

# Gaussian fluctuations of Young diagrams under the Plancherel measure

BY LEONID V. BOGACHEV<sup>1,\*</sup> AND ZHONGGEN SU<sup>2</sup>

<sup>1</sup>*Department of Statistics, University of Leeds, Leeds LS2 9JT, UK*

<sup>2</sup>*Department of Mathematics, Zhejiang University, Hangzhou 310 027,  
People's Republic of China*

We obtain the central limit theorem for fluctuations of Young diagrams around their limit shape in the bulk of the ‘spectrum’ of partitions  $\lambda \vdash n \in \mathbb{N}$  (under the Plancherel measure), thus settling a long-standing problem posed by Logan & Shepp. Namely, under normalization growing like  $\sqrt{\log n}$ , the corresponding random process in the bulk is shown to converge, in the sense of finite-dimensional distributions, to a Gaussian process with independent values, while local correlations in the vicinity of each point, measured on various power scales, possess certain self-similarity. The proofs are based on the Poissonization techniques and use Costin–Lebowitz–Soshnikov’s central limit theorem for determinantal random point processes. Our results admit a striking reformulation after the rotation of Young diagrams by  $45^\circ$ , whereby the normalization no longer depends on the location in the spectrum. In addition, we explain heuristically the link with an earlier result by Kerov on the convergence to a generalized Gaussian process.

**Keywords:** partitions; Young diagrams; Plancherel measure; limit shape;  
central limit theorem; determinantal random point process

## 1. Introduction

*Partitions* of natural numbers (i.e. decompositions  $\lambda \vdash n \in \mathbb{N}$  into sums of non-negative integer parts,  $n = \lambda_1 + \lambda_2 + \dots$ ,  $\lambda_1 \geq \lambda_2 \geq \dots \geq 0$ ) and their geometric counterpart, *Young diagrams* (i.e. planar figures formed by consecutive columns with  $\lambda_1, \lambda_2, \dots$  unit square cells), are a classical subject of discrete mathematics dating back to Euler, which has since grasped the imagination of many mathematicians including G. H. Hardy and S. Ramanujan (see [Andrews 1976](#)). More recently, interest in asymptotic properties of a ‘typical’ partition  $\lambda \vdash n$  as  $n \rightarrow \infty$  has led to the idea to treat the set of partitions  $\mathcal{P}_n = \{\lambda \vdash n\}$  as a random ensemble equipped with a uniform probability measure (so that each partition of  $n$  is ‘equally likely’). The principal technical device that makes this idea work is based on a suitable randomization of the parameter  $n$  (thus replacing  $\mathcal{P}_n$  by  $\mathcal{P} = \bigcup_n \mathcal{P}_n$ ), which can be accomplished in such a way that the parts of a random partition  $\lambda \in \mathcal{P}$  become statistically independent ([Fristedt 1993](#); [Vershik 1995, 1996](#)).

\* Author for correspondence (bogachev@maths.leeds.ac.uk).

In memory of Sergei Kerov (1946–2000).

The probabilistic approach has proved to be extremely powerful and efficient, offering a unified explanation of statistical properties obtained earlier and, more importantly, yielding many new, often surprising discoveries including the ‘limit shape’ of Young diagrams associated with uniform partitions (Vershik 1995, 1996). Similar methods have also led to deep results about the asymptotic structure of the constituting components (‘spectrum’) for many other classes of growing combinatorial objects (see Arratia & Tavaré 1994; Arratia *et al.* 2003).

Quite a different type of probability distribution on the space  $\mathcal{P}_n = \{\lambda \vdash n\}$  is the *Plancherel measure*, motivated by representation theory of the symmetric group  $\mathfrak{S}_n$  (i.e. the group of permutations of order  $n$ ). It is defined by (see Fulton 1997)

$$P_n(\lambda) := \frac{d_\lambda^2}{n!}, \quad \lambda \in \mathcal{P}_n, \tag{1.1}$$

where  $d_\lambda$  is the number of *standard tableaux* of a given shape  $\lambda$  (i.e. arrangements of the numbers  $1, \dots, n$  in the cells of the Young diagram  $\lambda$ , such that the numbers increase along the rows and up the columns). The well-known Robinson–Schensted–Knuth (RSK) correspondence (see Fulton 1997) establishes a bijection between permutations  $\sigma \in \mathfrak{S}_n$  and pairs of standard tableaux of the same shape  $\lambda(\sigma) \in \mathcal{P}_n$ . This implies that  $\sum_{\lambda \in \mathcal{P}_n} d_\lambda^2 = \#\mathfrak{S}_n = n!$ , and so equation (1.1) determines a probability distribution on  $\mathcal{P}_n$ .

Besides representation theory, the Plancherel measure arises rather unexpectedly in some combinatorial and probabilistic problems (see Deift 2000). For example, the RSK correspondence implies that the Plancherel distribution of the partition’s largest term  $\lambda_1 = \max\{\lambda_i \in \lambda \vdash n\}$  coincides with the distribution of the length  $L_n$  of the longest increasing subsequence in a random (uniformly distributed) permutation  $\sigma \in \mathfrak{S}_n$  (Baik *et al.* 1999; Deift 2000). The asymptotic behaviour of  $L_n$  was queried by Ulam’s Problem (Deift 2000), ultimately solved by Vershik & Kerov (1977) who showed that  $L_n$  grows like  $2\sqrt{n}$  (cf. (1.6)).

