Techniques from harmonic analysis and asymptotic results in probability theory

Pierre-Loïc Méliot

2018, July 3rd

University Paris-Sud (Orsay)

Objective: present some mathematical tools

- which allow us to study various random objets (stemming from combinatorics, arithmetics, random matrix theory, *etc.*), when their size n goes to infinity;
- which rely on several versions of the Fourier transform, or on related quantities (moments, cumulants).

Red line: a concrete problem on certain random variables stemming from the representation theory of symmetric groups.

0. Fine asymptotics of the Plancherel measures
1. Random objects chosen in a group or in its dual
2. Mod-Gaussian convergence and the method of cumulants
3. Perspectives

Plancherel measures

We denote $\mathfrak{P}(n)$ the set of integer partitions of n (Young diagrams). The **Plancherel measure** on $\mathfrak{P}(n)$ is the probability measure

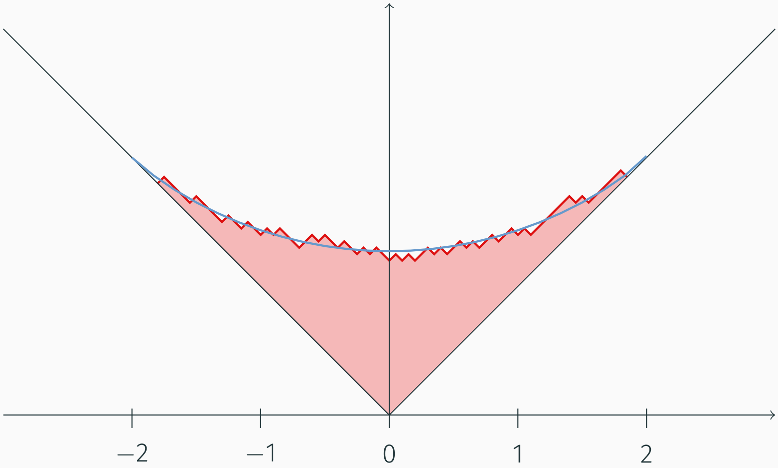
$$\mathbb{P}_n[\lambda] = \frac{(\dim \lambda)^2}{n!},$$

where $\dim \lambda$ is the number of standard tableaux with shape λ.

$$\lambda = (4,3,1) = \begin{array}{|c|c|c|c|} \cline{1-1} & \multicolumn{3}{c}{} \\ \cline{1-3} & & & \multicolumn{1}{c}{} \\ \hline & & & \\ \hline \end{array} \quad ; \quad T = \begin{array}{|c|c|c|c|} \cline{1-1} 6 & \multicolumn{3}{c}{} \\ \cline{1-3} 3 & 5 & 8 & \multicolumn{1}{c}{} \\ \hline 1 & 2 & 4 & 7 \\ \hline \end{array}\ .$$

The measure $\mathbb{P}_n$

- plays an essential role in the solution of Ulam's problem of the longest increasing subword,
- has its fluctuations related to the spectra of random matrices.



A random partition λ under the Plancherel measure $\mathbb{P}_{n=400}$.

Random character values and Kerov's central limit theorem

Each partition $\lambda \in \mathfrak{P}(n)$ corresponds to an irreducible representation $(V^\lambda, \rho^\lambda)$ of $\mathfrak{S}(n)$:

$$V^\lambda \text{ complex vector space with dimension } \dim \lambda,$$
$$\rho^\lambda : \mathfrak{S}(n) \to \mathrm{GL}(V^\lambda).$$

The random characters

$$X_k(\lambda) = \frac{\mathrm{tr}\,(\rho^\lambda(c_k))}{\mathrm{tr}\,(\rho^\lambda(\mathrm{id}))}, \quad c_k \text{ } k\text{-cycle}$$

are important random variables for studying the fluctuations of $\mathbb{P}_n$. Kerov's central limit theorem (1993) ensures that for any $k \geq 2$,

$$n^{k/2}\, X_k(\lambda) \rightharpoonup_{\substack{n\to+\infty \\ \lambda\sim\mathbb{P}_n}} \mathcal{N}_{\mathbb{R}}(0,k).$$

