Gaussian fluctuations for random partitions

Zhonggen Su*[1,2] & Leonid Bogachev[2]

[1] Department of Mathematics, University of Zhejiang
[2] Department of Statistics, University of Leeds

1 Introduction

A partition λ of an integer number $n \geq 1$ is any integer sequence $\lambda = (\lambda_1, \lambda_2, \dots)$ such that $\lambda_1 \geq \lambda_2 \geq \dots \geq 0$ and $n = \lambda_1 + \lambda_2 + \cdots$ (notation: $\lambda \vdash n$). In particular, $\lambda_1 = \max\{\lambda_i \in \lambda\}$. The standard geometric object associated to the partition $\lambda \vdash n$ is its *Young diagram* consisting of n unit square cells arranged in adjacent columns (placed on common ground) in such a way that the i-th column contains λ_i cells.

On the set $\mathcal{P}_n := \{\lambda \vdash n\}$ of all partitions of n, consider the Plancherel measure defined as

$$P_n(\lambda) := \frac{d_\lambda^2}{n!}\,, \qquad \lambda \in \mathcal{P}_n,$$

where d_λ is the number of standard tableaux of shape λ, that is, the number of all allocations of the numbers $1, \dots, n$ in n cells of the Young diagram corresponding to the partition λ, such that the numbers in the cells increase along each row (from left to right) and each column (bottom up). Note that P_n is a probability measure due to the Burnside identity, $\sum_{\lambda \in \mathcal{P}_n} d_\lambda^2 = n!$. The Plancherel measure arises naturally in representation-theoretic, combinatorial, and probabilistic problems (see Deift, 2000). For example, the Plancherel distribution of the largest term of a partition, λ_1, coincides with the distribution of the longest increasing subsequence in a random (uniformly distributed) permutation of $1, \dots, n$ (see Baik *et al.*, 1999).

The upper boundary of the Young diagram corresponding to partition $\lambda \in \mathcal{P}_n$ can be viewed as the graph of the stepwise (left-continuous) function $\lambda(x)$ defined as

$$\lambda(0) := \lambda_1, \qquad \lambda(x) := \sum_{i=1}^{\infty} \lambda_i \mathbf{1}_{\{i-1<x\leq i\}} \equiv \lambda_{\lceil x \rceil} \qquad (x > 0), \tag{1.1}$$

where $\lceil x \rceil := \min\{m \in \mathbb{Z} : m \geq x\}$ is the ceiling integer part of x. Logan and Shepp (1977) and, independently, Vershik and Kerov (1977) have discovered that, as $n \to \infty$, a typical Young diagram, suitably scaled, has a "limit shape", represented by the graph of some function $y = \omega(x)$. That is to say, for the overwhelming majority of partitions $\lambda \in \mathcal{P}_n$ (measured under the Plancherel law P_n), the scaled boundary of their Young diagrams is contained in an arbitrarily small vicinity of the graph of $\omega(x)$. More specifically, set

$$\tilde{\lambda}_n(x) := \frac{1}{\sqrt{n}}\, \lambda(\sqrt{n}x), \qquad x \geq 0.$$

and let $y = \omega(x)$ be the function defined by the parametric equations

$$x = \frac{2}{\pi}\,(\sin\theta - \theta\cos\theta), \qquad y = x + 2\cos\theta, \qquad 0 \leq \theta \leq \pi. \tag{1.2}$$

It follows that $\omega(x)$ is a decreasing function on $[0,2]$ and $\omega(0)=2,\ \omega(2)=0$. Then the limit shape result of Logan and Shepp (1977) and Vershik and Kerov (1977) states that $\tilde{\lambda}_n(x)$ satisfies the following functional law of large numbers:

$$\forall\varepsilon>0,\qquad \lim_{n\to\infty} P_n\Big\{\sup_{x\ge 0}\big|\tilde{\lambda}_n(x)-\omega(x)\big|>\varepsilon\Big\}=0.$$

The natural question about fluctuations of the random function $\tilde{\lambda}_n$ around the curve ω was also posed in those papers. However, it remained open until the remarkable progress was made by Baik *et al.* (1999), who proved that the limit distribution of the largest term of the partition is given by the Tracy–Widom law:

$$\frac{\lambda_1-2\sqrt{n}}{n^{1/6}}\xrightarrow{\ d\ }F_{TW}.$$

Shortly afterwards, Borodin *et al.* (2000), Johansson (2001), and Okounkov (2000) extended this result (independently and via different methods) to an arbitrary $\lambda_i\in\lambda\vdash n$ by showing that the sequence

$$\frac{\lambda_i-2\sqrt{n}}{n^{1/6}},\qquad i=1,2,\dots$$

converges in distribution to the Airy ensemble (in the sense of finite-dimensional distributions). On the other hand, Kerov (1993) proved the asymptotic normality of global (integral) fluctuations (with respect to Chebyshev polynomial kernels). However, no central limit theorem has been known (although largely expected to be true) for the partition "spectrum" *in the bulk*, that is, for $\lambda_i\in\lambda\vdash n$ with $i/n\sim x\in(0,2)$.