Logan & Shepp (1977) and, independently, Vershik & Kerov (1977) discovered that, under the Plancherel measure  $P_n$  as  $n \rightarrow \infty$ , a typical Young diagram  $\lambda \vdash n$ , appropriately scaled, has a certain limit shape  $y = \omega(x)$ . More precisely, set

$$\lambda(x) := \begin{cases} \lambda_1, & x = 0, \\ \lambda_{\lceil x \rceil}, & x > 0, \end{cases} \tag{1.2}$$

where  $\lceil x \rceil := \min\{m \in \mathbb{Z}: m \geq x\}$  is the ceiling integer part of  $x$ , and let the function  $y = \omega(x)$ ,  $x \geq 0$ , be defined on  $[0, 2]$  by the parametric equations

$$x = \frac{2}{\pi}(\sin \theta - \theta \cos \theta), \quad y = x + 2 \cos \theta, \quad 0 \leq \theta \leq \pi, \tag{1.3}$$

and continued as zero for all  $x > 2$ . Then, the random process

$$\Delta_n(x) := \lambda(\sqrt{n}x) - \sqrt{n}\omega(x), \quad x \geq 0, \tag{1.4}$$

satisfies the law of large numbers (Vershik & Kerov 1977)

$$\forall \varepsilon > 0, \quad \lim_{n \rightarrow \infty} P_n \left\{ \frac{1}{\sqrt{n}} \sup_{x \geq 0} |\Delta_n(x)| > \varepsilon \right\} = 0. \tag{1.5}$$

In particular, since  $\Delta_n(0) = \lambda_1 - 2\sqrt{n}$ , formula (1.5) implies the law of large numbers for the maximum term in a random partition,

$$\forall \varepsilon > 0, \quad \lim_{n \rightarrow \infty} P_n \left\{ \left| \frac{\lambda_1}{\sqrt{n}} - 2 \right| > \varepsilon \right\} = 0. \tag{1.6}$$

**Remark 1.1.** Owing to the invariance of the Plancherel measure under the transposition of Young diagrams  $\lambda \leftrightarrow \lambda'$  (when columns of the diagram  $\lambda$  become rows of the transposed diagram  $\lambda'$  and vice versa), the same law of large numbers holds for  $\lambda'_1 = \#\{\lambda_i \in \lambda\}$  (i.e. for the number of parts in the random partition  $\lambda$ ).

## 2. Fluctuations of Young diagrams

A natural question about the limit distribution for fluctuations of Young diagrams under the Plancherel measure was posed by Logan & Shepp (1977). One might expect that a central limit theorem should hold, at least *in the bulk* of the partition spectrum (i.e. for  $\lambda_i \in \lambda \vdash n$  such that  $i/\sqrt{n} \sim x \in (0, 2)$ ). However, the problem has remained open as yet, and it is the purpose of the present paper to tackle it. In fact, we will show that the random process  $\Delta_n(x)$ ,  $x \in (0, 2)$ , normalized by  $\sqrt{\log n}$ , converges in the sense of finite-dimensional distributions (FDDs) to a Gaussian process with zero mean and independent values (§3). Moreover, local correlations in the vicinity of each point, measured on various power scales, possess certain self-similarity properties (§§3a and 7). The randomization method (called ‘Poissonization’, see §4) will be crucial for our considerations, but unlike the uniform distribution case it does not remove the dependence between the parts. Instead, our argument is based on the fundamental fact that the correlation functions of the Poissonized point process  $\{\lambda_i - i\}$  have a *determinantal structure*, so one may apply a general central limit theorem for determinantal random point processes (§5). Unfortunately, the corresponding calculations are quite involved, so in order to keep this paper to a reasonable size we will only outline the proof, leaving most of the more technical details to a separate publication.

The FDD convergence to a Gaussian process that we establish in the present paper should be contrasted with an earlier result by Kerov (1993) who proved the asymptotic normality of *global* fluctuations (i.e. integrals of  $\Delta_n(x)$  with respect to a suitable class of test functions). That is to say, the random process  $\Delta_n(x)$  converges in distribution (without any normalization!) to a *generalized* Gaussian process, whose sample paths make sense only as generalized functions rather than ordinary functions (cf. (2.2)).

To state Kerov’s result more precisely, it is convenient to pass to the coordinates  $u = x - y$ ,  $v = x + y$ , corresponding to anticlockwise rotation by  $45^\circ$  and dilation by  $\sqrt{2}$ . In the new coordinates, the boundary of the Young diagram is determined by the corresponding piecewise-linear (continuous) function  $\tilde{\lambda}(u)$ , and the limit shape is given by (Vershik & Kerov 1977)

$$\Omega(u) := \begin{cases} \frac{2}{\pi} \left( u \arcsin \frac{u}{2} + \sqrt{4 - u^2} \right), & |u| \leq 2, \\ |u|, & |u| \geq 2. \end{cases}$$

Then (Kerov 1993; see also Ivanov & Olshanski 2002), the random process

$$\tilde{\Delta}_n(u) := \tilde{\lambda}(\sqrt{n}u) - \sqrt{n}\Omega(u), \quad u \in \mathbb{R}, \tag{2.1}$$

converges in distribution to a generalized Gaussian process  $\tilde{\Delta}(u)$ ,  $u \in [-2, 2]$ , defined by the formal random series

$$\tilde{\Delta}(2 \cos \theta) = \frac{2}{\pi} \sum_{k=2}^{\infty} \frac{X_k \sin(k\theta)}{\sqrt{k}}, \quad \theta \in [0, \pi], \tag{2.2}$$

where  $\{X_k\}$  are mutually independent random variables with standard normal distribution  $\mathcal{N}(0, 1)$ .

