Invent. math.
https://doi.org/10.1007/s00222-022-01123-y

Large genus asymptotic geometry of random square-tiled surfaces and of random multicurves

Vincent Delecroix[1] · **Élise Goujard**[2] · **Peter Zograf**[3] · **Anton Zorich**[4,5]

Received: 10 July 2020 / Accepted: 5 May 2022

Abstract We study the combinatorial geometry of a random closed multicurve on a surface of large genus g and of a random square-tiled surface of large genus g. We prove that primitive components $\gamma_1, \dots, \gamma_k$ of a ran-

This work was supported by the Agence Nationale de la Recherche grants ANR-19-CE40-0021 and ANR-19-CE40-0003, by the NSF grant DMS-1440140, and by the Ministry of Science and Higher Education of the Russian Federation, agreement no. 075-15-2019-162. In addition to that, Goujard's work was partially supported by a Projets Exploratoires Premier Soutien (PEPS) JCJC grant.

✉ Anton Zorich
anton.zorich@gmail.com

Vincent Delecroix
vincent.delecroix@u-bordeaux.fr

Élise Goujard
elise.goujard@gmail.com

Peter Zograf
zograf@pdmi.ras.ru

1 LaBRI, Domaine Universitaire, 351 cours de la Libération, 33405 Talence, France

2 Institut de Mathématiques de Bordeaux, Université de Bordeaux, 351 cours de la Libération, 33405 Talence, France

3 Euler International Mathematical Institute, Pesochnaya nab. 10, Saint Petersburg, Russia 197022

4 Center for Advanced Studies, Skoltech, Moscow, Russia

5 Institut de Mathématiques de Jussieu – Paris Rive Gauche, Case 7012, 8 Place Aurélie Nemours, 75205 Paris Cedex 13, France

Published online: 20 June 2022

dom multicurve $m_1\gamma_1 + \cdots + m_k\gamma_k$ represent linearly independent homology cycles with asymptotic probability 1 and that all its weights m_i are equal to 1 with asymptotic probability $\sqrt{2}/2$. We prove analogous properties for random square-tiled surfaces. In particular, we show that all conical singularities of a random square-tiled surface belong to the same leaf of the horizontal foliation and to the same leaf of the vertical foliation with asymptotic probability 1. We show that the number of components of a random multicurve and the number of maximal horizontal cylinders of a random square-tiled surface of genus g are both very well approximated by the number of cycles of a random permutation for an explicit non-uniform measure on the symmetric group of $3g-3$ elements. In particular, we prove that the expected value of these quantities has asymptotics $(\log(6g-6)+\gamma)/2+\log 2$ as $g \to \infty$, where γ is the Euler–Mascheroni constant. These results are based on our formula for the Masur–Veech volume $\operatorname{Vol}\mathcal{Q}_g$ of the moduli space of holomorphic quadratic differentials combined with deep large genus asymptotic analysis of this formula performed by A. Aggarwal and with the uniform asymptotic formula for intersection numbers of ψ-classes on $\overline{\mathcal{M}}_{g,n}$ for large g proved by A. Aggarwal in 2020.

Contents

1 Introduction and statements of main results

We aim to study random multicurves on surfaces of large genus g. Before proceeding to the statements of our main results we consider the classical setting of random integers and of random permutations which allow to set up the concept of a random compound object. For more information on the probabilistic analysis of decomposition of combinatorial objects into elementary components we recommend the monograph of Arratia et al. [8] and the survey of Pitman [72]. An enlightening introduction can be found in the blog post of Tao [78].

Prime decomposition of a random integer The Prime Number Theorem states that an integer number n taken randomly in a large interval $[1, N]$ is prime with asymptotic probability $\frac{\log N}{N}$. Denote by $\omega(n)$ the number of prime divisors of an integer n counted without multiplicities. In other words, if n has prime decomposition $n = p_1^{m_1} \dots p_k^{m_k}$, let $\omega(n) = k$. By the Erdős–Kac theorem [31], the centered and rescaled distribution prescribed by the counting function $\omega(n)$ tends to the normal distribution:

$$\lim_{N\to+\infty} \frac{1}{N} \operatorname{card} \left\{ n \le N \;\middle|\; \frac{\omega(n) - \log\log n}{\sqrt{\log\log n}} \le x \right\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt. \tag{1.1}$$

The subsequent papers of Rényi and Turán [74], and of Selberg [76] describe the rate of convergence.

Cycle decomposition of a uniform random permutation Denote by $\mathrm{K}_n(\sigma)$ the number of disjoint cycles in the cycle decomposition of a permutation σ in the symmetric group S_n. Consider the uniform probability measure on S_n. A random permutation σ of n elements has exactly k cycles in its cyclic decomposition with probability $\mathbb{P}\big(\mathrm{K}_n(\sigma) = k\big) = \frac{s(n,k)}{n!}$, where $s(n, k)$ is the unsigned Stirling number of the first kind. It is immediate to see that $\mathbb{P}\big(\mathrm{K}_n(\sigma) = 1\big) = \frac{1}{n}$. Goncharov proved in [37] the following expansions for the expected value and for the variance of K_n as $n \to +\infty$:

$$\mathbb{E}(\mathrm{K}_n) = \log n + \gamma + o(1), \qquad \mathbb{V}(\mathrm{K}_n) = \log n + \gamma - \zeta(2) + o(1), \tag{1.2}$$

as well as the following central limit theorem

$$\lim_{n\to+\infty} \frac{1}{n!} \operatorname{card} \left\{ \sigma \in S_n \;\middle|\; \frac{\mathrm{K}_n(\sigma) - \log n}{\sqrt{\log n}} \le x \right\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt. \tag{1.3}$$

Here γ in Formula (1.2) is the Euler–Mascheroni constant.

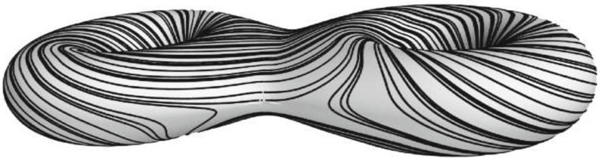

Fig. 1 Simple closed multicurve on a surface of genus two

As can be seen in (1.2) and in (1.3), the number of cycles in the cycle decomposition of a random permutation is of the order of $\log n$. In such a situation one expects the distribution to be close to a Poisson distribution. Recall that the *Poisson distribution with parameter* λ is

$$\mathrm{Poi}_\lambda(k) = e^{-\lambda}\,\frac{\lambda^k}{k!}, \qquad \text{where } k = 0, 1, 2, \dots \tag{1.4}$$

Hwang proved in [43] that the distribution of the random variable K_n is approximated by the Poisson distribution $\mathrm{Poi}_{\log n}$ in a very strong sense, which can be formalized as "*mod-Poisson convergence with parameter* $\log n$ *and limiting function* $1/\Gamma(t)$", using the terminology of Kowalski and Nikeghbali [51]. We discuss the notion of mod-Poisson convergence in Sect. 3.2. We emphasize that such approximation is much stronger than the central limit theorem.

The result of Hwang [45] (representing a particular case of results in [43, Chapter 5]) implies, that for large n and for any positive x, the distribution of the number of cycles is uniformly well-approximated in a neighborhood of $x \log n$ by the Poisson distribution with parameter $\log n + \mathrm{a}(x)$, where the explicit correctional constant term $\mathrm{a}(x)$ is completely determined by the limiting function and does not depend on n. Namely, for any $x > 0$ we have uniformly in $0 \le k \le x \log n$

$$\mathbb{P}\big(\mathrm{K}_n = k+1\big) = \frac{(\log n)^k}{n \cdot k!}\left(\frac{1}{\Gamma(1+\frac{k}{\log n})} + O\left(\frac{k}{(\log n)^2}\right)\right). \tag{1.5}$$

Shape of a random multicurve on a surface of fixed genus Consider a smooth oriented closed surface S of genus g. A *multicurve* $\gamma = \sum m_i \gamma_i$ (as the one in the picture by Calegari from [19] presented in Fig. 1) is a formal weighted sum of curves γ_i with strictly positive integer weights (also called "multiplicities") m_i where $\gamma_1, \dots, \gamma_k$ is a collection of non-contractible simple curves on S that are pairwise non-isotopic. Following the usual convention, we do not distinguish between the free homotopy class of a multicurve and the multicurve itself.

Every multicurve $\gamma = \sum m_i \gamma_i$ defines a *reduced* multicurve $\gamma_{reduced} = \sum \gamma_i$. Note that the number of reduced multicurves on a surface of a fixed genus g considered up to the action of the mapping class group Mod_g is finite. We say that two multicurves have the same topological type if they belong to the same orbit of Mod_g. For example, a simple closed curve has one of the following topological types: either it is non-separating, or it separates the surface into subsurfaces of genera g' and $g - g'$ for some $1 \leq g' \leq g/2$.

Multicurves on a closed surface of genus g (considered up to free homotopy) are parameterized by integer points $\mathcal{ML}_g(\mathbb{Z})$ in the space of measured laminations $\mathcal{ML}_g$ introduced by Thurston [79]. Any hyperbolic metric on S provides a length function ℓ that associates to a closed curve γ the length $\ell(\gamma)$ of its unique geodesic representative. The length function ℓ extends to multicurves as $\ell(\gamma) = m_1 \ell(\gamma_1) + \cdots + m_k \ell(\gamma_k)$. Fixing some upper bound L for the length of a multicurve, one can consider the finite set of multicurves of length at most L on S with respect to the length function ℓ. See also the paper of Mirzakhani [61] and works of Erlandsson, Parlier, Rafi and Souto [32,33,73] for alternative ways to measure the length of a multicurve.

Choosing the uniform measure on all integral multicurves of length at most L and letting L tend to infinity we define a "random multicurve" on a surface of fixed genus g in the same manner as we considered "random integers", see Sect. 2.4 for details. We emphasize that studying asymptotic statistical geometry of multicurves as the bound L tends to infinity we always keep the genus g, considered as a parameter, fixed. One can ask, for example, what is the probability that a random simple closed curve separates the surface of genus g in two components? Or, more generally, what is the probability that the reduced multicurve $\gamma_{reduced}$ associated to a random multicurve γ separates the surface of genus g into several components? With what probability a random multicurve $m_1\gamma_1 + m_2\gamma_2 + \cdots + m_k\gamma_k$ has $k = 1, 2, \ldots, 3g-3$ primitive connected components $\gamma_1, \ldots, \gamma_k$? What are the typical weights $m_1, \ldots, m_k$?

A beautiful answer to all these questions was found by Mirzakhani in [62]. She expressed the frequency of multicurves of any fixed topological type in terms of the intersection numbers $\int_{\overline{\mathcal{M}}_{g',n'}} \psi_1^{d_1} \ldots \psi_{n'}^{d_{n'}}$, where $2g' + n' \leq 2g$. (These intersection numbers are also called *correlators of Witten's two dimensional topological gravity*). For small genera g the formula of M. Mirzakhani provides explicit rational values for the quantities discussed above. For example, the reduced multicurve associated to a random multicurve on a surface of genus 2 without cusps as in Fig. 1 separates the surface with probability $\frac{67}{315}$ and has 1, 2 or 3 components with probabilities $\frac{7}{27}, \frac{5}{9}, \frac{5}{27}$ respectively.

The formulae of Mirzakhani are applicable to surfaces of any genera. The exact values of the intersection numbers can be computed through Witten–Kontsevich theory [50,84]. However, despite the fact that these intersection numbers were extensively studied, there were no uniform estimates for Witten's correlators for large g till the recent results of Aggarwal [3]. This is one of the reasons why the following question remained open.

Question 1 *What shape has a random multicurve on a surface of large genus?*

The present paper aims to answer this question to some extent. Denote by $K_g(\gamma)$ the number of components k of the multicurve $\gamma = \sum_{i=1}^{k} m_i \gamma_i$ counted without multiplicities.

Theorem 1.1 *Consider a random multicurve $\gamma = \sum_{i=1}^{k} m_i \gamma_i$ on a surface S of genus g. Let $\gamma_{reduced} = \gamma_1 + \cdots + \gamma_k$ be the underlying reduced multicurve. The following asymptotic properties hold as $g \to +\infty$.*

(a) *The multicurve $\gamma_{reduced}$ does not separate the surface (i.e. $S - \sqcup \gamma_i$ is connected) with probability which tends to 1.*
(b) *The probability that a random multicurve $\gamma = \sum_{i=1}^{k} m_i \gamma_i$ is primitive (i.e. that $m_1 = m_2 = \cdots = 1$) tends to $\frac{\sqrt{2}}{2}$. (Here the number k of components is not fixed and might take any value from 1 to $3g - 3$.)*
(c) *For any sequence of positive integers k_g with $k_g = o(\log g)$ the probability that a random multicurve $\gamma = \sum_{i=1}^{k_g} m_i \gamma_i$ is primitive (i.e. that $m_1 = \cdots = m_{k_g} = 1$) tends to 1.*

There is no contradiction between parts (b) and (c) of the above Theorem since in (c) we consider only those random multicurves for which the underlying primitive multicurve has an imposed number k_g of components, while in (b) we consider all multicurves. In other words, in part (c) we consider the conditional probability. Part (b) of the above Theorem admits the following generalization.

Theorem 1.2 *For any positive integer m, the probability that all weights m_i of a random multicurve $\gamma = m_1\gamma_1 + m_2\gamma_2 + \cdots$ on a surface of genus g are bounded by a positive integer m (i.e. that $m_1 \leq m, m_2 \leq m, \ldots$) tends to $\sqrt{\frac{m}{m+1}}$ as $g \to +\infty$.*

We describe the probability distribution of the random variable $K_g(\gamma)$ later in this section. However, to follow comparison with prime decomposition of random integers and with cycle decomposition of random permutations we present here the central limit theorem stated for random multicurves.

Theorem 1.3 *Choose a non-separating simple closed curve ρ_g on a surface of genus g. Denote by $\iota(\rho_g, \gamma)$ the geometric intersection number of ρ_g and γ. The centered and rescaled distribution defined by the counting function $K_g(\gamma)$*

tends to the normal distribution:

$$\lim_{g\to+\infty}\sqrt{\frac{3\pi g}{2}}\cdot 12g\cdot(4g-4)!\cdot\left(\frac{9}{8}\right)^{2g-2}\lim_{N\to+\infty}\frac{1}{N^{6g-6}}$$
$$\operatorname{card}\left(\left\{\gamma\in\mathcal{ML}_g(\mathbb{Z})\,\middle|\,\iota(\rho_g,\gamma)\le N\ \text{ and }\ \frac{K_g(\gamma)-\frac{\log g}{2}}{\sqrt{\frac{\log g}{2}}}\le x\right\}/\operatorname{Stab}(\rho_g)\right)$$
$$=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{t^2}{2}}\,dt.$$

Here $\operatorname{Stab}(\rho_g)\subset\operatorname{Mod}_g$ is the stabilizer of the simple closed curve ρ_g in the mapping class group Mod_g.

In plain words, the above theorems say that the components $\gamma_1,\dots,\gamma_k$ of a random multicurve $\gamma=\sum_{i=1}^{k}m_i\gamma_i$ on a surface of large genus g have all chances to go around k independent handles, where k is close to $\frac{1}{2}\log g$, and that with a high probability all the weights m_i of a random multicurve are uniformly small. In particular, with probability greater than 0.7 a random multicurve is primitive, i.e. all the weights m_i are equal to 1.

Our description of the asymptotic geometry of random multicurves on surfaces of large genus and of random square-tiled surfaces of large genus relies on fundamental recent results [3] of A. Aggarwal, who proved, in particular, the large genus asymptotic formulae for the Masur–Veech volume $\operatorname{Vol}\mathcal{Q}_g$ and for the intersection numbers of ψ-classes on $\overline{\mathcal{M}}_{g,n}$, conjectured by the authors in [26,29].

Random square-tiled surfaces of large genus A *square-tiled surface* is a closed oriented quadrangulated surface (i.e. a surface built by gluing identical squares along their sides), such that the quadrangulation satisfies the following properties. Consider the flat metric on the surface induced by the flat metric on the squares. We assume that sides of the squares are identified by isometries, which implies that the induced flat metric is non-singular on the complement of the vertices of the squares. We require that the parallel transport of a vector $\vec{v}$ tangent to the surface along any closed path avoiding conical singularities brings the vector $\vec{v}$ either to itself or to $-\vec{v}$. In other words, we require that the holonomy group of the metric is $\mathbb{Z}/2\mathbb{Z}$ (compared to $\mathbb{Z}/4\mathbb{Z}$ for a general quadrangulation). This holonomy assumption implies that defining some side to be "horizontal" or "vertical" we uniquely determine for each of the remaining sides, whether it is "horizontal" or "vertical". Speaking of square-tiled surfaced we always assume that the choice of horizontal and vertical sides is done.

Our holonomy assumption implies that the number of squares adjacent to any vertex is even. In this article we restrict ourselves to considering square-tiled surfaces with no conical singularities of angle π. In other words, vertices adjacent to exactly two squares are not allowed. Square-tiled surfaces satisfying the above restrictions can be seen as integer points in the total space $\mathcal{Q}_g$ of the vector bundle of holomorphic quadratic differentials over the moduli space of complex curves $\mathcal{M}_g$.

A stronger restriction on the quadrangulation imposing trivial linear holonomy to the induced flat metric defines *Abelian* square-tiled surfaces; they correspond to integer points in the total space $\mathcal{H}_g$ of the Hodge bundle of holomorphic Abelian differentials over the moduli space of complex curves $\mathcal{M}_g$. The subset of square-tiled surfaces having prescribed linear holonomy and prescribed cone angle at each conical singularity corresponds to the set of integer points in the associated *stratum* in the moduli space of quadratic or Abelian differentials respectively.

A square-tiled surface admits a natural decomposition into maximal horizontal cylinders. For example, the square-tiled surface in the left picture of Fig. 2 (which, for simplicity of illustration, contains conical points with cone angles π) has four maximal horizontal cylinders highlighted by different shades of grey. Two of these cylinders are composed of two horizontal bands of squares. Each of the remaining two cylinders is composed of a single horizontal band of squares.

For any positive integer N, the set $\mathcal{ST}_g(N)$ of square-tiled surfaces of genus g having no singularities of angle π and having at most N squares in the tiling is finite. Choosing the uniform measure on the set $\mathcal{ST}_g(N)$ and letting the bound N for the number of squares tend to infinity, we define a "random square-tiled surface" of fixed genus g in the same manner as we considered "random multicurves" on a fixed surface, see Sect. 2.5 for details. We emphasize that studying asymptotic statistical geometry of square-tiled surfaces as the bound N tends to infinity we always keep the genus g, considered as a parameter, fixed. One can study the decomposition of a random square-tiled surface into maximal horizontal cylinders in the same sense as we considered prime decomposition of random integers or cycle decomposition of random permutations.

For each stratum in the moduli space of Abelian differentials, we computed in [28] the probability that a random square-tiled surface in this stratum has a single cylinder in its horizontal cylinder decomposition. This result can be seen as an analog of the Prime Number Theorem for square-tiled surfaces. In particular, using results [2,23] we proved that for strata of Abelian differentials corresponding to large genera, this probability is asymptotically $\frac{1}{d}$, where d is the dimension of the stratum. However, more detailed description of statistics of square-tiled surfaces in individual strata of Abelian differentials is currently out of reach with the exception of several low-dimensional strata. Conjecturally, for any stratum of Abelian differentials of dimension d, the statistics of the number of maximal horizontal cylinders of a random square-tiled surface in the stratum becomes very well-approximated by the statistics of the number $\mathrm{K}_n(\sigma)$ of disjoint cycles in a random permutation of d elements as $d \to +\infty$; see Sect. 6.2 for details.

In the present paper we address more general question.

Question 2 *What shape has a random square-tiled surface of large genus assuming that it does not have conical points of angle π?*

Denote by $K_g(S)$ the number of maximal horizontal cylinders in the cylinder decomposition of a square-tiled surface S of genus g.

Theorem 1.4 *A random square-tiled surface S of genus g with no conical singularities of angle π has the following asymptotic properties as $g \to +\infty$.*

(a) *All conical singularities of S are located at the same leaf of the horizontal foliation and at the same leaf of the vertical foliation with probability which tends to 1.*
(b) *The probability that each maximal horizontal cylinder of S is composed of a single band of squares tends to $\frac{\sqrt{2}}{2}$.*
(c) *For any sequence of positive integers k_g with $k_g = o(\log g)$ the probability that each maximal horizontal cylinder of a random k_g-cylinder square-tiled surface of genus g is composed of a single band of squares tends to* 1.

Similarly to the case of multicurves, part (b) of the above Theorem admits the following generalization.

Theorem 1.5 *For any $m \in \mathbb{N}$, the probability that every maximal horizontal cylinder of a random square-tiled surface of genus g has at most m bands of squares tends to $\sqrt{\frac{m}{m+1}}$ as $g \to +\infty$.*

We state now the central limit theorem for square-tiled surfaces.

Theorem 1.6 *The centered and rescaled distribution defined by the counting function $K_g(S)$ tends to the normal distribution as $g \to +\infty$:*

$$\lim_{g\to+\infty} 3\pi g \cdot \left(\frac{9}{8}\right)^{2g-2} \lim_{N\to+\infty} \frac{1}{N^{6g-6}}$$
$$\operatorname{card}\left\{ S \in \mathcal{ST}_g(N) \,\middle|\, \frac{K_g(S) - \frac{\log g}{2}}{\sqrt{\frac{\log g}{2}}} \le x \right\}$$
$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt.$$

Approach to the study of random multicurves and of random square-tiled surfaces of large genera: from $p_g(k)$ to $q_g(k)$. It is time to admit that the parallelism between Theorems 1.1–1.3 and respectively Theorems 1.4–1.6 is not accidental.

Recall that we denote by $K_g(\gamma)$ the number of components k of the multicurve $\gamma = \sum_{i=1}^{k} m_i \gamma_i$ on a surface of genus g counted without multiplicities and by $K_g(S)$ the number of maximal horizontal cylinders in the cylinder decomposition of a square-tiled surface S of genus g. The following theorem is a direct corollary of Theorem 1.22 from Section 1.8 in [29]. (For the sake of completeness we reproduce the original Theorem in Sect. 2.5 below.)

Theorem 1.7 *For any genus $g \geq 2$ and for any $k \in \mathbb{N}$, the probability $p_g(k)$ that a random multicurve γ on a surface of genus g has exactly k components counted without multiplicities coincides with the probability that a random square-tiled surface S of genus g has exactly k maximal horizontal cylinders:*

$$p_g(k) = \mathbb{P}\big(K_g(\gamma) = k\big) = \mathbb{P}\big(K_g(S) = k\big). \tag{1.6}$$

In other words, $K_g(\gamma)$ and $K_g(S)$, considered as random variables, determine the same probability distribution $p_g(k)$, where $k = 1, 2, \ldots, 3g - 3$.

The above Theorem shows that Questions 1 and 2 are, basically, equivalent. The description of the large genus asymptotic properties of the resulting probability distribution $p_g(k)$ can be seen as the main unified goal of the present paper.

The starting point of our approach to the study of the probability distribution $p_g(k)$ is the formula for the Masur–Veech volume Vol $\mathcal{Q}_g$ of the moduli space of holomorphic quadratic differentials derived in our recent paper [29]. This formula represents Vol $\mathcal{Q}_g$ as a finite sum of contributions of genus g square-tiled surfaces of all possible topological types (Sect. 2.3 describes this in detail). However, the number of such topological types grows exponentially as genus grows. Moreover, the contribution of square-tiled surfaces of a fixed topological type to Vol $\mathcal{Q}_g$ is expressed in terms of the intersection numbers of ψ-classes (Witten's correlators) which are difficult to evaluate explicitly in large genera.

We conjectured in [26] that in large genera, the dominant part of the contribution to Vol $\mathcal{Q}_g$ comes from square-tiled surfaces having all conical singularities at the same horizontal level. The topological type (see Sect. 2.1 for the rigorous definition of the "topological type") of such square-tiled surfaces is completely determined by the number k of maximal horizontal cylinders which varies from 1 to g. This conjecture suggested a strategy for overcoming the first difficulty, reducing the study of all immense variety of topological types of square-tiled surfaces to the study of g distinguished topological types. We also conjectured in [26] that under certain assumptions on g and n, the intersection numbers $\int_{\overline{\mathcal{M}}_{g,n}} \psi_1^{d_1} \dots \psi_n^{d_n}$ are uniformly well-approximated by an explicit closed expression in the variables $d_1, \dots, d_n$, and that the error term becomes uniformly small with respect to all possible partitions $d_1+\dots+d_n = 3g-3+n$ for large values of g. This conjecture suggested a plan for overcoming the second difficulty reducing analysis of volume contributions of square-tiled surfaces of g distinguished topological types to analysis of closed expressions in multivariate harmonic sums. Such analysis led us, in particular, to the conjectural large genus asymptotics of the Masur–Veech volume Vol $\mathcal{Q}_g$.

In terms of the probability distributions, we replace the original distribution $p_g(k)$ with an auxiliary probability distribution $q_g(k)$ in this approach. The distribution $q_g(k)$ describes the contributions of square-tiled surfaces of g distinguished topological types (corresponding to the situation when all conical singularities are located at same horizontal layer and the surface has $k = 1, \dots, g$ maximal horizontal cylinders), where, moreover, we replace the Witten's correlators with the corresponding asymptotic expressions. The precise definition of $q_g(k)$ is given in Eq. (3.17) in Sect. 3.1. Informally, our conditional asymptotic result in [26] stated that for large genera g the auxiliary distribution $q_g(k)$ well-approximates the original probability distribution $p_g(k)$ modulo the conjectures mentioned above.

Deep analysis of volume contributions of square-tiled surfaces of different topological types was performed by Aggarwal in [3]. Moreover, in the same paper A. Aggarwal established uniform asymptotic bounds for Witten's correlators using elegant approach through asymmetric random walk. In particular, he proved all conjectures from [26] (in a stronger form) transforming conditional statements from [26] into proven results.

In the present paper we follow the original approach, approximating the probability distribution $p_g(k)$ with a slight modification of the probability distribution $q_g(k)$. The fine asymptotic analysis of A. Aggarwal allows to state that $q_g(k)$ "well-approximates" $p_g(k)$ in a much stronger sense than it was claimed in the original preprint [26]. Moreover, we realized that our "slight modification of the probability distribution $q_g(k)$" has combinatorial interpretation of independent interest and admits a detailed description based on the technique developed by Hwang in [43].

Having explained the scheme of our approach we can state now the main results concerning the probability distribution $p_g(k)$. We start with a formal definition of the "slight modification of the probability distribution $q_g(k)$" through random permutations. It plays an important role in the present paper.

Non-uniform random permutations and distribution $q_{n,\infty,1/2}$. Let θ be a sequence $\{\theta_k\}_{k\geq 1}$ of positive real numbers. Given a permutation $\sigma \in S_n$ with cycle type $(1^{\mu_1} 2^{\mu_2} \dots n^{\mu_n})$, where $1 \cdot \mu_1 + 2 \cdot \mu_2 + \dots + n \cdot \mu_n = n$, we define its *weight* $w_\theta(\sigma)$ by the following formula:

$$w_\theta(\sigma) = \theta_1^{\mu_1}\theta_2^{\mu_2}\cdots\theta_n^{\mu_n}.$$

To every collection of positive numbers $\theta = \{\theta_k\}_{k\geq 1}$, we associate a probability measure on the sym-met-ric group S_n by means of the weight function defined above:

$$\mathbb{P}_{\theta,n}(\sigma) := \frac{w_\theta(\sigma)}{n!\cdot W_{\theta,n}}, \quad \text{where} \quad W_{\theta,n} := \frac{1}{n!}\sum_{\sigma\in S_n} w_\theta(\sigma). \tag{1.7}$$

Denote by $\mathbb{P}_{n,\infty,1/2}$ the non-uniform probability measure on the symmetric group S_n associated to the collection of positive numbers $\theta_k = \zeta(2k)/2$, where $k = 1, 2, \ldots$ and ζ is the Riemann zeta function. Consider the random variable $\mathrm{K}_n(\sigma)$ on the symmetric group S_n endowed with the probability measure $\mathbb{P}_{n,\infty,1/2}$, where $\mathrm{K}_n(\sigma)$ is the number of disjoint cycles in the cycle decomposition of such a random permutation σ. The random variable $\mathrm{K}_n(\sigma)$ takes integer values in the range $[1, n]$. We introduce the notation:

$$q_{n,\infty,1/2}(k) = \mathbb{P}_{n,\infty,1/2}\big(\mathrm{K}_n(\sigma) = k\big) \tag{1.8}$$

for the distribution law of the random variable $\mathrm{K}_n(\sigma)$ with respect to the probability measure $\mathbb{P}_{n,\infty,1/2}$. We prove in Sect. 3 a series of results which can informally be summarized as follows: *the probability distribution* $q_{3g-3,\infty,1/2}$ *well-approximates the probability distribution* q_g. We admit that the approximating distribution q_g will be formally defined only later, namely, in Eq. (3.17) in Sect. 3.1 and, strictly speaking, would not be used explicitly. The above claim explains, however, our interest for the probability distribution $q_{3g-3,\infty,1/2}$ which would be actually used for approximation. An important step of comparison of distributions p_g and $q_{3g-3,\infty,1/2}$ is established in Lemma 3.6 stated and proved in Sect. 3.2. Theorems 1.8 and 1.10 below carry comprehensive information on the probability distribution $q_{3g-3,\infty,1/2}(k) = \mathbb{P}_{3g-3,\infty,1/2}\big(\mathrm{K}_{3g-3}(\sigma) = k\big)$.

