ISSN 1064–5624, Doklady Mathematics, 2007, Vol. 75, No. 3, pp. 381–384. © Pleiades Publishing, Ltd., 2007.
Published in Russian in Doklady Akademii Nauk, 2007, Vol. 414, No. 3, pp. 295–298.

MATHEMATICS

Central Limit Theorem for Random Partitions under the Plancherel Measure[1]

L. V. Bogachev[a] **and Z. G. Su**[b]

Presented by Academician Ya.G. Sinai August 9, 2006

Received January 9, 2007

DOI: 10.1134/S1064562407030143

In this paper, we obtain a central limit theorem for the Plancherel measure on the ensemble of partitions of asymptotically growing integers. It is proved that local fluctuations of the corresponding Young diagrams are asymptotically normal both in the bulk and near the edges of the limiting "spectrum" of partitions. The results of this work answer a question of Logan and Shepp (1977) and significantly complement Kerov's theorem (1993) on the convergence of integral fluctuations to a generalized Gaussian process.

1. INTRODUCTION

A **partition** of a positive integer n is any integer sequence $\lambda = \{\lambda_1, \lambda_2, \dots\}$ such that $\lambda_1 \geq \lambda_2 \geq \dots \geq 0$ and $\lambda_1 + \lambda_2 + \dots = n$ (notation: $\lambda \vdash n$). Every partition $\lambda \vdash n$ can be represented geometrically by the so-called *Young diagram* consisting of n unit squares (cells) in consecutive columns containing $\lambda_1, \lambda_2, \dots$ cells, respectively.

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

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

where d_λ is the number of *standard tableaux* of a given shape λ, i.e., all possible arrangements of the numbers $1, \dots, n$ in the cells of the Young diagram λ such that the numbers increase in each row (from left to right) and in each column (from bottom to top).

Note that P_n is a probability measure due to the Burnside identity $\sum_{\lambda \in \mathcal{P}_n} d_\lambda^2 = n!$ [5, 6]. The Plancherel measure is related in a natural way to the representation theory of the symmetric group [6] but also arises in some combinatorial and probabilistic problems (see [5]). For example, the distribution of the largest term $\lambda_1 = \max\{\lambda_i \in \lambda \vdash n\}$ under the measure P_n coincides with the distribution of the length of the longest increasing subsequence contained in a random (uniformly distributed) permutation of order n (see [3, 5]).

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

$$\lambda(x) := \begin{cases} \lambda_1, & x = 0 \\ \lambda_{\lceil x \rceil}, & x > 0, \end{cases} \tag{1}$$

where $\lceil x \rceil := \min\{m \in \mathbb{Z}: m \geq x\}$ is the ceiling integer part of x. Logan and Shepp [10] and, independently, Vershik and Kerov [1] have discovered that, as $n \to \infty$, a typical Young diagram, suitably scaled, has a limit shape determined by some function $y = \omega(x)$. This means that, for the overwhelming majority of partitions $\lambda \in \mathcal{P}_n$ (with respect to the Plancherel measure P_n), the boundary of their scaled Young diagrams is contained in an arbitrarily small vicinity of the graph $y = \omega(x)$. More specifically, consider the function $y = \omega(x)$, $x \geq 0$, defined for $x \in [0, 2]$ by the parametric equations

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

---

[1] The article was translated by the authors.

---

[a] *Department of Statistics, University of Leeds, Leeds LS2 9JT, United Kingdom*
e-mail: bogachev@maths.leeds.ac.uk
[b] *Department of Mathematics, Zhejiang University, Hangzhou, P.R. China*
e-mail: suzhonggen@zju.edu.cn

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

$$\Delta_n(x) := \lambda(\sqrt{n}x) - \sqrt{n}\omega(x), \quad x \geq 0 \tag{3}$$

satisfies the following law of large numbers [1]:

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

In particular, for $x = 0$, this implies a law of large numbers for the maximal term λ_1:

$$\forall \varepsilon > 0 \ \lim_{n \to \infty} P_n\left\{\left|\frac{\lambda_1}{\sqrt{n}} - 2\right| > \varepsilon\right\} = 0.$$

Remark 1. Due to the invariance of the Plancherel measure under the transposition of Young diagrams $\lambda \leftrightarrow \lambda'$ (when the columns of λ become rows of the transposed diagram λ' and vice versa), the same law of large numbers holds for λ'_1, i.e., for the number of terms in a partition λ.