Convergence in a generalized sense is a manifestation of increasing ‘roughness’ of Young diagrams, which is nevertheless suppressed after integration. Admittedly, this result might have cast doubt on the validity of the usual convergence, so it is quite surprising that an ordinary limit does exist (under a modest logarithmic normalization). Let us point out that the limiting Gaussian process appears to be a continuous-time ‘white noise’ (i.e. a continuum set of independent normal variables), which is very rough in its own right, with sample paths being nowhere measurable! Coexistence of both generalized and ordinary (though extremely irregular) versions of the limit must seem striking indeed, and we will comment more on this peculiar situation in §6.

Note that the asymptotic behaviour of fluctuations at the upper edge of the limiting spectrum (corresponding to  $x=0$ ) is different from Gaussian. As was shown by Baik *et al.* (1999) for  $\lambda_1$  and by Borodin *et al.* (2000), Okounkov (2000) and Johansson (2001) for any  $\lambda_k$  with fixed  $k=1, 2, \dots$ ,

$$\lim_{n \rightarrow \infty} P_n \left\{ \frac{\lambda_k - 2\sqrt{n}}{n^{1/6}} \leq z \right\} = F_k(z), \quad z \in \mathbb{R}, \tag{2.3}$$

where  $F_k(\cdot)$  is the cumulative distribution function (CDF) of the  $k$ -th order statistic in the Airy point process discovered earlier by Tracy & Widom (1994) in connection with the limit distribution of largest eigenvalues of random matrices in the Gaussian unitary ensemble (GUE). In particular,  $F_1(\cdot)$  is known as the *Tracy–Widom* CDF.

### 3. Main results

#### (a) *Limit theorems in the bulk*

Recall that the random variable  $\Delta_n(x)$  is given by equation (1.4), and set

$$Y_n(x) := \frac{2\theta(x)\Delta_n(x)}{\sqrt{\log n}}, \quad x \geq 0, \tag{3.1}$$

where

$$\theta(x) = \arccos \frac{\omega(x) - x}{2} \tag{3.2}$$

is the value of the parameter  $\theta$  in equations (1.3) corresponding to the coordinates  $x$  and  $y=\omega(x)$ .

**Theorem 3.1.** *Suppose that  $x_n \rightarrow x \in (0, 2)$  as  $n \rightarrow \infty$ . Then, the distribution of the random variable  $Y_n(x_n)$  with respect to the Plancherel measure  $P_n$  converges to the standard normal distribution  $\mathcal{N}(0, 1)$ .*

The asymptotics of the FDDs of the random process  $\Delta_n(x)$  are described by theorem 3.2. We write  $c_n \asymp 1$  if  $c_n n^\varepsilon \rightarrow \infty$ ,  $c_n n^{-\varepsilon} \rightarrow 0$  for any  $\varepsilon > 0$ , and  $a_n \asymp b_n$  if  $a_n/b_n \asymp 1$ .

**Theorem 3.2.** *Suppose that  $x_{n,i} \rightarrow x_i \in (0, 2)$  as  $n \rightarrow \infty$  ( $i=0, \dots, m$ ) and  $x_{n,i} - x_{n,i-1} \asymp n^{-s_i/2}$ ,  $0 \leq s_i \leq 1$  ( $i=1, \dots, m$ ). Then, the random vector  $(Y_n(x_{n,0}), \dots, Y_n(x_{n,m}))$  converges in distribution to a Gaussian vector  $(Z_{s_0}, \dots, Z_{s_m})$  with zero mean and covariance matrix  $\mathbf{K}$  with the elements  $K_{ij} = K_{ji} = \min_{i < k \leq j} s_k$  ( $i < j$ ),  $K_{ii} = 1$ .*

Theorem 3.2 has a number of insightful implications. First of all, by choosing  $s_1 = \dots = s_m = 0$ , we obtain an FDD generalization of theorem 3.1.

**Corollary 3.3.** *Suppose that  $x_{n,i} \rightarrow x_i \in (0, 2)$  as  $n \rightarrow \infty$  ( $i=0, \dots, m$ ), where  $x_i \neq x_j$  ( $i \neq j$ ). Then, the random vector  $(Y_n(x_{n,0}), \dots, Y_n(x_{n,m}))$  converges in distribution to a Gaussian vector with independent standard normal components.*

The local structure of correlations with regard to a given power scale in the vicinity of any point is described as follows.

**Corollary 3.4.** *Fix  $x \in (0, 2)$ ,  $s \in (0, 1]$ , and set  $x_n(t) := x + tn^{-s/2}$ . Then, the random process  $Y_n(x_n(t))$  converges, in the FDD sense, to a stationary Gaussian process  $Z^{(s)}(t)$  ( $t \in \mathbb{R}$ ) with zero mean and the covariance function  $K^{(s)}(t, t') = s$  ( $t \neq t'$ ),  $K^{(s)}(t, t) = 1$ .*

Another interesting consequence is obtained for a variable spatial scale.

**Corollary 3.5.** *Fix  $x \in (0, 2)$ ,  $a \neq 0$ , and set  $x_n(s) := x + an^{-s/2}$  ( $0 \leq s \leq 1$ ). Then, the random process  $Y_n(x_n(s))$  converges, in the FDD sense, to a Gaussian process  $Z(s)$  ( $0 \leq s \leq 1$ ) with zero mean and the covariance function  $K(s, s') = \min\{s, s'\}$  ( $s \neq s'$ ),  $K(s, s) = 1$ .*

**Remark 3.1.** Results similar to theorems 3.1 and 3.2 were obtained by Gustavsson (2005) in the bulk of the spectrum of random matrices in the GUE.