Theorem 1.8 *Let* $\mathbb{P}_{n,\infty,1/2}$ *be the probability distribution on* S_n *associated to the collection* $\theta_k = \zeta(2k)/2$. *Then for all* $t \in \mathbb{C}$ *we have as* $n \to +\infty$

$$\mathbb{E}_{n,\infty,1/2}\left(t^{\mathrm{K}_n}\right) = (2n)^{\frac{t-1}{2}} \cdot \frac{t\cdot\Gamma(\frac{3}{2})}{\Gamma(1+\frac{t}{2})}\left(1 + O\left(\frac{1}{n}\right)\right), \tag{1.9}$$

where the error term is uniform in t *on any compact subset of* $\mathbb{C}$.

For any $\lambda > 0$ and $k \in \mathbb{N}$, let $u_{\lambda,1/2}(k)$ be the coefficients of the following Taylor expansion

$$e^{\lambda(t-1)} \cdot \frac{t \cdot \Gamma\left(\frac{3}{2}\right)}{\Gamma\left(1+\frac{t}{2}\right)} = \sum_{k \geq 1} u_{\lambda,1/2}(k) \cdot t^k. \tag{1.10}$$

Recall that $\Gamma\left(\frac{3}{2}\right) = \frac{\Gamma\left(\frac{1}{2}\right)}{2} = \frac{\sqrt{\pi}}{2}$. We have

$$u_{\lambda,1/2}(k) = \sqrt{\pi} \cdot e^{-\lambda} \cdot \frac{1}{k!} \cdot \sum_{i=1}^{k} \binom{k}{i} \cdot \phi_i \cdot \left(\frac{1}{2}\right)^i \cdot \lambda^{k-i},$$

where ϕ_j is defined by the Taylor expansion

$$\frac{t}{\Gamma(1+t)} = \frac{1}{\Gamma(t)} = \sum_{j=1}^{+\infty} \phi_j \cdot \frac{t^j}{j!}. \tag{1.11}$$

The first few values of ϕ_j are given by

$$\phi_1 = 1; \quad \phi_2 = 2\gamma; \quad \phi_3 = 3\big(\gamma^2 - \zeta(2)\big),$$

where γ is the Euler–Mascheroni constant. Theorem 1.8 has the following consequence.

Corollary 1.9 *For* $\lambda_n = \frac{\log(2n)}{2}$ *we have*

$$q_{n,\infty,1/2}(k) = u_{\lambda_n,1/2}(k) + O\left(\frac{1}{n}\right) \text{ as } n \to \infty$$

uniformly in $k \geq 1$.

Theorem 1.8 and Corollary 1.9 are particular cases of, respectively, Corollary 3.11 and Corollary 3.12 stated and proved in Sect. 3.3. We also illustrate the numerical aspects of this approximation in Sect. 6.1.

Theorem 1.10 *Let* $\lambda_n = \log(2n)/2$. *Then, for any* $x > 0$, *we have the following asymptotic behavior that is uniform in* $k \in [0, x\lambda_n]$ *as* $n \to +\infty$:

$$\begin{aligned} q_{n,\infty,1/2}(k+1) &= \mathbb{P}_{n,\infty,1/2}(\mathrm{K}_n(\sigma) = k+1) \\ &= e^{-\lambda_n} \cdot \frac{(\lambda_n)^k}{k!} \cdot \left(\frac{\sqrt{\pi}}{2\Gamma\left(1+\frac{k}{2\lambda_n}\right)} + O\left(\frac{k+1}{(\log n)^2}\right)\right). \end{aligned} \tag{1.12}$$

For any $x > 1$ *such that* $x\lambda_n$ *is an integer we have*

$$\sum_{k=x\lambda_n+1}^{n} q_{n,\infty,1/2}(k+1) = \mathbb{P}_{n,\infty,1/2}\big(\mathrm{K}_n(\sigma) > x\lambda_n + 1\big)$$
$$= \frac{(2n)^{-\frac{x\log x - x + 1}{2}}}{\sqrt{2\pi\lambda_n x}} \cdot \frac{x}{x-1} \cdot \left(\frac{\sqrt{\pi}}{2\Gamma\left(1+\frac{x}{2}\right)} + O\left(\frac{1}{\log n}\right)\right), \quad (1.13)$$

where the error term is uniform in x *on compact subsets of* $(1, +\infty)$. *Similarly, for any* $0 < x < 1$ *such that* $x\lambda_n$ *is an integer we have*

$$\sum_{k=0}^{x\lambda_n} q_{n,\infty,1/2}(k+1) = \mathbb{P}_{n,\infty,1/2}\big(\mathrm{K}_n(\sigma) \le x\lambda_n + 1\big)$$
$$= \frac{(2n)^{-\frac{x\log x - x + 1}{2}}}{\sqrt{2\pi\lambda_n x}} \cdot \frac{x}{1-x} \cdot \left(\frac{\sqrt{\pi}}{2\Gamma\left(1+\frac{x}{2}\right)} + O\left(\frac{1}{\log n}\right)\right), \quad (1.14)$$

where the error term is uniform in x *on compact subsets of* $(0, 1)$.

Theorem 1.10 is a particular case of Corollary 3.17 stated and proved in Sect. 3.4. Note that for $x \neq 1$, we have $x\log x - x + 1 > 0$. Hence, Eqs. (1.13) and (1.14) provide explicit polynomial bounds in n for the tails of the distribution.

Remark 1.11 Let $G(x) = \dfrac{\sqrt{\pi}}{2\Gamma(1+\frac{x}{2})}$ and define

$$\mathrm{a}(x) = \frac{\log G(x)}{x-1}. \quad (1.15)$$

Since $\log G(1) = 0$, the function $\mathrm{a}(x)$ admits a continuous extension at $x = 1$

$$\lim_{x\to 1} \mathrm{a}(x) = G'(1) = \frac{\gamma}{2} + \log 2 - 1,$$

where γ is the Euler–Mascheroni constant. Now, for any $x > 0$, we have

$$(\lambda_n)^k \cdot G\left(\frac{k}{\lambda_n}\right) = e^{-\mathrm{a}\left(\frac{k}{\lambda_n}\right)}\left(\lambda_n + \mathrm{a}\left(\tfrac{k}{\lambda_n}\right)\right)^k \cdot \left(1 + O\left(\frac{k}{\lambda_n^2}\right)\right)$$

uniformly in $k \in [0, x\lambda_n]$ as $n \to +\infty$. We can, hence, rewrite the right-hand side of (1.12), namely, for any $x > 0$ we have the following asymptotic behavior:

$$q_{n,\infty,1/2}(k+1) = e^{-\left(\lambda_n + \mathrm{a}\left(\frac{k}{\lambda_n}\right)\right)} \cdot \frac{\left(\lambda_n + \mathrm{a}\left(\frac{k}{\lambda_n}\right)\right)^k}{k!} \cdot \left(1 + O\left(\frac{k}{(\log n)^2}\right)\right)$$

uniformly in $k \in [0, x\lambda_n]$ as $n \to +\infty$. In the latter expression, the right-hand side reads as the value of a Poisson random variable with parameter $\lambda_n + \mathrm{a}\left(\frac{k}{\lambda_n}\right)$.

The extended version of the above results as well as the closely related notion of mod-Poisson convergence are discussed in Sect. 3.2. The above theorems follow from singularity analysis at the boundary of the domain of definition of holomorphic functions representing the relevant generating series performed by Hwang in [43].

Properties of the probability distribution $p_g(k)$. The key theorems below strongly rely on the asymptotic analysis of the Masur–Veech volumes of the moduli spaces of quadratic differentials performed by Aggarwal in [3] and on the uniform asymptotic bounds for Witten's correlators obtained in the same paper.

Theorem 1.12 *Let K_g be the random variable satisfying the probability law* (1.6). *For all $t \in \mathbb{C}$ such that $|t| < \frac{8}{7}$ the following asymptotic relation is valid as $g \to +\infty$:*

$$\mathbb{E}\left(t^{K_g}\right) = (6g-6)^{\frac{t-1}{2}} \cdot \frac{\Gamma(\frac{1}{2})}{\Gamma(\frac{t}{2})}\, (1 + o(1))\,. \tag{1.16}$$

Moreover, for any compact set U in the open disk $|t| < \frac{8}{7}$ there exists $\delta(U) > 0$, such that for all $t \in U$ the error term has the form $O(g^{-\delta(U)})$.

Note that the right-hand side of expression (1.16) is very close to the right-hand side of the analogous expression (1.9) from Theorem 1.8 evaluated at $n = 3g - 3$.

We expect that the mod-Poisson convergence (1.16) holds in a larger domain than the disk $|t| < \frac{8}{7}$. If our guess is correct, the asymptotics (1.17) for the distribution p_g should hold for larger interval of x than described below. We also expect that the mod-Poisson convergence analogous to (1.16) holds for all non-hyperelliptic components of all strata of holomorphic quadratic differentials; see Conjecture 2 in Sect. 6.2 for more details.

Theorem 1.13 *Let* $\lambda_{3g-3} = \log(6g-6)/2$. *For any* $x \in \left[0, \frac{1}{\log\frac{9}{4}}\right)$ *we have*

$$\begin{aligned} p_g(k+1) &= \mathbb{P}\big(K_g(\gamma) = k+1\big) \\ &= e^{-\lambda_{3g-3}} \cdot \frac{\lambda_{3g-3}^k}{k!} \cdot \left(\frac{\sqrt{\pi}}{2\Gamma\left(1 + \frac{k}{2\lambda_{3g-3}}\right)} + O\left(\frac{k+1}{(\log g)^2}\right)\right) \end{aligned} \tag{1.17}$$

uniformly in $k \in [0, x\lambda_{3g-3}]$.

For any $x \in (1, 1.236\,]$ *such that* $x\lambda_{3g-3}$ *is an integer we have*

$$\begin{aligned} &\sum_{k=x\lambda_{3g-3}+1}^{3g-3} p_g(k+1) = \mathbb{P}\big(K_g(\gamma) > x\lambda_{3g-3} + 1\big) \\ &= \frac{(6g-6)^{-\frac{x\log x - x + 1}{2}}}{\sqrt{2\pi\lambda_{3g-3}x}} \cdot \frac{x}{x-1} \cdot \left(\frac{\sqrt{\pi}}{2\Gamma\left(1+\frac{x}{2}\right)} + O\left(\frac{1}{\log g}\right)\right), \end{aligned} \tag{1.18}$$

where the error term is uniform in x on compact subsets of $(1, 1.236\,]$. *Similarly for any* $x \in (0, 1)$ *such that* $x\lambda_{3g-3}$ *is an integer we have*

$$\begin{aligned} &\sum_{k=0}^{x\lambda_{3g-3}} p_g(k+1) = \mathbb{P}\big(K_g(\gamma) \le x\lambda_{3g-3} + 1\big) \\ &= \frac{(6g-6)^{-\frac{x\log x - x + 1}{2}}}{\sqrt{2\pi\lambda_{3g-3}x}} \cdot \frac{x}{1-x} \cdot \left(\frac{\sqrt{\pi}}{2\Gamma\left(1+\frac{x}{2}\right)} + O\left(\frac{1}{\log g}\right)\right), \end{aligned} \tag{1.19}$$

where the error term is uniform in x on compact subsets of $(0, 1)$. *Finally,*

$$\begin{aligned} \sum_{k=\lfloor 0.09\log g\rfloor}^{\lceil 0.62\log g\rceil} p_g(k) &= \mathbb{P}\Big(0.09\log g < K_g(\gamma) < 0.62\log g\Big) \\ &= 1 - O\left((\log g)^{24} g^{-1/4}\right). \end{aligned} \tag{1.20}$$

Similarly to Remark 1.11, Eq. (1.17) tells, in particular, that any x in the interval $[0, 1.236]$ (which carries, essentially, all but $O(g^{-1/4})$ part of the total mass of the distribution) and for large g, the values $p_g(k+1)$ for k in a neighborhood of $x\frac{\log g}{2}$ of size $o(\log g)$ are uniformly well-approximated by the Poisson distribution $\mathrm{Poi}_\lambda(k)$ with parameter $\lambda = \frac{\log(6g-6)}{2} + \mathrm{a}(x)$, where $\mathrm{a}(x)$ is defined in (1.15).

The approximation results given in Theorem 1.8 for $q_{n,\infty,1/2}$ and in Theorem 1.12 for p_g imply an asymptotic expansion of the moments that we present now. Recall that the *Stirling number of the second kind*, denoted by $S(i, j)$, is the number of ways to partition a set of i objects into j non-empty subsets.

Theorem 1.14 *For any fixed $k \in \mathbb{N}$ the difference between the i-th moments of random variables with the probability distributions p_g and $q_{3g-3,\infty,1/2}$ tends to zero as $g \to +\infty$.*

Furthermore, the i-th cumulant $\kappa_i(K_g(\sigma))$ of the random variable K_g associated to the probability distribution p_g admits the following asymptotic expansion:

$$\kappa_i(K_g) = \frac{\log(6g-6)}{2} + \frac{\gamma}{2} + \log 2 - \sum_{j=2}^{i} S(i, j) \cdot (-1)^j \cdot \zeta(j) \cdot (j-1)! \cdot \left(2^j - 1\right) \cdot \left(\frac{1}{2}\right)^j + O\left(\frac{1}{g}\right) \text{ as } g \to +\infty, \tag{1.21}$$

where $S(i, j)$ are the Stirling numbers of the second kind. In particular, the mean value $\mathbb{E}(K_g)$ and the variance $\mathbb{V}(K_g)$ satisfy

$$\mathbb{E}(K_g) = \kappa_1(K_g) = \frac{\log(6g-6)}{2} + \frac{\gamma}{2} + \log 2 + O\left(\frac{1}{g}\right),$$
$$\mathbb{V}(K_g) = \kappa_2(K_g) = \frac{\log(6g-6)}{2} + \frac{\gamma}{2} + \log 2 - \frac{3}{4}\zeta(2) + O\left(\frac{1}{g}\right),$$

where $\gamma = 0.5572\ldots$ denotes the Euler–Mascheroni constant. The third and the fourth cumulants $\kappa_3(K_g)$ and $\kappa_4(K_g)$ admit the following asymptotic expansions:

$$\kappa_3(K_g) = \frac{\log(6g-6)}{2} + \frac{\gamma}{2} + \log 2 - \frac{9}{4}\zeta(2) + \frac{7}{4}\zeta(3) + O\left(\frac{1}{g}\right),$$
$$\kappa_4(K_g) = \frac{\log(6g-6)}{2} + \frac{\gamma}{2} + \log 2 - \frac{21}{4}\zeta(2) + \frac{21}{2}\zeta(3) - \frac{45}{8}\zeta(4) + O\left(\frac{1}{g}\right).$$

Other approaches to random multicurves. One more interesting aspect of geometry of random multicurves is the lengths statistics of simple closed hyperbolic geodesics associated to components of multicurves of fixed topological type. Mirzakhani studied in [65] random pants decompositions of a hyperbolic surface of genus g. She considered the orbit $\mathrm{Mod}_g\,\gamma$ of a multicurve $\gamma = \gamma_1 + \cdots + \gamma_{3g-3}$ corresponding to a fixed pants decomposition. Choosing multicurves of hyperbolic length at most L in this orbit, she got a finite collection of multicurves. Letting $L \to +\infty$ she defined a *random pants decomposition*. M. Mirzakhani proved in Theorem 1.2 of [65] that under the normalization $x_i = \frac{\ell(f\cdot\gamma_i)}{\ell(f\cdot\gamma)}$ for $i = 1, \ldots, 3g-3$, and $f \in \mathrm{Mod}_g$, the lengths statistics of components of a random pair of pants $f\cdot\gamma$ has the limiting density function $const \cdot x_1 \ldots x_{3g-3}$ with respect to the Lebesgue measure on the unit simplex. F. Arana-Herrera and M. Liu independently proved in [9,11,54] a generalization of this result to arbitrary multicurves. In terms of square-tiled surfaces the resulting hyperbolic lengths statistics coincides with statistics of flat lengths of the waist curves of maximal horizontal cylinders of the square-tiled surface (see Section 1.9 in [29]).

One can define unconstrained random integral multicurves (considering all integral multicurves of bounded length at once) in an analogous way. In this setting the number of components of $f\cdot\gamma$ is not fixed anymore, so it is natural to list the normalized lengths of components $\frac{\ell(f\cdot\gamma_i)}{\ell(f\cdot\gamma)}$ in a decreasing order completing them with an infinite sequence of zeroes. M. Liu and V. Delecroix proved recently in [30] that in the large genus limit, the resulting vector converges (in distribution) to the asymptotic vector of normalized and ordered lengths of cycles in a cyclic decomposition of a random permutation of n elements under the probability measure $\mathbb{P}_{n,\infty,1/2}$ when $n \to +\infty$.

In the regime where one considers simple closed curves of length at most L for any fixed $L > 0$ and lets the genus tend to $+\infty$, a very precise description of the distribution of lengths was provided by Mirzakhani and Petri in [66]. A similar result but with discrete metrics instead of hyperbolic ones has been proved by Janson and Louf in [46].

It would be interesting to establish relations between random multicurves and the general framework of random partitions introduced by Vershik in [82].

Random quadrangulations versus random square-tiled surfaces. In this article we are concerned with random square-tiled surfaces, which are a particular case of random quadrangulations, which are themselves a particular case of random combinatorial maps (surfaces obtained from gluing polygons). The latter two families have a much longer mathematical history. The two important parameters are the number of polygons N and the genus g.

Surfaces obtained by random gluing of polygons have been studied for a long time. Their enumeration can be traced back to the works of Tutte [80] for $g = 0$ and of Walsh and Lehman [83] for arbitrary g. In particular, their results allow to compute the probability of getting a closed surface of genus g as a result of a random pairwise gluing of the sides of a $2n$-gon. Somewhat later Harer and Zagier [41] were able to enumerate genus g gluings of a $2n$-gon in a more explicit and effective way. This was a crucial ingredient in their computation of the orbifold Euler characteristic of the moduli space $\mathcal{M}_g$ of complex algebraic curves.

Surfaces obtained from randomly glued polygons have been studied since a long time in physics in relation to string theory and quantum gravity as in the paper of Kazakov et al. [47]. In this approach the main attention is paid to surfaces of genus zero (planar approximation) with several perturbative terms corresponding to surfaces of low genera.

In the case $g = 0$ and $N \to +\infty$, the Brownian map has been shown to be the scaling limit of various models of combinatorial maps, see the surveys of Miermont [60] and of Le Gall [52] and the references therein. Combinatorial maps also admit local limits, as proved, in particular, in the papers of Angel and Schramm [5], of Krikun [49], of Chassaing and Durhuus [20], of Ménard [59]. In higher but fixed genus, the scaling limits giving rise to higher genera Brownian maps have been investigated by Bettinelli in [13,14].

Surfaces obtained by gluing polygons without restriction on the genus have been studied by Brooks and Makover in [15], by Gamburd in [36], by Guth, Parlier ang Young in [40], by Chmutov and Pittel in [24], by Alexeev and Zograf in [7], and by Budzinski, Curien and Petri in [16,17]. In this approach the genus g of the resulting surface is a random variable whose expectation is proportional to the number of polygons N. See also the recent paper of Shresta [77] studying square-tiled surfaces in a similar context. Finally, in the regime $g = \theta N$ with $\theta \in [0, \frac{1}{2})$ a local limit has been conjectured by Curien in [25] and recently proved by Budzinski and Louf in [18], see also Louf [55].

Note that our approach is different from all approaches mentioned above. We fix the genus of the surface, and consider square-tiled surfaces build from at most N squares (or geodesic multicurves of length bounded by some large number L). We define asymptotic frequencies of square-tiled surfaces or of geodesic multicurves of a fixed combinatorial type by passing to the limit when N (respectively L) tends to infinity. Only when the resulting limiting frequencies (probabilities) are already defined in each individual genus we study their behavior in the regime when the genus becomes very large. This approach is natural in the context of dynamics of polygonal billiards, dynamics of interval exchange transformations and of translation surfaces, and in the context of geometry and dynamics on the moduli space of quadratic differentials.

Note also that all but negligible part of our square-tiled surfaces of genus g have $4g - 4$ vertices of valence 6, while all other vertices have valence 4, and the number of such vertices is incomparably larger than g. This is one more substantial difference between our random surface model and the random quadrangulations considered in the probability theory literature where, usually, there is no such degree constraint imposed and vertices, typically, have arbitrary degrees even if the resulting surface has genus 0. As a result, our square-tiled surfaces locally look like a tiling of $\mathbb{R}^2$ by squares except around $4g - 4$ conical singularities with cone angle 3π. This is not the case for a random planar quadrangulation.

A regime similar to ours was used by H. Masur, K. Rafi and A. Randecker who studied in [58] the covering radius of random translation surfaces (corresponding to Abelian differentials).

In the hyperbolic setting, another regime similar to ours was studied by M. Mirzakhani and by M. Mirzakhani and P. Zograf in [67]. M. Mirzakhani considered in [64] random hyperbolic surfaces of fixed large genus g, where "randomness" is defined by means of the Weil–Petersson measure. In particular, M. Mirzakhani proved that with probability which tends to 1 when $g \to \infty$, the spectrum of Laplacian of a random hyperbolic surface admits a uniform spectral gap (i.e., it is separated from zero). The pioneering ideas of M. Mirzakhani had very successful development in recent works of Anantharaman, Monk, and Thomas [4,68,69]. The lower bound for the spectral gap obtained by M. Mirzakhani was improved in independent recent papers of Lipnowski and Wright [53] and of Wu and Xue [85] to $\frac{3}{16} - \varepsilon$. Finally, the recent papers of W. Hide, Magee, Naud and Puder [42,56] investigated random covers of growing genus. In this settings Hide and Magee proved in [42] the optimal asymptotic spectral gap $\frac{1}{4} - \varepsilon$ awaited since more than half of a century.

Structure of the paper. To make the present paper self-contained, we reproduce in Sect. 2 all necessary background material. We start by recalling in Sect. 2.1 the definition of the Masur–Veech volume of the moduli space of quadratic differentials $\mathcal{Q}_g$. We sketch in Sect. 2.2 how Masur–Veech volumes are related to the count of square-tiled surfaces. In the same section we associate to every square-tiled surface a multicurve and we recall the notion of a stable graph, particularly important in the framework of the present paper. We present in Sect. 2.3 a formula for the Masur–Veech volume $\operatorname{Vol} \mathcal{Q}_g$ and the theorem of A. Aggarwal on the asymptotic value of this volume for large genus g. The reader interested in more ample information is addressed to the original papers [3,29] respectively. In Sect. 2.4 we recall Mirzakhani's count [62] of frequencies of multicurves. In Sect. 2.5 we explain why Questions 1 and 2 are equivalent and demystify Theorem 1.7. In Sect. 2.6 we recall the recent breakthrough results of Aggarwal [3] on large genus asymptotics of Witten's correlators.

In Sect. 3 we recall general background from the works of Hwang [43], and of Kowalski, Méliot, Nikeghbali, Zeindler [34,51,71] on random permutations and on mod-Poisson convergence and apply this general technique to the probability distribution $q_{3g-3,\infty,1/2}$. In particular, we prove Theorems 1.8 and 1.10.

We then introduce a probability distribution $p_g^{(1)}(k)$ of the random variable $K_g(\gamma) = K_g(S)$ restricted to non-separating random multicurves γ on a surface of genus g (equivalently, restricted to random square-tiled surfaces of genus g having a single horizontal critical level). Using the results of Aggarwal [3] on asymptotics of Witten's correlators we prove that the distribution $q_{3g-3,\infty,1/2}$ very well-approximates the distribution $p_g^{(1)}$ (namely, that they share the same mod-Poisson convergence, but $p_g^{(1)}$ has smaller radius of convergence). This allows us to extend all the results obtained for random permutations to these special random multicurves (special random square-tiled surfaces).

It remains, however, to pass from the special multicurves (and square-tiled surfaces) to the general ones. The necessary estimates are prepared in Sect. 4. In a sense, this step was already performed by Aggarwal in [3], who proved a generalization of our conjecture from [26] claiming that random multicurves (random square-tiled surfaces) which do not contribute to the distribution $p_g^{(1)}$ become rare for large genera. This justifies the fact that the distribution $p_g^{(1)}$ well-approximates the distribution p_g. However, to prove this statement in a much stronger form stated in the present paper we have to adjust certain estimates from Sections 9 and 10 from the original paper [3] to our current needs.

We recommend the readers interested in all the details of Sect. 4 to read it in parallel with Sections 9 and 10 of the original paper [3]. (Actually, we recommend reading the entire paper [3] of A. Aggarwal. We have no doubt that the reader looking for a deep understanding of the subject will appreciate the beauty, strength and originality of the ideas and proofs in [3] as we do.)

Having obtained all the necessary estimates in Sect. 4 we prove in Sect. 5 that the distribution $p_g^{(1)}$ well-approximates the distribution p_g. By transitivity this implies that the distribution $q_{3g-3,\infty,1/2}$ well-approximates the distribution p_g. We show in Sect. 5 how the properties of $q_{3g-3,\infty,1/2}$ derived in Sect. 3 imply all our main results.

In Sect. 6.1 we compare our theoretical results with experimental and exact numerical data. We complete by suggesting in Sect. 6.2 a conjectural description of the combinatorial geometry of random Abelian square-tiled surfaces of large genus and of random square-tiled surfaces restricted to any non-hyperelliptic component of any stratum in the moduli space of Abelian or quadratic differentials of large genus.

This article is born out of Appendices D–F of the original preprint [26]. The latter contained several conjectures, and all other results were derived from them as "conditional theorems". All these conjectures were proved by

A. Aggarwal; see Theorems 2.3, 2.6, 2.7, 2.8, and Corollary 5.3 in the present paper or, respectively, Theorem 1.7 and Propositions 1.2, 4.1, 4.2, 10.7 in the original paper [3]. Moreover, most of the results are proved in [3] in a much stronger form than we initially conjectured. Combining our initial approach with these recent results of A. Aggarwal and elaborating close ties with random permutations we radically strengthened the initial assertions from [26].

2 Background material

2.1 Masur–Veech volume of the moduli space of quadratic differentials

Consider the moduli space $\mathcal{M}_{g,n}$ of complex curves of genus g with n distinct labeled marked points. The total space $\mathcal{Q}_{g,n}$ of the cotangent bundle over $\mathcal{M}_{g,n}$ can be identified with the moduli space of pairs (C, q), where $C \in \mathcal{M}_{g,n}$ is a smooth complex curve with n (labeled) marked points and q is a meromorphic quadratic differential on C with at most simple poles at the marked points and no other poles. In the case $n = 0$ the quadratic differential q is holomorphic. Thus, the *moduli space of quadratic differentials* $\mathcal{Q}_{g,n}$ is endowed with the canonical real symplectic structure. The induced volume element $d\mathrm{Vol}$ on $\mathcal{Q}_{g,n}$ is called the *Masur–Veech volume element*. In the next Section we provide alternative more common definition of the Masur–Veech volume element. The two definitions are equivalent up to a global multiplicative constant.