2. FLUCTUATIONS OF YOUNG DIAGRAMS

A natural question about the limit distribution of $\Delta_n(x)$ was posed by Logan and Shepp [10]. Kerov [9] gave a partial answer by establishing the asymptotic normality of integral fluctuations with respect to a suitable class of test functions (i.e., in the sense of generalized convergence). More precisely, in the coordinates $u = x - y$, $v = x + y$, the boundary of the Young diagram is represented by a piecewise linear (continuous) function $\tilde{\lambda}(u)$ and the limit shape takes the form (see [1])

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

Then, according to [9], the random process

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

converges in distribution (without any further normalization!) to a generalized Gaussian process $\tilde{\Delta}(u)$, $u \in [-2, 2]$ defined by the formal random series

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

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

However, a localized version of the central limit theorem (i.e., for fluctuations at a given point) has not been known as yet. On the one hand, the existence of such a theorem would seem quite natural, at least in the bulk of the partition "spectrum,"[2] i.e., for $\lambda_i \in \lambda \vdash n$ such that $\frac{i}{\sqrt{n}} \sim x \in (0, 2)$. On the other hand, Kerov's result on the generalized convergence cast some doubt on the validity of the usual convergence. The main aim of the present work is to obtain such a theorem (see Section 3).

Note that the asymptotic behavior of fluctuations at the upper edge of the limiting spectrum (corresponding to $x = 0$) is different from a Gaussian distribution. As was shown in [3] for λ_1 and in [4, 8, 11] for any λ_k with fixed index $k \in \mathbb{N}$,

$$\lim_{n \to \infty} P_n\left\{\frac{\lambda_k - 2\sqrt{n}}{n^{1/6}} \leq z\right\} = F_k(z), \quad z \in \mathbb{R}, \tag{4}$$

where $F_k(\cdot)$ is the distribution function of the kth order statistic in the Airy ensemble, which was discovered earlier in connection with the limit distribution of the largest eigenvalues for random matrices from the Gaussian unitary ensemble (GUE) (see [13]). In particular, the function $F_1(\cdot)$ determines the Tracy–Widom distribution.

From the point of view of Kerov's limit theorem, the extreme values $\lambda_1, \lambda_2, \ldots$ might present a danger, since, according to formula (4), fluctuations of the process $\Delta_n(x)$ in the zone of size $O(n^{-1/2})$ near the edge $x = 0$ are quite large (on the order of $n^{1/6}$). In fact, this theorem shows that the edge of the spectrum does not make any significant contribution to the integral fluctuations. Let us stress, however, that the situation in the bulk of the spectrum (i.e., for $0 < x < 2$) remained unclear.

3. MAIN RESULTS

Note that the value of θ in Eqs. (2) that corresponds to the coordinates x and $y = \omega(x)$ is given by

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

Recall that $\Delta_n(x)$ is defined by formula (3). The following theorem is the main result of this paper.

Theorem 1. *Suppose that* $x_n \in (0, 2)$ *and*

$$\lim_{n \to \infty} n\sin^6\theta(x_n) = \infty. \tag{5}$$

Then the distribution of the random variable $\Delta_n(x_n)$ *with respect to the Plancherel measure* P_n *is asymptotically normal, namely,*

$$\frac{2\theta(x_n)\Delta_n(x_n)}{\sqrt{\ln(n\sin^6\theta(x_n))}} \xrightarrow{d} \mathcal{N}(0, 1) \quad (n \to \infty). \tag{6}$$

---

[2] We use the term "spectrum" informally to refer to the variety of partition's terms $\{\lambda_i \in \lambda\}$ (cf. [2], where this term is used in a general context of combinatorial structures characterized by their components).