**Remark 3.2.** By a well-known criterion of convergence in distribution (Billingsley 1999, theorem 2.3), the condition  $x_{n,i} \rightarrow x_i \in (0, 2)$  in theorems 3.1 and 3.2 may be relaxed to  $0 < \liminf_{n \rightarrow \infty} x_{n,i} \leq \limsup_{n \rightarrow \infty} x_{n,i} < 2$ .

(b) ‘Rotated’ versions

It is instructive to reformulate theorems 3.1 and 3.2 in the ‘rotated’ coordinates  $u$  and  $v$  (see §2). To this end, one needs to find the sliding projection (divided by  $\sqrt{2}$ ) of the deviation  $\tilde{\Delta}_n(u)$  (see (2.1)) onto the line  $u + v = 0$  along the tangent to the graph of  $\mathcal{Q}(u)$  at the point  $u$ . Differentiating equations (1.3), we get

$$\frac{dv}{du} = \frac{x'_\theta + y'_\theta}{x'_\theta - y'_\theta} = \frac{2\theta}{\pi} - 1,$$

which implies

$$\tilde{\Delta}_n(u) = \frac{2\theta(x)}{\pi} \Delta_n(\tilde{x}_n)(1 + \eta_n),$$

where  $\tilde{x}_n \rightarrow x$  and  $\eta_n \rightarrow 0$  (in probability). Hence, setting

$$\tilde{Y}_n(u) := \frac{\pi \tilde{\Delta}_n(u)}{\sqrt{\log n}}, \quad u \in \mathbb{R}, \tag{3.3}$$

we obtain the following elegant versions of theorems 3.1 and 3.2, whereby—quite surprisingly—the normalization does not depend on the location in the spectrum.

**Theorem 3.6.** *If  $u_n \rightarrow u \in (-2, 2)$  as  $n \rightarrow \infty$ , then the distribution of the random variable  $\tilde{Y}_n(u_n)$  with respect to the Plancherel measure  $P_n$  converges to  $\mathcal{N}(0, 1)$ .*

**Theorem 3.7.** *Suppose that  $u_{n,i} \rightarrow u_i \in (-2, 2)$  as  $n \rightarrow \infty$  ( $i=0, \dots, m$ ) and  $u_{n,i} - u_{n,i-1} \asymp n^{-s_i/2}$ ,  $0 \leq s_i \leq 1$  ( $i=1, \dots, m$ ). Then, the random vector  $(\tilde{Y}_n(u_{n,0}), \dots, \tilde{Y}_n(u_{n,m}))$  converges in distribution to a Gaussian vector  $(\tilde{Z}_{s_0}, \dots, \tilde{Z}_{s_m})$  with zero mean and the same covariance matrix  $\mathbf{K}$  as in theorem 3.2.*

Of course, implications parallel to corollaries 3.3–3.5 are also valid here.

(c) *Limit theorems at the edges*

Let us describe the asymptotics of the random function  $\Delta_n(x)$  at the ends of the limit spectrum, i.e. for  $x=0$  and  $x=2$ . By the definition (1.4),  $\Delta_n(0) = \lambda_1 - 2\sqrt{n}$ , and according to equation (2.3)

$$\lim_{n \rightarrow \infty} P_n \left\{ \frac{\Delta_n(0)}{n^{1/6}} \leq z \right\} = F_1(z), \quad z \in \mathbb{R},$$

where  $F_1(\cdot)$  is the Tracy–Widom CDF (see §2). However, the limit distribution of  $\Delta_n(2) = \lambda(2\sqrt{n})$  appears to be discrete.

**Theorem 3.8.** *Under the Plancherel measure  $P_n$ , for any  $z \geq 0$*

$$\lim_{n \rightarrow \infty} P_n \{ \Delta_n(2) \leq z \} = F_k(0), \quad k := [z] + 1,$$

where  $[z]$  is the integer part of  $z$  and the CDF  $F_k(\cdot)$  is the same as in equation (2.3).

Indeed, using the invariance of the measure  $P_n$  under the transposition  $\lambda \leftrightarrow \lambda'$  (see remark 1.1), we have, due to equation (2.3),

$$P_n \{ \Delta_n(2) \leq z \} = P_n \{ \lambda'_k < 2\sqrt{n} \} = P_n \left\{ \frac{\lambda_k - 2\sqrt{n}}{n^{1/6}} < 0 \right\} \rightarrow F_k(0).$$

In the rotated coordinates  $u$  and  $v$ , a similar result holds for both edges,

$$\lim_{n \rightarrow \infty} P_n \left\{ \frac{1}{2} \tilde{\Delta}_n(\pm 2) \leq z \right\} = F_k(0), \quad k = [z] + 1. \tag{3.4}$$

### 4. Poissonization

The proof of our results is based on the *Poissonization* technique (see Baik *et al.* 1999). Denote  $\mathcal{P} := \bigcup_{n=0}^{\infty} \mathcal{P}_n$  (by convention,  $\mathcal{P}_0$  consists of the ‘empty’ partition of zero) and for  $\lambda \in \mathcal{P}$ , set  $|\lambda| := \sum_{\lambda_i \in \lambda} \lambda_i$ . The Poissonization  $P^t$  ( $t > 0$ ) of the

measure  $P_n$  is defined as follows:

$$P^t(\lambda) := e^{-t} t^{|\lambda|} \left( \frac{d_\lambda}{|\lambda|!} \right)^2, \quad \lambda \in \mathcal{P}. \tag{4.1}$$

Formula (4.1) defines a probability measure on the set  $\mathcal{P}$ , since for  $\lambda \in \mathcal{P}_n$  we have  $|\lambda| = n$  and hence

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

We first establish the Poissonized versions of theorems 3.1 and 3.2. Let  $Y_t(\cdot)$  be given by formula (3.1) with  $n$  replaced by  $t$ .