2 Main results and sketch of proofs

The main result of this paper is the following central limit theorem for the function $\tilde{\lambda}_n(x)$.
Theorem 2.1. *For each* $0<x<2$*, under the Plancherel measure*

$$\frac{2\theta_x\sqrt{n}\,\big(\tilde{\lambda}_n(x)-\omega(x)\big)}{\sqrt{\log n}}\xrightarrow{\ d\ }\mathcal{N}(0,1)\qquad(n\to\infty),$$

where $\theta_x=\arccos\frac{\omega(x)-x}{2}$ *is the value of the parameter* θ *in equations* (1.2) *corresponding to* x *and* $y=\omega(x)$*, and* $\mathcal{N}(0,1)$ *is the standard normal distribution* (*with zero mean and unit variance*)*.*

The proof of Theorem 2.1 is based on the standard Poissonization and de-Poissonization techniques (see, e.g., Baik *et al.*, 1999). Let $\mathcal{P}=\cup_{n=0}^{\infty}\mathcal{P}_n$ be the set of all partitions of all integers (it is convenient to include the case $n=0$, where there is only one, "empty" partition, $\lambda_\emptyset\vdash 0$). Set $|\lambda|:=\sum_{i=1}^{\infty}\lambda_i$, and for $t>0$ define the Poissonization P^t of P_n as follows:

$$P^t(\lambda):=e^{-t^2}t^{2|\lambda|}\left(\frac{d_\lambda}{|\lambda|!}\right)^2,\qquad \lambda\in\mathcal{P}. \tag{2.3}$$

This is a probability measure on $\mathcal{P}$, because for $\lambda\in\mathcal{P}_n$ we have $|\lambda|=n$ and so

$$\sum_{\lambda\in\mathcal{P}}P^t(\lambda)=\sum_{n=0}^{\infty}\frac{e^{-t^2}t^{2n}}{n!}\sum_{\lambda\in\mathcal{P}_n}\frac{d_\lambda^2}{n!}=\sum_{n=0}^{\infty}\frac{e^{-t^2}t^{2n}}{n!}=1.$$

The key fact is that the correlation functions of P^t have *determinantal* structure. Set

$$D(\lambda) := \cup_{i=1}^{\infty} \{\lambda_i - i\}, \qquad \lambda \in \mathcal{P},$$

and for a given (finite) set $X \subset \mathbb{Z}$ define the correlation function $\rho_t(X)$ as the P^t-probability that the random set $D(\lambda)$ contains X:

$$\rho_t(X) := P^t\bigl\{\lambda \in \mathcal{P} : X \subset D(\lambda)\bigr\}.$$

Then, as was proved by Borodin *et al.* (2000) and Johansson (2001), ρ_t is represented as

$$\rho_t(X) = \det[J(x_i, x_j; t)]_{1\le i,j\le s}, \qquad X = \{x_1, x_2, \cdots, x_s\}, \tag{2.4}$$

with the kernel J given by

$$J(x,y;t) = \begin{cases} t\,\dfrac{J_x J_{y+1} - J_{x+1} J_y}{x-y}\,, & x \ne y, \\ t\left(J'_x J_{x+1} - J_x J'_{x+1}\right), & x = y, \end{cases}$$

where $J_m = J_m(2t)$ is the Bessel function of integral order m.

We can then apply a powerful theorem due to Soshnikov (2000) (generalizing an earlier result by Costin and Lebowitz, 1995), stating that the central limit theorem holds for determinantal random point processes under some mild conditions. This yields

Theorem 2.2. *Under the poissonized Plancherel measure P^t, for any $0 < x < 2$*

$$\frac{2\theta_x\bigl(\lambda(tx) - t\omega(x)\bigr)}{\sqrt{2\log t}} \xrightarrow{d} \mathcal{N}(0,1) \qquad (t \to \infty), \tag{2.5}$$

where θ_x is as in Theorem 2.1.

