Limit Theorems for Combinatorial Structures

Hsien-Kuei Hwang

LIX, École Polytechnique

November 8, 1993

[summary by Michèle Soria]

Abstract

This presentation concerns limiting distributions of parameters—like the number of components—in a variety of combinatorial objects. Under general analytic assumptions on the moment generating function, Hwang obtains complete asymptotic expansions for central and local limit theorems (expressing convergence to a Gaussian law), as well as quantitative estimates for probabilities of large deviations.

1. Introduction

It has been well-known since Gončarov that the number of cycles in a random permutation of size n has a Gaussian limiting distribution, with mean and variance both asymptotic to $\log n$. A similar asymptotic normality result (with a scaling factor of $\log\log n$ instead of $\log n$) was obtained by Erdős and Kac for the number of distinct prime factors of a random integer $\leq n$. These two results belong to different areas and were first proved by different techniques. It is shown here that they are in fact different facets of a common analytic structure.

A recent trend in asymptotic combinatorics is to explain similarity of distributions by similarity of "structure" (see e.g. [1, 3, 6]). In various types of combinatorial schemas, Flajolet and Soria [4, 5] proved a series of central limit theorems of the form

$$\Pr\left\{\frac{\Omega_n - \mu_n}{\sigma_n} < x\right\} \underset{n\to\infty}{\to} \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-t^2/2}\,dt.$$

There Ω_n is the number of components in a random object of size n, with mean μ_n and variance σ_n^2. The methods of proof rely on complex analysis for evaluating characteristic functions combined with continuity theorems for establishing convergence to the normal law. Using analytic techniques of probability theory, Hwang [7] gives a precise quantification of asymptotic normality. He obtains full asymptotic expansions for distribution functions and densities (this implies well-quantified convergence rates), together with estimates on probabilities of large deviations from the mean.

2. Central and local limit theorems

The starting point is a general condition for the moment generating function $M_n(s)$ of a sequence $\{\Omega_n\}$ of discrete random variables.

Condition 1. Assume that, uniformly for $|s| \leq \rho$ with $\rho > 0$,

$$M_n(s) \equiv \sum_m \Pr(\Omega_n = m)\, e^{ms} = e^{\phi(n)u(s)+v(s)}\left(1 + O\left(\frac{1}{\kappa_n}\right)\right), \qquad n \to \infty$$

where $u(s)$ and $v(s)$ are analytic for $|s| \leq \rho$, $u''(0) \neq 0$, and where $\phi(n)$ and κ_n tend to ∞ as $n \to \infty$.

The rate of convergence to the normal distribution results from applying Esseen's Theorem, a standard tool of probability theory (see e.g. [2]), that relates the distance between two distribution functions to the distance between corresponding characteristic functions.

THEOREM 1 (CONVERGENCE RATES). *Under condition 1,*

$$F_n(x) \equiv \Pr\left(\frac{\Omega_n - \mu_n}{\sigma_n} < x\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\, dt + O\left(\frac{1}{\kappa_n} + \frac{1}{\sqrt{\phi(n)}}\right),$$

uniformly with respect to x as $n \to +\infty$, where $\mu_n = u'(0)\phi(n)$ and $\sigma_n^2 = u''(0)\phi(n)$.

When convergence is slow (typically in applications the rate is of the order of $n^{-1/2}$, $(\log n)^{-1/2}$ or $(\log\log n)^{-1/2}$), it is useful to have a full asymptotic expansion. In fact, a slightly stronger condition on the function $u(s)$ of Condition 1 ("well-behavedness" of [7]) leads to a precise estimate on the characteristic function around $t = 0$, namely

$$\chi_n(t) = e^{-t^2/2}(1 + \sum P_k(it)/\sigma_n^k).$$

This permits in turn to obtain the asymptotic expansion for densities by means of Fourier inversion.