**Theorem 4.1.** *Suppose that  $x_t \rightarrow x \in (0, 2)$  as  $t \rightarrow \infty$ . Then, the distribution of the random variable  $Y_t(x_t)$  with respect to the measure  $P^t$  converges to  $\mathcal{N}(0, 1)$ .*

**Theorem 4.2.** *In the notations of theorem 3.2 (with  $n$  replaced by  $t$ ), the random vector  $(Y_t(x_{t,0}), \dots, Y_t(x_{t,m}))$  converges in distribution (with respect to  $P^t$ ) to a Gaussian vector  $(Z_{s_0}, \dots, Z_{s_m})$  with zero mean and the same covariance matrix  $\mathbf{K}$ .*

To derive theorems 3.1 and 3.2 from theorems 4.1 and 4.2, we use a *de-Poissonization* method. According to equations (1.1) and (4.1),  $P^t$  can be viewed as the expectation of the random measure  $P_N$ , where  $N$  is a Poisson random variable with parameter  $t$ ,

$$P^t(A) = E(P_N(A)) = e^{-t} \sum_{k=0}^{\infty} \frac{t^k}{k!} P_k(A). \tag{4.2}$$

Since  $N$  has mean  $t$  and standard deviation  $\sqrt{t}$ , formula (4.2) suggests that the asymptotics of the probability  $P_n(A)$  as  $n \rightarrow \infty$  may be recovered from that of  $P^t(A)$  as  $t \sim n \rightarrow \infty$ . More precisely, one can prove that  $P_n(A) \sim P^t(A)$  as  $t \sim n \rightarrow \infty$ , provided that variations of  $P_k(A)$  are small in the zone  $k - n = O(\sqrt{n})$ . In the context of random partitions, such a result was obtained by Johansson (see Baik *et al.* 1999).

### 5. Sketch of the proof of theorem 4.1

Using equations (1.3) and (3.2), the statement of theorem 4.1 is equivalent to saying that for any  $z \in \mathbb{R}$

$$\lim_{t \rightarrow \infty} P^t \left\{ \lambda \in \mathcal{P} : \lambda(\sqrt{t}x_t) - \lceil \sqrt{t}x_t \rceil < a_t \right\} = \Phi(z), \tag{5.1}$$

where

$$a_t \equiv a_t(z) := 2\sqrt{t} \cos \theta(x_t) + \frac{z\sqrt{\log t}}{2\theta(x_t)}$$

and  $\Phi(\cdot)$  is the CDF of the standard normal law  $\mathcal{N}(0, 1)$ . Set  $I_t := [a_t, \infty)$  and let  $\#I_t$  be the number of points in the random set  $\mathcal{D}(\lambda) := \cup_{i=1}^{\infty} \{\lambda_i - i\}$  ( $\lambda \in \mathcal{P}$ ) that fall in the interval  $I_t$ . Then, recalling the definition (1.2) of the function  $\lambda(\cdot)$  and using that the sequence  $\{\lambda_i - i\}$  is strictly decreasing, it is easy to see that equation (5.1) is reduced to

$$\lim_{t \rightarrow \infty} P^t \left\{ \lambda \in \mathcal{P} : \#I_t < \lceil \sqrt{t}x \rceil \right\} = \Phi(z). \tag{5.2}$$

For  $k=1, 2, \dots$ , define the  $k$ -point correlation functions by

$$\rho_k^t(x_1, \dots, x_k) := P^t\{\lambda \in \mathcal{P} : x_1, \dots, x_k \in \mathcal{D}(\lambda)\} \quad (x_i \in \mathbb{Z}, x_i \neq x_j). \tag{5.3}$$

The key fact is that the functions  $\rho_k^t$  have a *determinantal structure* (Borodin *et al.* 2000; Johansson 2001),

$$\rho_k^t(x_1, \dots, x_k) = \det[\mathbf{J}(x_i, x_j; t)]_{1 \leq i, j \leq k}, \tag{5.4}$$

with the kernel  $\mathbf{J}$  of the form

$$\mathbf{J}(x, x'; t) = \begin{cases} \sqrt{t} \frac{J_x J_{x'+1} - J_{x+1} J_{x'}}{x - x'}, & x \neq x', \\ \sqrt{t}(J'_x J_{x+1} - J'_{x+1} J_x), & x = x', \end{cases} \tag{5.5}$$

where  $J_m = J_m(2\sqrt{t})$  is the Bessel function of integral order  $m$ . Then, by a central limit theorem due to Soshnikov (2000), generalizing an earlier result by Costin & Lebowitz (1995) (see further generalization set in a purely probabilistic framework in Ben Hough *et al.* 2006), the distribution of  $(\#I_t - E(\#I_t))/\sqrt{\text{Var}(\#I_t)}$  converges to  $\mathcal{N}(0, 1)$ , provided that  $\text{Var}(\#I_t) \rightarrow \infty$ . Thus, in order to derive the limit (5.2), it remains to obtain the asymptotics of the first two moments of  $\#I_t$  with respect to the measure  $P^t$ . The lemma 5.1 is the main technical (and the most difficult) part of the work.