If X is a real-valued random variable, its **cumulants** are the coefficients $\kappa^{(r)}(X)$ of

$$\log(\mathbb{E}[\mathrm{e}^{zX}]) = \sum_{r=1}^{\infty} \frac{\kappa^{(r)}(X)}{r!}\, z^r.$$

A possible proof of Kerov's CLT relies on Śniady's estimate (2006)

$$\left|\kappa^{(r)}(n^{k/2}\, X_k(\lambda))\right| = O_{k,r}\left(n^{1-\frac{r}{2}}\right).$$

Idea: with a better control, one can make the CLT more precise and obtain Berry–Esseen estimates, concentration inequalities, large deviation principles, *etc.*

Conjecture: (2011) there exist constants $C = C_k$ such that

$$\forall (n,r),\ \left|\kappa^{(r)}(n^{k/2}\, X_k(\lambda))\right| \leq (Cr)^r\, n^{1-\frac{r}{2}}.$$

Random objects chosen in a group or in its dual

Groups and observables

G = a group (finite, or compact, or reductive);

$\widehat{G}$ = set $\{(V^\lambda, \rho^\lambda)\}$ of the irreducible representations of G.

We are interested in two kinds of random objects:

1. **random variables** $g \in G$, or objects constructed from such variables (examples: $g = g_t$ random walk on G; Γ random graph connecting random elements g).
2. **random representations** $\lambda \in \widehat{G}$, which encode combinatorial problems (examples: Plancherel measure on partitions; systems of interacting particles).

Goal: study g or λ when the size of the group or of the random object goes to infinity.

One obtains relevant information by considering the real- or matrix-valued random variables

$$\rho^\lambda(g) \qquad ; \qquad \mathrm{ch}^\lambda(g) = \mathrm{tr}\,(\rho^\lambda(g))$$

and by computing their moments (Poincaré; Diaconis; Kerov–Vershik).

Ingredients:

1. The set $\widehat{G}$ (or $\widehat{G}^K$ for random variables in G/K) is explicit (set of partitions, or subset of a lattice).
2. *Idem* for the dimensions d^λ and the characters (or the spherical functions), which satisfy orthogonality formulas.
3. The matrices $\rho^\lambda(g)$ are much more complicated to describe (problem of the choice of a basis; crystal theory).

Example: for Plancherel measures, $\mathbb{E}_n[X_k(\lambda)] = \mathbb{E}_n[\frac{\mathrm{ch}^\lambda(c_k)}{d^\lambda}] = 1_{(k=1)}$.

Cutoff phenomenon for Brownian motions

$$X = \text{compact Lie group } (\mathrm{SU}(n), \mathrm{SO}(n),\ \textit{etc.})$$
$$\text{or compact symmetric quotient } (\mathbb{S}^n, \mathrm{Gr}(n,d,\Bbbk),\ \textit{etc.}).$$

If $(x_t)_{t\in\mathbb{R}_+}$ is the Brownian motion on X (diffusion associated to the Laplace–Beltrami operator, starting from a fixed base point x_0), we know that

$$(\mu_t = \text{law of } x_t) \rightharpoonup_{t\to+\infty} \mathrm{Haar}.$$

The speed of convergence is given by the L^p-distances

$$d_p(\mu_t, \mathrm{Haar}) = \left(\int_X \left|\frac{d\mu_t(x)}{dx} - 1\right|^p dx\right)^{1/p},$$

in particular the L^1-distance (total variation).