A non-zero quadratic differential q in $\mathcal{Q}_{g,n}$ defines a flat metric $|q|$ on the complex curve C. The resulting metric has conical singularities at zeroes and simple poles of q. The total area of (C, q)

$$\mathrm{Area}(C, q) = \int_C |q|$$

is positive and finite. For any real $a > 0$, consider the following subset in $\mathcal{Q}_{g,n}$:

$$\mathcal{Q}_{g,n}^{\mathrm{Area}\le a} := \left\{(C, q) \in \mathcal{Q}_{g,n} \mid \mathrm{Area}(C, q) \le a\right\}.$$

Since $\mathrm{Area}(C, q)$ is a norm in each fiber of the bundle $\mathcal{Q}_{g,n} \to \mathcal{M}_{g,n}$, the set $\mathcal{Q}_{g,n}^{\mathrm{Area}\le a}$ is a ball bundle over $\mathcal{M}_{g,n}$. In particular, it is non-compact. However, by the independent results of Masur [57] and Veech [81], the total mass of $\mathcal{Q}_{g,n}^{\mathrm{Area}\le a}$ with respect to the Masur–Veech volume element is finite.

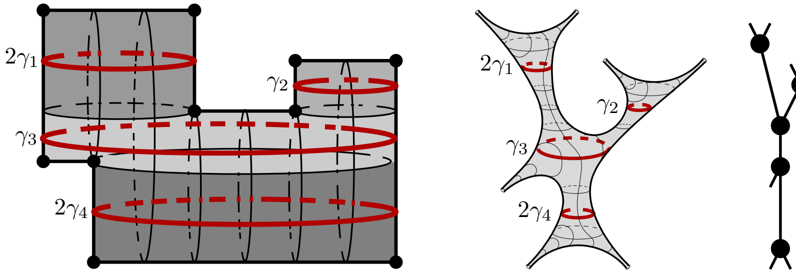

Fig. 2 Square-tiled surface in $\mathcal{Q}_{0,7}$, and associated multicurve and stable graph

2.2 Square-tiled surfaces, simple closed multicurves and stable graphs

We have already mentioned that a non-zero meromorphic quadratic differential q on a complex curve C defines a flat metric with conical singularities. One can construct a discrete collection of quadratic differentials of this kind by assembling together identical flat squares in the following way. Take a finite set of copies of the oriented $1/2 \times 1/2$-square for which two opposite sides are chosen to be horizontal and the remaining two sides are declared to be vertical. Identify pairs of sides of the squares by isometries in such way that horizontal sides are glued to horizontal ones and vertical sides to vertical ones. We get a topological surface S without boundary. We consider only those surfaces obtained in this way which are connected and oriented. The form dz^2 on each square is compatible with the gluing and endows S with a complex structure and with a non-zero quadratic differential q with at most simple poles. The total area $\text{Area}(S, q)$ is $\frac{1}{4}$ times the number of squares. We call such surface a *square-tiled surface*.

Suppose that the resulting closed *square-tiled surface* has genus g and n conical singularities of angle π, i.e. n vertices shared by only two squares. For example, the square-tiled surfaces in Fig. 2 has genus $g = 0$ and $n = 7$ conical singularities of angle π. Consider the complex coordinate z in each square and a quadratic differential $(dz)^2$. It is easy to check that the resulting square-tiled surface inherits the complex structure and globally defined meromorphic quadratic differential q having simple poles at n conical singularities of angle π and no other poles. Thus, any square-tiled surface of genus g having n conical singularities of angle π canonically defines a point $(C, q) \in \mathcal{Q}_{g,n}$. Fixing the size of the square once and forever and considering all resulting square-tiled surfaces in $\mathcal{Q}_{g,n}$ we get a discrete subset $\mathcal{ST}_{g,n}$ in $\mathcal{Q}_{g,n}$.

Given a sequence of integers $\mu = [\mu_1 \dots \mu_m, \mu_{m+1} \dots \mu_{m+n}]$, where $[\mu_1 \dots \mu_m]$ is a partition of $4g - 4 + n$ and $\mu_{m+1} = \cdots = \mu_{n+m} = -1$, the corresponding *stratum of quadratic differentials* $\mathcal{Q}(\mu)$ is the space of equivalence classes of pairs consisting of a complex curve C with $m + n$ distinct marked points $z_1, \ldots, z_m, p_1, \ldots, p_n$ and a quadratic differential q with the divisor $\sum_{i=1}^{m} \mu_i z_i - \sum_{j=1}^{n} p_j$ (both zeroes and poles of q are considered to be labeled).

For any pair of nonnegative integers (g, n) satisfying $2g + n > 3$, the *principal stratum* is $\mathcal{Q}(1^{4g-4+n}, -1^n)$ (that is, $\mu = [1^{4g-4+n}, -1^n]$). For example, the square-tiled surface in Fig. 2 belongs to the principal stratum $\mathcal{Q}(1^3, -1^7)$. The natural morphism $\mathcal{Q}(1^{4g-4+n}, -1^n) \to \mathcal{Q}_{g,n}$ that forgets the labeling of zeroes of q is a $(4g-4+n)!$-sheeted ramified cover of its image in $\mathcal{Q}_{g,n}$. Moreover, this image is open and dense in $\mathcal{Q}_{g,n}$.

In the two special cases $(g, n) = (0, 3)$ and $(g, n) = (1, 1)$ which correspond to $2g + n = 3$ the moduli space $\mathcal{Q}_{g,n}$ does not admit any natural interpretation in terms of meromorphic quadratic differentials with simple zeros and simple poles.

Denote by $\mathcal{ST}_{g,n}(N) \subset \mathcal{ST}_{g,n}$ the subset of square-tiled surfaces in $\mathcal{Q}_{g,n}$ made of at most N identical squares. The strata have a natural linear structure and the square-tiled surfaces form a covolume one lattice in associated period coordinates in every stratum. This justifies the following conventional definition of the Masur–Veech volume of $\mathcal{Q}_{g,n}$ (for (g, n) different from $(0, 3)$ and $(1, 1)$):

$$\operatorname{Vol} \mathcal{Q}_{g,n} := \operatorname{Vol} \mathcal{Q}(1^{4g-4+n}, -1^n) = 2d \cdot \lim_{N \to +\infty} \frac{\operatorname{card}\left(\mathcal{ST}_{g,n}(2N)\right)}{N^d}, \quad (2.1)$$

where $d = \dim_{\mathbb{C}} \mathcal{Q}(1^{4g-4+n}, -1^n) = \dim_{\mathbb{C}} \mathcal{Q}_{g,n} = 6g - 6 + 2n$. We emphasize that in the above formula we assume that all conical singularities of square-tiled surfaces are labeled.

The cardinality of the subset of square-tiled surfaces in $\mathcal{ST}_{g,n}(2N)$ which belong to strata different from the principal one is negligible as $n \to +\infty$, so restricting the count to square-tiled surfaces in the principal stratum $\mathcal{Q}(1^{4g-4+n}, -1^n)$ does not change the above limit.

We admit that certain conventions used in the definition (2.1) might seem unexpected. For example, the square-tiled surfaces in $\mathcal{ST}_{g,n}(2N)$ are made of at most $2N$ squares, while we normalize the cardinality of this set by N^d. Also, as we already mentioned, the principal stratum $\mathcal{Q}(1^{4g-4+n}, -1^n)$ is a $(4g-4+n)!$-sheeted cover over an open and dense subspace in $\mathcal{Q}_{g,n}$. However, the normalization in (2.1) follows the one used in the literature including [3,6,12,21,29,38].

Multicurve associated to a cylinder decomposition. Any square-tiled surface admits a decomposition into maximal horizontal cylinders filled with isometric simple closed flat geodesics. Every such maximal horizontal cylinder has at least one conical singularity on each of the two boundary components. The square-tiled surface in Fig. 2 has four maximal horizontal cylinders which are represented in the picture by different shades. For every maximal horizontal cylinder choose the corresponding waist curve γ_i.

By construction each resulting simple closed curve γ_i is non-peripheral (i.e. it does not bound a topological disk without punctures or with a single puncture) and different γ_i, γ_j are not freely homotopic on the underlying n-punctured topological surface. In other words, pinching simultaneously all waist curves γ_i we get a stable curve representing a point in the Deligne–Mumford compactification $\overline{\mathcal{M}}_{g,n}$.

We encode the number of circular horizontal bands of squares contained in the corresponding maximal horizontal cylinder by the integer weight H_i associated to the curve γ_i. The above observation implies that the resulting formal linear combination $\gamma = \sum H_i \gamma_i$ is a simple closed integral multicurve in the space $\mathcal{ML}_{g,n}(\mathbb{Z})$ of measured laminations. For example, the simple closed multicurve associated to the square-tiled surface as in Fig. 2 has the form $2\gamma_1 + \gamma_2 + \gamma_3 + 2\gamma_4$.

Given a simple closed integral multicurve γ in $\mathcal{ML}_{g,n}(\mathbb{Z})$, consider the subset $\mathcal{ST}_{g,n}(\gamma) \subset \mathcal{ST}_{g,n}$ of those square-tiled surfaces, for which the associated horizontal multicurve is in the same $\operatorname{Mod}_{g,n}$-orbit as γ, where $\operatorname{Mod}_{g,n}$ preserves the labelling of punctures. Denote by $\operatorname{Vol}(\gamma)$ the contribution to $\operatorname{Vol} \mathcal{Q}_{g,n}$ of square-tiled surfaces from the subset $\mathcal{ST}_{g,n}(\gamma) \subset \mathcal{ST}_{g,n}$:

$$\operatorname{Vol}(\gamma) = 2(6g - 6 + 2n) \cdot \lim_{N \to +\infty} \frac{\operatorname{card}(\mathcal{ST}_{g,n}(2N) \cap \mathcal{ST}_{g,n}(\gamma))}{N^d}.$$

The results in [27] imply that for any γ in $\mathcal{ML}_{g,n}(\mathbb{Z})$ the above limit exists, is strictly positive, and that

$$\operatorname{Vol} \mathcal{Q}_{g,n} = \sum_{[\gamma] \in \mathcal{O}} \operatorname{Vol}(\gamma), \tag{2.2}$$

where the sum is taken over representatives $[\gamma]$ of all orbits $\mathcal{O}$ of the mapping class group $\operatorname{Mod}_{g,n}$ in $\mathcal{ML}_{g,n}(\mathbb{Z})$.

Definition 2.1 Formula (2.2) allows to interpret the ratio $\operatorname{Vol}(\gamma)/\operatorname{Vol} \mathcal{Q}_{g,n}$ as the *asymptotic probability* to get a square-tiled surface in $\mathcal{ST}_{g,n}(\gamma)$ taking a random square-tiled surface in $\mathcal{ST}_{g,n}(N)$ as $N \to +\infty$. We will also call the same quantity by *frequency* of square-tiled surfaces of type $\mathcal{ST}_{g,n}(\gamma)$ among all square-tiled surfaces.

Stable graph associated to a multicurve Following Kontsevich [50] we assign to any multicurve γ a *stable graph* $\Gamma(\gamma) = \Gamma(\gamma_{reduced})$. The stable graph $\Gamma(\gamma)$ is a decorated graph dual to $\gamma_{reduced}$. It consists of vertices, edges, and "half-edges", also called "legs". Vertices of $\Gamma(\gamma)$ represent the connected components of the complement $S_{g,n} \setminus \gamma_{reduced}$. Each vertex is decorated with the integer number encoding the genus of the corresponding connected component of $S_{g,n} \setminus \gamma_{reduced}$. By convention, when this number is not explicitly indicated, it equals to zero. Edges of $\Gamma(\gamma)$ are in the natural bijective correspondence with curves γ_i; an edge joins a vertex to itself when on both sides of the corresponding simple closed curve γ_i we have the same connected component of $S_{g,n} \setminus \gamma_{reduced}$. Finally, the n punctures are represented by n *legs*. The right picture in Fig. 2 provides an example of the stable graph associated to the multicurve $\gamma = 2\gamma_1 + \gamma_2 + \gamma_3 + 2\gamma_4$.

Pinching all components of a reduced multicurve $\gamma_{reduced}$ on a complex curve of genus g with n marked points by we get a stable complex curve representing a point in the Deligne–Mumford compactification $\overline{\mathcal{M}}_{g,n}$. In this way stable graphs parameterize the boundary cycles of $\overline{\mathcal{M}}_{g,n}$. In particular, the set $\mathcal{G}_{g,n}$ of all stable graphs is finite. It is in the natural bijective correspondence with the set of boundary cycles of $\overline{\mathcal{M}}_{g,n}$ or, equivalently, with $\mathrm{Mod}_{g,n}$-orbits of reduced multicurves in $\mathcal{ML}_{g,n}(\mathbb{Z})$.

2.3 Formula for the Masur–Veech volumes

In this section we introduce polynomials $N_{g,n}(b_1, \ldots, b_n)$ that appear in different contexts, in particular, in the formula for the Masur–Veech volume.

Let g be a non-negative integer and n a positive integer. Let the pair (g, n) be different from $(0, 1)$ and $(0, 2)$. Let $d_1, \ldots, d_n$ be an ordered partition of $3g-3+n$ into a sum of non-negative integers, $|d| = d_1+\cdots+d_n = 3g-3+n$, let $\boldsymbol{d}$ be a multiindex $(d_1, \ldots, d_n)$ and let $b^{2\boldsymbol{d}}$ stand for $b_1^{2d_1} \cdot \cdots \cdot b_n^{2d_n}$.

Define the following homogeneous polynomial $N_{g,n}(b_1, \ldots, b_n)$ of degree $6g - 6 + 2n$ in variables $b_1, \ldots, b_n$ in the following way.

$$N_{g,n}(b_1, \ldots, b_n) = \sum_{|d|=3g-3+n} c_{\boldsymbol{d}} b^{2\boldsymbol{d}}, \tag{2.3}$$

where

$$c_{\boldsymbol{d}} = \frac{1}{2^{5g-6+2n}\,\boldsymbol{d}!} \langle \tau_{d_1} \ldots \tau_{d_n} \rangle_{g,n}, \tag{2.4}$$

$$\langle \tau_{d_1} \ldots \tau_{d_n} \rangle_{g,n} = \int_{\overline{\mathcal{M}}_{g,n}} \psi_1^{d_1} \ldots \psi_n^{d_n}, \tag{2.5}$$

and $\boldsymbol{d}! = d_1! \cdots d_n!$. Note that $N_{g,n}(b_1, \ldots, b_n)$ contains only even powers of b_i, where $i = 1, \ldots, n$.

Following [12] we consider the following linear operators $\mathcal{Y}(\boldsymbol{H})$ and $\mathcal{Z}$ on the spaces of polynomials in variables $b_1, b_2, \ldots$, where $H_1, H_2, \ldots$ are positive integers. The operator $\mathcal{Y}(\boldsymbol{H})$ is defined on monomials as

$$\mathcal{Y}(\boldsymbol{H}) \ : \quad \prod_{i=1}^{k} b_i^{m_i} \longmapsto \prod_{i=1}^{k} \frac{m_i!}{H_i^{m_i+1}}, \tag{2.6}$$

and extends to arbitrary polynomials by linearity. The operator $\mathcal{Z}$ is defined on monomials as

$$\mathcal{Z} \ : \quad \prod_{i=1}^{k} b_i^{m_i} \longmapsto \prod_{i=1}^{k} \big(m_i! \cdot \zeta(m_i+1)\big), \tag{2.7}$$

and extends to arbitrary polynomials by linearity.

Given a stable graph Γ, denote by $V(\Gamma)$ the set of its vertices and by $E(\Gamma)$ the set of its edges. To each stable graph $\Gamma \in \mathcal{G}_{g,n}$ we associate the following homogeneous polynomial P_Γ of degree $6g-6+2n$. To every edge $e \in E(\Gamma)$ we assign a formal variable b_e; we associate 0 to each leg. Given a vertex $v \in V(\Gamma)$ denote by g_v the integer number decorating v and denote by n_v the valency of v, where the legs incident to v are counted towards the valency of v. Take a small neighborhood of v in Γ. We denote by $\boldsymbol{b}_v$ the resulting collection of size n_v. If some edge e is a loop joining v to itself, b_e would be present in $\boldsymbol{b}_v$ twice; if an edge e joins v to a distinct vertex, b_e would be present in $\boldsymbol{b}_v$ once; all the other entries of $\boldsymbol{b}_v$ correspond to legs; they are represented by zeroes. To each vertex $v \in V(\Gamma)$ we associate the polynomial $N_{g_v,n_v}(\boldsymbol{b}_v)$, where $N_{g,n}$ is defined in (2.3). We associate to the stable graph Γ the polynomial P_Γ defined as follows:

$$P_\Gamma(\boldsymbol{b}) = \frac{2^{6g-5+2n} \cdot (4g-4+n)!}{(6g-7+2n)!} \times \frac{1}{2^{|V(\Gamma)|-1}} \cdot \frac{1}{|\operatorname{Aut}(\Gamma)|} \cdot \prod_{e \in E(\Gamma)} b_e \cdot \prod_{v \in V(\Gamma)} N_{g_v,n_v}(\boldsymbol{b}_v). \tag{2.8}$$

Theorem ([29]) *The Masur–Veech volume* $\operatorname{Vol}\mathcal{Q}_{g,n}$ *of the principal stratum of quadratic differentials with* $4g-4+n$ *labelled simple zeros and* n *labelled simple poles has the following value:*

$$\operatorname{Vol}\mathcal{Q}_{g,n} = \sum_{\Gamma \in \mathcal{G}_{g,n}} \operatorname{Vol}(\Gamma), \tag{2.9}$$

where the contribution of an individual stable graph Γ *has the form*

$$\operatorname{Vol}(\Gamma) = \mathcal{Z}(P_\Gamma). \tag{2.10}$$

Remark 2.2 The contribution (2.10) of any individual stable graph has the following natural interpretation. We have seen that stable graphs Γ in $\mathcal{G}_{g,n}$ are in natural bijective correspondence with $\mathrm{Mod}_{g,n}$-orbits of *reduced* multicurves $\gamma_{reduced} = \gamma_1 + \gamma_2 + \cdots$, where simple closed curves γ_i and γ_j are not isotopic for any $i \neq j$. Take $\Gamma \in \mathcal{G}_{g,n}$, and put $k = |E(\Gamma)|$. Let $\gamma_{reduced} = \gamma_1 + \cdots + \gamma_k$ be the reduced multicurve associated to Γ. Let $\gamma_{\boldsymbol{H}} = \gamma(\Gamma, \boldsymbol{H}) = H_1\gamma_1 + \cdots + H_k\gamma_k$, where $\boldsymbol{H} = (H_1, \ldots, H_k) \in \mathbb{N}^k$. We have

$$\mathrm{Vol}(\Gamma) = \sum_{H \in \mathbb{N}^k} \mathrm{Vol}\big(\Gamma, \boldsymbol{H}\big), \tag{2.11}$$

where the contribution $\mathrm{Vol}\big(\Gamma, \boldsymbol{H}\big)$ of square-tiled surfaces with the horizontal cylinder decomposition of type $(\Gamma, \boldsymbol{H})$ to $\mathrm{Vol}\,\mathcal{Q}_{g,n}$ is given by the formula:

$$\mathrm{Vol}\big(\Gamma, \boldsymbol{H}\big) = \mathcal{Y}(\boldsymbol{H})(P_\Gamma). \tag{2.12}$$

In other words, we can rearrange the sum in (2.2) as

$$\mathrm{Vol}\,\mathcal{Q}_{g,n} = \sum_{[\gamma] \in \mathcal{O}} \mathrm{Vol}(\gamma) = \sum_{\Gamma \in \mathcal{G}_{g,n}} \sum_{\{[\gamma]\,|\,\Gamma(\gamma) = \Gamma\}} \mathrm{Vol}(\gamma), \tag{2.13}$$

where

$$\sum_{\{[\gamma]\,|\,\Gamma(\gamma) = \Gamma\}} \mathrm{Vol}(\gamma) = \mathrm{Vol}(\Gamma).$$

In this way we can extend Definition 2.1 and speak of *asymptotic probability* of getting a square-tiled surface in $\mathcal{ST}_{g,n}(\Gamma) = \cup_{\{[\gamma]\,|\,\Gamma(\gamma)=\Gamma\}} \mathcal{ST}_{g,n}(\gamma)$ taking a random square-tiled surface in $\mathcal{ST}_{g,n}(N)$ as $N \to +\infty$. In the same way we define *frequency* of square-tiled surfaces having exactly k maximal horizontal cylinders among all square-tiled surfaces of genus g.

In particular, we define the probability $\mathbb{P}\big(K_g(S) = k\big)$ from Eq. (1.6) as

$$\mathbb{P}\big(K_g(S) = k\big) = \frac{1}{\mathrm{Vol}\,\mathcal{Q}_g} \cdot \sum_{\substack{\Gamma \in \mathcal{G}_g \\ |E(\Gamma)| = k}} \mathrm{Vol}(\Gamma). \tag{2.14}$$

We complete this section with a theorem that is one of the two keystone results on which rely all further asymptotic results of the present paper. Morally, it serves to establish explicit normalization allowing to pass from a finite measure with unspecified total mass to a specific probability measure. This statement was conjectured in [26,29] and proved in [3, Theorem 1.7].

Theorem 2.3 (Aggarwal [3]) *The Masur–Veech volume of the moduli space of holomorphic quadratic differentials has the following large genus asymptotics:*

$$\operatorname{Vol}\mathcal{Q}_g = \frac{4}{\pi}\cdot\left(\frac{8}{3}\right)^{4g-4}\cdot\big(1+o(1)\big) \quad \text{as } g\to+\infty. \tag{2.15}$$

Remark 2.4 The exact values of $\operatorname{Vol}\mathcal{Q}_g$ for $g \le 250$ (and more) can be obtained by combining results of Chen et al. [21] with the results of Kazarian [48] or of Yang et al. [86]. Supported by serious data analysis, the authors of [86] conjecture that the error term in (2.15) admits an asymptotic expansion in g^{-1} with the leading term $-\frac{\pi^2}{144}\cdot\frac{1}{g}$ and with terms of order g^{-2} and g^{-3} with explicit coefficients. In Theorem 5.1, using a refinement of the estimates from [3] we prove that the error term $o(1)$ in 2.15 can be improved to a finer estimate $O(g^{-1/4})$.

Conjectural generalization of formula (2.15) to all strata of meromorphic quadratic differentials and numerical evidence beyond this conjecture are presented in [6]. Actually, [3, Theorem 1.7] provides the volume asymptotics in the more general setting for $\operatorname{Vol}\mathcal{Q}_{g,n}$ under assumption that the number n of simple poles satisfies the relation $20n < \log g$.

2.4 Frequencies of multicurves (after M. Mirzakhani)

Recall that two integral multicurves on the same smooth surface of genus g with n punctures *have the same topological type* if they belong to the same orbit of the mapping class group $\mathrm{Mod}_{g,n}$.

We change now flat setting to hyperbolic setting. Following M. Mirzakhani, given an integral multicurve γ in $\mathcal{ML}_{g,n}(\mathbb{Z})$ and a hyperbolic surface X in the Teichmüller space $\mathcal{T}_{g,n}$ consider the function $s_X(L,\gamma)$ counting the number of simple closed geodesic multicurves on X of length at most L of the same topological type as γ. M. Mirzakhani proves in [62] the following Theorem.

Theorem (M. Mirzakhani) *For any rational multicurve γ and any hyperbolic surface $X \in \mathcal{T}_{g,n}$,*

$$s_X(L,\gamma) \sim B(X)\cdot\frac{c(\gamma)}{b_{g,n}}\cdot L^{6g-6+2n}, \tag{2.16}$$

as $L\to+\infty$.

The factor $B(X)$ in the above formula has the following geometric meaning. Consider the unit ball $B_X = \{\gamma\in\mathcal{ML}_{g,n} \mid \ell_X(\gamma)\le 1\}$ defined by means of the length function ℓ_X. Then $B(X)$ is Thurston's measure of B_X in $\mathcal{ML}_{g,n}$:

$$B(X) = \mu_{\mathrm{Th}}(B_X).$$

The factor $b_{g,n}$ is defined as the average of $B(X)$ over $\mathcal{M}_{g,n}$ viewed as the moduli space of hyperbolic metrics, where the average is taken with respect to the Weil–Petersson volume form on $\mathcal{M}_{g,n}$:

$$b_{g,n} = \int_{\mathcal{M}_{g,n}} B(X)\, dX. \tag{2.17}$$

Mirzakhani showed that

$$b_{g,n} = \sum_{[\gamma]\in\mathcal{O}(g,n)} c(\gamma), \tag{2.18}$$

where the sum of $c(\gamma)$ taken with respect to representatives $[\gamma]$ of all orbits $\mathcal{O}(g,n)$ of the mapping class group $\mathrm{Mod}_{g,n}$ in $\mathcal{ML}_{g,n}(\mathbb{Z})$. This allows to interpret the ratio $\frac{c(\gamma)}{b_{g,n}}$ as the probability to get a multicurve of type γ taking a "large random" multicurve (in the same sense as the probability that coordinates of a "random" point in $\mathbb{Z}^2$ are coprime equals $\frac{6}{\pi^2}$).

In particular, we define the quantity $\mathbb{P}\big(K_g(\gamma)=k\big)$ from Eq. (1.6) as

$$\mathbb{P}\big(K_g(\gamma)=k\big) = \frac{1}{b_g}\cdot \sum_{[\gamma]\in\mathcal{O}_k(g)} c(\gamma), \tag{2.19}$$

where, $b_g = b_{g,0}$ and $\mathcal{O}_k(g) \subset \mathcal{O}(g) = \mathcal{O}(g,0)$ is the subcollection of orbits of those multicurves γ, for which $\gamma_{reduced}$ has exactly k connected components.

M. Mirzakhani found an explicit expression for the coefficient $c(\gamma)$ and for the global normalization constant $b_{g,n}$ in terms of the intersection numbers of ψ-classes.

2.5 Frequencies of square-tiled surfaces of fixed combinatorial type

The following Theorem establishes a bridge between flat and hyperbolic count.

Theorem ([29]) *For any integral multicurve* $\gamma \in \mathcal{ML}_{g,n}(\mathbb{Z})$*, the volume contribution* $\mathrm{Vol}(\gamma)$ *to the Masur–Veech volume* $\mathrm{Vol}\,\mathcal{Q}_{g,n}$ *coincides with the Mirzakhani's asymptotic frequency* $c(\gamma)$ *of simple closed geodesic multicurves of topological type* γ *up to the explicit factor:*

$$\mathrm{Vol}(\gamma) = const_{g,n}\cdot c(\gamma), \tag{2.20}$$

where

$$const_{g,n} = 2\cdot(6g-6+2n)\cdot(4g-4+n)!\cdot 2^{4g-3+n}. \tag{2.21}$$

Proof of Theorem 1.7 Definitions (2.14) and (2.19) and Formulae (2.13) and (2.18) combined with relation (2.21) imply that $\mathbb{P}\big(K_g(\gamma) = k\big) = \mathbb{P}\big(K_g(S) = k\big)$. □

Corollary ([29]) *For any admissible pair of non-negative integers* (g, n) *different from* $(1, 1)$ *and* $(2, 0)$*, the Masur–Veech volume* $\operatorname{Vol}\mathcal{Q}_{g,n}$ *and the average Thurston measure of a unit ball* $b_{g,n}$ *are related as follows:*

$$\operatorname{Vol}\mathcal{Q}_{g,n} = 2\cdot(6g-6+2n)\cdot(4g-4+n)!\cdot 2^{4g-3+n}\cdot b_{g,n}. \tag{2.22}$$

Remark 2.5 In [63, Theorem 1.4] M. Mirzakhani established the relation

$$\operatorname{Vol}\mathcal{Q}_g = const_g \cdot b_g,$$

where b_g was computed in [62, Theorem 5.3]. However, Mirzakhani does not give any formula for the value of the normalization constant $const_g$ presented in (2.22). This constant was recently computed by Arana-Herrera [10] and by Monin and Telpukhovskiy [70] simultaneously and independently of us by different methods. The same value of $const_{g,n}$ was obtained by Erlandsson and Souto in [33] through an approach different from all the ones mentioned above.

2.6 Uniform large genus asymptotics of correlators (after A. Aggarwal)

We denote by $\Pi(m, n)$ the set of nonnegative compositions of an integer m as sum of n non-negative integers. For any nonnegative composition $\boldsymbol{d} \in \Pi(3g-3+n, n)$ define $\varepsilon(\boldsymbol{d})$ through the following equation:

$$\langle\tau_{d_1}\dots\tau_{d_n}\rangle_{g,n} = \frac{(6g-5+2n)!!}{(2d_1+1)!!\cdots(2d_n+1)!!}\cdot\frac{1}{g!\cdot 24^g}\cdot\big(1+\varepsilon(\boldsymbol{d})\big). \tag{2.23}$$

By construction, the intersection numbers are nonnegative rational numbers, so $\varepsilon\big(\boldsymbol{d}\big) \ge -1$ for any $\boldsymbol{d} \in \Pi(3g-3+n, n)$. We conjectured in [26] that $\varepsilon(\boldsymbol{d})$ tends to zero uniformly for all nonnegative compositions $\boldsymbol{d} \in \Pi(3g-3+n, n)$ as soon as $n \le 2\log g$ and $g \to +\infty$. This conjecture was proved in much stronger form in the recent paper of Aggarwal [3].