**Lemma 5.1.** *As  $t \rightarrow \infty$ ,*

$$E(\#I_t) = \sqrt{t}x - \frac{z}{2\pi} \sqrt{\log t} + O(1),$$

$$\text{Var}(\#I_t) = \frac{\log t}{4\pi^2} (1 + o(1)).$$

For the proof of lemma 5.1, we first derive suitable expressions for the expectation and variance of  $\#I_t$ . Recalling the definition of the correlation functions (5.3) together with the determinantal property (5.4) and using an identity from Borodin *et al.* (2000, proposition 2.9), we obtain

$$\begin{aligned} E(\#I_t) &= \sum_{i \geq a_t} \rho_1^t(i) = \sum_{i \geq a_t} \mathbf{J}(i, i; t) = \sum_{i \geq a_t} \sum_{j \geq 1} J_{i+j}(2\sqrt{t})^2 \\ &= \sum_{m > [a_t]} (m - [a_t]) J_m(2\sqrt{t})^2. \end{aligned} \tag{5.6}$$

Similarly, for the variance we get (cf. Johansson 2001; Gustavsson 2005)

$$\begin{aligned} \text{Var}(\#I_t) &= \sum_{i \geq a_t} (\rho_1^t(i) - \rho_1^t(i)^2) + 2 \sum_{i > j \geq a_t} (\rho_2^t(i, j) - \rho_1^t(i)\rho_1^t(j)) \\ &= \sum_{i \geq a_t} \mathbf{J}(i, i; t) - \sum_{i, j \geq a_t} \mathbf{J}(i, j; t)^2 = \sum_{i \geq a_t, j > i} \mathbf{J}(i, j; t)^2, \end{aligned} \tag{5.7}$$

which again can be expressed in terms of the Bessel function  $J_m(2\sqrt{t})$  using equation (5.5). The rest of the proof is based on a direct asymptotic analysis of equations (5.6) and (5.7), heavily exploiting the asymptotics of the Bessel function in various regions of variation of its parameters. Calculations are straightforward but quite laborious, and we omit them here.

**Remark 5.1.** The proof of theorem 4.2 follows the similar ideas using a central limit theorem by Soshnikov (2002) for linear statistics of the form  $\sum_i \alpha_i \# I_i^i$ .

### 6. Link with Kerov’s result

Let us discuss the link between our results and the limit theorem by Kerov (1993) (see §2). In particular, our goal is to explain heuristically the mechanism of the effects that take place for the process  $\tilde{\Delta}_n(u)$ .

Note that if  $|u - u'| = n^{-s/2}$ , then  $s = -2 \log|u - u'| / \log n$ . That is to say, ‘time’  $s$  indexing the components of the limit Gaussian vector in theorem 3.7 has the meaning of the logarithmic distance between points  $u$  and  $u'$ , normalized by  $\log n$ . From this point of view, theorem 3.7 suggests that

$$\text{Cov}(\tilde{Y}_n(u), \tilde{Y}_n(u')) \approx s_n(u, u') := \min\left\{\frac{-2 \log|u - u'|}{\log n}, 1\right\}.$$

In fact, in the course of the proof of theorems 3.2 and 3.7, we obtain that for any  $\varepsilon > 0$ , the following estimates hold uniformly in  $u, u' \in (-2, 2)$ :

$$\text{Cov}(\tilde{\Delta}_n(u), \tilde{\Delta}_n(u')) = \begin{cases} O(1) \log|u - u'|, & |u - u'| \leq \varepsilon, \\ O(1), & |u - u'| \geq \varepsilon. \end{cases} \tag{6.1}$$

Now, setting  $\tilde{\Delta}_n[\varphi] := \int_{-2}^2 \tilde{\Delta}_n(u) \varphi(u) du$ , from equation (6.1) we have

$$\begin{aligned} \text{Var}(\tilde{\Delta}_n[\varphi]) &= O(1) \iint_{|u-u'| \leq \varepsilon} \varphi(u) \varphi(u') \log|u - u'| du du' \\ &\quad + O(1) \iint_{|u-u'| \geq \varepsilon} \varphi(u) \varphi(u') du du' = O(1), \end{aligned}$$

since the function  $\log|u - u'|$  is integrable at zero. Hence,  $\tilde{\Delta}_n[\varphi]$  is bounded in distribution as  $n \rightarrow \infty$ , which helps to understand why Kerov’s (1993) result holds without any normalization (see §2).

Conversely, the limiting process  $\tilde{\Delta}(u)$  defined in equation (2.2) can be used to render information contained in theorems 3.6 and 3.7 (at least heuristically). To this end, observe that since the number of parts in a typical partition is close to  $2\sqrt{n}$  (see remark 1.1), it is reasonable to think that the number of ‘degrees of freedom’ of a random partition  $\lambda \vdash n$  is of the order of  $m \asymp \sqrt{n}$ , and hence  $\tilde{\Delta}_n(u)$ ,  $u \in (-2, 2)$ , may be represented by a partial sum of the series (2.2)

$$S_m(2 \cos \theta) := \frac{2}{\pi} \sum_{k=2}^m \frac{X_k \sin(k\theta)}{\sqrt{k}}, \quad \theta \in (0, \pi).$$