Theorem (M., 2013)

In each class, there exists an explicit positive constant α such that

$$\begin{aligned} \forall \varepsilon > 0, \quad d_{\mathrm{TV}}(\mu_t, \mathrm{Haar}) &\geq 1 - \frac{C}{n^{c\varepsilon}} \quad && \text{if } t = \alpha(1-\varepsilon)\log n, \\ d_{\mathrm{TV}}(\mu_t, \mathrm{Haar}) &\leq \frac{C}{n^{c\varepsilon}} && \text{if } t = \alpha(1+\varepsilon)\log n. \end{aligned}$$

The same cutoff phenomenon occurs at the same time for the distances d_p, $p \in (1, +\infty)$.

Sketch of proof:

1. After the cutoff, one can compute $d_2(\mu_t, \mathrm{Haar}) \geq d_{\mathrm{TV}}(\mu_t, \mathrm{Haar})$.
2. Before the cutoff, one can find discriminating functions which behave differently under μ_t and under $\mu_\infty = \mathrm{Haar}$.

Random geometric graphs

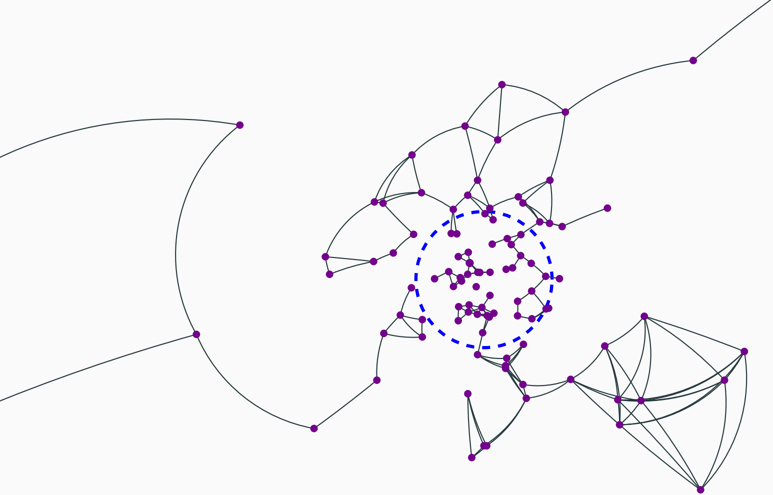
$(X, d) =$ compact symmetric space endowed with the geodesic distance.

The **geometric graph** with level $L > 0$ and size $N \in \mathbb{N}$ on X is the graph $\Gamma^X_{\mathrm{geom}}(N, L)$ obtained:

- by taking N independent points $x_1, \ldots, x_N$ under the Haar measure of X;
- by connecting x_i to x_j if $d(x_i, x_j) \leq L$.

We are interested in the spectra of the adjacency matrix of $\Gamma^X_{\mathrm{geom}}(N, L)$, in two distinct regimes:

1. **Gaussian regime:** L is fixed and $N \to +\infty$.
2. **Poisson regime:** $L = (\frac{\ell}{N})^{\frac{1}{\dim X}}$ with $\ell > 0$, and $N \to +\infty$.



Random geometric graph on the sphere $\mathbb{S}^2$, with $N = 100$ points and level $L = \frac{\pi}{8}$ (stereographic projection).

Gaussian regime

We assume that $X = G$ is a compact Lie group, and we denote d the rank of G, W its Weyl group and $\lambda \in \widehat{G}$ the dominant weights. We denote the spectrum of $\Gamma_{\mathrm{geom}}(N,L)$

$$e_{-1}(N,L) \leq e_{-2}(N,L) \leq \cdots \leq 0 \leq \cdots \leq e_2(N,L) \leq e_1(N,L) \leq e_0(N,L).$$