The following Theorem corresponds to [3, Proposition 1.2].

Theorem 2.6 (A. Aggarwal) *Let* $n \in \mathbb{Z}_{\ge 1}$ *and* $\boldsymbol{d} \in \mathbb{Z}^n_{\ge 0}$ *satisfy* $|\boldsymbol{d}| = 3g + n - 3$*, for some* $g \in \mathbb{Z}_{\ge 0}$*. Then,*

$$1+\varepsilon(\boldsymbol{d}) \le \left(\frac{3}{2}\right)^{n-1}. \tag{2.24}$$

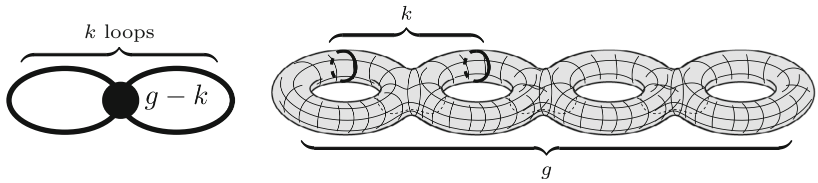

Fig. 3 The stable graph $\Gamma_k(g)$ (on the left) corresponds to the reduced multicurve $\gamma_{reduced} = \gamma_1 + \cdots + \gamma_k$ represented by k linearly independent homology cycles on a surface of genus g (on the right)

The next Theorem corresponds to [3, Proposition 4.1].

Theorem 2.7 (A. Aggarwal) *Let $g > 2^{15}$ and $n \geq 1$ be integers such that $g > 30n$, and let $\boldsymbol{d} \in \Pi(3g-3+n, n)$. Then we have*

$$\varepsilon(\boldsymbol{d}) \geq -20 \cdot \frac{(n + 4\log g)}{g}. \tag{2.25}$$

Finally, the following Theorem corresponds to [3, Proposition 4.2].

Theorem 2.8 (A. Aggarwal) *Let $g > 2^{30}$ and $n \geq 1$ be integers such that $g > 800n^2$, and let $\boldsymbol{d} \in \Pi(3g-3+n, n)$. Then we have*

$$1 + \varepsilon(\boldsymbol{d}) \leq \exp\left(625 \cdot \frac{(n + 2\log g)^2}{g}\right). \tag{2.26}$$

Remark 2.9 We proved in [29] explicit sharp upper and lower bounds for 2-correlators. Guo and Yang suggested in [39] an alternative proof of our conjecture [26] on the large genus asymptotics on the intersection numbers of psi-classes and stated an interesting polynomiality conjecture.

3 Random non-separating multicurves and non-uniform random permutations

Consider the stable graph $\Gamma_k(g)$ having a single vertex decorated with genus $g-k$, and having k loops, see the left picture in Fig. 3. This stable graph corresponds to multicurves on a closed surface of genus g, for which the components $\gamma_1, \ldots, \gamma_k$ of the underlying reduced multicurve $\gamma_{reduced} = \gamma_1 + \cdots + \gamma_k$ represent k linearly independent homology cycles. The square-tiled surfaces associated to this stable graph have single horizontal singular layer and k maximal horizontal cylinders.

Recall from Sect. 2.3 that by $\mathrm{Vol}(\Gamma_k(g))$ we denote the volume contribution from all square-tiled surfaces corresponding to the stable graph $\Gamma_k(g)$.

By $\mathrm{Vol}\left(\Gamma_k(g), (m_1, \ldots, m_k)\right)$ we denote the volume contribution from those square-tiled surfaces corresponding to the stable graph $\Gamma_k(g)$ for which one maximal horizontal cylinder is filled with m_1 bands of squares, another cylinder is filled with m_2 bands of squares, and so on up to the kth maximal horizontal cylinder, which is filled with m_k bands of squares. The corresponding multicurve has the form $m_1\gamma_1 + \cdots + m_k\gamma_k$, where $\gamma_1, \ldots, \gamma_k$ are as described above. By (2.11) we have

$$\mathrm{Vol}(\Gamma_k(g)) = \sum_{m_1,\ldots,m_k} \mathrm{Vol}\left(\Gamma_k(g), (m_1, \ldots, m_k)\right).$$

In this section we prove the following result, which relies on the uniform asymptotics of Witten's correlators proved by A. Aggarwal (see Theorems 2.6–2.8 in the present paper or, respectively, Propositions 1.2, 4.1, 4.2 in the original paper [3] of A. Aggarwal).

Theorem 3.1 *Let $m \in \mathbb{N} \cup \{+\infty\}$. For any complex number t in the disk $|t| < 2$ we have as $g \to +\infty$*

$$\sum_{k=1}^{g} \sum_{\substack{m_1,\ldots,m_k \\ 1\le m_i\le m \text{ for } i=1,\ldots,k}} \mathrm{Vol}(\Gamma_k(g), (m_1, \ldots, m_k)) \cdot t^k$$
$$= \frac{2\sqrt{2}\left(\frac{2m}{m+1}\right)^{t/2}}{\sqrt{\pi}\cdot\Gamma(\frac{t}{2})} \cdot (3g-3)^{\frac{t-1}{2}} \cdot \left(\frac{8}{3}\right)^{4g-4} \left(1 + O\left(\frac{(\log g)^2}{g}\right)\right), \tag{3.1}$$

where for every compact subset U of the complex disk $|t| < 2$ the error term is uniform in $m \in \mathbb{N} \cup \{+\infty\}$ and $t \in U$. In particular, for $m = +\infty$ and $t = 1$ we obtain

$$\sum_{k=1}^{g} \mathrm{Vol}(\Gamma_k(g)) = \frac{4}{\pi}\left(\frac{8}{3}\right)^{4g-4} \left(1 + O\left(\frac{(\log g)^2}{g}\right)\right). \tag{3.2}$$

We prove Theorem 3.1 in Sect. 3.7. We note that asymptotics (3.2) was first obtained by A. Aggarwal in [3, Proposition 8.3]. Our refinement consists in the bound $O\left(\frac{(\log g)^2}{g}\right)$ for the error term. Conjecturally, the bound can be further improved to $O\left(\frac{1}{g}\right)$; see Remark 2.4.

3.1 Volume contribution of stable graphs with a single vertex

In this section, we show how to express an approximate value of the contribution $\mathrm{Vol}(\Gamma_k(g))$ of square-tiled surfaces corresponding to the stable graph $\Gamma_k(g)$ to the Masur–Veech volume $\mathrm{Vol}\,\mathcal{Q}_g$ in terms of the following normalized weighted multi-variate harmonic sum.

Definition 3.2 Let $m \in \mathbb{N} \cup \{+\infty\}$ and let α be a positive real number. For integers k, n such that $1 \leq k \leq n$, define

$$\widetilde{H}_{n,m,\alpha}(k) = \frac{\alpha^k}{k!} \sum_{j_1+\dots+j_k=n} \frac{\zeta_m(2j_1)\cdot\zeta_m(2j_2)\cdots\zeta_m(2j_k)}{j_1\cdot j_2\cdots j_k}, \tag{3.3}$$

where the sum is taken over all k-tuples $(j_1, j_2, \dots, j_k) \in \mathbb{N}^k$ of positive integers summing up to n and

$$\zeta_m(s) = 1 + \frac{1}{2^s} + \frac{1}{3^s} + \ldots + \frac{1}{m^s}$$

is the partial zeta function.

Remark 3.3 The particular cases of the above numbers, namely,

$$H_k(n) = \sum_{j_1+\dots+j_k=n} \frac{1}{j_1\cdot j_2\cdots j_k} = k!\cdot\widetilde{H}_{n,1,1}(k),$$
$$Z_k(n) = \sum_{j_1+\dots+j_k=m} \frac{\zeta(2j_1)\cdots\zeta(2j_k)}{j_1\cdot j_2\cdots j_k} = k!\cdot\widetilde{H}_{n,\infty,1}(k)$$

appeared in the preprint [26]; the asymptotic expansions for these quantities were obtained by A. Aggarwal in Sections 6 and 7 of [3]. The framework which we develop here allows to treat all normalized weighted multi-variate harmonic sums $\widetilde{H}_{n,m,\alpha}(k)$ in a unified way.

Theorem 3.4 *There exists a constant C_1 such that for sufficiently large $g \in \mathbb{N}$ the following property holds. For any couple m, k, such that $m \in \mathbb{N} \cup \{+\infty\}$, $k \in \mathbb{N}$, $800k^2 \leq g$, we have*

$$\sum_{\substack{m_1,\dots,m_k \\ 1\leq m_i\leq m \text{ for } i=1,\dots,k}} \mathrm{Vol}\left(\Gamma_k(g), (m_1, \dots, m_k)\right)$$
$$= \frac{2\sqrt{2}}{\sqrt{\pi}}\cdot\sqrt{3g-3}\cdot\left(\frac{8}{3}\right)^{4g-4}\cdot\widetilde{H}_{3g-3,m,\frac{1}{2}}(k)\cdot\left(1+\varepsilon_1(g,k)\right), \tag{3.4}$$

where $|\varepsilon_1(g,k)| \le C_1 \cdot \dfrac{(k+2\log g)^2}{g}$.

There exists a constant C_2 *such that for all triples* (g,k,m), *where* $g \in \mathbb{N}$, $g \ge 2$; $k \in \mathbb{N}$, $k \le g$; $m \in \mathbb{N} \cup \{+\infty\}$, *we have*

$$\sum_{\substack{m_1,\dots,m_k \\ 1\le m_i \le m \text{ for } i=1,\dots,k}} \operatorname{Vol}\left(\Gamma_k(g), (m_1,\dots,m_k)\right) \le C_2 \cdot \sqrt{g} \cdot \left(\frac{8}{3}\right)^{4g-4} \cdot \widetilde{H}_{3g-3,m,\frac{1}{2}}(k) \cdot \left(\frac{9}{4}\right)^k, \tag{3.5}$$

where $\widetilde{H}_{3g-3,m,\frac{1}{2}}(k)$ *is the normalized weighted multi-variate harmonic sum defined in* (3.3).

In order to prove Theorem 3.4 we first state and prove Lemma 3.5 below. Let $\mathbf{D} = (D_1,\dots,D_k) \in \Pi(3g-3+2k,k)$. Define $c_{g,k}(\boldsymbol{D})$ as

$$\begin{aligned} c_{g,k}(\boldsymbol{D}) := {} & \frac{g!\cdot(3g-3+2k)!}{(6g+4k-5)!} \cdot \frac{3^g}{2^{3g-6+5k}} \\ & \times \sum_{d_{1,1}+d_{1,2}=D_1} \cdots \sum_{d_{k,1}+d_{k,2}=D_k} \int_{\overline{\mathcal{M}}_{g,2k}} \psi_1^{d_{1,1}} \psi_2^{d_{1,2}} \dots \psi_{2k-1}^{d_{k,1}} \psi_{2k}^{d_{k,2}} \\ & \times \prod_{j=1}^{k} \frac{(2D_j+2)!}{d_{j,1}!\cdot d_{j,2}!}. \end{aligned} \tag{3.6}$$

The following result is a corollary of the uniform asymptotics of Witten's correlators proved by A. Aggarwal (see Theorems 2.6–2.8 in the present paper or, respectively, Propositions 1.2, 4.1, 4.2 in the original paper [3] of A. Aggarwal).

Lemma 3.5 *There exists a constant* C_3 *such that for sufficiently large* $g \in \mathbb{N}$ *and for* $k \in \mathbb{N}$ *satisfying* $800k^2 \le g$ *we have*

$$|c_{g,k}(\boldsymbol{D}) - 1| \le C_3 \cdot \frac{(k+2\log g)^2}{g}. \tag{3.7}$$

For any positive integers $g, k \in \mathbb{N}$ *satisfying* $1 \le k \le g$ *and* $g \ge 2$, *we have*

$$c_{g,k}(\boldsymbol{D}) \le \left(\frac{9}{4}\right)^k. \tag{3.8}$$

Proof Passing to double factorials and applying definition (2.23) of $\varepsilon(\boldsymbol{d})$ we get

$$\begin{aligned} c_{g,k}(\boldsymbol{D}) = & \frac{g!}{2^{3g-3+2k}\cdot(6g+4k-5)!!}\cdot\frac{3^g}{2^{3g-6+5k}} \\ & \times \sum_{d_{1,1}+d_{1,2}=D_1}\cdots\sum_{d_{k,1}+d_{k,2}=D_k}\langle\tau_{d_{1,1}}\tau_{d_{1,2}}\ldots\tau_{d_{k,1}}\tau_{d_{k,2}}\rangle_{g,2k} \\ & \times\prod_{j=1}^{k}\left(\frac{(2d_{j,1}+1)!}{d_{j,1}!}\cdot\frac{(2d_{j,2}+1)!}{d_{j,2}!}\cdot\binom{2D_j+2}{2d_{j,1}+1}\right) \\ = & \frac{1}{2^{6g-6+5k}}\sum_{d_{1,1}+d_{1,2}=D_1}\cdots\sum_{d_{k,1}+d_{k,2}=D_k}\left(\big(1+\varepsilon(\boldsymbol{d})\big)\cdot\prod_{j=1}^{k}\binom{2D_j+2}{2d_{j,1}+1}\right). \end{aligned}$$

Applying the combinatorial identity

$$\sum_{m=0}^{n-1}\binom{2n}{2m+1}=2^{2n-1}$$

we get

$$\begin{aligned} & \sum_{d_{1,1}+d_{1,2}=D_1}\cdots\sum_{d_{k,1}+d_{k,2}=D_k}\prod_{j=1}^{k}\binom{2D_j+2}{2d_{j,1}+1} \\ & =\left(\prod_{j=1}^{k}\sum_{d_{j,1}=0}^{D_j}\binom{2D_j+2}{2d_{j,1}+1}\right)=\left(\prod_{j=1}^{k}2^{2D_j+1}\right)=2^{6g-6+5k}. \end{aligned}$$

The claim that bound (3.7) is valid for sufficiently large g now follows from combination of bounds (2.25) and (2.26) from Theorems 2.7 and 2.8 of A. Aggarwal (respectively, Propositions 4.1, 4.2 in the original paper [3]).

For $k\geq 2$ the universal bound (3.8) follows from the universal bound (2.24) from Theorem 2.6 of A. Aggarwal (see Proposition 1.2 in the original paper [3]), using the fact that for $k\geq 2$ we have $1+(3/2)^{2k-1}\leq(3/2)^{2k}$.

We proved in [29, Proposition 4.1] that $\varepsilon(\boldsymbol{d})\leq 0$ for any $\boldsymbol{d}\in\Pi(3g-1,2)$. This implies bound (3.8) for $k=1$, which completes the proof of the Lemma. □

Proof of Theorem 3.4 Let us denote

$$V_{m,k}(g)=\sum_{\substack{m_1,\dots,m_k\\ 1\le m_i\le m\\ \text{for } i=1,\dots,k}}\mathrm{Vol}(\Gamma_k(g),(m_1,\dots,m_k)) \quad\text{and}\quad V_m(g)=\sum_{k=1}^{g}V_{m,k}(g). \tag{3.9}$$

The automorphism group $\mathrm{Aut}(\Gamma_k(g))$ consists of all possible permutations of loops composed with all possible flips of individual loops, so

$$|\,\mathrm{Aut}(\Gamma)|=2^k\cdot k!.$$

The graph $\Gamma_k(g)$ has a single vertex, so $|V(\Gamma_k(g)|=1$. Thus, applying (2.12) to $\Gamma_k(g)$ we get

$$\begin{aligned}V_{m,k}(g)&=\frac{2^{6g-5}\cdot(4g-4)!}{(6g-7)!}\cdot 1\cdot\frac{1}{2^k\cdot k!}\\ &\times\sum_{\substack{\boldsymbol{H}=(m_1,\dots,m_k)\\ 1\le m_i\le m\\ \text{for } i=1,\dots,k}}\mathcal{Y}\big(\boldsymbol{H}, b_1b_2\dots b_k\cdot N_{g-k,2k}(b_1,b_1,b_2,b_2,\dots,b_k,b_k)\big)\\ &=\frac{(4g-4)!}{(6g-7)!}\cdot\frac{2^{6g-5}}{2^k\cdot k!}\cdot\frac{1}{2^{5(g-k)-6+4k}}\cdot\sum_{\boldsymbol{d}\in\Pi(3g-3-k,2k)}\frac{\langle\tau_{\mathbf{d}}\rangle_{g-k,2k}}{\mathbf{d}!}\\ &\times\prod_{i=1}^{k}\Big((2d_{2i-1}+2d_{2i}+1)!\cdot\zeta_m(2d_{2i-1}+2d_{2i}+2)\Big).\end{aligned}\tag{3.10}$$

Rewrite the latter sum using notation $\mathbf{D}=(D_1,\dots,D_k)\in\Pi(3g-3-k,k)$ and $c_{g-k,k}(\boldsymbol{D})$ defined by (3.6). Adjusting expression (3.6) from genus g to genus $g-k$ we get

$$\begin{aligned}&\sum_{\boldsymbol{d}\in\Pi(3g-3-k,2k)}\frac{\langle\tau_{\mathbf{d}}\rangle_{g-k,2k}}{\mathbf{d}!}\\ &\quad\times\prod_{i=1}^{k}\Big((2d_{2i-1}+2d_{2i}+1)!\cdot\zeta_m(2d_{2i-1}+2d_{2i}+2)\Big)\\ &=\sum_{\boldsymbol{D}\in\Pi(3g-3-k,k)}c_{g-k,k}(\boldsymbol{D})\cdot\frac{(6(g-k)+4k-5)!}{(g-k)!\cdot(3(g-k)-3+2k)!}\end{aligned}$$

$$\times \frac{2^{3(g-k)-6+5k}}{3^{g-k}} \cdot \prod_{j=1}^{k} \frac{\zeta_m(2D_j+2)}{2D_j+2},$$

which allows us to rewrite (3.10) as

$$\begin{aligned} &V_{m,k}(g) \\ &= \left(\frac{(4g-4)!}{(6g-7)!} \cdot 2^{g+1} \cdot \frac{1}{k!}\right) \cdot \left(\frac{(6g-2k-5)!}{(g-k)! \cdot (3g-3-k)!} \cdot \frac{2^{3g-6+2k}}{3^{g-k}}\right) \\ &\quad \times \sum_{\boldsymbol{D} \in \Pi(3g-3-k,2k)} \frac{c_{g-k,k}(\boldsymbol{D})}{2^k} \cdot \prod_{j=1}^{k} \frac{\zeta_m(2D_j+2)}{D_j+1}. \end{aligned} \tag{3.11}$$

Let us define

$$c_{g-k,k}^{min} := \min_{\boldsymbol{D}} c_{g-k,k}(\boldsymbol{D}) \quad \text{and} \quad c_{g-k,k}^{max} := \max_{\boldsymbol{D}} c_{g-k,k}(\boldsymbol{D}).$$

Rearranging factors with factorials, collecting powers of 2 and 3, and passing to notation $\widetilde{H}_{m,3g-3,\frac{1}{2}}(k)$ for the multivariate harmonic sum (3.3) we get the following bounds:

$$\begin{aligned} &c_{g-k,k}^{min} \le V_{m,k}(g) \\ &\times \left(\frac{(6g-2k-5)!}{(6g-7)!} \cdot \frac{(4g-4)!}{(g-k)! \cdot (3g-3-k)!} \cdot \frac{2^{4g-5+2k}}{3^{g-k}} \cdot \widetilde{H}_{3g-3,m,\frac{1}{2}}(k)\right)^{-1} \\ &\quad \le c_{g-k,k}^{max}. \end{aligned} \tag{3.12}$$

We start by proving the first assertion of the theorem represented by relation (3.4). We rewrite the product of factorials in (3.12) as

$$\begin{aligned} &\frac{(6g-2k-5)! \cdot (4g-4)!}{(6g-7)! \cdot (g-k)! \cdot (3g-3-k)!} \\ &\quad = (6g-6) \cdot \binom{4g-4}{g-1} \cdot \frac{\frac{(3g-3)!}{(3g-3-k)!} \cdot \frac{(g-1)!}{(g-k)!}}{\frac{(6g-6)!}{(6g-2k-5)!}} \\ &\quad = (6g-6) \cdot \binom{4g-4}{g-1} \cdot \frac{(3g-3)^k \cdot (g-1)^{k-1}}{(6g-6)^{2k-1}} \big(1+\varepsilon_3(g,k)\big) \\ &\quad = \sqrt{3g-3} \cdot \sqrt{\frac{2}{\pi}} \cdot \frac{2^{8g-6-2k}}{3^{3g-4+k}} \big(1+\varepsilon_4(g,k)\big). \end{aligned} \tag{3.13}$$

Note that there exist constants C'_3 and a_0 such that for any integer a satisfying $a \geq a_0$ and for any $b \in \mathbb{N}$, satisfying $b \leq \sqrt{a}$, we have

$$a^b \left(1 - C'_3 \cdot \frac{b^2}{a}\right) \leq \frac{a!}{(a-b)!} \leq a^b \left(1 + C'_3 \cdot \frac{b^2}{a}\right).$$

This implies that there exists g_0 such that for any $g \in \mathbb{N}$ satisfying $g \geq g_0$ and for any $k \in \mathbb{N}$ satisfying $800k^2 \leq g$ we have the bound

$$|\varepsilon_3(g,k)| \leq C_3 \cdot \frac{k^2}{g}$$

for the error term in the third line of (3.13). Let

$$\binom{4g-4}{g-1} = \sqrt{\frac{2}{\pi(3g-3)}} \cdot \frac{2^{8g-8}}{3^{3g-3}} \cdot \big(1 + \varepsilon_5(g)\big).$$

There exist constants C_5 and g_1 such that for any $g \in \mathbb{N}$ satisfying $g \geq g_1$ we have

$$|\varepsilon_5(g)| \leq C_5 \cdot \frac{1}{g}.$$

The latter two bounds imply that there exist a constant C_4 and a constant g_2 such that for any $g \in \mathbb{N}$ satisfying $g \geq g_2$ we have the bound

$$|\varepsilon_4(g)| \leq C_4 \cdot \frac{k^2}{g},$$

for the error term on the right-hand side of the fourth line of (3.13). Using the latter bound and collecting together the powers of 2, 3 and g, we can rewrite (3.12) in the following way:

$$\begin{aligned} c^{min}_{g-k,k}\left(1 - C_4 \cdot \frac{k^2}{g}\right) &\leq \frac{V_{m,k}(g)}{\frac{2\sqrt{2}}{\sqrt{\pi}} \cdot \left(\frac{8}{3}\right)^{4g-4} \cdot \sqrt{3g-3} \cdot \widetilde{H}_{3g-3,m,\frac{1}{2}}(k)} \\ &\leq c^{max}_{g-k,k}\left(1 + C_4 \cdot \frac{k^2}{g}\right). \end{aligned} \tag{3.14}$$

Now, using the bound (3.7) from the first part of Lemma 3.5 we get (3.4).

The proof of the upper bound (3.5) is similar. For the product of factorials we use the bound

$$\frac{\frac{(3g-3)!}{(3g-3-k)!}\cdot\frac{(g-1)!}{(g-k)!}}{\frac{(6g-6)!}{(6g-2k-5)!}} = \prod_{i=0}^{k-1}\frac{3g-3-i}{6g-6-2i}\prod_{i=1}^{k-1}\frac{g-i}{6g-5-2i}$$
$$= \frac{6}{12^k}\prod_{i=1}^{k-1}\frac{6g-6i}{6g-5-2i}$$
$$= \frac{6}{12^k}\left(1+\frac{1}{6g-7}\right)\prod_{i=2}^{k-1}\frac{6g-6i}{6g-5-2i}$$
$$\leq \frac{36}{5}\cdot\frac{1}{12^k}. \tag{3.15}$$

valid for any couple (g,k) of positive integers satisfying $g \geq 2$ and $k \leq g$. Here we used the inequality $6g-6i < 6g-5-2i$ valid for any integer g, i such that $g \geq 2$ and $i \geq 2$. We also used the inequality $1/(6g-7) \leq 1/5$ valid for any integer $g \geq 2$. The upper bound for $c^{max}_{g-k,k}$ was established in Eq. (3.8) in the second part of Lemma 3.5. Plugging this bound for $c^{max}_{g-k,k}$ in (3.12) and the bound for the product of factorials in (3.13) and proceeding as before we obtain the result. □

Define

$$\tilde{V}_{m,k}(g) = \left(\frac{(4g-4)!}{(6g-7)!}\cdot 2^{g+1}\cdot\frac{1}{k!}\right)\cdot\left(\frac{(6g-2k-5)!}{(g-k)!\cdot(3g-3-k)!}\cdot\frac{2^{3g-6+2k}}{3^{g-k}}\right)$$
$$\times \sum_{\boldsymbol{D}\in\Pi(3g-3-k,2k)}\frac{1}{2^k}\cdot\prod_{j=1}^{k}\frac{\zeta_m(2D_j+2)}{D_j+1}, \tag{3.16}$$

where the partial ζ-function ζ_m is defined below expression (3.3). Define $\tilde{V}_{\infty,k}(g)$ by an analogous expression, where ζ_m is replaced by ζ.

This expression is obtained by replacing $c_{g-k,k}(\boldsymbol{D})$ with 1 in (3.11). We have seen that this is equivalent to replacing the Witten's correlators in the right-hand side of formula (3.10) for $V_{m,k}(g)$ by the asymptotic expression (2.23) from Sect. 2.6. We are now ready to give the formal definition of the approximating distribution $q_g(k)$ informally described in Sect. 1.

Define the probability distribution $q_g(k)$ as

$$q_g(k) := \frac{\tilde{V}_{\infty,k}(g)}{\tilde{V}_{\infty}(g)}, \quad \text{where} \quad \tilde{V}_{\infty}(g) := \sum_{k=1}^{g}\tilde{V}_{\infty,k}(g). \tag{3.17}$$

It follows from the proof of Theorem 3.4 that for sufficiently large $g \in \mathbb{N}$ and for $k \in \mathbb{N}$ satisfying $k^2 \leq g$ we have the bounds for $\tilde{V}_{m,k}(g)$ analogous to (3.14), where $c_{g-k,k}^{max}$ and $c_{g-k,k}^{max}$ are replaced with 1. This implies that $\tilde{V}_{m,k}(g)$ satisfies the lower bound

$$\tilde{V}_{m,k}(g) \geq \frac{2\sqrt{2}}{\sqrt{\pi}} \cdot \left(\frac{8}{3}\right)^{4g-4} \cdot \sqrt{3g-3} \cdot \widetilde{H}_{3g-3,m,\frac{1}{2}}(k) \cdot \left(1 + O\left(\frac{k^2}{g}\right)\right), \tag{3.18}$$

where the constant in the error term is uniform in $k \in \mathbb{N}$ satisfying $k^2 \leq g$.

The upper bound for the expression in factorials on the left-hand side of (3.15) can be expressed for large g as $\frac{1}{12^k} \cdot \left(1 + O\left(g^{-1}\right)\right)$. Thus, analog of (3.12) for $\tilde{V}_{m,k}(g)$, where $c_{g-k,k}^{max}$ and $c_{g-k,k}^{max}$ are replaced with 1 implies that for sufficiently large $g \in \mathbb{N}$ and for any $k \in \mathbb{N}$ we have the upper bound

$$\tilde{V}_{m,k}(g) \leq \frac{2\sqrt{2}}{\sqrt{\pi}} \cdot \left(\frac{8}{3}\right)^{4g-4} \cdot \sqrt{3g-3} \cdot \widetilde{H}_{3g-3,m,\frac{1}{2}}(k) \cdot \left(1 + O\left(\frac{1}{g}\right)\right). \tag{3.19}$$

3.2 Multi-variate harmonic sums and non-uniform random permutations

In this section we analyze the normalized weighted multi-variate harmonic sum from Definition 3.2 and Theorem 3.4. We show how these kind of sums naturally appear in the study of random permutations in the symmetric group.