Then, for  $u=2 \cos \theta$  and  $u'=2 \cos \theta'$ , we get, using some trigonometry, that

$$\begin{aligned} \text{Cov}(S_m(u), S_m(u')) &= \frac{4}{\pi^2} \sum_{k=2}^m \frac{\sin(k\theta) \sin(k\theta')}{k} \\ &= \frac{2}{\pi^2} \sum_{k=2}^m \frac{\cos(k(\theta - \theta'))}{k} - \frac{2}{\pi^2} \sum_{k=2}^m \frac{\cos(k(\theta + \theta'))}{k}. \end{aligned} \tag{6.2}$$

As  $m \rightarrow \infty$ , the second sum on the right-hand side of equation (6.2) converges for all  $\theta, \theta' \in (0, \pi)$ . For  $\theta = \theta'$ , from equation (6.2) we get

$$\text{Var}[S_m(u)] \sim \frac{2}{\pi^2} \sum_{k=2}^m \frac{1}{k} \sim \frac{2 \log m}{\pi^2} \sim \frac{\log n}{\pi^2},$$

and it follows that the distribution of  $\pi S_m(u)/\sqrt{\log n}$  converges to  $\mathcal{N}(0, 1)$ , in agreement with theorem 3.6. Moreover, if  $u - u' \asymp n^{-s/2}$  (hence  $\theta - \theta' \asymp n^{-s/2}$ ), then the first sum on the right-hand side of equation (6.2) is approximated by the integral

$$\int_2^m \frac{\cos(x(\theta - \theta'))}{x} dx \sim \int_{2|\theta - \theta'|}^\epsilon \frac{dy}{y} \sim -\log|\theta - \theta'| \sim \frac{s}{2} \log n.$$

Inserting this into equation (6.2), we obtain

$$\text{Cov}(S_m(u), S_m(u')) \sim \frac{s \log n}{\pi^2} \quad (n \rightarrow \infty),$$

as predicted by theorem 3.7.

### 7. Concluding remarks

The limiting random process  $Z^{(s)}(t)$  in corollary 3.4 can be represented (in the FDD sense) as

$$Z^{(s)}(t) \stackrel{d}{=} \begin{cases} \zeta_0, & t = 0, \\ \sqrt{s} \zeta_0 + \sqrt{1-s} \zeta_t, & t \neq 0, \end{cases}$$

where  $\{\zeta_t\}$  are mutually independent random variables with normal distribution  $\mathcal{N}(0, 1)$ . A similar representation for the process  $Z(s)$  of corollary 3.5 reads

$$Z(s) \stackrel{d}{=} B(s) + \sqrt{1-s} \zeta_s, \quad 0 \leq s \leq 1, \tag{7.1}$$

where  $B(s)$  is a standard Brownian motion (i.e. with covariance function given by  $\min\{s, s'\}$  and  $B(0)=0$ ), independent of the family  $\{\zeta_s\}$ . These decompositions show that the random processes  $Z^{(s)}(t)$  and  $Z(s)$  are highly irregular (e.g. stochastically discontinuous unless  $s=1$ ), which is a manifestation of asymptotically fast oscillations of the process  $\mathcal{A}_n(x)$  (as well as  $\tilde{\mathcal{A}}_n(u)$ ) in the vicinity of each point  $0 < x < 2$  (respectively,  $-2 < u < 2$ ).

Furthermore, decomposition (7.1) implies that the reverse-time increment process  $Z^*(s) := Z(1-s) - Z(1)$  is represented as

$$Z^*(s) \stackrel{d}{=} B(s) + \sqrt{s} \zeta_s, \quad 0 \leq s \leq 1,$$

and it follows that  $Z^*(s)$  is a *self-similar process* with the Hurst parameter  $H=1/2$  (see Embrechts & Maejima 2002),

$$Z^*(\alpha s) \stackrel{d}{=} \sqrt{\alpha} Z^*(s), \quad 0 \leq s \leq 1 \quad (0 \leq \alpha \leq 1).$$

Finally, let us briefly comment on the nature of convergence we have established. Let  $\tilde{Y}(u)$ ,  $u \in [-2, 2]$ , be a random process with independent values, such that  $\tilde{Y}(u)$  has a standard normal distribution for each  $u \in (-2, 2)$ , whereas  $\tilde{Y}(\pm 2) = 0$  (with probability 1). Theorem 3.7 (cf. corollary 3.3) and equation (3.4) imply that the random process  $\tilde{Y}_n(u)$  converges to  $\tilde{Y}(u)$  in the FDD sense, and it is of interest to ask whether this can be extended to weak convergence (cf. Logan & Shepp 1977, p. 211). However, the answer appears to be negative, at least under the natural choice of the space of continuous functions  $C[-2, 2]$ , because the necessary condition of tightness (Billingsley 1999) breaks down. Indeed, for any  $\delta > 0$ ,  $\varepsilon > 0$  and all  $u, u' \in (-2, 2)$  such that  $|u - u'| \leq \delta$ , we have

$$\lim_{n \rightarrow \infty} P_n\{|\tilde{Y}_n(u) - \tilde{Y}_n(u')| \geq \varepsilon\} = P\{|\tilde{Y}(u) - \tilde{Y}(u')| \geq \varepsilon\} = 2\left(1 - \Phi(\varepsilon/\sqrt{2})\right) > 0.$$

The analogous remark applies to the process  $Y_n(x)$ ,  $x \in [0, 2]$ , considered in the space  $D[0, 2]$  of *cadlag* functions (i.e. right continuous with left limits).

