combinatorial argument which estimates the factorial moments of X_M. The method also works for other families of maps. As an example, we also derive similar results for 2-connected triangulations and quadrangulations with no multiple edges.

Although the simple argument used in this paper gives the asymptotic distribution for the number of copies of a submap, it does not provide any bound on the rate of convergence. To some extent this complements the results of [2], where the limiting distribution is not obtained but an exponentially small bound is found for the probabilty that the number of submaps lies in a certain range which is well away from the expected number. On the other hand, it is possible to bound the upper tail of the distribution more sharply from the factorial moment than is done in Theorem 1.

In the following we use $\eta_n = \eta_n(M)$ to denote the number of copies of a subtriangulation M in a random rooted 3-connected triangulation with $n+2$ vertices. Any such M is necessarily 3-connected. *For simplicity, we only count copies that do not contain the root face of the random triangulation.* Note that $\eta_n(M)$ differs from the total number of copies by at most 1, since no two copies of M can share an interior face.

Lemma 1 *For each n, let M be a planar 3-connected triangulation with $j+3$ vertices. Suppose there are r distinct ways to root M. Then for $kj = o(n)$,*

$$\mathbf{E}\left([\eta_n]_k\right) = r^k[2(n-kj)-1]_k \left(\frac{27}{256}\right)^{kj} \left(1 + \frac{5kj}{2n} + O(k^2j^2/n^2)\right).$$

Proof: Let $D_n(M)$ be the number of rooted triangulations with $n+2$ vertices and with a copy of M distinguished. Removing the distinguished copy of M from such a rooted triangulation yields a 3-connected triangulation which has $n+2-j$ vertices and has a face distinguished. We can reverse this process by inserting M back in r different ways. This gives an r-to-1 mapping. Hence

$$D_n(M) = r(2(n-j)-1)T_{n-j},$$

and by (4.1)

$$\begin{aligned} \mathbf{E}(\eta_n(M)) &= \frac{D_n(M)}{T_n} = 2rn\left(1 - \frac{2j+1}{2n}\right)\frac{T_{n-j}}{T_n} && (4.4)\\ &= 2rn\left(\frac{27}{256}\right)^j \left(1 + \frac{3j-1}{2n} + O(j^2/n^2)\right). && (4.5) \end{aligned}$$

Similarly for each integer k with $kj = o(n)$, we can consider 3-connected triangulations with k distinguished copies of M, where the distinguished copies are given a linear ordering. It is important to note that different copies of M cannot overlap since M is 3-connected. Hence the above argument also gives

$$\begin{aligned} \mathbf{E}\left([\eta_n]_k\right) &= r^k[2(n-kj)-1]_k \frac{T_{n-kj}}{T_n} \\ &= r^k[2(n-kj)-1]_k \left(\frac{27}{256}\right)^{kj} \left(1 + \frac{5kj}{2n} + O(k^2j^2/n^2)\right). \quad \blacksquare \end{aligned}$$

Theorem 3 *Let r be as defined in Lemma 1, put*

$$\mu_n = 2rn\left(\frac{27}{256}\right)^j, \quad \sigma_n = \sqrt{\mu_n - \frac{(4j+1)\mu_n^2}{2n}},$$

and assume $\mu_n \to \infty$. Then $(\eta_n - \mu_n)/\sigma_n$ tends in distribution to the standard normal as $n \to \infty$.

Proof: We first note that

$$\mu_n - \frac{(4j+1)\mu_n^2}{2n} = \mu_n \left(1 - r(4j+1)\left(\frac{27}{256}\right)^j\right).$$

We also note that $r = 1$ when $j = 0, 1$, and hence

$$\left(1 - r(4j+1)\left(\frac{27}{256}\right)^j\right) > 0.$$

When $j \geq 3$, we use $r \leq 4 \times 3(j+1)$, and hence

$$r(4j+1)\left(\frac{27}{256}\right)^j \leq 48(j+1)^2\left(\frac{27}{256}\right)^j.$$

It is easy to see that the right hand side of the above inequality is decreasing for $j \geq 3$ and is less than 1 at $j = 3$. Hence

$$c_1\sqrt{\mu_n} < \sigma_n < c_2\sqrt{\mu_n}$$

for some positive constants c_1 and c_2, and $\mu_n/\sigma_n = O(\sqrt{\mu_n}) = O(\sqrt{n})$. Since $\mu_n \to \infty$, this implies $j = O(\log n)$, and hence

$$[2(n-kj)-1]_k = (2n)^k \exp\left(-\frac{k^2(4j+1)}{4n}\right)(1+o(1))$$

for $k = O(\sqrt{n})$, and (using Lemma 1)

$$\mathbf{E}([\eta_n]_k) = \mu_n^k \exp\left(-\frac{k^2(4j+1)}{4n}\right)(1+o(1)).$$

The theorem now follows from Theorem 1. ∎

It is clear that the above argument works equally well for other families of maps with specified face degrees, provided that their asymptotical expressions similar to (4.1) are known and the submap considered has the required non-overlapping property. For example, using (4.2) and (4.3), we obtain the following.