Let us recall the setting of Sect. 1. Let $\theta = \{\theta_k\}_{k \geq 1}$ be a sequence of non-negative real numbers. From now on we assume for simplicity that $\theta_1 > 0$. Recall that given a permutation $\sigma \in S_n$ with cycle type $(1^{\mu_1} 2^{\mu_2} \ldots n^{\mu_n})$ we define its *weight* as

$$w_\theta(\sigma) := \theta_1^{\mu_1} \theta_2^{\mu_2} \cdots \theta_n^{\mu_n}.$$

To every sequence $\theta = \{\theta_k\}_{k \geq 1}$ we associate a probability measure on the symmetric group S_n as in (1.7) by setting

$$\mathbb{P}_{\theta,n}(\sigma) := \frac{w_\theta(\sigma)}{n! \cdot W_{\theta,n}}, \qquad \text{where} \qquad W_{\theta,n} := \frac{1}{n!} \sum_{\sigma \in S_n} w_\theta(\sigma).$$

Constant weights $\theta_i = 1$ correspond to the uniform measure on S_n. More generally, the probability measures on S_n obtained from constant weights $\theta_i = \alpha$ are called *Ewens measures*. The following lemma identifies our normalized weighted multi-variate harmonic sums from Definition 3.2 as total contribution of permutations having exactly k cycles to the sum $W_{\theta,n}$.

Lemma 3.6 *Let $\theta = \{\theta_k\}_{k\geq 1}$ be non-negative real numbers and consider the associated probability measure $\mathbb{P}_{\theta,n}$ on the symmetric group S_n for some n. Then*

$$\frac{1}{n!} \cdot \sum_{\substack{\sigma\in S_n \\ \mathrm{K}_n(\sigma)=k}} w_\theta(\sigma) = \frac{1}{k!} \cdot \sum_{i_1+\cdots+i_k=n} \frac{\theta_{i_1}\theta_{i_2}\cdots\theta_{i_k}}{i_1\cdots i_k}, \tag{3.20}$$

where $\mathrm{K}_n(\sigma)$ is the number of cycles in the cycle decomposition of σ and the sum in the right hand-side is taken over positive integers $i_1,\ldots,i_k$. In other words, we have the following identity in the ring $\mathbb{Q}[[t,z]]$ of formal power series in t and z:

$$\sum_{n\geq 1}\sum_{\sigma\in S_n} w_\theta(\sigma) t^{\mathrm{K}_n(\sigma)} \frac{z^n}{n!} = \exp\left(t\sum_{k\geq 1}\theta_k \frac{z^k}{k}\right). \tag{3.21}$$

The first few terms of the expansion of (3.21) in z have the following form:

$$\begin{aligned}\exp\left(t\sum_{k\geq 1}\theta_k\frac{z^k}{k}\right) &= 1 \\ &+ (\theta_1 t)\, z \\ &+ \left(\theta_2 t + \theta_1^2 t^2\right)\frac{z^2}{2!} \\ &+ \left(2\theta_3 t + 3\theta_1\theta_2 t^2 + \theta_1^3 t^3\right)\frac{z^3}{3!} \\ &+ \cdots\end{aligned}$$

Proof of Lemma 3.6 From each permutation σ in S_n and a composition $(i_1,\ldots,i_k)$ of n we build the following permutation $\widetilde{\sigma}$ with k cycles (in cycle notation)

$$\begin{aligned}&\big(\sigma(1),\sigma(2),\ldots,\sigma(i_1)\big)\big(\sigma(i_1+1),\ldots,\sigma(i_1+i_2)\big) \\ &\quad\cdots\big(\sigma(i_1+\cdots+i_{k-1}+1),\ldots,\sigma(n)\big).\end{aligned}$$

Here the cycles of $\widetilde{\sigma}$ are ordered from 1 to k so that the first cycle has length i_1, the second has length i_2, etc. Since each cycle is defined up to cyclic ordering, for each fixed $(i_1,\ldots,i_k)$ we obtain the same permutation (with ordered cycles) $i_1\cdot i_2\cdots i_k$ times. Hence the number

$$n!\frac{\theta_{i_1}\theta_{i_2}\cdots\theta_{i_k}}{i_1 i_2\cdots i_k}.$$

is the weighted count of permutations with k labelled cycles of lengths i_1, ..., i_k. Now summing over all possible compositions $(i_1, \dots, i_k)$ of n and dividing by $k!$ gives the weighted sum of permutations having exactly k cycles. □

We see that the normalized weighted multi-variate harmonic sums $\widetilde{H}_{n,m,\alpha}(k)$ defined in (3.3) represent the total weight of permutations having exactly k disjoint cycles in their cycle decomposition, where the weights $w_\theta(\sigma)$ correspond to the sequence $\theta_k = \alpha\zeta_m(2k)$, $k \in \mathbb{N}$. Thus, Lemma 3.6 implies the following relation for the generalization of the quantities $q_{n,\infty,\alpha}$ defined in (1.8) for arbitrary $m \in \mathbb{N} \cup \{+\infty\}$:

$$q_{n,m,\alpha}(k) = \mathbb{P}_{n,m,\alpha}\big(\mathrm{K}_n(\sigma) = k\big) = \frac{\widetilde{H}_{n,m,\alpha}(k)}{W_{n,m,\alpha}}, \tag{3.22}$$

where

$$W_{n,m,\alpha} = \sum_{k=1}^{n} \widetilde{H}_{n,m,\alpha}(k) = \sum_{k=1}^{n} \frac{1}{n!} \cdot \sum_{\substack{\sigma \in S_n \\ \mathrm{K}_n(\sigma)=k}} w_\theta(\sigma). \tag{3.23}$$

Theorem 3.4 relates the contributions $\mathrm{Vol}\big(\Gamma_k(g)\big)$ of the stable graphs $\Gamma_k(g)$ to the Masur–Veech volume $\mathrm{Vol}\,\mathcal{Q}_g$ to the total weight of permutations having exactly k disjoint cycles in their cycle decomposition, where the weights $w_\theta(\sigma)$ corresponding to the sequence $\theta_k = \frac{1}{2}\zeta(2k)$, $k \in \mathbb{N}$, that is to the normalized weighted multi-variate harmonic sums with $m = +\infty$ and $\alpha = \frac{1}{2}$.

The unsigned Stirling numbers of the first kind $s(n, k)$ corresponding to the uniform distribution on S_n satisfy $s(n, k) = n! \cdot \widetilde{H}_{n,1,1}(k)$.

Theorem 3.7 *Let t be a complex number and $m \in \mathbb{N} \cup \{+\infty\}$. Then*

$$\sum_{k=1}^{n} \widetilde{H}_{n,m,\alpha}(k) t^k = \frac{\left(\frac{2m}{m+1}\right)^{\alpha t} n^{\alpha t - 1}}{\Gamma(\alpha t)} \left(1 + O\left(\frac{1}{n}\right)\right), \tag{3.24}$$

where the error term is uniform in $m \in \mathbb{N} \cup \{+\infty\}$ and in t over compact subsets of complex numbers.

Here we use the convention

$$\left.\frac{2m}{m+1}\right|_{m=+\infty} = 2.$$

A version of Theorem 3.7 stated for the values $m = 1$ and $m = +\infty$, $\alpha = \frac{1}{2}$, $t = 1$, which are particularly important in the context of the present paper, was stated as a conjecture in the preprint [26] and was first proved by A. Aggarwal in [3, Proposition 7.2]. We suggest here a proof of Theorem 3.7 based on technique of Hwang [43] applied to the generating function in the right-hand side of Eq. (3.21). We discovered this approach for ourselves after the paper [3] became available.

We will use the following elementary facts in the proof Theorem 3.7.

Lemma 3.8 *Let* $m \in \mathbb{N} \cup \{+\infty\}$. *The series*

$$g_m(z) = \sum_{k \geq 1} \zeta_m(2k) \frac{z^k}{k}$$

converges in the unit disk $|z| < 1$. *Considered as a holomorphic function, it extends to* $\mathbb{C} \setminus [1, +\infty)$. *Moreover, as* $z \to 1$ *inside* $\mathbb{C} \setminus [1, +\infty)$ *we have*

$$g_m(z) = -\log(1 - z) + \log\left(\frac{2m}{m+1}\right) + O(1 - z). \tag{3.25}$$

Proof Expanding the definition of the partial zeta function ζ_m and changing the order of summation we find the alternative formula

$$g_m(z) = -\sum_{n=1}^{m} \log\left(1 - \frac{z}{n^2}\right),$$

which proves the first assertion of the Lemma.

Now, we have

$$g_m(z) = -\log(1 - z) - \sum_{n=2}^{m} \log\left(1 - \frac{1}{n^2}\right) + O(z - 1).$$

For finite m, we can rewrite the constant term as

$$-\sum_{n=2}^{m} \log(1 - \frac{1}{n^2}) = \sum_{n=2}^{m} (2\log(n) - \log(n-1) - \log(n+1)) = \log\left(\frac{2m}{m+1}\right).$$

The case $m = +\infty$ is obtained by passing to the limit. □

Proof of Theorem 3.7 Theorem 3.7 can be derived as a corollary of Theorem 12 of [43] (see also [71, Lemma 2.13]). To make the proof tractable we provide here a complete argument based on the asymptotic analysis performed in the classical book by Flajolet and Sedgewick [35].

By Lemma 3.6 we have

$$\sum_{k=1}^{n} \widetilde{H}_{n,m,\alpha}(k)t^k = [z^n]\exp(t\alpha g_m(z)), \tag{3.26}$$

where $g_m(z)$ is the function defined in Lemma 3.8 and $[z^n]$ stands for the coefficient at z^n in the expansion of $\exp(t\alpha g_m(z))$. Plugging the asymptotic expansion (3.25) into (3.26) we obtain the following expansion as $z \to 1$ inside $\mathbb{C} \setminus [1, +\infty)$:

$$\exp(\alpha t g_m(z)) = \left(\frac{1}{1-z}\right)^{\alpha t} \cdot \left(\frac{2m}{m+1}\right)^{\alpha t} \cdot (1 + O(z-1)), \tag{3.27}$$

where the error term is uniform in t on compact subsets of complex numbers. Now by [35, Theorem VI.I] we have

$$[z^n]\left(\frac{1}{1-z}\right)^{\alpha t} = \frac{n^{\alpha t-1}}{\Gamma(\alpha t)}\left(1 + O\left(\frac{1}{n}\right)\right),$$

where the error term is uniform in t over compact subsets of complex numbers. The term $\left(\frac{2m}{m+1}\right)^{\alpha t}$ in (3.27) does not depend on z. In order to bound the contribution of the error term $\left(\frac{1}{1-z}\right)^{\alpha t} \cdot O(z-1)$ in (3.27) we use the following estimate [35, Theorem VI.3]:

$$[z^n]O\left(\left(\frac{1}{1-z}\right)^{s-1}\right) = O\left(n^{s-2}\right).$$

Hence

$$[z^n]\exp(tg_{\alpha,m}(z)) = \frac{n^{\alpha t-1}}{\Gamma(\alpha t)} \cdot \left(\frac{2m}{m+1}\right)^{\alpha t} \cdot \left(1 + O\left(\frac{1}{n}\right)\right)$$

and the theorem is proved. □

3.3 Mod-Poisson convergence

In this section we recall some facts about mod-Poisson convergence of probability distributions. As a direct corollary of Theorem 3.7 we derive mod-Poisson convergence of the probability distribution $q_{n,m,\alpha} = \mathbb{P}_{n,m,\alpha}\big(\mathrm{K}_n(\sigma) = k\big)$ of the number of cycles associated by Lemma 3.6 to the normalized weighted multi-variate harmonic sums $\widetilde{H}_{n,m,\alpha}(k)$. For details we refer to the monograph of Féray et al. [34] and, for the particular case of uniformly distributed random permutations, to the original article of Nikehgbali and Zeindler [71].

Given a probability distribution $p(k)$ of a random variable X taking values in non-negative integers, we associate to it the *generating series*

$$F_p(t) = \sum_{k=1}^{+\infty} p(k)t^k. \tag{3.28}$$

The generating series of the Poisson distribution defined in (1.4) is $e^{\lambda(t-1)}$.

Recall that given two independent discrete random variables with non-negative integer values X and Y with distributions $p_X(k)$ and $p_Y(k)$ respectively, the distribution of their sum $X+Y$ is the convolution

$$p_{X+Y}(k) = \sum_{i+j=k} p_X(i) \cdot p_Y(j).$$

The generating series of p_{X+Y} is the product of the generating series of p_X and p_Y:

$$F_{X+Y} = F_X F_Y. \tag{3.29}$$

We are particularly interested in the situations when we have a sequence of distributions that are close to the convolution of the Poisson distribution with a varying parameter λ_n which tends to $+\infty$ as $n \to +\infty$ and an additional fixed distribution.

Definition 3.9 Let p_n be a sequence of probability distributions on the non-negative integers, let λ_n be a sequence of positive real numbers tending to $+\infty$ as $n \to +\infty$. Moreover, let $R \in (1, +\infty]$, let $G(t)$ be an analytic function on the disk $|t| < R$ in $\mathbb{C}$ and let ε_n be a sequence of positive real numbers converging to zero. We say that p_n *converges mod-Poisson with parameters* λ_n*, limiting function* G*, radius* R *and speed* ε_n if for all $t \in \mathbb{C}$ such that $|t| < R$ we have

$$F_{p_n}(t) = e^{\lambda_n(t-1)} \cdot G(t) \cdot (1 + O(\varepsilon_n)), \tag{3.30}$$

where the error term $O(\varepsilon_n)$ is uniform in t on compact subsets of the complex disk $|t| < R$.

We say that a sequence X_n of random variables taking values in non-negative integers *converges mod-Poisson* if the sequence of the associated probability distributions p_n converges mod-Poisson, where $p_n(k) = \mathbb{P}(X_n = k)$ for $k = 0, 1, \dots$.

The term $e^{\lambda_n(t-1)} \cdot G(t)$ in the right hand side of (3.30) is the product of the generating series of Poi_{λ_n} with $G(t)$. In other words, it looks like (3.29). However, we emphasize that $G(t)$ is not necessarily the generating series of a probability distribution.

Note that Eq. (3.28) implies that for any n we have $F_{p_n}(1) = 1$. Thus, condition (3.30) from the definition of mod-Poisson convergence implies that

$$G(1) = 1. \tag{3.31}$$

Remark 3.10 Let us emphasize that our definition of mod-Poisson convergence differs from [34] in that we take generating series $\mathbb{E}(t^X)$ of random variables instead of the moment generating function $\mathbb{E}(e^{zX})$. One can pass from one to the other by setting $t = e^z$. In particular, our assumption that G is analytic at $t = 0$ is not a requirement in the definition of [34]. This extra assumption allows us to control the asymptotics of $p_n(k)$ when k is in the range $k \ll \log n$.

Let $\mathbb{P}_{n,m,\alpha}$ be the discrete probability measure on the symmetric group S_n corresponding to the weights $w_\theta(\sigma)$ associated to the sequence $\theta_i = \alpha \cdot \zeta_m(2i)$ for $i = 1, 2 \dots$ as defined in (1.7). Recall that Lemma 3.6 and, more specifically, Eq. (3.22) express the probability distribution $q_{n,m,\alpha}(k) = \mathbb{P}_{n,m,\alpha}\big(\mathrm{K}_n(\sigma) = k\big)$ through multivariate harmonic sums $\widetilde{H}_{n,m,\alpha}(k)$ defined in (3.3). The corollary below is a more general version of Theorem 1.8 from the introduction.

Corollary 3.11 *Let $\mathrm{K}_n(\sigma)$ be the number of cycles of a permutation σ in the symmetric group S_n. Let $\mathbb{E}_{n,m,\alpha}$ be the expectation with respect to the probability measure $\mathbb{P}_{n,m,\alpha}$ on S_n as in* (3.22)*. Then for all $t \in \mathbb{C}$ we have as $n \to +\infty$*

$$\mathbb{E}_{n,m,\alpha}\left(t^{\mathrm{K}_n}\right) = \left(\frac{2m}{m+1} \cdot n\right)^{\alpha(t-1)} \cdot \frac{\Gamma(\alpha)}{\Gamma(\alpha t)} \left(1 + O\left(\frac{1}{n}\right)\right), \tag{3.32}$$

Moreover, the convergence in (3.32) *is uniform in t on any compact subset of $\mathbb{C}$.*

In other words the sequence of random variables K_n with respect to the probability measures $\mathbb{P}_{n,m,\alpha}$ converges mod-Poisson with parameters $\lambda_n = \alpha \log\left(\frac{2m}{m+1} \cdot n\right)$, limiting function $\Gamma(\alpha)/\Gamma(\alpha t)$, radius $R = +\infty$ and speed $1/n$.

Proof of Corollary 3.11 Let us define

$$G_n(t) := \sum_{k=1}^{n} \widetilde{H}_{n,m,\alpha}(k) t^k.$$

Formula (3.28) for an abstract generating function combined with Formula (3.22) for $\mathbb{P}_{n,m,\alpha}\big(\mathrm{K}_n(\sigma) = k\big)$ gives the following expression for the generating series in the left-hand side of (3.32):

$$\mathbb{E}_{n,m,\alpha}\left(t^{\mathrm{K}_n(\sigma)}\right) = \frac{G_n(t)}{G_n(1)}$$

and the corollary now directly follows from Formula (3.24) from Theorem 3.7. □

Generalizing $u_{\lambda,1/2}$ given by (1.10) let us define

$$e^{\lambda(t-1)} \cdot \frac{t \cdot \Gamma(1+\alpha)}{\Gamma(1+t\alpha)} = \sum_{k\geq 1} u_{\lambda,\alpha}(k) \cdot t^k. \tag{3.33}$$

Corollary 3.12 *Let $m \in \mathbb{N} \cup \{\infty\}$ and let α be a positive real. We have*

$$q_{n,m,\alpha}(k) = u_{\lambda_n,\alpha}(k) + O\left(\frac{1}{n}\right) \quad \text{as } n \to +\infty$$

uniformly in $k \geq 1$, where $\lambda_n = \alpha \log\left(\frac{2m}{m+1} n\right)$.

Proof of Corollary 3.12 Note that

$$\frac{\Gamma(\alpha)}{\Gamma(\alpha t)} = \frac{\alpha t \cdot \Gamma(\alpha)}{\alpha t \cdot \Gamma(\alpha t)} = \frac{t \cdot \Gamma(1+\alpha)}{\Gamma(1+t\alpha)}.$$

Let

$$F_1(t) := \sum_{k\geq 1} q_{n,m,\alpha}(k) t^k ,$$

$$F_2(t) := e^{\lambda_n(t-1)} \cdot \frac{t \cdot \Gamma(1+\alpha)}{\Gamma(1+t\alpha)} = \sum_{k\geq 1} u_{\lambda_n,\alpha}(k) t^k.$$

Both $F_1(t)$ and $F_2(t)$ are holomorphic in $\mathbb{C}$. Since $F_2(t)$ does not vanish we have

$$F_1(t) - F_2(t) = O\left(\frac{1}{n}\right) \quad \text{as } n \to +\infty$$

uniformly in $t \in D(0, 1 + \varepsilon)$. Using the saddle-point bound (18) from [35, Proposition IV.1] with radius $R = 1$ we obtain that

$$q_{n,m,\alpha}(k) - u_{\lambda_n,\alpha}(k) = O\left(\frac{1}{n}\right) \quad \text{as } n \to +\infty$$

uniformly in $k \geq 1$. □

3.4 Large deviations and central limit theorem

Having proved the mod-Poisson convergence in Corollary 3.11, we could have derived most of the following large deviation results by referring to Theorem 3.2.2 from the monograph of Feray et al. [34] (see also Example 3.2.6 of the same monograph providing more details in the case of uniform random permutations). However, as we mentioned in Remark 3.10, the monograph [34] uses slightly weaker definition of mod-Poisson convergence which does not allow to study the probability distribution in the range of values of the random variable of the order $o(\lambda_n)$. To overcome this difficulty we rely on Theorem 14 in [43] and on Theorem 2 in [44] due to H. Hwang.

Theorem 3.13 (Hwang [43,44]) *Let $\{X_n\}_n$ be a sequence of random variables taking values in non-negative integers that converges mod-Poisson with parameters λ_n, limiting function $G(t)$, radius R and speed at least λ_n^{-1}. Assume furthermore that $G(0) \neq 0$. Then for any $x \in (0, R)$, we have*

$$\mathbb{P}(X_n = k) = e^{-\lambda_n} \frac{\lambda_n^k}{k!} \cdot (G(k/\lambda_n) + O((k+1)/(\lambda_n)^2)) \text{ as } n \to +\infty \quad (3.34)$$

uniformly in $k \in [0, \lambda_n]$.

For all $x \in (1, R)$ such that $x\lambda_n$ is an integer

$$\mathbb{P}(X_n > x\lambda_n) = \frac{e^{-\lambda_n(x \log x - x + 1)}}{\sqrt{2\pi\lambda_n x}} \cdot \frac{x}{x-1} \cdot (G(x) + O(\lambda_n^{-1})) \quad (3.35)$$

where the error term is uniform in x on compact subsets of $(1, R)$. Similarly, for all $x \in (0, 1)$ such that $x\lambda_n$ is an integer

$$\mathbb{P}(X_n \leq x\lambda_n) = \frac{e^{-\lambda_n(x \log x - x + 1)}}{\sqrt{2\pi\lambda_n x}} \cdot \frac{x}{1-x} \cdot (G(x) + O(\lambda_n^{-1})) \quad (3.36)$$

where the error term is uniform in x on compact subsets of $(0, 1)$.

Remark 3.14 Note that by Stirling formula, for $x = \frac{k}{\lambda_n}$ we have

$$\frac{e^{-\lambda_n(x\log x - x + 1)}}{\sqrt{2\pi x\lambda_n}} = e^{-\lambda_n}\frac{(\lambda_n)^k}{k!}(1 + O((\log n)^{-1})).$$

Note also that $x\log x - x + 1$ is convex and attains its minimum at $x = 1$ for which it has value zero. Hence both quantities in the right-hand sides of (3.35) and (3.36) are exponentially decreasing in n.

Remark 3.15 If the limiting function G vanishes at 0, we can apply the following trick. Let $a \in \mathbb{N}$ be the order of zero. Then the sequence of shifted variables $X_n - a$ converges mod-Poisson with the same parameters and radius but with the limiting function $t^{-a}G(t)$ which does not vanish anymore at zero. We can then apply Theorem 3.13 to $X_n - a$.

Remark 3.16 Since [34, Theorem 3.2.2] is stated for the more general mod-ϕ convergence let us explain how their notation translates in our context. Because we use Poisson variables, we have $\eta(t) = e^t - 1$ whose Legendre–Fenchel transform is $F(x) = x\log x - x - 1$. Because of this $h = \log x$. The limiting function is $\phi(e^z) = G(z)$. This difference of notation for the limiting functions is due to the fact that we used generating series $\mathbb{E}(t^X)$ instead of the moment generating functions $\mathbb{E}(e^{zX})$.

The statement below is a generalization of Theorem 1.10 from Sect. 1 to an arbitrary probability measure $\mathbb{P}_{n,m,\alpha}$.

Corollary 3.17 *Let $\alpha > 0$, let $m \in \mathbb{N} \cup \{+\infty\}$. Let $\mathbb{P}_{n,m,\alpha}$ be the probability measure as in* (3.22) *and let $\lambda_n := \alpha \log\left(\frac{2m}{m+1}n\right)$. Then for any $x > 0$ we have*

$$\begin{aligned} q_{n,m,\alpha}(k+1) &= \mathbb{P}_{n,m,\alpha}(\mathrm{K}_n = k+1) \\ &= e^{-\lambda_n}\frac{(\lambda_n)^k}{k!}\left(\frac{\Gamma(1+\alpha)}{\Gamma\left(1+\alpha\frac{k}{\lambda_n}\right)} + O\left(\frac{k+1}{(\log n)^2}\right)\right) \text{ as } n \to +\infty \end{aligned} \tag{3.37}$$

uniformly in $k \in [0, x\lambda_n]$.

For $x \in (1, +\infty)$ such that $x\lambda_n$ is an integer we have

$$\sum_{k=x\lambda_n+1}^{n} q_{n,m,\alpha}(k+1) = \mathbb{P}_{n,m,\alpha}\big(\mathrm{K}_n > x\lambda_n + 1\big)$$
$$= \frac{e^{-\lambda_n(x\log x - x + 1)}}{\sqrt{2\pi x\lambda_n}} \cdot \frac{x}{x-1}\left(\frac{\Gamma(1+\alpha)}{\Gamma(1+\alpha x)} + O\left(\frac{1}{\log n}\right)\right) \text{ as } n \to +\infty, \tag{3.38}$$

where the error term is uniform in x on compact subsets of $(1, +\infty)$.

For $x \in (0, 1)$ such that $x\lambda_n$ is an integer we have

$$\sum_{k=0}^{x\lambda_n} q_{n,m,\alpha}(k+1) = \mathbb{P}_{n,m,\alpha}\big(\mathrm{K}_n \le x\lambda_n + 1\big)$$
$$= \frac{e^{-\lambda_n(x\log x - x + 1)}}{\sqrt{2\pi x\lambda_n}} \cdot \frac{x}{1-x}\left(\frac{\Gamma(1+\alpha)}{\Gamma(1+\alpha x)} + O\left(\frac{1}{\log n}\right)\right) \text{ as } n \to +\infty, \tag{3.39}$$

where the error term is uniform in x on compact subsets of $(0, 1)$.

Proof By Corollary 3.11, the sequence of random variables $\mathrm{K}_n(\sigma)$ with respect to the probability measures $\mathbb{P}_{n,m,\alpha}$ on the symmetric group S_n converges mod-Poisson with parameters $\lambda_n = \alpha \log\left(\frac{2m}{m+1}n\right)$, limiting function $\Gamma(\alpha)/\Gamma(\alpha t)$, radius $R = +\infty$ and speed $1/n$. The limiting function $\Gamma(\alpha)/\Gamma(\alpha t)$ has zero of the first order at $t = 0$, so we have to apply the trick described in Remark 3.15. The sequence of random variables $\mathrm{K}_n - 1$ converges mod-Poisson with the same radius $R = +\infty$ and speed $1/n$ and has the limiting function $\Gamma(\alpha)/(t \cdot \Gamma(\alpha t))$. Applying the identity $\Gamma(z+1) = z\Gamma(z)$ we conclude that the new limiting function

$$G(t) = \frac{\Gamma(\alpha)}{t \cdot \Gamma(\alpha t)} = \frac{\alpha\Gamma(\alpha)}{\Gamma(1+\alpha t)} = \frac{\Gamma(1+\alpha)}{\Gamma(1+\alpha t)}$$

does not vanish at $t = 0$ and Theorem 3.13 becomes applicable to the sequence of random variables $X_n - 1$. □

Corollary 3.18 *Let α be a positive real number and let $m \in \mathbb{N} \cup \{+\infty\}$. Let $\widetilde{H}_{n,m,\alpha}$ be the normalized weighted multi-variate harmonic sum (3.3), and let $\{k_n\}_n$ be a sequence of integers such that $k_n = O(\log n)$. Then as $n \to +\infty$*

we have

$$\widetilde{H}_{n,m,\alpha}(k_n) = \frac{\alpha^{k_n}}{n}\frac{\left(\log n + \log\left(\frac{2m}{m+1}\right)\right)^{k_n-1}}{(k_n-1)!}\cdot\left(\frac{1}{\Gamma\left(1+\alpha\frac{k_n-1}{\lambda_n}\right)} + O\left(\frac{k_n-1}{(\log n)^2}\right)\right). \tag{3.40}$$

If, moreover, $k_n = o(\log n)$*, then as* $n \to +\infty$ *we have*

$$\widetilde{H}_{n,m,\alpha}(k_n) = \frac{\alpha^{k_n}}{n}\frac{\left(\log n + \log\left(\frac{2m}{m+1}\right)\right)^{k_n-1}}{(k_n-1)!}\left(1 + \frac{\gamma\cdot(k_n-1)}{\log n} + O\left(\left(\frac{k_n}{\log n}\right)^2\right)\right). \tag{3.41}$$

Proof of Corollary 3.18 Applying (3.37) with $\lambda_n = \alpha\log\left(\frac{2m}{m+1}n\right)$ we get

$$\begin{aligned}
q_{n,m,\alpha}(k_n) &= e^{-\lambda_n}\frac{(\lambda_n)^{k_n-1}}{(k_n-1)!}\left(\frac{\Gamma(1+\alpha)}{\Gamma\left(1+\alpha\frac{k_n-1}{\lambda_n}\right)} + O\left(\frac{k_n-1}{(\log n)^2}\right)\right)\\
&= \left(\frac{2m}{m+1}\right)^{-\alpha} n^{-\alpha}\frac{\left(\alpha\log\left(\frac{2m}{m+1}n\right)\right)^{k_n-1}}{(k_n-1)!}\\
&\quad\times\left(\frac{\Gamma(1+\alpha)}{\Gamma\left(1+\alpha\frac{k_n-1}{\lambda_n}\right)} + O\left(\frac{k_n-1}{(\log n)^2}\right)\right).
\end{aligned}$$

Applying Eq. (3.24) with $t = 1$, we get

$$W_{n,m,\alpha} = \sum_{k=1}^{n}\widetilde{H}_{n,m,\alpha}(k) = \frac{\alpha\left(\frac{2m}{m+1}\right)^{\alpha} n^{\alpha-1}}{\Gamma(1+\alpha)}\left(1 + O\left(\frac{1}{n}\right)\right),$$

where we used the identity $\alpha\Gamma(\alpha) = \Gamma(1+\alpha)$. By definition (3.22) of $q_{n,m,\alpha}(k)$ we have

$$\widetilde{H}_{n,m,\alpha}(k) = q_{n,m,\alpha}(k)\cdot W_{n,m,\alpha}$$

Multiplying the two expressions computed above we get (3.40).