Let us sketch the proof of Theorem 2. Note that (2.5) is equivalent to saying that

$$\forall y \in \mathbb{R}, \qquad \lim_{t\to\infty} P^t \left\{ \lambda(tx) - \lceil tx \rceil \le 2t\cos\theta_x + \frac{y\sqrt{2\log t}}{2\theta_x} \right\} = \Phi(y), \tag{2.6}$$

where $\Phi(\cdot)$ is the standard normal distribution function. Set

$$I_t := [2t\cos\theta_x + y\sqrt{2\log t}/2\theta_x, \infty), \qquad D_t(\lambda) := D(\lambda) \cap I_t,$$

and let $\#D_t(\lambda)$ denote the cardinality of $D_t(\lambda)$, that is, the number of points $\lambda_i - i \in D(\lambda)$ falling into the interval I_t. Using that the sequence $\lambda_i - i$ is strictly decreasing and recalling the definition (1.1) of $\lambda(\cdot)$, it is easy to see that the inequality in (2.6) amounts to

$$\#D_t(\lambda) \le \lceil tx \rceil.$$

It remains to work out the asymptotics of the first two moments of $\#D_t(\lambda)$ (with respect to P^t).

Lemma 2.1. *Fix $0 < x < 2$ and $y \in \mathbb{R}$. Then, as $t \to \infty$,*

$$E\bigl(\#D_t(\lambda)\bigr) = tx - \frac{y\sqrt{2\log t}}{2\pi} + O(1),$$
$$Var\bigl(\#D_t(\lambda)\bigr) = \frac{\log t}{2\pi^2}\,(1 + o(1)).$$

Calculations in the proof of the lemma are straightforward but laborious, heavily using the asymptotics of Bessel functions in various regions.

Finally, in order to derive Theorem 1 from Theorem 2, we use the de-Poissonization method. According to (2.3), P^t can be viewed as the expectation of P_N, where N is a Poisson random variable with mean t^2. In particular, denoting the event under P^t in (2.6) by A_t, we have

$$P^t(A_t) = \sum_{k=0}^{\infty} \frac{e^{-t^2} t^{2k}}{k!} P_k(A_t).$$

This suggests that the asymptotics of the probability $P_n(A_{\sqrt{n}})$ as $n \to \infty$ can be recovered from that of $P^t(A_t)$ as $t \sim \sqrt{n} \to \infty$. More precisely, one can prove that

$$P_n(A_{\sqrt{n}}) \sim P^t(A_t), \qquad t \sim \sqrt{n}, \quad n \to \infty,$$

provided that the variations of $P_k(A_t)$ in the region $|k - n| \le \text{const}\sqrt{n},\ t \sim \sqrt{n}$, are small. In the setting of random partitions, such a de-Poissonization lemma was established by Johansson (see Baik *et al.*, 1999).

Acknowledgements

This work was done when Zhonggen Su was visiting the University of Leeds. His research was supported by a Royal Society K.C. Wong International Fellowship.

References

Baik, J., Deift, P. and Johansson, K. (1999). On the distribution of the length of the longest increasing subsequence of random permutations. *J. Amer. Math. Soc.* **12**, 1119-1178.

Borodin, A., Okounkov, A. and Olshanski, G. (2000). Asymptotics of Plancherel measures for symmetric groups. *J. Amer. Math. Soc.* **13**, 481-515.

Costin, O. and Lebowitz, J.L. (1995). Gaussian fluctuation in random matrices. *Phys. Rev. Lett.* **75**, 69-72.

Deift, P. (2000). Integrable systems and combinatorial theory. *Notices Amer. Math. Soc.* **47**, 631-640.

Johansson, K. (2001). Discrete orthogonal polynomials ensembles and the Plancherel measure. *Ann. Math.* **153**, 259-296.

Kerov, S. (1993). Gaussian limit for the Plancherel measure of the symmetric group. *C. R. Acad. Sci. Paris Sér. I Math.* **316**, 303-308.

Logan, B.F. and Shepp, L.A. (1977). A variational problem for random Young tableaux. *Adv. Math.* **26**, 206-222.

Okounkov, A. (2000). Random matrices and random permutations. *Internat. Math. Res. Notices* **2000**, No. 20, 1043-1095.

Soshnikov, A.B. (2000). Gaussian fluctuation for the number of particles in Airy, Bessel, sine, and other determinantal random point fields. *J. Statist. Phys.* **100**, 491-522.

Vershik, A.M. and Kerov, S. (1977). Asymptotics of the Plancherel measure of the symmetric group and the limiting form of Young tables. *Soviet Math. Doklady* **18**, 527-531.