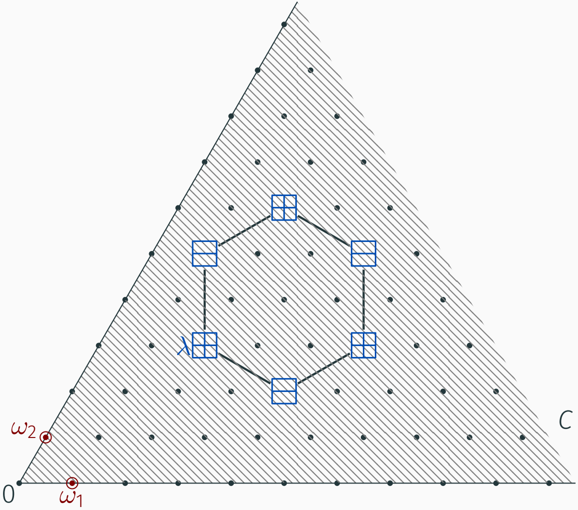
Theorem (M., 2017)

If L is fixed and N goes to infinity, there exist a.s. limits

$$e_i(L) = \lim_{N\to\infty} \frac{e_i(N,L)}{N}$$

for any $i \in \mathbb{Z}$. This limiting spectrum is the spectrum of a compact operator on $\mathrm{L}^2(G)$, and it consists in $(d^\lambda)^2$ values c_λ for each $\lambda \in \widehat{G}$, with

$$c_\lambda = \frac{1}{d^\lambda \,\mathrm{vol}(\mathfrak{t}/\mathfrak{t}_\mathbb{Z})} \left(\frac{L}{\sqrt{2\pi}}\right)^d \sum_{w\in W} \varepsilon(w)\, \widetilde{J}_{\frac{d}{2}}\left(L\,\|\lambda+\rho-w(\rho)\|\right).$$



A limiting eigenvalue c_λ for each dominant weight, given by an alternate sum of values of Bessel functions ($G = \mathrm{SU}(3)$).

Poisson regime

In the Poisson regime, $L = L_N = \left(\frac{\ell}{N}\right)^{\frac{1}{\dim X}}$ is chosen so that each vertex of $\Gamma^X_{\mathrm{geom}}(N, L_N)$ has $O(1)$ neighbors.

Theorem (M., 2018)

One has a Benjamini–Schramm local convergence

$$\Gamma^X_{\mathrm{geom}}(N, L_N) \to \Gamma_\infty,$$

where Γ_∞ is the geometric graph with level 1 obtained from a Poisson point process on $\mathbb{R}^{\dim X}$ with intensity $\frac{\ell}{\mathrm{vol}(X)}$, rooted at the point 0. This implies the convergence in probability

$$\mu_N = \frac{1}{N} \sum_{i=1}^{N} \delta_{e_i(N,L)} \rightharpoonup_{N \to +\infty} \mu_\infty;$$

the limiting measure μ_∞ is determined by its moments.

The moments $M_r = \int_{\mathbb{R}} x^r \, \mu_\infty(dx)$ have a combinatorial expansion involving certain graphs (circuits and reduced circuits), and one can give an explicit formula if $r \leq 7$.

Example: $M_5 = e$  $+ 5\, e$  $+ 5\, e$ 

$$= I_5\, (\ell')^4 + 5\, I_3\, I_2\, (\ell')^3 + 5\, I_3\, (\ell')^2$$

with $\ell' = \frac{\ell}{\mathrm{vol}(\mathfrak{t}/\mathfrak{t}_{\mathbb{Z}})}$ and $I_k = \int_C \left(\frac{(\partial_\Phi \_ \widetilde{J}_{\mathbb{R}\Omega})(x)}{(2\pi)^{d/2}} \right)^k \frac{dx}{(\delta(x))^{k-2}}$.

A general formula for M_r is related to a conjecture on certain functionals of the representations $\lambda \in \widehat{G}$, and to an interpretation of these functionals in the Kashiwara–Lusztig crystal theory. This conjecture is true for tori, which should allow us to find the support of μ_∞.

Mod-Gaussian convergence and the method of cumulants

Mod-ϕ convergence

The previous problems have been solved by computing the moments of observables of the random objects. To get more precise information, one can look at the fine asymptotics of the Fourier or Laplace transform of the observables.