THEOREM 2 (LOCAL LIMIT THEOREM). *Under Condition 1, and if additionally $u(s)$ is "well-behaved" on $[-i\pi, +i\pi]$,*

$$\sigma_n \ \Pr\left(\frac{\Omega_n - \mu_n}{\sigma_n} = x\right) = \sum_{0 \leq k \leq \nu} \frac{p_k(x)}{\sigma_n^k} + O\left(\frac{1}{\kappa_n} + \frac{1}{\phi(n)^{(\nu+1)/2}}\right),$$

uniformly with respect to x as $n \to +\infty$, where

$$p_k(x) = \frac{d}{dx} P_k(-\Phi)(x) \qquad \text{and} \qquad \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\, dt.$$

Obtaining an asymptotic expansion for a central limit theorem is trickier, since the distribution function of Ω_n is a step function as Ω_n is discrete. The jumps at a discrete set of points are reflected by the "saw-tooth" function $\frac{1}{2} - \{x\}$ (with $\{x\}$ denoting the fractional part of x) and its repeated integrals. This leads to an oscillating component in expansions. The proof uses the method of Kubilius [8]. We only quote here a simplified version of the theorem, and refer to [7] for a complete statement.

THEOREM 3 (CENTRAL LIMIT THEOREM). *Under Condition 1, if $u(s)$ is "well-behaved" on the interval $[-i\pi, +i\pi]$,*

$$F_n(x) \sim \Phi(x) + \frac{e^{-x^2/2}}{\sqrt{2\pi}} \sum_{k=1}^{\infty} \frac{\pi_k(x) + \varpi_k(x)}{\sigma_n^k} \qquad (n \to \infty),$$

where $\pi_k(x)$ are polynomials of degree $3k + 1$ and $\varpi_k(x)$ are periodic functions.

3. Large deviations

Results of the preceding section deal with the behaviour of Ω_n at a distance $O(\sigma_n)$ from the mean. Probabilities of large deviations predict extreme cases, i.e., the situation when x is allowed to be at a distance $\gg \sigma_n$ from the mean.

It is known that analyticity of a moment generating function around 0 is associated with the occurrence of exponential tails for the corresponding probability distribution by Markov's inequality. Like in [5], the argument may be adapted to the $M_n(s)$, providing $\Pr(|\Omega_n - \mu_n| < x\sigma_n) = O(e^{-cx})$, with $c > 0$, for all $x > 0$. Actually, a much more precise formula can be obtained by using a method of Cramer–Kubilius [8]. It consists of two steps: the integral transform technique of associated distributions, and a saddle-point estimate. The next theorem generalizes Cramer's classical result on large deviations for sums of independent, identically distributed random variables.

THEOREM 4 (LARGE DEVIATIONS FOR CENTRAL LIMIT THEOREM). *Under Condition 1,*

$$\frac{1 - F_n(x)}{1 - \Phi(x)} = e^{\phi(n)Q(x/\sigma_n)}\left(1 + O\left(\frac{x}{\kappa_n} + \frac{x}{\sqrt{\phi(n)}}\right)\right), \qquad x = o(\min\{\kappa_n, \sqrt{\phi(n)}\}),$$

for $x > 0$, where $Q(t)$ is a function analytic at 0 whose coefficients are explicitly computable. A similar estimate holds for the symmetric side of the distribution corresponding to $x < 0$.

Hwang also proves an asymptotic expansion for $\Pr(\Omega_n = m)$, for m lying in the interval $\mu_n \pm o(\sigma_n^2)$. In this case the proof uses the saddle-point method applied to Laplace-type integrals.

THEOREM 5 (LARGE DEVIATIONS FOR LOCAL LIMIT THEOREM). *Let $m = \mu_n + x\sigma_n$ with $x = o(\sqrt{\phi(n)})$. Under the hypotheses of Theorem 4 and with the further assumption that $e^{u(r+it)}/e^{u(r)}$ "behaves like" a characteristic function, one has*

$$\Pr(\Omega_n = m) = \frac{e^{-\frac{x^2}{2} + \phi(n)Q(x/\sigma_n)}}{\sqrt{2\pi\phi(n)u''(0)}}\left(1 + \sum_{1 \le k \le \nu} \frac{P_k(x)}{(\phi(n)u''(0))^{k/2}} + O\left(\frac{x^{\nu+1} + 1}{\phi(n)^{(\nu+1)/2}} + \frac{1}{\kappa_n}\right)\right),$$

where $P_k(x)$ is a polynomial of degree k.

4. Application to combinatorial schemas

Decomposable combinatorial objects are built from *sets*, *sequences* or *cycles* of components. This is known to correspond to functional schemas of the form $P(w, z) = F(w, C(z))$. Here, $C(z)$ is the usual counting generating function of the component objects, and $P(w, z)$ is the bivariate generating function of the composite objects with w marking the number of components.