Theorem 4 **(i)** *Let M be a 3-connected triangulation which has $j+3$ vertices and r distinct rootings. Let η_n be the number of copies of M in a random rooted 2-connected triangulation with $n+2$ vertices. Define*

$$\mu_n = 2rn(27/2)^{-j}, \;\; \sigma_n = \sqrt{\mu_n - \frac{(4j+1)\mu_n^2}{2n}},$$

and suppose $\mu_n \to \infty$. Then $(\eta_n - \mu_n)/\sigma_n$ tends in distribution to the standard normal as $n \to \infty$.

(ii) *Let M be a quadrangulation which has $j+4$ vertices, r distinct rootings, and contains no separating quadrangles. Let η_n be the number of copies of M in a random rooted quadrangulation with $n+2$ vertices and with no multiple edges. Define*

$$\mu_n = rn(27/4)^{-j}, \;\; \sigma_n = \sqrt{\mu_n - \frac{(2j+1)\mu_n^2}{n}},$$

and suppose $\mu_n \to \infty$. Then $(\eta_n - \mu_n)/\sigma_n$ tends in distribution to the standard normal as $n \to \infty$.

We remark that the precise range of j in which $\mu_n \to \infty$ can be determined when the behavior of $r = r(j)$ is known. Suppose $\mu_n = 2rn\rho^{-j}$, and $c_1 j^\beta \le r \le c_2 j^\beta$ for some constants $c_1 > 0$, $c_2 > 0$, and $\beta \ge 0$. Then $\mu_n \to \infty$ when $j < (\log n + \beta \log\log n)/\log \rho - \Omega_n$, and $\mu_n \to 0$ when $j > (\log n + \beta \log\log n)/\log \rho + \Omega_n$, where Ω_n is any function which goes to ∞ as $n \to \infty$. Moreover, $\mu_n = O(1)$ when $|j - (\log n + \beta \log\log n)/\log \rho| = O(1)$, and hence Lemma 1 implies that η_n is asymptotically Poisson with mean μ_n.

Acknoweldgment The authors would like to thank Boris Pittel for some suggestions, and an anonymous referee for pointing out how to significantly shorten the proof of Theorems 1 and 2 and incidentally remove the condition $\sigma_n(\log^2 \sigma_n) = o(\mu_n)$.

References

[1] A.D. Barbour, L. Holst and S. Janson, *Poisson Approximation*, Clarendon Press, Oxford, 1992.

[2] E.A. Bender, Z.C. Gao and L.B. Richmond, Submaps of maps I: General 0-1 laws, *J. Combin. Theory Ser. B* **55** (1992) 104–117.

[3] P. Billingsley, *Probability and Measure*, 3nd ed., John Wiley & Sons, New York, 1995.

[4] W.G. Brown, Enumeration of quadrangular dissections of the disk, *Canad. J. Math.* **17**(3) (1965), 302–317.

[5] P. Erdős and A. Rényi, On the evolution of random graphs, *Publ. Math. Inst. Hungar. Acad. Sci.* **5** (1960) 17–61.

[6] Z.C. Gao and N.C. Wormald, Sharp concentration of the number of submaps in random planar triangulations, *Combinatorica* **23** (2003), 467–486.

[7] S. Janson, T. Łuczak and A. Ruciński, *Random graphs*, Wiley, New York, 2000.

[8] B. Pittel and R.S. Weishaar, The random bipartite nearest neighbor graphs, *Random Structures and Algorithms* **15** (1999), 279–310.

[9] M. Régnier and W. Szpankowski, On pattern frequency occurrences in a Markovian sequence. *Algorithmica* **22** (1998), 631–649.

[10] L.B. Richmond, R.W. Robinson and N.C. Wormald, On Hamilton cycles in 3-connected cubic maps, in "Cycles in Graphs" (B. Alspach and C.D. Godsil, eds.), *Annals of Discrete Mathematics* **27** (1985), 141–150.

[11] L.B. Richmond and N.C. Wormald, Random triangulations of the plane, *European J. Combinatorics* **9** (1988), 61–71.

[12] L.B. Richmond and N.C. Wormald, Almost all maps are asymmetric, *J. Combin. Theory, Ser. B* **63** (1995), 1–7.

[13] A. Ruciński, Proving normality in combinatorics. In *Random Graphs Vol. 2*, A. Frieze and T. Łuczak eds, Wiley, New York (1992), 215–231.

[14] W.T. Tutte, The enumerative theory of planar maps, in "A survey of Combinatorial Theory" (J.N. Srivastava *et al.*, Eds.), 437–448, North-Holland, Amsterdam, 1973.