To prove (3.41) we use the asymptotic expansion

$$\frac{1}{\Gamma(1+t)} = 1 + \gamma t + O(t^2) \quad \text{as} \quad t \to 0,$$

where $\gamma = 0.5572\ldots$ denotes the Euler–Mascheroni constant. □

Note that for the values of parameters $m = \alpha = 1$ and for the constant sequence $k_n = 2$ for $n = 1, 2, \ldots$, the expansion (3.41) gives

$$\widetilde{H}_{n,1,1}(k_n) = \frac{1}{n}\log n \left(1 + \frac{\gamma}{\log n} + O\left(\frac{1}{(\log n)^2}\right)\right)$$

in agreement with the classical formula

$$\frac{1}{2}\sum_{j=1}^{n-1} \frac{1}{j \cdot (n-j)} = \frac{1}{n}\left(\log n + \gamma + O\left(\frac{1}{n}\right)\right).$$

The following strong form of the central limit theorem corresponds to Theorem 3.3.1 of [34].

Theorem 3.19 (Féray et al. [34]) *Let $\{X_n\}$ be a sequence of random variables on the non-negative integers that converges mod-Poisson with parameters λ_n. Let x_n be a sequence of real numbers with $x_n = o((\lambda_n)^{1/6})$. Then as $n \to +\infty$*

$$\mathbb{P}\left(\frac{X_n - \lambda_n}{\sqrt{\lambda_n}} \le x_n\right) = \left(\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x_n} e^{-\frac{t^2}{2}}\,dt\right)(1 + o(1)).$$

Note that, contrarily to the large deviations, the radius R, the limiting function G and the speed ε_n of the mod-Poisson convergence are irrelevant in the above Theorem.

Corollary 3.20 *Let $\alpha > 0$, $m \in \mathbb{N} \cup \{+\infty\}$ and let $\mathbb{P}_{n,m,\alpha}$ be the probability distribution on the symmetric group defined in (3.22). Let $\lambda_n := \alpha \log\left(\frac{2m}{m+1}n\right)$ and x_n be a sequence of real numbers with $x_n = o((\lambda_n)^{1/6})$. Then as $n \to +\infty$*

$$\mathbb{P}_{n,m,\alpha}\left(\frac{\mathrm{K}_n - \lambda_n}{\sqrt{\lambda_n}} \le x_n\right) = \left(\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x_n} e^{-\frac{t^2}{2}}\,dt\right)(1 + o(1)).$$

Proof By Corollary 3.11, the sequence of random variables K_n converges mod-Poisson, so Theorem 3.19 is applicable to this sequence. □

3.5 Moments of the Poisson distribution

Recall that given a non-negative integer n and a positive real number λ, the n-th moment $P_n(\lambda)$ of a random variable corresponding to the Poisson distribution Poi_λ with parameter λ is defined as

$$P_n(\lambda) := e^{-\lambda} \cdot \sum_{k=0}^{+\infty} k^n \frac{\lambda^k}{k!}. \tag{3.42}$$

Recall that given two integers n, k satisfying $1 \le k \le n$, the *Stirling number of the second kind*, denoted $S(n, k)$, is the number of ways to partition a set of n objects into k non-empty subsets. It is well-known that the Stirling numbers of the second kind satisfy the following recurrence relation:

$$S(n+1, k) = k \cdot S(n, k) + S(n, k-1), \tag{3.43}$$

and are uniquely determined by the initial conditions, where by convention we set: $S(0, 0) = 1$ and $S(n, -1) = S(n, 0) = S(0, n) = S(n, n+1) = 0$ for $n \in \mathbb{N}$.

Though the following statement is well-known, see, for example, [75], its proof is so short that we present it for the sake of completeness.

Lemma 3.21 *For any $n \in \mathbb{N}$, the expression $P_n(\lambda)$ defined in (3.42) coincides with the following monic polynomial in λ of degree n:*

$$P_n(\lambda) = \sum_{k=0}^{n} S(n, k)\lambda^k, \tag{3.44}$$

where $S(n, k)$ are the Stirling numbers of the second kind.

The polynomials $P_n(\lambda)$ are sometimes called *Touchard polynomials*, *exponential polynomials* or *Bell polynomials*. For $n \le 4$ the polynomials $P_n(\lambda)$ have the following explicit form:

$$\begin{aligned}
P_0(\lambda) &= 1, \\
P_1(\lambda) &= \lambda, \\
P_2(\lambda) &= \lambda^2 + \lambda, \\
P_3(\lambda) &= \lambda^3 + 3\lambda^2 + \lambda, \\
P_4(\lambda) &= \lambda^4 + 6\lambda^3 + 7\lambda^2 + \lambda.
\end{aligned} \tag{3.45}$$

Proof of Lemma 3.21 Let X be a random variable with distribution Poi_λ and let

$$\phi(z) = \mathbb{E}(e^{zX}) = \sum_{n=0}^{+\infty} \mathbb{E}(X^n)\frac{z^n}{n!}$$

be its moment generating series. Then

$$\phi(z) = \sum_{k\geq 0} e^{-\lambda}\frac{\lambda^k}{k!}e^{zk} = e^{-\lambda}\sum_{k\geq 0}\frac{(\lambda e^z)^k}{k!} = e^{\lambda(e^z-1)}.$$

By definition, $P_n(\lambda) = \frac{d^n}{dz^n}\phi(z)|_{z=0}$. We claim that for any $n = 0, 1, \ldots$ the following identity holds:

$$\frac{d^n}{dz^n}\phi(z) = \sum_{k=0}^{n} S(n,k)\cdot(\lambda e^z)^k\cdot\phi(z). \tag{3.46}$$

Indeed, $\frac{d}{dz}\phi(z) = \lambda e^z\phi(z)$ and the identity holds for $n = 0$ and $n = 1$. Taking the derivative of the expression in the right hand side of (3.46) we obtain

$$\begin{aligned}
&\frac{d}{dz}\sum_{k=0}^{n} S(n,k)\cdot\left(\lambda e^z\right)^k\cdot\phi(z)\\
&\quad=\sum_{k=0}^{n} S(n,k)\cdot\left(k\cdot\left(\lambda e^z\right)^k+\left(\lambda e^z\right)^{k+1}\right)\cdot\phi(z)\\
&\quad=\sum_{k=0}^{n+1}\Big(S(n,k-1)+k\cdot S(n,k)\Big)\cdot(\lambda e^z)^k\cdot\phi(z).
\end{aligned}$$

We recognize the recurrence relations (3.43) for Stirling numbers of the second kind, which proves identity (3.46). Taking $z = 0$ in (3.46) we obtain (3.44). $\square$

3.6 Moment expansion

In this section we analyze the asymptotic expansions of cumulants of probability distributions that satisfy mod-Poisson convergence. We then apply it to the probability distribution $q_{n,m,\alpha}(k) = \frac{\widetilde{H}_{n,m,\alpha}(k)}{W_{n,m,\alpha}}$ (see Definition 3.2 and (3.22)).

The *cumulants* $\kappa_i(X)$ of a random variable X are the coefficients of the expansion

$$\log \mathbb{E}(e^{tX}) = \sum_{i\geq 1} \kappa_i(X) \frac{t^i}{i!}.$$

The first cumulant $\kappa_1(X) = \mathbb{E}(X)$ is the mean and the second cumulant $\kappa_2(X) = \mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$ is the variance. The cumulants are combinations of moments, but contrarily to moments, cumulants are additive: if X and Y are independent then $\kappa_i(X+Y) = \kappa_i(X) + \kappa_i(Y)$.

If X is a Poisson random variable with parameter λ then

$$\log \mathbb{E}(e^{tX}) = \lambda(e^t - 1).$$

This implies that all cumulants of a Poisson random variable are equal to λ. Theorem 3.22 below proves that when a sequence of random variables converges mod-Poisson, the main contribution to the cumulants comes from the Poisson part while an explicit correction comes from the logarithmic derivative of the limiting function.

Theorem 3.22 *Let X_n be a sequence of probability distributions that converges mod-Poisson with parameters λ_n limiting function G and speed ε_n as $n \to +\infty$. Then for all $i \geq 1$ we have the following asymptotic formula for the i-th cumulant*

$$\kappa_i(X_n) = \lambda_n + \sum_{k=1}^{i} S(i,k) \cdot \delta_k + O(\varepsilon_n) \quad \text{as } n \to +\infty, \tag{3.47}$$

where $S(i,k)$ are the Stirling numbers of the second kind and $\delta_k = \frac{d^k}{dt^k} \log G(t)|_{t=1}$ are the values of the logarithmic derivatives of the limiting function G at $t = 1$.

We warn the reader that the error term $O(\varepsilon_n)$ in (3.47) is uniform in n but not in i.

Remark 3.23 Note that the Theorem 3.22 above is valid for any radius of convergence R as soon as $R > 1$. This condition is a part of Definition 3.9 of mod-Poisson convergence.

Proof By Definition 3.9 of the mod-Poisson convergence we have

$$\mathbb{E}(e^{zX_n}) = e^{\lambda_n(e^z - 1)} G(e^z)(1 + O(\varepsilon_n)),$$

see (3.30). We have seen in (3.31) that our definition of mod-Poisson convergence implies that $G(1) = 1$. Hence, there exists a radius R' such that for $|z| < R'$ we have $G(e^z) \notin [-\infty, 0)$. On the disk $|z| < R'$ we can take the principal branch of the logarithm to obtain

$$\log \mathbb{E}(e^{zX_n}) = \lambda_n(e^z - 1) + \log G(e^z) + O(\varepsilon_n).$$

which can be rewritten as

$$\sum_{i \geq 1} (\kappa_i(X_n) - \lambda_n - \Delta_i) \cdot \frac{z^i}{i!} = O(\varepsilon_n) \tag{3.48}$$

where $\Delta_i := \frac{d^i}{dz^i} \log G(e^z)|_{z=0}$. Let $g^{(i)}(t) = \frac{d^i}{dt^i} \log G(t)$. The rest of the proof is similar to the proof of Lemma 3.21. Namely, we claim that for $i \geq 1$ we have

$$\frac{d^i}{dz^i} \log G(e^z) = \sum_{k=1}^{i} S(i,k) g^{(k)}(e^z) e^{kz}. \tag{3.49}$$

It is indeed the case for $i = 1$ and when differentiating once the formula in the right hand side we obtain

$$\begin{aligned} &\frac{d}{dz} \sum_{k=1}^{i} S(i,k) g^{(k)}(e^z) e^{kz} \\ &\quad = \sum_{k=1}^{i} (S(i,k) g^{(k+1)}(e^z) e^{(k+1)z} + kS(i,k) g^{(k)}(e^z) e^{kz}) \\ &\quad = \sum_{k=1}^{i+1} (S(i,k-1) + kS(i,k)) g^{(k)}(e^z) e^{kz}. \end{aligned}$$

We recognize the recurrence relation (3.43) for the unsigned Stirling numbers of the second kind. This proves the claim. Now let $\delta_i = g^{(i)}(1)$. Specializing (3.49) at $z = 0$ we obtain $\Delta_i = \sum_{k=1}^{i} S(i,k)\delta_k$.

Now, since the radius of convergence in (3.48) is positive, we obtain

$$\kappa_i(X_n) - \lambda_n - \delta_i(\alpha) = O(\varepsilon_n)$$

(where the error term depends on i). This concludes the proof. □

Recall that for $m \geq 0$, the m-th polygamma function is defined as

$$\psi^{(m)}(z) = \frac{d^{m+1}}{dz^{m+1}} \log \Gamma(z). \tag{3.50}$$

Corollary 3.24 *Let $m \in \mathbb{N} \cup \{+\infty\}$, let $\alpha \geq 0$ and let K_n be the random variable corresponding to the probability law $q_{n,m,\alpha}$ defined in (3.22). Then for any $i \in \mathbb{N}$ we have the following asymptotics for the i-th cumulant of K_n as $n \to +\infty$:*

$$\kappa_i(\mathrm{K}_n) = \alpha \log\left(\frac{2m}{m+1} \cdot n\right) - \sum_{k=1}^{i} S(i,k) \cdot \psi^{(k-1)}(\alpha) \cdot \alpha^k + O\left(\frac{1}{n}\right), \quad (3.51)$$

where $S(i,k)$ is the Stirling number of the second kind and $\psi^{(j)}$ is the polygamma function.

Proof By Corollary 3.11, the random variables K_n with respect to $\mathbb{P}_{n,m,\alpha}$ converges mod-Poisson with parameters $\lambda_n = \alpha \log\left(\frac{2m}{m+1} n\right)$, limiting function $G(t) = \Gamma(\alpha)/\Gamma(\alpha t)$ and rate $O(1/n)$. The logarithmic derivatives of the limiting function are expressed in terms of the polygamma function by the following relation:

$$\frac{d^k}{dt^k} \log \frac{\Gamma(\alpha)}{\Gamma(\alpha t)} = -\frac{d^k}{dt^k} \log \Gamma(\alpha t) = -\Gamma(\alpha) \cdot \alpha^k \cdot \psi^{(k-1)}(\alpha t).$$

Applying Eq. (3.47) from Theorem 3.22 we obtain the desired relation (3.51). □

The derivatives of the polygamma functions at rational points have explicit expressions in terms of ζ-values. The following lemma provides the values of these derivatives at 1 and at 1/2 relevant for the purposes of the present paper. These formulae reproduce Formulae 6.4.2 and 6.4.4, at page 260 of [1]. The proofs can be found, for example, in the paper [22] of Choi and Cvijović.

Lemma 3.25 *We have*

$$\psi^{(0)}(1) = -\gamma, \qquad \psi^{(0)}(1/2) = -\gamma - 2\log 2$$

and for $m \geq 1$

$$\begin{aligned} \psi^{(m)}(1) &= (-1)^{m+1} \zeta(m+1)\, m!, \\ \psi^{(m)}(1/2) &= (-1)^{m+1} \zeta(m+1)\, m! \left(2^{m+1} - 1\right). \end{aligned}$$

Remark 3.26 In the special case $m = 1$ and $\alpha = 1$ which corresponds to the uniform distribution on S_n we obtain

$$\begin{aligned}
\kappa_1(q_{n,1,1}) &= \log n + \gamma + O(1/n),\\
\kappa_2(q_{n,1,1}) &= \log n + \gamma - \zeta(2) + O(1/n),\\
\kappa_3(q_{n,1,1}) &= \log n + \gamma - 3\zeta(2) + 2\zeta(3) + O(1/n),\\
\kappa_4(q_{n,1,1}) &= \log n + \gamma - 7\zeta(2) + 12\zeta(3) - 4\zeta(4) + O(1/n).
\end{aligned}$$

We recover the expression (1.2) obtained by Goncharov [37] for the expectation and variance of the cycle count of uniform random permutations in S_n.

3.7 From $q_{3g-3,\infty,1/2}$ to $p_g^{(1)}$

Recall that Vol $\left(\Gamma_k(g), (m_1, \ldots, m_k)\right)$ denotes the volume contribution from those square-tiled surfaces corresponding to the stable graph $\Gamma_k(g)$ for which the first maximal horizontal cylinder is filled with m_1 bands of squares, the second cylinder is filled with m_2 bands of squares, and so on up to the k-th maximal horizontal cylinder, which is filled with m_k bands of squares. Recall also that for any $m \in \mathbb{N} \cup \{\infty\}$ we defined in (3.9) the quantities

$$\begin{aligned}
V_{m,k}(g) &= \sum_{\substack{m_1,\ldots,m_k \\ 1\le m_i \le m \text{ for } i=1,\ldots,k}} \mathrm{Vol}(\Gamma_k(g), (m_1, \ldots, m_k)) \quad \text{and}\\
V_m(g) &= \sum_{k=1}^{g} V_{m,k}(g).
\end{aligned}$$

We define the probability distribution $p^{(1)}_{g,m}(k)$ for $k = 1, \ldots, g$ as

$$p^{(1)}_{g,m}(k) := \frac{V_{m,k}(g)}{V_m(g)}. \tag{3.52}$$

We will sometimes denote $p^{(1)}_{g,\infty}$ just by $p^{(1)}_g$. In this section we use estimates (3.4) and (3.5) obtained in Theorem 3.4 for $V_{m,k}(g)$ and our study of the normalized weighted harmonic sums $\widetilde{H}_{n,m,\alpha}$ performed in the previous sections to deduce properties of the probability distribution $p^{(1)}_{g,m}$. We now state and prove a lemma that we will use in the proof of Theorem 3.1.

Proposition 3.27 *Let m be in $\mathbb{N} \cup \{+\infty\}$. For any $t \in \mathbb{C}$ satisfying $|t| < 2$ we have the following estimates as $n \to +\infty$*

$$\sum_{k=\lceil 22\cdot\log n\rceil+1}^{n} \widetilde{H}_{n,m,9/8}(k)|t|^k = \left|\sum_{k=1}^{n} \widetilde{H}_{n,m,1/2}(k)t^k\right| \cdot o\left(n^{-1}\right), \tag{3.53}$$

$$\sum_{k=\lceil 22\cdot\log n\rceil+1}^{n} \widetilde{H}_{n,m,1/2}(k)|t|^k = \left|\sum_{k=1}^{n} \widetilde{H}_{n,m,1/2}(k)t^k\right| \cdot o\left(n^{-1}\right), \tag{3.54}$$

where the error term is uniform in t on compact subsets of the complex disk $|t| < 2$.

Proof It follows from definition (3.3) of the weighted multi-variate harmonic sum $\widetilde{H}_{n,m,\alpha}(k)$ that for any n, m, k we have $\widetilde{H}_{n,m,1/2}(k) \le \widetilde{H}_{n,m,9/8}(k)$. Thus, estimate (3.53) implies estimate (3.54) and it is sufficient to prove (3.53).

We consider separately the cases $|t| \le 1/e$ and $1/e \le |t| < 2$. We start with the case $|t| \le 1/e$. Using the fact that we have a generating series of a probability distribution, and that for $|t| \in [0, 1/e]$ and positive k the power $|t|^k$ is bounded from above by e^{-k}, we get the following estimate valid for any $|t| \in [0, 1/e]$ and any $m \in \mathbb{N} \cup \{+\infty\}$:

$$\begin{aligned}
&\frac{1}{W_{n,m,9/8}} \sum_{k=\lceil 22\cdot\log n\rceil+1}^{n} \widetilde{H}_{n,m,9/8}(k)\ |t|^k \\
&\le \frac{1}{W_{n,m,9/8}} \sum_{k=\lceil 22\log n\rceil+1}^{n} \widetilde{H}_{n,m,9/8}(k)\ |t|^{1+22\log n} \\
&\le \left(\frac{1}{W_{n,m,9/8}} \sum_{k=1}^{n} \widetilde{H}_{n,m,9/8}(k)\right) \cdot |t| \cdot e^{-22\log n} \\
&= |t| \cdot n^{-22},
\end{aligned} \tag{3.55}$$

where $W_{n,m,9/8}$ is the sum over k of $\widetilde{H}_{n,m,9/8}(k)$ as defined in (3.23). On the other hand, using the identity $z\Gamma(z) = \Gamma(1+z)$ and applying Eq. (3.24) from Theorem 3.7 for $\alpha = 1/2$ and $\alpha = 9/8$, respectively, we have

$$\sum_{k=1}^{n} \widetilde{H}_{n,m,1/2}(k)\ t^k = t \cdot n^{-t/2} \cdot \frac{\left(\frac{2m}{m+1}\right)^{t/2}}{2\Gamma(1+t/2)} \left(1 + O\left(\frac{1}{n}\right)\right),$$

$$\sum_{k=1}^{n} \widetilde{H}_{n,m,9/8}(k)\ t^k = t \cdot n^{t/8} \cdot \frac{9\left(\frac{2m}{m+1}\right)^{9t/8}}{8\Gamma(1+9t/8)} \left(1 + O\left(\frac{1}{n}\right)\right), \tag{3.56}$$

as $n \to +\infty$, where the error terms are uniform in t over the compact complex disk $|t| \le 2$. In particular, for $\alpha = 9/8$ and $t = 1$ we get

$$W_{n,m,9/8} \sim \frac{\left(\frac{2m}{m+1}\right)^{9/8} n^{1/8}}{\Gamma(9/8)}.$$

The latter asymptotics combined with (3.55) imply that uniformly for t, such that $|t| \in [0, 1/e]$, we have:

$$\sum_{k=\lceil 22\cdot\log n\rceil+1}^{n} \widetilde{H}_{n,m,9/8}(k)\ |t|^k = O(n^{1/8}) \cdot |t| \cdot n^{-22}. \tag{3.57}$$

Recall that $1/\Gamma(1+z)$ is an entire function having zeroes precisely at the negative integers. Thus, for all $m \in \mathbb{N} \cup \{\infty\}$ and for all t satisfying $|t| \le 1/e$ we have

$$\min_{|t|\le 1/e} \left| \frac{\left(\frac{2m}{m+1}\right)^{t/2}}{2\Gamma(1+t/2)} \right| = C > 0.$$

Together with (3.56) this bound implies that for $|t| \le 1/e$ and sufficiently large n we have

$$\left| \sum_{k=1}^{n} \widetilde{H}_{n,m,1/2}(k)\ t^k \right| \ge 2C \cdot |t| \cdot n^{|t|/2} \ge 2C \cdot |t| \cdot n^{1/(2e)}.$$

The inequality above combined with asymptotic estimate (3.57) implies the desired relation (3.53) for $|t| \le 1/e$.

We now prove (3.53) for $|t| \ge 1/e$. In this case the proof is based on Corollary 3.17.

Choose any real parameter R satisfying $1/e < R < 2$. From now on we assume that the complex variable t belongs to the closed annulus $1/e \le |t| \le R$. All the estimates below are uniform in $t \in [1/e, R]$, but the constants might depend on the choice of R.

We start with several preparatory remarks. We consider the function

$$f(y) = \frac{1}{2}(y \log y - y + 1),$$

where $y > 0$. We note that the function f is strictly convex with a minimum at $y = 1$, where $f(1) = 0$. We will need the following inequalities for $f(44/9)$:

$$f(44/9) > 1, \tag{3.58}$$

and

$$\max_{1/e \le |t| \le 2} \left(\frac{13}{8}|t| - \frac{9}{4} f(44/9) \cdot |t| \right) = \left(\frac{13}{8}|t| - \frac{9}{4} f(44/9) \cdot |t| \right) \Bigg|_{|t|=1/e} < -1. \tag{3.59}$$

We denote $\lambda_{m,n} = \frac{\log\left(\frac{2m}{m+1} \cdot n\right)}{2}$. For $n \ge 3$ and any $m \in \mathbb{N} \cup \{\infty\}$ we have

$$\frac{\log n}{2} \le \frac{\log\left(\frac{2m}{m+1} \cdot n\right)}{2} < \log n, \tag{3.60}$$

which implies, in particular, that for real positive y we have

$$e^{-\lambda_{m,n}(y \log y - y + 1)} \le e^{-\frac{\log n}{2}(y \log y - y + 1)} = n^{-f(y)}. \tag{3.61}$$

The next remark is particularly important for the proof. It follows directly from definition (3.3) of the weighted multi-variate harmonic sum $\widetilde{H}_{n,m,\alpha}(k)$ that

$$\widetilde{H}_{n,m,\alpha}(k)\, t^k = \widetilde{H}_{n,m,\alpha t}(k). \tag{3.62}$$

We also get

$$W_{n,m,\alpha t} = \sum_{k=1}^{n} \widetilde{H}_{n,m,\alpha t}(k) = \sum_{k=1}^{n} \widetilde{H}_{n,m,\alpha}(k)\, t^k.$$

Using this remark we can rewrite the asymptotic estimate (3.53) (which we aim to prove for $|t| \in [1/e, R]$) in the following equivalent form:

$$\frac{1}{|W_{n,m,t/2}|} \sum_{k=\lceil 22 \log n \rceil + 1}^{n} \widetilde{H}_{n,m,|9t/8|}(k) \stackrel{?}{=} o(n^{-1}). \tag{3.63}$$

Now everything is ready for the proof of Proposition 3.27 for $|t| \in [1/e, R]$. By Theorem 3.7 for $\alpha = 1/2$ and $\alpha = 9/8$ we have

$$W_{n,m,\alpha t} = \frac{\alpha t \left(\frac{2m}{m+1}\right)^{\alpha t} n^{\alpha t - 1}}{\Gamma(1+\alpha t)} \left(1 + O\left(\frac{1}{n}\right)\right) = O\left(n^{\alpha t - 1}\right) \quad \text{as } n \to +\infty \tag{3.64}$$

uniformly in t in the annulus $1/e \leq |t| \leq R$. Recall definition (3.22) of $q_{n,m,\alpha}(k)$ and apply estimate (3.38) from Corollary (3.17), where we let $\alpha = 9|t|/8$. Under such choice of α, the variable λ_n, present in (3.38), takes the value $\lambda_n = \frac{9|t|}{8} \log\left(\frac{2m}{m+1} n\right) = (9|t|/4)\lambda_{m,n}$. For any $y > 1$ we have

$$\begin{aligned}\frac{1}{W_{n,m,9|t|/8}} \sum_{k=\lceil y\lambda_n \rceil+1}^{n} \widetilde{H}_{n,m,9|t|/8}(k) &= \sum_{k=\lceil y\cdot(9|t|/4)\cdot\lambda_{m,n}\rceil+1}^{n} q_{n,m,9|t|/8}(k) \\ &= \frac{e^{-(9|t|/4)\cdot\lambda_{m,n}(y\log y - y + 1)}}{\sqrt{2\pi\cdot y\cdot(9|t|/4)\lambda_{m,n}}} \frac{y}{(y-1)} \cdot O(1) = o\left(n^{-\frac{9|t|}{4}\cdot f(y)}\right),\end{aligned} \tag{3.65}$$

where we used (3.61) for the rightmost equality.

Note that $|t| \leq R < 2$. This implies that

$$\min_{1/e \leq |t| \leq R} \left| \frac{t\left(\frac{2m}{m+1}\right)^{\alpha t}}{2\Gamma(1+t/2)} \right| = C'(R) > 0.$$

This observation together with (3.64) imply that

$$\frac{W_{n,m,9|t|/8}}{|W_{n,m,t/2}|} = O\left(n^{\frac{9|t|}{8} + \frac{|t|}{2}}\right) = O\left(n^{\frac{13|t|}{8}}\right).$$

uniformly in t such that $|t| \in [1/e, R]$. Combining the latter estimate with (3.65) we obtain

$$\begin{aligned}&\frac{1}{|W_{n,m,t/2}|} \sum_{k=\lceil y\lambda_n\rceil+1}^{n} \widetilde{H}_{n,m,9|t|/8}(k) \\ &= \frac{W_{n,m,9|t|/8}}{|W_{n,m,t/2}|} \cdot \frac{1}{W_{n,m,9|t|/8}} \sum_{k=\lceil y\lambda_n\rceil+1}^{n} \widetilde{H}_{n,m,9|t|/8}(k) \\ &= o\left(n^{\frac{13|t|}{8}}\right) \cdot n^{-\frac{9|t|}{4}\cdot f(y)}.\end{aligned}$$

We now choose $y = 44/9$. Applying (3.59), we conclude that we have

$$\frac{1}{|W_{n,m,t/2}|} \sum_{k=\left\lceil \frac{44}{9}\lambda_n \right\rceil + 1}^{n} \widetilde{H}_{n,m,9|t|/8}(k) = o(n^{-1}) \tag{3.66}$$

uniformly in t such that $|t| \in [1/e, R]$ It remains to note that for $|t| \le R < 2$ and for $n \ge 3$ we have

$$\frac{44}{9}\lambda_n = \frac{44}{9} \cdot \frac{9}{4} \cdot |t| \cdot \frac{\log\left(\frac{2m}{m+1} \cdot n\right)}{2} < 11 \cdot |t| \cdot \log n < 22 \cdot \log n.$$

This implies that the sum on the left-hand side of (3.63) is contained in the sum on the left-hand side of (3.66). Thus, (3.66) implies (3.63) and, hence, it implies the equivalent estimate (3.53) in the case $1/e \le |t| \le R < 2$. □

Now everything is ready to prove Theorem 3.1.

