# Functionals of random partitions and the generalised Erdős-Turán laws for permutations

Alexander Gnedin (QMUL, London)
Alexander Iksanov (Kiev University)
Alexander Marynych (Kiev University)

# Order of permutation

$$\sigma = (1\ 9\ 6\ 2)(3\ 7\ 5)(4\ 8), \qquad \sigma^{12} = id$$

$$\text{l.c.m.}(4, 3, 2) = 12$$

- For permutation of $[n] := \{1, 2, \ldots, n\}$
  $K_{n,r} := \#$ cycles of length $r$, $(K_{n,r}; r \in [n])$ cycle partition
- $O_n := \text{l.c.m.}\{r : K_{n,r} > 0\}$.

# Erdős-Turán laws

- Erdős-Turán (1967):
  *For uniformly random permutation of* $[n]$

$$\frac{\log O_n - \frac{1}{2}\log^2 n}{\sqrt{\frac{1}{3}\log^3 n}} \xrightarrow{d} \mathcal{N}(0,1)$$

# Ewens' permutations

Ewens' distribution on permutations of $[n]$

$$P(\sigma) = \frac{\theta^{K_n}}{\theta(\theta+1)\dots(\theta+n-1)}, \quad \theta > 0$$

$$K_n := \sum_r K_{n,r} \quad \text{\# of cycles}$$

The distribution of $(K_{n,r}; r \in [n])$ is the Ewens sampling formula.

- Arratia and Tavaré 1992: *For Ewens' permutation of* $[n]$

$$\frac{\log O_n - \frac{\theta}{2}\log^2 n}{\sqrt{\frac{\theta}{3}\log^3 n}} \xrightarrow{d} \mathcal{N}(0,1)$$

- $\log O_n$ approximable by $\log T_n = \sum_r \log r \; K_{n,r}$
- $K_{n,r}$'s asymptotically independent, Poisson$(\theta/r)$

# Poisson-Dirichlet/GEM connection

$W \overset{d}{=} \mathrm{Beta}(\theta, 1)$, $\mathbb{P}(W \in dx) = \theta x^{\theta-1} dx$, $x \in (0,1)$
$W_1, W_2, \ldots$ i.i.d. copies of $W$

- PD/GEM random discrete distribution
  $P_j = W_1 \cdots W_{j-1}(1 - W_j)$, $j \in \mathbb{N}$
- For sample of size $n$ from $(P_j)$,
  $K_{n,r}$ is the number of values $j \in \mathbb{N}$ represented $r$ times
  (so $\sum_r r K_{n,r} = n$)
- From random partition to permutation: conditionally on the cycle partition $(K_{n,r}; r \in [n])$ the permutation is uniformly distributed.
- LLN: $P_j$'s are asymptotic frequencies of 'big' components of the partition

# General stick-breaking factor $W$

- $P_j = W_1 \cdots W_{j-1}(1 - W_j), \quad j \in \mathbb{N}$, with i.i.d. $W_j \stackrel{d}{=} W$, where $W$ is a 'stick-breaking factor' with general distribution on $[0, 1]$
- generate partition/permutation of $[n]$ by sampling $n$ elements from $(P_j)$ and letting $K_{n,r}$ to be the number of integer values represented $r$ times in the sample.
- Problem: What is the limit distribution of

$$\frac{\log O_n - b_n}{a_n}$$

for suitable centering/scaling constants $b_n, a_n$?

Permutations with distribution of the Gibbs form

$$p(\lambda_1, \dots, \lambda_k) = c_{n,k} \prod_{i=1}^{k} \theta_{\lambda_i}$$

(Betz/Ueltschi/Velenik, Nikeghbali/Zeindler, ...) are not permutations derived by the stick-breaking, unless they belong to Ewens's family.

- Regenerative property: the collection of cycle-sizes coincides with the set of jumps of a decreasing Markov chain with transition matrix

$$q(n, m) = \binom{n}{m} \frac{\mathbb{E}[W^{n-m}(1-W)^m]}{1 - \mathbb{E}W^n}, \quad 0 \le m \le n-1.$$

  starting state $n$ and absorbing state 0.

Example: for Ewens' permutations

$$q(n, m) = \binom{n}{m} \frac{(\theta)_m \, (n-m)!}{n \, (\theta+1)_{n-1}}.$$

For Ewens' permutations, general *separable* (additive) functionals

$$\sum_r h(r) K_{n,r}$$

have been studied by Babu and Manstavicius (2002, 2009) for unbounded functions $h$ (we need $h(r) = \log r$).