Theorem 1 embraces several particular cases corresponding to the location of the points x_n in the bulk of the spectrum or near its edges.

Corollary 1. *Let* $x_n \to x \in (0, 2)$ *as* $n \to \infty$. *Then*

$$\frac{2\theta(x)\Delta_n(x_n)}{\sqrt{\ln n}} \xrightarrow{d} \mathcal{N}(0, 1). \tag{7}$$

If $x_n \to 0$ *and* $nx_n^2 \to \infty$ *as* $n \to \infty$, *then*

$$\frac{(12\pi x_n)^{1/3}\Delta_n(x_n)}{\sqrt{\ln(nx_n^2)}} \xrightarrow{d} \mathcal{N}(0, 1). \tag{8}$$

Finally, if $x_n \to 2$ *and* $n(2 - x_n)^3 \to \infty$ *as* $n \to \infty$, *then*

$$\frac{2\pi\Delta_n(x_n)}{\sqrt{\ln(n(2-x_n)^3)}} \xrightarrow{d} \mathcal{N}(0, 1). \tag{9}$$

Indeed, if $x_n \to x \in (0, 2)$, then $\theta(x_n) \to \theta(x) \in (0, \pi)$ and condition (5) is automatically satisfied. Furthermore, Eqs. (2) imply that $\theta(x_n) \sim \left(\frac{3\pi x_n}{2}\right)^{1/3}$ as $x_n \to 0$; therefore, relation (6) is reduced to (8). Similarly, for $x_n \to 2$, due to (2), we have $\pi - \theta(x_n) \sim (2 - x_n)^{1/2}$, and (9) follows from (6).

Remark 2. Results similar to Corollary 1 were obtained in [7] for eigenvalues of random matrices in the GUE.

4. POISSONIZATION

The proof of Theorem 1 is based on a standard poissonization technique (see, e.g., [3]). Let $\mathcal{P} := \bigcup_{n=0}^{\infty}\mathcal{P}_n$ be the set of partitions of all positive integers (formally, $\mathcal{P}_0$ contains only the "empty" partition of zero). For any $\lambda \in \mathcal{P}$, we set $|\lambda| := \sum_{\lambda_i \in \lambda}\lambda_i$ and define the measure P^t $(t > 0)$ as

$$P^t(\lambda) := e^{-t}t^{|\lambda|}\left(\frac{d_\lambda}{|\lambda|!}\right)^2, \quad \lambda \in \mathcal{P}. \tag{10}$$

Formula (10) defines a probability measure on $\mathcal{P}$, 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 prove the poissonized version of Theorem 1 obtained by replacing the measure P_n by P^t and the parameter n by t.

Theorem 2. *Suppose that* $x_t \in (0, 2)$ *and* $t\sin^6\theta(x_t) \to \infty$ *as* $t \to \infty$. *Then, with respect to the measure* P^t,

$$\frac{2\theta(x_t)\Delta_t(x_t)}{\sqrt{\ln(t\sin^6\theta(x_t))}} \xrightarrow{d} \mathcal{N}(0, 1) \quad (t \to \infty).$$

Theorem 1 can be derived from Theorem 2 by using depoissonization (see, e.g., [3]). According to (10), P^t can be viewed as the expectation of the random measure P_N, where N is a Poisson random variable with parameter t:

$$P^t(A) = \mathrm{E}(P_N(A)) = e^{-t}\sum_{k=0}^{\infty}\frac{t^k}{k!}P_k(A). \tag{11}$$

Since N has the mean t and the standard deviation $\sqrt{t}$, Eq. (11) suggests that the asymptotics of the probability $P_n(A)$ as $n \to \infty$ can be recovered from that of $P^t(A)$ as $t \sim n \to \infty$. More precisely, one can prove that $P_n(A) \sim P^t(A)$ as $t \sim n \to \infty$, provided that the 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 in [3].