Proof of Theorem 3.1 The main ingredients of the proof are the asymptotic equivalence (3.4) and the upper bound (3.5) from Theorem 3.4 combined with Proposition 3.27. We use abbreviation (3.9). We split the sum (3.1) into two parts $\sum_{k=1}^{g} V_{m,k}(g) \cdot t^k = \Sigma_1 + \Sigma_2$, where

$$\Sigma_1 = \sum_{k=1}^{\lceil 22\cdot\log(3g-3)\rceil} V_{m,k}(g) \cdot t^k, \qquad \Sigma_2 = \sum_{k=\lceil 22\cdot\log(3g-3)\rceil+1}^{g} V_{m,k}(g) \cdot t^k$$

and evaluate the two sums separately. Using (3.4) from Theorem 3.4 we get

$$\Sigma_1 = \frac{2\sqrt{2}}{\sqrt{\pi}} \cdot \sqrt{3g-3}\left(\frac{8}{3}\right)^{4g-4} \times \left(\sum_{k=1}^{\lceil 22\cdot\log(3g-3)\rceil} \widetilde{H}_{3g-3,m,1/2}(k) t^k\right)\left(1 + O\left(\frac{(\log g)^2}{g}\right)\right). \tag{3.67}$$

Applying (3.54) from Proposition 3.27 combined with (3.24) from Theorem 3.7, where we let $\alpha = 1/2$ and $n = 3g - 3$, we get

$$\begin{aligned}
&\sum_{k=1}^{\lceil 22\cdot\log(3g-3)\rceil} \widetilde{H}_{3g-3,m,1/2}(k)t^k \\
&= \sum_{k=1}^{3g-3} \widetilde{H}_{3g-3,m,1/2}(k)t^k - \sum_{k=\lceil 22\cdot\log(3g-3)\rceil+1}^{3g-3} \widetilde{H}_{3g-3,m,1/2}(k)t^k \\
&= \left(\sum_{k=1}^{3g-3} \widetilde{H}_{3g-3,m,1/2}(k)t^k\right)\left(1 - O\left(g^{-1}\right)\right) \\
&= \frac{\left(\frac{2m}{m+1}\right)^{t/2}(3g-3)^{t/2-1}}{\Gamma(t/2)} \cdot \left(1 + O\left(g^{-1}\right)\right). \qquad (3.68)
\end{aligned}$$

Plugging the latter expression in (3.67) we get

$$\begin{aligned}
\Sigma_1 &= \sum_{k=1}^{\lceil 22\cdot\log(3g-3)\rceil} V_{m,k}(g)\cdot t^k \\
&= \frac{2\sqrt{2}\left(\frac{2m}{m+1}\right)^{t/2}}{\sqrt{\pi}\cdot\Gamma(\frac{t}{2})} \cdot (3g-3)^{\frac{t-1}{2}} \cdot \left(\frac{8}{3}\right)^{4g-4}\left(1 + O\left(\frac{(\log g)^2}{g}\right)\right),
\end{aligned} \qquad (3.69)$$

where for every compact subset U of the complex disk $|t| < 2$ the error term is uniform in both $m \in \mathbb{N} \cup \{+\infty\}$ and $t \in U$. Note that the expression on the right-hand side of the latter equation coincides with the right-hand side of (3.1) from Theorem 3.1.

Using (3.5), from Theorem 3.4, we get the following bound for the second sum:

$$|\Sigma_2| \le C_2 \cdot \sqrt{g} \cdot \left(\frac{8}{3}\right)^{4g-4} \sum_{k=\lceil 22\cdot\log(3g-3)\rceil+1}^{g} \widetilde{H}_{3g-3,m,9/8}(k)\cdot |t|^k. \qquad (3.70)$$

Combining (3.53) from Proposition 3.27 with (3.70) and comparing the resulting expressions for Σ_1 from (3.69) and for $|\Sigma_2|$ from (3.70) we conclude that $\Sigma_2 = \Sigma_1 \cdot o\left(g^{-1}\right)$ uniformly in both $m \in \mathbb{N} \cup \{\infty\}$ and $t \in U$, where U is any compact of the complex disk $|t| < 2$. This completes the proof of Theorem 3.1. □

We show now that Theorem 3.1 implies the following result.

Corollary 3.28 *For any* $m \in \mathbb{N} \cup \{+\infty\}$ *the family of probability distributions* $\{p^{(1)}_{g,m}\}_{g \geq 2}$ *defined in* (3.52) *converges mod-Poisson with radius* $R = 2$*, parameters* $\lambda_{3g-3} = \frac{\log\left(\frac{2m}{m+1}\cdot(3g-3)\right)}{2}$*, limiting function* $\frac{\Gamma\left(\frac{1}{2}\right)}{\Gamma\left(\frac{t}{2}\right)}$ *and speed* $O\left(\frac{(\log g)^2}{g}\right)$.

Proof Let

$$\Phi_g(t) = \sum_{k=1}^{g} V_{m,k}(g) t^k.$$

The above sum coincides with expression (3.1) from Theorem (3.1). By definition (3.28), the generating series $F(t)$ of the probability distribution $p^{(1)}_{g,m}$ is $\Phi_g(t)/\Phi_g(1)$. Applying Eq. (3.1) we get

$$F(t) = \frac{\Phi_g(t)}{\Phi_g(1)} = \left(\frac{2m}{m+1} \cdot (3g-3)\right)^{\frac{t-1}{2}} \frac{\Gamma(1/2)}{\Gamma(t/2)} \cdot \left(1 + O\left(\frac{(\log g)^2}{g}\right)\right).$$

We conclude that the generating series satisfies all the conditions of Definition 3.9 of mod-Poisson convergence, with parameters $\lambda_{3g-3} = \log(\frac{2m}{m+1} \cdot (3g-3))/2$, limiting function $\frac{\Gamma(1/2)}{\Gamma(t/2)}$, radius $R = 2$ and speed $\frac{(\log g)^2}{g}$. □

We complete this section with a proof of the assertion stated in Sect. 1 claiming that *the probability distribution* $q_{3g-3,\infty,1/2}$ *well-approximates the probability distribution* $q_g(k)$. We admit that we will not use this statement in this particular form, so we provide this justification just for the sake of completeness.

Consider the asymptotic relation (3.54) from Proposition 3.27 in which we let $t = 1$, $n = 3g-3$ and $m = +\infty$. We get

$$\sum_{k=\lceil 10 \log n \rceil + 1}^{3g-3} \widetilde{H}_{3g-3,\infty,1/2}(k) = \left(\sum_{k=1}^{3g-3} \widetilde{H}_{3g-3,\infty,1/2}(k)\right) \cdot o\left(n^{-1}\right).$$

The latter relation combined with (3.18) and (3.19) imply the following asymptotic relations for the probability distribution q_g defined in (3.17). For $k \in \mathbb{N}$, $k^2 \leq g$, we have

$$q_g(k) = q_{3g-3,\infty,\frac{1}{2}}(k) \cdot \left(1 + O\left(\frac{k^2}{g}\right)\right).$$

The following asymptotic bound is valid as $g \to +\infty$:

$$\sum_{k=\lceil 10\log g\rceil+1}^{g} q_g(k) = O\left(\frac{(\log g)^3}{g}\right).$$

Analogous considerations imply that for $k \in \mathbb{N}$, $800k^2 \le g$, we have

$$p^{(1)}_{g,\infty}(k) = q_{3g-3,\infty,\frac{1}{2}}(k) \cdot \left(1 + O\left(\frac{(k+2\log g)^2}{g}\right)\right), \tag{3.71}$$

and

$$\sum_{k=\lceil 10\log g\rceil+1}^{g} p^{(1)}_{g,\infty}(k) = O\left(\frac{(\log g)^3}{g}\right).$$

4 Contribution of separating multicurves

In Sect. 3 we studied the volume contributions $\mathrm{Vol}(\Gamma_k(g))$ of stable graphs $\Gamma_k(g)$ with a single vertex and with k loops. In particular, Theorem 3.1 provides precise asymptotics for the generating series $\sum_{k\ge 1} \mathrm{Vol}\,\Gamma_k(g)t^k$ as $g \to +\infty$. In this section we study the volume contribution of the remaining stable graphs.

In Sect. 4.1 we provide some simple estimates of tails of certain series related to Poisson distribution. In Sects. 4.2 and 4.3 we prove the necessary minor refinements of estimates from [3] to bound respectively the volume contributions of stable graphs with 2 vertices and the volume contributions of stable graphs with at least 3 vertices. We emphasize that the main asymptotic analysis of volume contributions of various stable graphs was already performed by Aggarwal in [3]. Our presentation in Sects. 4.2 and 4.3 closely follows original presentation in Sections 9 and 10 of [3], where we perform more or less straightforward adjustment of the original bounds for the sums to bounds for the associated generating series which we need in the context of the present paper.

Following A. Aggarwal let us introduce the following notation for the contributions of stable graphs with a given number of vertices.

Definition 4.1 Let g be an integer satisfying $g \ge 2$. Given a stable graph $\Gamma \in \mathcal{G}_g$ we denote by $V(\Gamma)$ and $E(\Gamma)$ respectively the set of vertices and the set of edges of Γ. We denote by $|V(\Gamma)|$ and $|E(\Gamma)|$ the cardinalities of these sets. We define

$$\Upsilon_g^{(V)} := \sum_{\substack{\Gamma \in \mathcal{G}_g \\ |V(\Gamma)|=V}} \mathrm{Vol}(\Gamma)\,; \qquad \Upsilon_g^{(V;E)} := \sum_{\substack{\Gamma \in \mathcal{G}_g \\ |V(\Gamma)|=V \\ |E(\Gamma)|=E}} \mathrm{Vol}(\Gamma), \tag{4.1}$$

where $\mathrm{Vol}(\Gamma)$ is the contribution of the stable graph Γ to the Masur-Veech volume $\mathrm{Vol}\,\mathcal{Q}_g$ as given in (2.10).

Note that by (2.9) we have

$$\mathrm{Vol}\,\mathcal{Q}_g = \sum_{V=1}^{2g-2} \Upsilon_g^{(V)} \quad \text{and} \quad \Upsilon_g^{(V)} = \sum_{E=1}^{3g-3} \Upsilon_g^{(V;E)}.$$

We also have $\Upsilon_g^{(1;E)} = \mathrm{Vol}\,\Gamma_E(g)$.

The following propositions are refinements of Propositions 8.4 and 8.5 respectively from the original paper [3] of Aggarwal.

Proposition 4.2 *There exists constants B_2 and g_2 such that for any couple g, t, satisfying $g \in \mathbb{N}$, $g \geq g_2$, and $0 \leq t \leq \frac{44}{19}$, we have*

$$\left(\frac{8}{3}\right)^{-4g} \cdot \sum_{E=1}^{3g-3} \Upsilon_g^{(2;E)} t^E \leq B_2 \cdot t \cdot (\log g)^{14} \cdot g^{\frac{2t-3}{2}}. \tag{4.2}$$

We prove Proposition 4.2 in Sect. 4.2.

Proposition 4.3 *There exist constants B_3 and g_3 such that for any triple V, g, t, satisfying $V \in \mathbb{N}$, $V \geq 3$, $g \in \mathbb{N}$, $g \geq g_3$, and $0 \leq t \leq \frac{44}{19}$, we have*

$$\left(\frac{8}{3}\right)^{-4g} \cdot \sum_{V=3}^{2g-2} \sum_{E=1}^{3g-3} \Upsilon_g^{(V;E)} t^E \leq B_3 \cdot t \cdot (\log g)^{24} \cdot g^{\frac{9t-10}{4}}. \tag{4.3}$$

We prove Proposition 4.3 in Sect. 4.3.

4.1 Bound for tail contribution to the moments of Poisson distribution

Recall that given a non-negative integer n and a positive real number λ, the n-th moment $P_n(\lambda)$ of a random variable associated to the Poisson distribution

Poi_λ with parameter λ is defined as

$$P_n(\lambda) := e^{-\lambda} \cdot \sum_{k=0}^{+\infty} k^n \frac{\lambda^k}{k!}.$$

In the next two sections we will use several times the following upper bound for the tail of the above expression.

Lemma 4.4 *Let λ and x be strictly positive real numbers and let n be an integer satisfying $n \geq 0$. Then*

$$\sum_{k=\lceil x\lambda \rceil}^{+\infty} \frac{k^n \cdot \lambda^k}{k!} \leq P_n(x\lambda) \cdot \exp\big(-\lambda(x \log x - x)\big). \tag{4.4}$$

We are interested in the case where x is fixed and λ tends to infinity. Note that the term $x \log x - x$ is positive for $x > e = 2.712\ldots$, so for $x > e$ we prove an exponential decay in λ of the expression in the left hand side of (4.4).

Proof of Lemma 4.4 Let $\theta \geq 0$. Then for $k \geq \lceil \lambda x \rceil$ we have $\exp(\theta(k-\lambda x)) \geq 1$. Hence

$$\sum_{k=\lceil x\lambda \rceil}^{+\infty} \frac{k^n \cdot \lambda^k}{k!} \leq \sum_{k=0}^{+\infty} \frac{\exp(\theta(k-\lambda x)) \cdot k^n \cdot \lambda^k}{k!} = e^{-\theta\lambda x} \sum_{k=0}^{+\infty} \frac{k^n \cdot (e^\theta \cdot \lambda)^k}{k!}.$$

By definition of the polynomial P_n in (3.42) we have

$$\sum_{k=0}^{+\infty} \frac{k^n \cdot (e^\theta \lambda)^k}{k!} = \exp(e^\theta \lambda) \cdot P_n(e^\theta \lambda).$$

The tail bound is obtained by taking $\theta = \log x$. □

We note that analogous Lemma 2.4 of Aggarwal [3] provides a similar upper bound for the case $n = 0$ given as

$$\sum_{k=\lfloor (1+2\delta)R \rfloor}^{+\infty} \frac{R^k}{k!} \leq \exp\left(-R\left(\delta \log(1+\delta) + \frac{\log(\delta)}{R} - 1\right)\right). \tag{4.5}$$

We will need a slightly stronger estimate. Bound (4.5) above implies exponential decay as soon as $\log(1+\delta) > 1/\delta$ which corresponds to

$\delta > 1.23997\ldots$. In comparison, bound (4.4) reads as

$$\sum_{k=\lceil(1+2\delta)R\rceil}^{+\infty} \frac{R^k}{k!} \leq \exp\big(-R((1+2\delta)\log(1+2\delta)-(1+2\delta)\big),$$

which implies exponential decay as soon as $1+2\delta > e$, that is $\delta > 0.8591409\ldots$.

4.2 Volume contribution of stable graphs with 2 vertices

Following the approach of Aggarwal [3], we consider a refinement of the quantity $\Upsilon_g^{(V;E)}$ introduced in Definition 4.1.

Given a stable graph $\Gamma \in \mathcal{G}_g$, denote by $S(\Gamma)$ the number of loops of Γ (edges with both ends at the same vertex). Denote by $T(\Gamma)$ the set of edges of Γ with their two ends at distinct vertices. The set $E(\Gamma)$ decomposes into the disjoint union $E(\Gamma) = S(\Gamma) \sqcup T(\Gamma)$. Following [3, Definition 8.6] let

$$\Upsilon_g^{(V;S,T)} := \sum_{\substack{\Gamma\in\mathcal{G}_g \\ |V(\Gamma)|=V \\ |S(\Gamma)|=S \\ |T(\Gamma)|=T}} \mathrm{Vol}(\Gamma).$$

The number $\Upsilon_g^{(V;S,T)}$ is non-zero only when the following three conditions are simultaneously satisfied: $V-1 \leq T$ (connectedness of the graph), $S+T-V+1 \leq g$ (bound on the genus) and $V \leq 2g-2$ ("stability" of the graph). In particular, the number $\Upsilon_g^{(V;S,T)}$ is non-zero only when we have simultaneously $0 \leq S \leq g$ and $V-1 \leq T \leq 3g-3$.

After [3, Lemma 9.5] we split the set of stable graphs with 2 vertices into three subsets corresponding to the following ranges of parameters S and T. We have $S \geq g-1$ for the first collection of stable graphs. We have $T > 13$ and $S \leq g-2$ for the second collection. We have $T \leq 13$ and $S \leq g-2$ for the third collection. Lemmas 4.5, 4.6, 4.7 provide upper bounds for the respective contributions to the sum (4.2) in Proposition 4.2. As we already mentioned, our proofs of Lemmas 4.5, 4.6, 4.7 consist of elementary adjustments of bounds obtained by A. Aggarwal in [3, Section 9].

Lemma 4.5 *For any real t and any integer g satisfying $t \geq 0$ and $g \geq 2$ we have*

$$\left(\frac{8}{3}\right)^{-4g} \sum_{\substack{T \geq 1 \\ S \geq g-1}} \Upsilon_g^{(2;S,T)} \cdot t^{S+T} \leq 2^{11} \cdot \left(t^g(1+t)\right) \cdot g^{3/2} \cdot \left(\frac{9}{8}\right)^g \frac{(\log g + 7)^g}{(g-1)!}. \tag{4.6}$$

Proof All possible stable graphs with 2 vertices and with $S \geq g-1$ split into the following three types:

(I) $1 + \lfloor (g-1)/2 \rfloor$ stable graphs with $S = g-1$, $T = 2$ and genera (decorations) $g_1 = g_2 = 0$ at the two vertices;
(II) $g-1$ graphs with $S = g-1$, $T = 1$ and $g_1 = 1$, $g_2 = 0$;
(III) $1 + \lfloor (g-1)/2 \rfloor$ graphs with $S = g$, $T = 1$ and $g_1 = g_2 = 0$.

By Equation (9.14) in [3] for any of these graphs Γ we have

$$\left(\frac{8}{3}\right)^{-4g} \operatorname{Vol}(\Gamma) \leq 2^{10}(S^2 + T^2) g^{-3/2} \xi_1 \xi_2 \frac{(\log g + 7)^{S+T-1}}{2^S S! T!}. \tag{4.7}$$

Here ξ_i, $i = 1, 2$, are defined by Equation (9.1) in [3] as

$$\xi_i = \max_{\boldsymbol{d} \in \Pi(3g_i + 2s_i + T + 3, 2s_i + T)} \left(1 + \varepsilon(\boldsymbol{d})\right),$$

where g_i and s_i are respectively the genera and the number of loops at the i-th vertex, for $i = 1, 2$ and $\varepsilon(\boldsymbol{d})$ is defined in (2.23). The bound (2.24) from Theorem 2.6 of A. Aggarwal (see Proposition 1.2 in the original paper [3]) implies that for the stable graphs with $V = 2$ we have

$$\xi_1 \xi_2 \leq \left(\frac{3}{2}\right)^{2E-2} \leq \left(\frac{3}{2}\right)^{2g},$$

where $E = S + T$, and (4.7) provides the following bound for any stable graph with $V = 2$ and $S \geq g - 1$:

$$\left(\frac{8}{3}\right)^{-4g} \operatorname{Vol}(\Gamma) \leq 2^{10}(g^2 + 1) g^{-3/2} \cdot \left(\frac{3}{2}\right)^{2g} \frac{(\log g + 7)^g}{2^g (g-1)!}. \tag{4.8}$$

We have seen that when $V = 2$ and $S \geq g - 1$ there are $g - 1$ stable graphs of type (II) for which $S + T = g$ and there are at most $g + 1$ stable graphs of types (I) and (III) counted together for which $S + T = g + 1$. Thus we get

$$\begin{aligned}
&\left(\frac{8}{3}\right)^{-4g} \cdot \sum_{\substack{T \geq 1 \\ S \geq g-1}} \Upsilon_g^{(2;S,T)} \cdot t^{S+T} \\
&\leq \Big((g-1)t^g + (g+1)t^{g+1}\Big) \cdot 2^{10} \cdot (g^2+1) \cdot g^{-3/2} \\
&\quad \times \left(\frac{3}{2}\right)^{2g} \frac{(\log g + 7)^g}{2^g (g-1)!} \\
&\leq 2^{11} \cdot \left(t^g + t^{g+1}\right) \cdot g^{3/2} \cdot \left(\frac{9}{8}\right)^g \frac{(\log g + 7)^g}{(g-1)!}.
\end{aligned}$$

□

Lemma 4.6 *For any real t and any integer g satisfying $t \geq 0$ and $g \geq 2$ there exists a constant C_7 such that*

$$\begin{aligned}
&\left(\frac{8}{3}\right)^{-4g} \sum_{\substack{14 \leq T \leq 3g-3 \\ 0 \leq S \leq g-2}} \Upsilon_g^{(2;S,T)} \cdot t^{S+T} \\
&\quad \leq C_7 \cdot t^{14} \cdot \exp(189t/8) \cdot (\log g + 7)^{13} \cdot g^{(27t-56)/8}. \qquad (4.9)
\end{aligned}$$

Proof By Equation (9.17) from Lemma 9.5 in [3], in the case $T > 13$ and $S \leq g - 2$ we have for g large enough

$$\left(\frac{8}{3}\right)^{-4g} \Upsilon_g^{(2;S,T)} \leq 2^{55} g^{-7} \left(\frac{9}{4}\right)^E \frac{(\log g + 7)^{E-1}}{2^S\, S!\, T!}.$$

By the binomial expansion we have

$$\sum_{\substack{S+T=E \\ S,T \geq 0}} \frac{1}{2^S\, S!\, T!} = \frac{1}{E!}\left(\frac{1}{2} + 1\right)^E = \frac{1}{E!}\left(\frac{3}{2}\right)^E.$$

Note that the set $\mathcal{G}_g$ of stable graphs of any fixed genus g is finite. Hence, there exists a constant C_7' such that for any $g \geq 2$ and for any fixed E we have

$$\left(\frac{8}{3}\right)^{-4g} \cdot \sum_{\substack{14 \le T \le 3g-3 \\ 0 \le S \le g-2 \\ S+T=E}} \Upsilon_g^{(2;S,T)}$$
$$\le C'_7 \cdot g^{-7} \cdot (\log g + 7)^{-1} \cdot \left(\frac{9}{4}\right)^E \cdot (\log g + 7)^E \sum_{\substack{S+T=E \\ S,T \ge 0}} \frac{1}{2^S\, S!\, T!}$$
$$= C'_7 \cdot g^{-7} \cdot (\log g + 7)^{-1} \cdot \left(\frac{27}{8}\right)^E \cdot (\log g + 7)^E \cdot \frac{1}{E!}.$$

Multiplying each term by $t^E = t^{S+T}$ and taking the sum with respect to E we obtain

$$\left(\frac{8}{3}\right)^{-4g} \cdot \sum_{\substack{14 \le T \le 3g-3 \\ 0 \le S \le g-2}} \Upsilon_g^{(2;S,T)} \cdot t^{S+T}$$
$$\le C'_7 \cdot g^{-7} \cdot (\log g + 7)^{-1} \sum_{E=14}^{+\infty} \left(\frac{27}{8}\right)^E \cdot (\log g + 7)^E \cdot t^E \cdot \frac{1}{E!}$$
$$\le C'_7 \cdot g^{-7} \cdot (\log g + 7)^{-1} \cdot \left(\frac{27t}{8} \cdot (\log g + 7)\right)^{14}$$
$$\times \exp\left(\frac{27t}{8} \cdot (\log g + 7)\right)$$
$$= C_7 \cdot t^{14} \cdot \exp(189t/8) \cdot (\log g + 7)^{13} \cdot g^{(27t-56)/8},$$

where we used the inequality $\sum_{k=n}^{+\infty} \frac{x^k}{k!} \le x^n \exp(x)$, which is valid for any non-negative x, and where we let $C_7 = C'_7 \cdot \left(\frac{27}{8}\right)^{14}$. □

Lemma 4.7 *There exists a constant C_8 such that for any non-negative real number t and for any integer g satisfying $g \ge 2$ we have*

$$\left(\frac{8}{3}\right)^{-4g} \cdot \sum_{\substack{1 \le T \le 13 \\ 0 \le S \le g-2}} \Upsilon_g^{(2;S,T)} \cdot t^{S+T} \le C_8 \cdot t(1+t)^{14} \cdot \exp(63t/4)$$
$$\times (\log g + 7)^{14} \cdot \left(g^{(2t-3)/2} + g^{(9t-28)/4}\right). \tag{4.10}$$

Proof It follows from Equation (9.18) from Lemma 9.5 from [3] that there exists a constant C_8 such that in the case $T \le 13$ and $S \le g-2$ the following bound is valid for any integer g satisfying $g \ge 2$:

$$\left(\frac{8}{3}\right)^{-4g} \Upsilon_g^{(2;S,T)} \leq C_8 \cdot \frac{(\log g + 7)^{S+12}}{S!} \left(g^{-3/2}(S^2+1) + g^{-7} \left(\frac{9}{4}\right)^S \right).$$

Multiplying by $t^E = t^{S+T}$ and taking the sum over $1 \leq T \leq 13$ and over $0 \leq S \leq g-2$ we obtain the following bound:

$$\begin{aligned} &\left(\frac{8}{3}\right)^{-4g} \cdot \sum_{\substack{1 \leq T \leq 13 \\ 0 \leq S \leq g-2}} \Upsilon_g^{(2;S,T)} \cdot t^{S+T} \\ &\leq C_8 \cdot \left(t + t^2 + \cdots + t^{13}\right) \cdot (\log g + 7)^{12} \cdot \left(g^{-3/2}\Sigma_1 + g^{-7}\Sigma_2\right) \\ &\leq C_8 \cdot t \cdot (1+t)^{12} \cdot (\log g + 7)^{12} \cdot \left(g^{-3/2}\Sigma_1 + g^{-7}\Sigma_2\right), \end{aligned} \tag{4.11}$$

where

$$\Sigma_1 = \sum_{S=0}^{+\infty} (S^2+1) \frac{(t \cdot (\log g + 7))^S}{S!},$$

$$\Sigma_2 = \sum_{S=0}^{+\infty} \frac{\left(\frac{9t}{4}(\log g + 7)\right)^S}{S!} = \exp(63t/4) \cdot g^{9t/4}.$$

The sum Σ_1 can be decomposed into two sums of the form (3.42), where we take $n = 2$ and $n = 0$ respectively and where we let $\lambda = t \cdot (\log g + 7)$. Applying Lemma 3.21 and taking into consideration that $P_2(\lambda) = \lambda^2 + \lambda$ by (3.45), we get

$$\begin{aligned} \Sigma_1 &= \sum_{S=0}^{+\infty} (S^2+1) \cdot \frac{(t \cdot (\log g + 7))^S}{S!} = e^{\lambda} \cdot \left(P_2(\lambda) + 1\right) \\ &= \exp(7t) \cdot g^t \cdot \left(1 + t \cdot (\log g + 7) + t^2 \cdot (\log g + 7)^2\right) \\ &\leq \exp(7t) \cdot g^t \cdot (1+t)^2 (\log g + 7)^2. \end{aligned}$$

Plugging the resulting bounds for the sums Σ_1 and Σ_2 into (4.11) we obtain the bound

$$\begin{aligned} \left(\frac{8}{3}\right)^{-4g} \cdot \sum_{\substack{1 \leq T \leq 13 \\ 0 \leq S \leq g-2}} \Upsilon_g^{(2;S,T)} \cdot t^{S+T} &\leq C_8 \cdot t \cdot (1+t)^{14} \cdot (\log g + 7)^{14} \\ &\times \left(g^{-3/2} \cdot \exp(7t) \cdot g^t + g^{-7} \cdot \exp(63t/4) \cdot g^{9t/4}\right) \end{aligned}$$

$$= C_8 \cdot t(1+t)^{14} \cdot (\log g + 7)^{14} \\ \times \Big(\exp(7t) \cdot g^{(2t-3)/2} + \exp(63t/4) \cdot g^{(9t-28)/4} \Big).$$

Now it remains to notice that $7t \le 63t/4$ to get the desired bound. □

Proof of Proposition 4.2 By taking the sum of the bounds (4.6), (4.6) and (4.7) from respectively Lemmas 4.5, 4.6 and 4.7 covering all possible combinations of S and T we obtain

$$\left(\frac{8}{3}\right)^{-4g} \cdot \sum_{E=1}^{3g-3} \Upsilon_g^{(2;E)} t^E \le 2^{11} \cdot \big(t^g(1+t)\big) \cdot g^{3/2} \cdot \left(\frac{9}{8}\right)^g \frac{(\log g + 7)^g}{(g-1)!} \\ + C_7 \cdot t^{14} \cdot \exp(189t/8) \cdot (\log g + 7)^{13} \cdot g^{(27t-56)/8} \\ + C_8 \cdot t \cdot (1+t)^{14} \cdot \exp(63t/4) \cdot (\log g + 7)^{14} \cdot \big(g^{(2t-3)/2} + g^{(9t-28)/4}\big). \tag{4.12}$$

Now note that

$$\frac{9t-28}{4} \le \frac{2t-3}{2} \qquad \text{for } t \le \frac{22}{5} = 4.4 \quad \text{and}$$
$$\frac{27t-56}{8} \le \frac{2t-3}{2} \qquad \text{for } t \le \frac{44}{19} \approx 2.3.$$

Note also that for the particular value $t = \frac{44}{19}$ of t for which the powers of g in the second and in the third term in (4.12) coincide, the power of $(\log g + 7)$ is larger in the third term. Since by construction $C_8 > 0$, there exists a constant g_0 such that for any $g \ge g_0$ and any t in the interval $\left[0, \frac{44}{19}\right]$ the expression

$$C_8 \cdot t \cdot (1+t)^{14} \cdot (\log g + 7)^{14} \cdot \exp(63t/4) \cdot g^{(2t-3)/2}$$

dominates the sum (4.12). This completes the proof of Proposition (4.2). □

4.3 Volume contribution of stable graphs with 3 and more vertices

In this section we adjust the bound for the sum of contributions of stable graphs with 3 and more vertices to the Masur–Veech volume Vol $\mathcal{Q}_g$ elaborated by A. Aggarwal in [3, Section 10] to bounds for the associated generating series in a variable t.