This work was done when Z. Su was visiting the University of Leeds under a Royal Society K. C. Wong Education Foundation International Incoming Fellowship. His research was also supported by the National Science Funds of China (grant no. 10 371109). The authors are grateful to Sir J. F. C. Kingman, A. Okounkov and Ya. G. Sinai for their interest in this work, and to A. V. Gnedin, J. T. Kent and Yu. V. Yakubovich for useful discussions. The authors also thank the unknown referees for their careful reading of the manuscript and helpful remarks.

## References

- Andrews, G. E. 1976 *The theory of partitions*. Reading, MA: Addison-Wesley.
- Arratia, R. & Tavaré, S. 1994 Independent process approximations for random combinatorial structures. *Adv. Math.* **104**, 90–154. (doi:10.1006/aima.1994.1022)
- Arratia, R., Barbour, A. D. & Tavaré, S. 2003 *Logarithmic combinatorial structures: a probabilistic approach*. Zürich, Switzerland: European Mathematical Society Publishers.
- Baik, J., Deift, P. & Johansson, K. 1999 On the distribution of the length of the longest increasing subsequence of random permutations. *J. Am. Math. Soc.* **12**, 1119–1178. (doi:10.1090/S0894-0347-99-00307-0)
- Ben Hough, J., Krishnapur, M., Peres, Y. & Virág, B. 2006 Determinantal processes and independence. *Prob. Surv.* **3**, 206–229. (doi:10.1214/154957806000000078)
- Billingsley, P. 1999 *Convergence of probability measures*, 2nd edn. New York, NY: Wiley.
- Borodin, A., Okounkov, A. & Olshanski, G. 2000 Asymptotics of Plancherel measures for symmetric groups. *J. Am. Math. Soc.* **13**, 481–515. (doi:10.1090/S0894-0347-00-00337-4)
- Costin, O. & Lebowitz, J. L. 1995 Gaussian fluctuation in random matrices. *Phys. Rev. Lett.* **75**, 69–72. (doi:10.1103/PhysRevLett.75.69)
- Deift, P. 2000 Integrable systems and combinatorial theory. *Notices Am. Math. Soc.* **47**, 631–640.

- Embrechts, P. & Maejima, M. 2002 *Selfsimilar processes*. Princeton, NJ: Princeton University Press.
- Fristedt, B. 1993 The structure of random partitions of large integers. *Trans. Am. Math. Soc.* **337**, 703–735. (doi:10.2307/2154239)
- Fulton, W. 1997 *Young diagrams: with applications to representation theory and geometry*. Cambridge, UK: Cambridge University Press.
- Gustavsson, J. 2005 Gaussian fluctuations of eigenvalues in the GUE. *Ann. Inst. H. Poincaré Prob. Stat.* **41**, 151–178. (doi:10.1016/j.anihpb.2004.04.002)
- Ivanov, V. & Olshanski, G. 2002 Kerov's central limit theorem for the Plancherel measure on Young diagrams. In *Symmetric functions 2001: surveys of developments and perspectives* (ed. S. Fomin), pp. 93–151. Dordrecht: Kluwer Academic Publishers. (<http://arxiv.org/abs/math.CO/0304010>)
- Johansson, K. 2001 Discrete orthogonal polynomial ensembles and the Plancherel measure. *Ann. Math.* **153**, 259–296. (<http://projecteuclid.org/Dienst/UI/1.0/Summarize/euclid.anm/1026916789>)
- Kerov, S. 1993 Gaussian limit for the Plancherel measure of the symmetric group. *C.R. Acad. Sci. Paris, Sér. I Math.* **316**, 303–308. (<http://www.pdmi.ras.ru/kerov/textfiles/gausslim.ps.gz>)
- Logan, B. F. & Shepp, L. A. 1977 A variational problem for random Young tableaux. *Adv. Math.* **26**, 206–222. (doi:10.1016/0001-8708(77)90030-5)
- Okounkov, A. 2000 Random matrices and random permutations. *Int. Math. Res. Notices* **2000**, 1043–1095. (doi:10.1155/S1073792800000532)
- Soshnikov, A. B. 2000 Gaussian fluctuation for the number of particles in Airy, Bessel, sine, and other determinantal random point fields. *J. Stat. Phys.* **100**, 491–522. (doi:10.1023/A:1018672622921)
- Soshnikov, A. B. 2002 Gaussian limit for determinantal random point fields. *Ann. Probab.* **30**, 171–187. (doi:10.1214/aop/1020107764)
- Tracy, C. A. & Widom, H. 1994 Level-spacing distributions and the Airy kernel. *Comm. Math. Phys.* **159**, 151–174. (doi:10.1007/BF02100489)
- Vershik, A. M. 1995 Asymptotic combinatorics and algebraic analysis. In *Proc. International Congress of Mathematicians, Zürich, 3–11 August 1994*, vol. 2 (ed. S. D. Chatterji), pp.1384–1394. Basel, Switzerland: Birkhäuser.
- Vershik, A. M. 1996 Statistical mechanics of combinatorial partitions, and their limit configurations. *Funkts. Anal. i ego Prilozh.* **30** (2), 19–39. (Transl. *Funct. Anal. Appl.* **30**, 90–105.) (doi:10.1007/BF02509449)
- Vershik, A. M. & Kerov, S. V. 1977 Asymptotics of the Plancherel measure of the symmetric group and the limiting form of Young tables. *Dokl. Akad. Nauk SSSR* **233**, 1024–1027. (Transl. *Soviet Math. Doklady* **18**, 527–531.)