General framework:

ϕ infinitely divisible law with $\int_{\mathbb{R}} \mathrm{e}^{zx}\,\phi(dx) = \mathrm{e}^{\eta(z)}$;

$(X_n)_{n\in\mathbb{N}}$ sequence of real random variables;

$(t_n)_{n\in\mathbb{N}}$ sequence of parameters growing to $+\infty$

with

$$\lim_{n\to+\infty} \frac{\mathbb{E}[\mathrm{e}^{zX_n}]}{\mathrm{e}^{t_n \eta(z)}} = \psi(z) \text{ locally uniformly in } z \in \mathbb{C}.$$

One obtains: an extended CLT, speed of convergence estimates, large deviation estimates, local limit theorems, concentration inequalities.

The method of cumulants

In the Gaussian case ($\eta(z) = \frac{z^2}{2}$), consider $(S_n)_{n\in\mathbb{N}}$ which satisfies the hypotheses of the **method of cumulants** with parameters (A, D_n, N_n):

(MC1) There exist $A > 0$ and two sequences $N_n \to +\infty$ and $D_n = o(N_n)$ such that

$$\forall r \geq 1, \quad \left|\kappa^{(r)}(S_n)\right| \leq N_n \,(2D_n)^{r-1}\, r^{r-2}\, A^r.$$

(MC2) There exist $\sigma^2 \geq 0$ and L such that

$$\frac{\kappa^{(2)}(S_n)}{N_n D_n} = \sigma^2 \,\left(1 + o\Big((D_n/N_n)^{\frac{1}{3}}\Big)\right);$$

$$\frac{\kappa^{(3)}(S_n)}{N_n (D_n)^2} = L\,(1 + o(1)).$$

1. If $\sigma^2 > 0$, then $\frac{S_n - \mathbb{E}[S_n]}{\sqrt{\mathrm{var}(S_n)}} = Y_n \rightharpoonup \mathcal{N}_{\mathbb{R}}(0,1)$, and more precisely,

$$d_{\mathrm{Kol}}(Y_n, \mathcal{N}_{\mathbb{R}}(0,1)) = O\left(\frac{A^3}{\sigma^3}\sqrt{\frac{D_n}{N_n}}\right).$$

2. The normality zone of $(Y_n)_{n\in\mathbb{N}}$ is a $o((\frac{N_n}{D_n})^{\frac{1}{6}})$. If $y_n \to +\infty$ and $y_n \ll (\frac{N_n}{D_n})^{\frac{1}{4}}$, then

$$\mathbb{P}[Y_n \geq y_n] = \frac{\mathrm{e}^{-\frac{(y_n)^2}{2}}}{y_n\sqrt{2\pi}}\ \exp\left(\frac{L}{6\sigma^3}\sqrt{\frac{D_n}{N_n}}\,(y_n)^3\right)\,(1+o(1)).$$

3. If $|S_n| \leq N_n\,A$ almost surely, then

$$\forall x \geq 0,\ \ \forall n \in \mathbb{N},\ \ \mathbb{P}[|S_n - \mathbb{E}[S_n]| \geq x] \leq 2\,\exp\left(-\frac{x^2}{9\,AD_nN_n}\right).$$

4. For any $\varepsilon \in (0,\frac{1}{2})$, and any Jordan-measurable subset B with $m(B) \in (0,+\infty)$,

$$\left(\frac{N_n}{D_n}\right)^{\varepsilon}\,\mathbb{P}\left[Y_n - y \in \left(\frac{D_n}{N_n}\right)^{\varepsilon}\,B\right] = \frac{\mathrm{e}^{-\frac{y^2}{2}}}{\sqrt{2\pi}}\,m(B)\,(1+o(1)).$$

(results obtained with V. Féray and A. Nikeghbali, 2013-17).

Dependency graph and applications

These theoretical results are obtained by using classical techniques from real harmonic analysis. We have also identified mathematical structures which imply the upper bound on cumulants (or a mod-ϕ convergence).