Let therefore $P(w, z) = \sum_{n,k} p_{nk} w^k z^n$ be such a function, so that p_{nk} is the number of structures of size n with k components; we are concerned with the asymptotic behaviour of the number of components in a random structure of size n whose probability distribution is given by $\Pr(\Omega_n = k) = p_{nk}/\sum_k p_{nk}$. Two major types of schemas are studied.

4.1. Exp-log schemas. These schemas are related to *Set* and *Multiset* constructions and they are already known to lead to Gaussian limit distributions [4]. The general form is $P(w, z) = \exp(wC(z) + S(w, z))$, where $C(z)$ is a function of logarithmic type ($a > 0$ and K a constant)

$$C(z) = a \log \frac{1}{1 - z/\rho} + K + o\left(\frac{1}{\log(1 - z/\rho)}\right) \qquad (z \to \rho, z \notin [\rho, \infty[),$$

and $S(w, z)$ is an analytic function for $|z| < \rho + \epsilon$ and $|w| < 1 + \epsilon'$, for some $\epsilon, \epsilon' > 0$.

By singularity analysis, one gets for the moment generating function

$$M_n(s) = e^{\phi(n)u(s)+v(s)} \left(1 + o\left(\frac{1}{\log n}\right)\right),$$

uniformly for small s when $n \to \infty$, with $u(s) = e^s - 1$, $\phi(n) = a \log n$ and $v(s) = K(e^s - 1) + S(e^s, \rho) - S(1, \rho) + \log(\Gamma(a)/\Gamma(ae^s))$. Hence all the conditions of Theorems 1-5 are fulfilled in this case.

Asymptotic normality of the number of cycles in a permutation or of the number of components in a random mapping provide an illustration of exp-log schemas. Hwang notes that the same process applies to Dirichlet series instead of power series (see Hwang's *"Factorisatio Numerorum"* in this volume). Thus the number of distinct prime factors of a random integer $\leq n$ also fits into this analytic schema.

4.2. Alg-log schemas. Another type of schema from [5],

$$P(w, z) = \frac{1}{(1 - wC(z))^{\alpha}} \left(\log \frac{1}{1 - wC(z)}\right)^k,$$

is related to *Sequence* and *Cycle* constructions and is again known to lead to Gaussian laws under the conditions: k is a non-negative integer, $\alpha > 0$, and $C(z)$ attains 1 before becoming singular.

By singularity analysis, the moment generating function is shown to have the right form for Theorems 1-5, with

$$u(s) = -\log \frac{\rho(e^s)}{\rho(1)}, \quad \phi(n) = n, \quad v(s) = -\alpha \log \frac{\rho(e^s)C'(\rho(e^s))}{\rho(1)C'(\rho(1))}.$$

Bender's schema for meromorphic functions [1] also fits within this framework.

In conclusion very precise quantitative asymptotic normality results hold for many types of combinatorial objects and number-theoretic functions.

Bibliography

[1] Bender (Edward A.). – Central and local limit theorems applied to asymptotic enumeration. *Journal of Combinatorial Theory, Series A*, vol. 15, 1973, pp. 91–111.

[2] Billingsley (Patrick). – *Probability and Measure.* – John Wiley & Sons, 1986, 2nd edition.

[3] Canfield (E. Rodney). – Central and local limit theorems for the coefficients of polynomials of binomial type. *Journal of Combinatorial Theory, Series A*, vol. 23, 1977, pp. 275–290.

[4] Flajolet (Philippe) and Soria (Michèle). – Gaussian limiting distributions for the number of components in combinatorial structures. *Journal of Combinatorial Theory, Series A*, vol. 53, 1990, pp. 165–182.

[5] Flajolet (Philippe) and Soria (Michèle). – General combinatorial schemas: Gaussian limit distributions and exponential tails. *Discrete Mathematics*, vol. 114, 1993, pp. 159–180.

[6] Gao (Zhicheng) and Richmond (L. Bruce). – Central and local limit theorems applied to asymptotic enumerations IV: Multivariate generating functions. *Journal of Computational and Applied Mathematics*, vol. 41, 1992, pp. 177–186.

[7] Hwang (Hsien-Kuei). – *Théorèmes limites pour les constructions combinatoires et les fonctions arithmétiques.* – Thèse de Doctorat, École Polytechnique, 1994.

[8] Kubilius (J.). – *Probabilistic Methods in the Theory of Numbers.* – American Mathematical Society, Providence, Rhode Island, 1964.