5. SKETCH OF THE PROOF OF THEOREM 2

In view of (2), the statement of Theorem 2 is equivalent to saying that for any $z \in \mathbb{R}$,

$$\lim_{t\to\infty} P^t\{\lambda \in \mathcal{P}: \lambda(\sqrt{t}x_t) - \lceil\sqrt{t}x_t\rceil < a_z(t)\} = \Phi(z), \tag{12}$$

where

$$a_z(t) := 2\sqrt{t}\cos\theta(x_t) + \frac{z}{2\theta(x_t)}\sqrt{\ln(t\sin^6\theta(x_t))}$$

and $\Phi(\cdot)$ is the distribution function of the standard normal law $\mathcal{N}(0, 1)$. Let $\#I_z(t)$ be the number of points from the random set $\mathcal{D}(\lambda) := \bigcup_{i=1}^{\infty}(\lambda_i - i)$ $(\lambda \in \mathcal{P})$ contained in the interval $I_z(t) = [a_z(t), \infty)$. Recalling definition (1) of the function $\lambda(\cdot)$ and taking into account that the sequence $\{\lambda_i - i\}$ is strictly decreasing, we can rewrite relation (12) as

$$\lim_{t\to\infty} P^t\{\lambda \in \mathcal{P}: \#I_z(t) < \lceil\sqrt{t}x_t\rceil\} = \Phi(z). \tag{13}$$

The key fact is that the correlation functions of the random point process $\{\lambda_i - i\}$ defined by

$$\rho_k^t(x_1, x_2, \dots, x_k) := P^t\{\lambda \in \mathcal{P}: x_1, x_2, \dots, x_k \in \mathcal{D}(\lambda)\},$$

$$x_i \in \mathbb{Z}, \quad x_i \neq x_j$$

have a determinantal structure [4, 8]:

$$\rho_k^t(x_1, x_2, \dots, x_k) = \det[J(x_i, x_j; t)]_{1 \le i, j \le k}, \quad k = 1, 2, \dots$$

with a kernel J of the form

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

where $J_m = J_m(2\sqrt{t})$ is the Bessel function of integral order m. Then, by Soshnikov's theorem [12], the random variable $\#I_z(t)$ satisfies the central limit theorem:

$$\frac{\#I_z(t) - \mathrm{E}(\#I_z(t))}{\sqrt{\mathrm{Var}(\#I_z(t))}} \xrightarrow{d} \mathcal{N}(0, 1), \quad t \to \infty,$$

provided that $\mathrm{Var}(\#I_z(t)) \to \infty$. Therefore, in order to derive (13) from this relation, it remains to obtain the asymptotics of the first two moments of the random variable $\#I_z(t)$ under the measure P^t. The following lemma is the main technical (and most difficult) part of this work.

Lemma 1. *As* $t \to \infty$,

$$\mathrm{E}(\#I_z(t)) = \sqrt{t}x_t - \frac{z}{2\pi}\sqrt{\ln(t\sin^6\theta(x_t))} + O(1),$$

$$\mathrm{Var}(\#I_z(t)) = \frac{\ln(t\sin^6\theta(x_t))}{4\pi^2}(1 + o(1)).$$

The proof of the lemma is based on a direct asymptotic analysis of the expressions for $\mathrm{E}(\#I_z(t))$ and $\mathrm{Var}(\#I_z(t))$. The calculations are quite laborious and rely heavily on the asymptotics of the Bessel function $J_m(2\sqrt{t})$ in various regions of variation of the parameters.