Theorem (FMN, 2013)

Let $S = \sum_{v \in V} A_v$ be a sum of random variables bounded by A and such that there is a graph $G = (V, E)$ with:

- *$N = |V|$, and $\max_{v \in V} \deg v \leq D$;*
- *if $V_1, V_2 \subset V$ are disjoint and not connected by an edge, then $(A_v)_{v \in V_1}$ and $(A_v)_{v \in V_2}$ are independent families.*

For any $r \geq 1$,

$$\left|\kappa^{(r)}(S)\right| \leq N\,(2D)^{r-1}\,r^{r-2}\,A^r.$$

Examples with the method of cumulants:

- count of subgraphs in a random graph (Erdös–Rényi: 2013; graphons: 2017);
- count of motives in a random permutation (2017);
- linear functionals of a Markov chain (2015);
- magnetisation of the Ising model (2014, 2016);
- random integer partitions under the central measures (2013, 2017).

Other examples of mod-ϕ convergence:

- characteristic polynomials of random matrices in compact groups (mod-Gaussian, 2013);
- arithmetic functions of random integers (mod-Poisson, 2013);
- random combinatorial objects whose generating series has an algebraic-logarithmic singularity (mod-Poisson, 2014).

Central measures on partitions

The method of cumulants enables one to identity in a family of random models the models which have additional symmetries and whose fluctuations are not of typical size.

$$\begin{aligned}\mathscr{T} &= \text{Thoma simplex} \\ &= \text{positive extremal characters of} \\ &\quad\ \text{the infinite symmetric group } \mathfrak{S}(\infty).\end{aligned}$$

Given $\omega \in \mathscr{T}$,

$$(\chi_\omega)_{|\mathfrak{S}(n)} = \sum_{\lambda \in \mathfrak{P}(n)} \mathbb{P}_{n,\omega}[\lambda]\, \frac{\mathrm{ch}^\lambda}{d^\lambda}$$

and the weights $\mathbb{P}_{n,\omega}[\lambda]$ form a central measure on $\mathfrak{P}(n)$, the Plancherel measure being a particular case.

For $k \geq 2$, we set $S_{n,k} = n^{\downarrow k} \frac{\mathrm{ch}^\lambda(c_k)}{d^\lambda}$ with $\lambda \sim \mathbb{P}_{n,\omega}$.

Theorem (FMN, 2013-17)

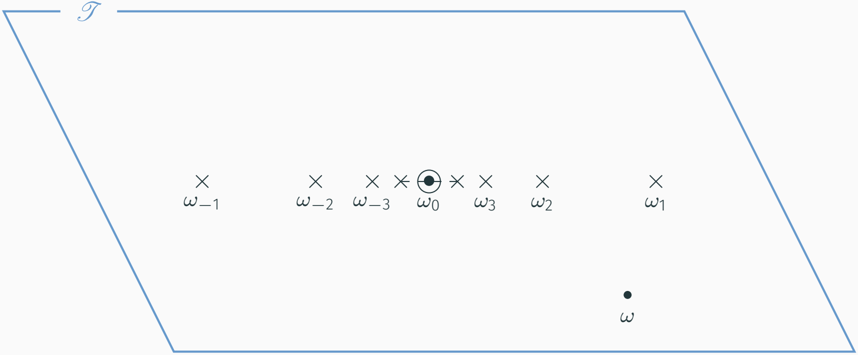
The variables $S_{n,k}$ satisfy the hypotheses of the method of cumulants with $A = 1$, $D_n = O(n^{k-1})$ and $N_n = O(n^k)$. The limiting parameters (σ^2, L) are explicit continuous functions of $\omega \in \mathscr{T}$.

Generically, $\sigma^2 = \sigma^2(k,\omega) > 0$ and the variables $\frac{S_{n,k}}{n^{k-1/2}}$ satisfy a CLT and all the other estimates.