For the permutations derived from stick-breaking:

- $K_{n,r}$'s are not asymptotically independent,
- $K_{n,r}$'s converge (if $\mathbb{E}|\log W| < \infty$) to some multivariate discrete distribution, which is intractable (G.,Iksanov and Roesler 2008)

- Density assumption $\mathbb{P}(W \in dx) = f(x)dx, \quad x \in (0,1)$,
- Define
  $\mu := \mathbb{E}|\log W|, \ \sigma^2 := \mathrm{Var}(\log W), \nu := \mathbb{E}|\log(1-W)|$;
  we shall assume $\mu < \infty$, $\sigma^2 \leq \infty$, $\nu \leq \infty$.

# Normal limit I

*Suppose*

$$(\mathbf{I}): \quad \sup_{x\in[0,1]} x^{\beta}(1-x)^{\beta} f(x) < \infty \text{ for some } \beta \in [0,1).$$

*Then*

$$\frac{\log O_n - b_n}{a_n} \xrightarrow{d} \mathcal{N}(0,1),$$

*with constants*

$$b_n = \frac{\log^2 n}{2\mu}$$

$$a_n = \sqrt{\frac{\sigma^2 \log^3 n}{3\mu^3}}$$

Example: $f = \text{Beta}(\theta, \zeta)$; $\theta, \zeta > 0$.

# Normal limit IIa

**(II)** *: Suppose (for some small $\delta$) $f$ is nonincreasing in $[0, \delta]$, nondecreasing in $[1-\delta, 1]$ and bounded on $[\delta, 1-\delta]$.*

*If $\sigma^2 < \infty$ then*

$$\frac{\log O_n - b_n}{a_n} \xrightarrow{d} \mathcal{N}(0,1),$$

*for*

$$b_n = \frac{1}{\mu}\left[\frac{\log^2 n}{2} - \int_0^{\log^2 n}\int_0^z \mathbb{P}(\log|1-W| > x)dxdz\right]$$

$$a_n = \sqrt{\frac{\sigma^2 \log^3 n}{3\mu^3}}.$$

# Normal limit IIb

*If* $\sigma^2 = \infty$ *and*

$$\int_0^x y^2 \mathbb{P}(|\log W| \in dy) \sim \ell(x)$$

*for function* $\ell$ *of slow variation at* $\infty$*, then the normal limit holds with*

$$a_n = \sqrt{\frac{c_{\lfloor \log n \rfloor} \log n}{3\mu^3}},$$

*where* $c_n$ *is any sequence satisfying*

$$\frac{n\ell(c_n)}{c_n^2} \to 1.$$

# Stable limit IIc

*If for some* $\alpha \in (1, 2)$ *and* $\ell$ *of slow variation at* $\infty$

$$\mathbb{P}(|\log W| > x) \sim x^{-\alpha}\ell(x),$$

*then the limit is* $\alpha$*-stable with characteristic function*

$$u \mapsto \exp\left[-|u|^{\alpha}\Gamma(1-\alpha)\left(\cos\frac{\pi\alpha}{2} + i\sin\frac{\pi\alpha}{2}\right)\mathrm{sgn}u\right].$$

*The centering* $b_n$ *is as in* $\mathrm{IIa}$ *and scaling*

$$a_n = \frac{c_{\lfloor \log n \rfloor}\log n}{((\alpha+1)\mu^{\alpha+1})^{1/\alpha}}$$

# Reduction to $T_n$

For $T_n = \prod_{r=1}^{n} r^{K_{n,r}}$

$$\mathbb{E}|\log O_n - \log T_n| = O(\log n \log\log n),$$

under any of the assumptions I, IIa, IIb, IIc.

# Perturbed random walk

$\xi > 0, \eta \geq 0$ any dependent random variables,
$(\xi_j, \eta_j)$ i.i.d. copies of $(\xi, \eta)$

$$S_k = \xi_1 + \cdots + \xi_k$$

- Perturbed random walk $\widetilde{S}_k = S_{k-1} + \eta_k$
- For $\xi = -\log W, \quad \eta = -\log(1 - W), \quad$ the log-frequencies $(\log P_k, k \geq 1)$ is a perturbed RW
- Number of 'renewals'

$$N(x) := \#\{k \geq 0 : S_k \leq x\}, \quad \widetilde{N}(x) := \#\{k \geq 1 : \widetilde{S}_k \leq x\}$$

$$\varphi(x) := \int_0^x \mathbb{P}(\eta > y)dy$$

*Assume that* $\mu = \mathbb{E}\xi < \infty$ *and for some* $c(x)$

$$\frac{N(x) - \frac{x}{\mu}}{c(x)} \xrightarrow{d} Z, \quad \text{as } x \to \infty.$$

*Then* $Z$ *is a stable random variable (Bingham 1973), and*

$$\frac{\int_0^x \left(N(y) - \frac{y-\varphi(y)}{\mu}\right) dy}{xc(x)} \xrightarrow{d} \int_0^1 Z(y)dy, \quad \text{as } x \to \infty,$$

*where* $(Z(t), t \geq 0)$ *is a stable Lévy process corresponding to* $Z \stackrel{d}{=} Z(1)$.