6. CONCLUDING REMARKS

Condition (5) in Theorem 1 (see also Corollary 1) means that the points x_n must not approach the edges $x = 0$ and $x = 2$ of the limit spectrum too closely. For instance, (5) is violated if $x_n \sim \frac{c}{\sqrt{n}}$ $(c > 0)$, which corresponds to $\lambda(\sqrt{n}x_n) = \lambda_{\lceil \sqrt{n}x_n \rceil}$ getting to the zone of the extreme values $\lambda_1 \ge \lambda_2 \ge \dots$. In this case, the normalizing coefficient in (8) is on the order of $n^{-1/6}$, which coincides with the normalization in the limit law (4). Moreover, if $x_n = \frac{k}{\sqrt{n}}$ $(k = 1, 2, \dots)$, then

$$\frac{\Delta_n(x_n)}{n^{1/6}} = \frac{\lambda_k - 2\sqrt{n}}{n^{1/6}} + \left(\frac{3\pi k}{2}\right)^{2/3} + o(1), \quad n \to \infty. \tag{14}$$

Therefore, the domain of asymptotically Gaussian fluctuations, which is described by Theorem 1 and Corollary 1, extends up to the domain of extreme values where the limit distribution is characterized by the Airy ensemble (see (4)).

Conversely, formally sending the parameter k to infinity (i.e., moving away from the edge $x = 0$ inwards the spectrum), in view of (8), it is natural to expect that the limit distribution of random variable (14), which is expressed in terms of the distribution function F_k (see (4)), will converge to a Gaussian law.

Conjecture. *Let a random variable* Υ_k *have the distribution corresponding to the kth order statistic of the Airy ensemble* (*see* (4)). *Then, as* $k \to \infty$,

$$\frac{(12\pi k)^{1/3}}{\sqrt{2\ln k}}\left(\Upsilon_k + \left(\frac{3\pi k}{2}\right)^{2/3}\right) \xrightarrow{d} \mathcal{N}(0, 1). \tag{15}$$

Note that, for the opposite edge of the spectrum, $x = 2$, similar arguments (using the invariance of the Plancherel measure under the transposition of Young diagrams, see Remark 1) lead to the same limit relation (15).

To the best of our knowledge, this fact has not been mentioned in the literature, and we intend to study this issue in another paper.

ACKNOWLEDGMENTS

This work was done when Z.G. Su was visiting the University of Leeds under a grant from the Royal Society. His research was also supported by the National Science Funds of China, Grant No. 10371109.

REFERENCES

1. A. M. Vershik and S. V. Kerov, Sov. Math. Dokl. **18**, 527–531 (1977) [Dokl. Akad. Nauk **233**, 1024–1027 (1977)].
2. R. Arratia, A. D. Barbour, and S. Tavaré, *Logarithmic Combinatorial Structures: A Probabilistic Approach* (Eur. Math. Soc. Publ., Zurich, 2003).
3. J. Baik, P. Deift, and K. Johansson, J. Am. Math. Soc. **12**, 1119–1178 (1999).
4. A. Borodin, A. Okounkov, and G. Olshanski, J. Am. Math. Soc. **13**, 481–515 (2000).
5. P. Deift, Notices Am. Math. Soc. **47**, 631–640 (2000).
6. W. Fulton, *Young Diagrams* (Cambridge Univ. Press, Cambridge, 1997).
7. J. Gustavsson, Ann. Inst. H. Poincaré B **41** (2), 151–178 (2005).
8. K. Johansson, Ann. Math. **153** (1), 259–296 (2001).
9. S. Kerov, C. R. Acad. Sci. Ser. I Math. **316** (4), 303–308 (1993).
10. B. F. Logan and L. A. Shepp, Adv. Math. **26** (2), 206–222 (1977).
11. A. Okounkov, Int. Math. Res. Notices **2000**, 1043–1095 (2000).
12. A. B. Soshnikov, J. Stat. Phys. **100** (3/4), 491–522 (2000).
13. C. A. Tracy and H. Widom, Commun. Math. Phys. **159** (1), 151–174 (1994).