The singular set of parameters $\omega \in \mathscr{T}$ such that $\sigma^2(k,\omega) = 0$ for any k consists in:

- ω_0 corresponding to the **Plancherel measures**;
- $\omega_{d\geq 1}$ corresponding to the Schur–Weyl measures, and $\omega_{d\leq -1}$ corresponding to their duals.

Mod-Gaussian moduli spaces



$$\begin{aligned}
\omega \text{ (generic)} &: |\kappa^{(r)}(S_{n,k})| \leq (Cr)^r\, n^{k+(r-1)(k-1)};\\
\omega_{d\neq 0} \text{ (Schur–Weyl)} &: |\kappa^{(r)}(S_{n,k})| \leq (Cr)^r\, n^{r(k-1)};\\
\omega_0 \text{ (Plancherel)} &: |\kappa^{(r)}(S_{n,k})| \leq (Cr)^r\, n^{1+\frac{r(k-1)}{2}} \quad ???
\end{aligned}$$

Other examples of **mod-Gaussian moduli spaces** (2017, 2018):

- space $\mathscr{G}$ of graphons (random graphs — observables: counts of subgraphs — singular models: Erdös–Rényi + ?);
- space $\mathscr{P}$ of permutons (random permutations — observables: counts of motives — singular models: ?);
- space $\mathscr{M}$ of measured metric spaces (discrete random metric spaces — topology: Gromov–Hausdorff–Prohorov — singular models: compact homogeneous spaces — work of J. De Catelan).

Perspectives

Fluctuations of dynamical systems

We want to investigate the fluctuations of sums

$$S_n(f) = \sum_{k=1}^{n} f(T^n(x)),$$

where $T : X \to X$ is a mixing dynamical system and $x \in X$ is chosen randomly. These fluctuations are classically studied with the Nagaev–Guivarc'h spectral method.

Objective: understand and improve these results by using the method of cumulants, in the framework of mod-Gaussian moduli spaces.

Ingredients: an extension of the theory of dependency graphs using weighted graphs (already involved in the study of functionals of Markov chains).

Cumulants of the Plancherel measures

With the technology of dependency graphs, one can show the following bound for the Plancherel measure (2018):

$$\left|\kappa^{(r)}(S_{n,k})\right| \leq \frac{(\sqrt[3]{3}\,k)^r}{r}\, C_{k,r,n}\, n^{1+\frac{r(k-1)}{2}}$$

with

$$C_{k,r,n} = \sum_{\substack{\pi \in \mathfrak{Q}(r) \\ g_1,\ldots,g_{\ell(\pi)} \geq 0}} \left(\frac{rk^2}{4}\right)^{\ell(\pi)-1} \prod_{a=1}^{\ell(\pi)} \left(\frac{\mathrm{C}^*\left(k, |\pi_a|, 1+\frac{|\pi_a|(k-1)}{2} - g_a\right)}{n^{g_a}\, n_a!}\right),$$

where $\mathfrak{Q}(r)$ is the set of set partitions $[\![1, r]\!] = \pi_1 \sqcup \pi_2 \sqcup \cdots \sqcup \pi_\ell$, and $\mathrm{C}^*(k, l, N)$ is the number of transitive factorisations of the identity of $[\![1, N]\!]$ as a product of l cycles with length k.

Objective: get upper bounds on the numbers of factorisations.

- On can interpret the numbers $\mathrm{C}^*(k,l,N)$ as structure coefficients of an **algebra of split permutations** $\mathscr{A}$.
- The sum $C_{k,r,n}$ can be rewritten as a trace $\tau((\Omega_k)^r)$ in a q-deformation $\mathscr{A}_q$ of the algebra $\mathscr{A}$, specialised at $q = \frac{1}{n}$.
- The main contribution to $C_{k,r,n}$ corresponds to the specialisation $q = 0$, and to the **algebras of planar factorisations** which are not semisimple, and whose representation theory is of particular interest.

The end