The Lemma below is based on Proposition 10.4 in [3] and is parallel to [3, Lemma 10.5].

Lemma 4.8 *For any couple of integers g and V satisfying $g \geq 2$, $V \geq 2$, and for any non-negative real t we have*

$$\begin{aligned}
&\left(\frac{8}{3}\right)^{-4g} \sum_{S=0}^{g} \Upsilon_g^{(V;S,T)} \cdot t^{S+T} \\
&\leq T^{2V+Y+1} \cdot \frac{A_{g,t}^T}{T!} \\
&\quad \times 2^{12} \cdot 2^{23V} \cdot \frac{1}{V^{3V}} \cdot (\log g + 7)^{-1} \cdot g^{1/2-V} \\
&\quad \times \Big(1 + 2^V\big(A_{g,t} + A_{g,t}^V\big) + A_{g,t}^V (2^V + A_{g,t}) \exp(A_{g,t})\Big). \qquad (4.13)
\end{aligned}$$

where we use the notation $Y = \min(2T, 3V)$ *and* $A_{g,t} = \frac{9t}{8} \cdot (\log g + 7)$.

Proof It follows from Proposition 10.4 in [3] that

$$\left(\frac{8}{3}\right)^{-4g} \sum_{S=0}^{g} \Upsilon_g^{(V;S,T)} \cdot t^{S+T} \leq 2^{11} \cdot B_{T,V} \cdot (\Sigma_1 + \Sigma_2)\,, \qquad (4.14)$$

where

$$B_{T,V} = 2T \cdot g^{1/2-V} \cdot 2^{20V} \cdot \left(\frac{9t}{4}\right)^T \left(\frac{T}{V}\right)^{2V} \frac{(\log g + 7)^{T-1} \cdot (2T-1)!!}{V^V (2T-Y)!},$$

and

$$\Sigma_1 = 1 + \sum_{S=1}^{V-1} S \cdot A_{g,t}^S \quad \text{and} \quad \Sigma_2 = \sum_{S=V}^{g} S \frac{A_{g,t}^S}{(S-V)!}.$$

In order to transform the bound in [3, Proposition 10.4] to the above form we used the following trivial observations. Since $V \geq 2$ we have $T \geq 1$. The case $S = 0$ corresponds to the constant summand "1" in Σ_1. In the remaining cases, that is when $S \geq 1$, we used the bound $S + T \leq 2TS$ for the factor $(S + T)$ present in [3, Proposition 10.4] which is valid for all $S, T \in \mathbb{N}$.

Using that $\sum_{i=1}^n nx^n \leq n^2(x + x^n)$, for any $x > 0$, and that $(V-1)^2 \leq 2^V$, for any $V \geq 0$, we obtain the following bound on the first sum:

$$\Sigma_1 \leq 1 + (V-1)^2\big(A_{g,t} + A_{g,t}^V\big) \leq 1 + 2^V\big(A_{g,t} + A_{g,t}^V\big).$$

The sum Σ_2 is a part of the infinite sum (3.42) taken with parameters $n = 1$ and $\lambda = A_{g,t}$. Applying Lemma 3.21 and using the fact that $P_1(\lambda) = \lambda$

by (3.45), we get the following bound:

$$\Sigma_2 \leq \sum_{S=V}^{+\infty} S \frac{A_{g,t}^S}{(S-V)!} = A_{g,t}^V \sum_{S=0}^{+\infty} (V+S) \frac{A_{g,t}^S}{S!} = A_{g,t}^V (V + A_{g,t}) \exp(A_{g,t}).$$

Applying the trivial bound $V \leq 2^V$ we obtain

$$\Sigma_1 + \Sigma_2 \leq 1 + 2^V \left(A_{g,t} + A_{g,t}^V\right) + A_{g,t}^V (2^V + A_{g,t}) \exp(A_{g,t}). \qquad (4.15)$$

Using $(2T)^Y (2T-Y)! \geq (2T)! = 2^T T!(2T-1)!!$ and $Y \leq 3V$ we obtain the following bound for $B_{T,V}$,

$$\begin{aligned} B_{T,V} &= 2T \cdot g^{1/2-V} \cdot 2^{20V} \cdot \left(\frac{9t}{4}\right)^T \frac{T^{2V}}{V^{2V}} \cdot \frac{(\log g + 7)^{T-1}}{V^V} \cdot \frac{(2T-1)!!}{(2T-Y)!} \\ &\leq 2^{1+20V+Y} \cdot (\log g + 7)^{-1} \cdot g^{1/2-V} \cdot T^{2V+Y+1} \cdot \frac{A_{g,t}^T}{V^{3V} T!} \\ &\leq 2 \cdot 2^{23V} \cdot (\log g + 7)^{-1} \cdot g^{1/2-V} \cdot \frac{T^{2V+Y+1}}{V^{3V}} \cdot \frac{A_{g,t}^T}{T!}. \end{aligned} \qquad (4.16)$$

Putting together (4.15) and (4.16) into (4.14) we obtain (4.13). □

We perform now summation over the variable T. The following statement is an adjustment of [3, Lemma 10.6] to the generating series in t.

Lemma 4.9 *There exist constants C_9 and C_{10} such that for any couple of integers g and V satisfying $g \geq 2$, $V \geq 2$, and for any non-negative real t we have*

$$\begin{aligned} &\left(\frac{8}{3}\right)^{-4g} \sum_E \Upsilon_g^{(V;E)} t^E \\ &\quad \leq t \cdot C_9 \cdot \exp(63t/4) \cdot g^{\frac{9t+2}{4}} \\ &\qquad \times \left(\left(\frac{C_{10} \cdot V^{1/2}}{g}\right)^V + \left(\frac{C_{10} \cdot t^8 \cdot V^{1/2} \cdot (\log g + 7)^8}{g}\right)^V \right). \end{aligned} \qquad (4.17)$$

Proof First note that the third and the fourth lines of expression (4.13) do not depend on the variable T. To bound the sum

$$\sum_{T=V-1}^{3g-3} T^{2V+Y+1} \cdot \frac{A_{g,t}^T}{T!},$$

where $Y = \min(2T, 3V)$, we bound separately the following three partial sums:

$$\Sigma_1 = \sum_{T=V-1}^{\lfloor 3V/2 \rfloor} T^{2V+2T+1} \frac{A_{g,t}^T}{T!},$$

$$\Sigma_2 = \sum_{T=\lceil 3V/2 \rceil}^{6V+1} T^{5V+1} \frac{A_{g,t}^T}{T!},$$

$$\Sigma_3 = \sum_{T=6V+2}^{+\infty} T^{5V+1} \frac{A_{g,t}^T}{T!},$$

where we use the same notation $A_{g,t} = \frac{9t}{8}(\log g + 7)$ as in Lemma 4.8.

To bound Σ_1 and Σ_2 we use twice the inequality $T! \geq e^{-T} T^T$ to obtain

$$\begin{aligned}
\Sigma_1 &\leq \sum_{T=1}^{\lfloor 3V/2 \rfloor} (e \cdot A_{g,t})^T \cdot T^{2V+T+1} \leq \left(\frac{3V}{2}\right)^{2V+1+3V/2} \cdot \sum_{T=1}^{\lfloor 3V/2 \rfloor} (e \cdot A_{g,t})^T \\
&\leq \left(\frac{3V}{2}\right)^{7V/2+1} \left(\frac{3V}{2}\right) \left(eA_{g,t} + (eA_{g,t})^{3V/2}\right) \\
&= \left(eA_{g,t} + (eA_{g,t})^{3V/2}\right) \left(\frac{3V}{2}\right)^{7V/2+2}.
\end{aligned}$$

Similarly,

$$\begin{aligned}
\Sigma_2 &\leq \sum_{T=\lceil 3V/2 \rceil}^{6V+1} T^{5V+1-T} (e \cdot A_{g,t})^T \\
&\leq (6V+1)^{5V+1-3V/2} \cdot \sum_{T=\lceil 3V/2 \rceil}^{6V+1} (e \cdot A_{g,t})^T \\
&\leq (6V+1)^{7V/2+1} \cdot (6V+1) \cdot \left((e \cdot A_{g,t})^{3V/2} + (e \cdot A_{g,t})^{6V+1}\right) \\
&\leq \left((e \cdot A_{g,t})^{3V/2} + (e \cdot A_{g,t})^{6V+1}\right) \cdot (7V)^{7V/2+2}.
\end{aligned}$$

To bound the third sum, we use the inequality

$$\frac{T^{5V+1}}{T!} \leq \frac{6^{5V+1}}{(T-5V-1)!},$$

valid for $T > 6V+1$, and the inequality $\sum_{k=n}^{+\infty} \frac{x^k}{k!} \le x^n \exp(x)$, valid for any non-negative x:

$$\begin{aligned}\Sigma_3 &= \sum_{T=6V+2}^{+\infty} T^{5V+1} \frac{A_{g,t}^T}{T!} \le 6^{5V+1} \sum_{T=6V+2}^{+\infty} \frac{A_{g,t}^T}{(T-5V-1)!}\\ &= (6A_{g,t})^{5V+1} \sum_{T=V+1}^{+\infty} \frac{A_{g,t}^T}{T!} \le (6A_{g,t})^{5V+1} \sum_{T=V}^{+\infty} \frac{A_{g,t}^T}{T!}\\ &\le (6A_{g,t})^{5V+1} A_{g,t}^V \exp(A_{g,t}) = (6A_{g,t})^{6V+1} \cdot \exp(A_{g,t}).\end{aligned}$$

Collecting the terms and applying the bounds $V^2 \le 4^V$ and $A_{g,t}^{3V/2} \le A_{g,t} + A_{g,t}^{7V/2}$ we get

$$\begin{aligned}\Sigma_1 + \Sigma_2 + \Sigma_3 \le{}& \left(eA_{g,t} + (eA_{g,t})^{3V/2}\right) \cdot \left(\frac{3V}{2}\right)^{7V/2+2}\\ &+\left((e \cdot A_{g,t})^{3V/2} + (e \cdot A_{g,t})^{6V+1}\right) \cdot (7V)^{7V/2+2}\\ &+(6A_{g,t})^{6V+1} \cdot \exp(A_{g,t})\\ \le{}& 3e \cdot A_{g,t} \cdot \Big(\left(1 + (e \cdot A_{g,t})^{6V}\right) \cdot (7V)^{7V/2+2}\\ &+(6A_{g,t})^{6V} \cdot \exp(A_{g,t})\Big)\\ \le{}& 10t \cdot (\log g + 7) \cdot \Big(49 \cdot 4^V \left(1 + (e \cdot A_{g,t})^{6V}\right) \cdot (7V)^{7V/2}\\ &+(6A_{g,t})^{6V} \cdot \exp(A_{g,t})\Big).\end{aligned}$$

Combining the resulting bound with (4.13) we get

$$\begin{aligned}&\left(\frac{8}{3}\right)^{-4g} \sum_E \Upsilon_g^{(V;E)} t^E\\ &\le t \cdot g^{1/2-V} \cdot 10 \cdot 2^{12} \cdot \left(\frac{2^{23}}{V^3}\right)^V\\ &\quad\times\left(49 \cdot 4^V \left(1 + (e \cdot A_{g,t})^{6V}\right) \cdot (7V)^{7V/2} + (6A_{g,t})^{6V} \exp(A_{g,t})\right) \cdot\\ &\quad\times\left(1 + 2^V \left(A_{g,t} + A_{g,t}^V\right) + A_{g,t}^V (2^V + A_{g,t}) \exp(A_{g,t})\right).\end{aligned}$$

Expanding the product of the terms located in the third and in the fourth lines of the expression above we get a sum of 15 non-negative terms, where every term has the form

$$a \cdot b^V \cdot (A_{g,t})^{cV+d} \cdot V^{\alpha V} \cdot \exp(kA_{g,t})$$

with constants a, b, c, d, α, k specific for each summand, but satisfying, however, the following common conditions. We always have $a, b > 0$, $c \in \{0, 1, 6, 7\}$, $d \in \{0, 1\}$, $\alpha \le \frac{1}{2}$, and $k \in \{0, 1, 2\}$. It remains to notice that since $A_{g,t} \ge 0$, we have $\exp(2A_{g,t}) \ge \exp(A_{g,t}) \ge \exp(0)$. Note also, that since $V \ge 2$, our restrictions on c and d imply that $A_{g,t}^{cV+d} \le \max(1, A_{g,t}^{8V})$. These observations imply that each of the terms can be bounded from above by the expression

$$a \cdot b^V \cdot \left(1 + A_{g,t}^{8V}\right) \cdot V^{V/2} \cdot \exp(2A_{g,t}).$$

Recall that $A_{g,t} = \frac{9t}{8}(\log g + 7)$, so $\exp(2A_{g,t}) = g^{\frac{9}{4}t} \exp(63t/4)$. The above observations imply that letting $C_9 = 15a'$, where a' is the maximal value of the parameter a over 15 terms, and letting $C_{10} = \left(\frac{9}{8}\right)^8 b'$, where b' is the maximal value of the parameter b over 15 terms, we complete the proof of (4.17). □

Proof of Proposition 4.3 By Lemma 4.9, we have that for all non-negative real t we have

$$\left(\frac{8}{3}\right)^{-4g} \sum_{V=3}^{2g-2} \sum_{E \ge 1} \Upsilon_g^{(V;E)} t^E$$
$$\le t \cdot C_9 \cdot \exp(63t/4) \cdot g^{\frac{9t+2}{4}} \cdot$$
$$\times \left(\sum_{V=3}^{2g-2} \left(\frac{C_{10} \cdot V^{1/2}}{g} \right)^V + \sum_{V=3}^{2g-2} \left(\frac{C_{10} \cdot t^8 \cdot V^{1/2} \cdot (\log g + 7)^8}{g} \right)^V \right).$$

Let us denote by

$$a_V := \left(\frac{C_{10} \cdot t^8 \cdot V^{1/2} \cdot (\log g + 7)^8}{g} \right)^V$$

the term in the second sum. Then

$$\frac{a_{V+1}}{a_V} = \frac{C_{10} \cdot t^8 \cdot (1 + 1/V)^{V/2} \cdot (V+1)^{1/2} \cdot (\log g + 7)^8}{g}$$
$$\le \frac{C_{10} \cdot t^8 \cdot e^{1/2} \cdot (2g-1)^{1/2} \cdot (\log g + 7)^8}{g}.$$

In particular, since t is bounded, there exists g_3 such that for $g \geq g_3$ we have $\frac{a_{V+1}}{a_V} \leq 1/2$ for all V. Hence

$$\sum_{V=3}^{2g-2} \left(\frac{C_{10} \cdot t^8 \cdot V^{1/2} \cdot (\log g + 7)^8}{g} \right)^V \leq 2 \left(\frac{C_{10} \cdot t^8 \cdot 3^{1/2} \cdot (\log g + 7)^8}{g} \right)^3.$$

Applying analogous bound for the first sum and collecting the estimates we get

$$\begin{aligned} &\left(\frac{8}{3}\right)^{-4g} \sum_{V=3}^{2g-2} \sum_{E\geq 1} \Upsilon_g^{(V;E)} t^E \\ &\leq t \cdot C_9 \cdot \exp(63t/4) \cdot g^{\frac{9t+2}{4}} \cdot g^{-3} \cdot 2 \\ &\quad \times \left(C_{10} \cdot 3^{1/2}\right)^3 \left(1 + t^{24} \cdot (\log g + 7)^{24}\right) \\ &\leq B_3 \cdot t \cdot g^{\frac{9t-10}{4}} \cdot (\log g + 7)^{24}, \end{aligned}$$

where

$$B_3 = C_9 \cdot \exp\left(\frac{63}{4} \cdot \frac{44}{19}\right) \cdot 2 \cdot \left(C_{10} \cdot 3^{1/2}\right)^3 \cdot 2 \cdot \left(\frac{44}{19}\right)^{24}.$$

□

5 Proofs

We proved in Sect. 3.7 the mod-Poisson convergence of the distribution $p_g^{(1)}$ corresponding to volume contributions of stable graphs with a single vertex. In Sect. 5.1 we apply the results collected in Sect. 4 on volume contributions of stable graphs with two and more vertices to prove that the distribution $p_g^{(1)}$ well-approximates the distribution p_g. In Sect. 5.2 we present the remaining proofs of Theorems stated in Sect. 1.

5.1 From $p_g^{(1)}$ to p_g

For any k denote

$$\mathrm{Vol}_k\, \mathcal{Q}_g := \sum_{\substack{\Gamma \in \mathcal{G}_g \\ |E(\Gamma)|=k}} \mathrm{Vol}(\Gamma), \tag{5.1}$$

the contribution to Vol $\mathcal{Q}_g$ of stable graphs with exactly k edges. Using this notation, the probability distribution $p_g(k)$ defined in Theorem 1.7 can be rewritten as

$$p_g(k) = \frac{\mathrm{Vol}_k\, \mathcal{Q}_g}{\mathrm{Vol}\, \mathcal{Q}_g}.$$

Recall that we also have a probability distribution $q_{3g-3,\infty,1/2}(k)$ defined in (1.8) and evaluated in (3.22) that corresponds to the number of cycles in a random permutation of $3g-3$ elements according to the probability distribution $\mathbb{P}_{3g-3,\infty,1/2}$ on S_{3g-3}, see Lemma 3.6.

We gather the results from Sects. 3 and 4 in the following two statements.

Theorem 5.1 *For $t \in \mathbb{C}$ satisfying $|t| \le 4/5$ we have as $g \to +\infty$*

$$\sum_{k=1}^{3g-3} \mathrm{Vol}_k\, \mathcal{Q}_g\, t^k = \frac{2t}{\sqrt{\pi}\,\Gamma\left(1+\frac{t}{2}\right)} (6g-6)^{\frac{t-1}{2}} \left(\frac{8}{3}\right)^{4g-4} \times \left(1 + O\left(g^{\frac{t}{2}-1} (\log g)^{24}\right)\right),$$

where the error term is uniform in t on the disk $|t| \le 4/5$.

For $t \in \mathbb{C}$ satisfying $4/5 \le |t| < 8/7$ we have as $g \to +\infty$

$$\sum_{k=1}^{3g-3} \mathrm{Vol}_k\, \mathcal{Q}_g\, t^k = \frac{2t}{\sqrt{\pi}\,\Gamma\left(1+\frac{t}{2}\right)} (6g-6)^{\frac{t-1}{2}} \left(\frac{8}{3}\right)^{4g-4} \times \left(1 + O\left(g^{\frac{7t}{4}-2} (\log g)^{24}\right)\right),$$

where the constant in the error term is uniform in t on any compact subset of the annulus $4/5 \le |t| < 8/7$. In particular, for $t = 1$ we get

$$\mathrm{Vol}(\mathcal{Q}_g) = \sum_{k=1}^{3g-3} \mathrm{Vol}_k\, \mathcal{Q}_g \; = \frac{4}{\pi} \cdot \left(\frac{8}{3}\right)^{4g-4} \cdot \left(1 + O\left(g^{-1/4} \cdot (\log g)^{24}\right)\right). \tag{5.2}$$

We note that the asymptotic formula for $\mathrm{Vol}(\mathcal{Q}_g)$ without explicit error term as in (5.2) was conjectured in [26,29] and proved in [3, Theorem 1.7]. See also Remark 2.4 for the discussion of the expected optimal error term.

Theorem 5.2 *For any k satisfying $k \le \dfrac{\log g}{\log \frac{9}{4}}$ we have*

$$\mathrm{Vol}_k\, \mathcal{Q}_g = \mathrm{Vol}\,\Gamma_k(g)\left(1 + O\left((\log g)^{25} \cdot g^{-1+\frac{k\log 2}{\log g}}\right)\right).$$

For any x satisfying $x < \frac{2}{\log \frac{9}{2}}$ and for all k satisfying $\frac{\log g}{\log \frac{9}{4}} \le k \le x \log g$ we have

$$\mathrm{Vol}_k\, \mathcal{Q}_g = \mathrm{Vol}\,\Gamma_k(g)\left(1 + O\left((\log g)^{25} \cdot g^{-2+\frac{k\log \frac{9}{2}}{\log g}}\right)\right).$$

For $k \le \frac{3}{4\log 2}\log g$ we have

$$p_g(k) = q_{3g-3,\infty,\frac{1}{2}}(k) \cdot \left(1 + O\left(g^{-1/4} \cdot (\log g)^{24}\right)\right). \tag{5.3}$$

For k in the range $\frac{3\log g}{4\log 2} \le k \le \frac{\log g}{\log \frac{9}{4}}$ we have

$$\begin{aligned} p_g(k) &= q_{3g-3,\infty,\frac{1}{2}}(k) \cdot \left(1 + O\left(g^{-1} \cdot 2^k \cdot (\log g)^{25}\right)\right) \\ &= q_{3g-3,\infty,\frac{1}{2}}(k)\left(1 + O\left((\log g)^{25} \cdot g^{-1+\frac{k\log 2}{\log g}}\right)\right). \end{aligned}$$

For k in the range $\frac{\log g}{\log \frac{9}{4}} \le k \le x \log g$ we have

$$\begin{aligned} p_g(k) &= q_{3g-3,\infty,\frac{1}{2}}(k) \cdot \left(1 + O\left(g^{-1} \cdot \left(\frac{9}{2}\right)^k \cdot (\log g)^{25}\right)\right) \\ &= q_{3g-3,\infty,\frac{1}{2}}(k)\left(1 + O\left((\log g)^{25} \cdot g^{-2+\frac{k\log \frac{9}{2}}{\log g}}\right)\right), \end{aligned}$$

where all above estimates are uniform in the corresponding ranges of k.

Proof of Theorem 5.1 Using the notation from Definition 4.1 we decompose

$$\mathrm{Vol}_k\, \mathcal{Q}_g = \Upsilon_g^{(1;k)} + \Upsilon_g^{(2;k)} + \sum_{V \ge 3} \Upsilon_g^{(V;k)}.$$

Here $\Upsilon_g^{(1;k)} = \operatorname{Vol}\Gamma_k(g)$. Note that $8/7 < 2$, so Theorem 3.1 gives the uniform asymptotic equivalence for the first term. Applying the identity $\frac{t}{2}\Gamma\left(\frac{t}{2}\right) = \Gamma\left(1+\frac{t}{2}\right)$ we set $m = +\infty$ and rewrite (3.1) as

$$\sum_{k\geq 1}\Upsilon_g^{(1;k)}t^k = \frac{2t}{\sqrt{\pi}\,\Gamma\left(1+\frac{t}{2}\right)}(6g-6)^{\frac{t-1}{2}}\left(\frac{8}{3}\right)^{4g-4}\left(1+O\left(\frac{(\log g)^2}{g}\right)\right). \tag{5.4}$$

The bounds for the contributions of the second and third terms are provided by Propositions 4.2 and 4.3 respectively. We have for $|t| \in [0, 44/19]$ and, hence, for $|t| \in [0, 8/7]$:

$$\left(\frac{8}{3}\right)^{-4g}\sum_{k\geq 1}\Upsilon_g^{(2;k)}|t|^k \leq g^{\frac{|t|-1}{2}}\cdot O\left(g^{\frac{|t|-2}{2}}(\log g)^{14}\right),$$

$$\left(\frac{8}{3}\right)^{-4g}\sum_{k\geq 1}\sum_{V\geq 3}\Upsilon_g^{(V;k)}|t|^k \leq g^{\frac{|t|-1}{2}}\cdot O\left(g^{\frac{7|t|-8}{4}}(\log g)^{24}\right).$$

Hence

$$\sum_{k\geq 1}\operatorname{Vol}_k \mathcal{Q}_g t^k = \left(\sum_{k\geq 1}\Upsilon_g^{(1;k)}t^k\right)\cdot\left(1+O\left((\log g)^{14}\cdot g^{(|t|-2)/2}\right)\right.$$
$$\left. +O\left((\log g)^{24}\cdot g^{(7|t|-8)/4}\right)\right). \tag{5.5}$$

Note that $\frac{|t|-2}{2} \geq -1$, so the error term $O\left((\log g)^{14}\cdot g^{(|t|-2)/2}\right)$ dominates the error term $O\left((\log g)^2\cdot g^{-1}\right)$ coming from (5.4). Note also that

$$\frac{|t|-2}{2} \geq \frac{7|t|-8}{4} \quad \text{for} \quad |t| \leq \frac{4}{5} \qquad \text{and} \qquad \frac{|t|-2}{2} \leq \frac{7|t|-8}{4} \quad \text{for} \quad |t| \geq \frac{4}{5}.$$

This shows which of the two error terms in (5.5) dominates on which interval of the values $|t|$. Plugging (5.4) into (5.5), taking into consideration the observation concerning the domination of the error terms, and taking the maximum of powers 14 and 24 of logarithms to cover the case $|t| = \frac{4}{5}$, we complete the proof of Theorem 5.1. Note that passing to (5.2) we used that $\Gamma(3/2) = \sqrt{\pi}/2$. □

In the proof of Theorem 5.2 we use the following saddle point bound which corresponds to Equation (18) from [35, Proposition IV.1].

Proposition ([35], Proposition IV.1) *Let $f(z)$ be analytic in the disk $|z| < R$ with $0 < R \leq \infty$. Define $M(f;r)$ for $r \in (0,R)$ by $M(f;r) :=$*

$\sup_{|z|=r} |f(z)|$. *Then, one has for any* r *in* $(0, R)$*, the family of saddle-point upper bounds*

$$\left[z^n\right] f(z) \leq \frac{M(f; r)}{r^n} \qquad \textit{implying} \qquad \left[z^n\right] f(z) \leq \inf_{r \in (0,R)} \frac{M(f; r)}{r^n}. \quad (5.6)$$

Proof of Theorem 5.2 Let $\delta = \frac{44}{19}$. From Propositions 4.2 and 4.3 we have for all t in the interval $[0, \delta)$ the bounds

$$\sum_{k \geq 1} \Upsilon_g^{(2;k)} t^k \leq B_2 \cdot \left(\frac{8}{3}\right)^{4g} \cdot t \cdot g^{\frac{2t-3}{2}} \cdot (\log g)^{14},$$

$$\sum_{k \geq 1} \sum_{V \geq 3} \Upsilon_g^{(V;k)} t^k \leq B_3 \cdot \left(\frac{8}{3}\right)^{4g} \cdot t \cdot g^{\frac{9t-10}{4}} \cdot (\log g)^{24}.$$

Combining these bounds with (5.6) we obtain for all non-negative integer k

$$\Upsilon_g^{(2;k)} \leq B_2 \cdot \left(\frac{8}{3}\right)^{4g} \cdot (\log g)^{14} \cdot g^{-3/2} \cdot \inf_{t \in (0,\delta)} \left(t^{1-k} g^t\right),$$

$$\sum_{V \geq 3} \Upsilon_g^{(V;k)} \leq B_3 \cdot \left(\frac{8}{3}\right)^{4g} \cdot (\log g)^{24} \cdot g^{-5/2} \cdot \inf_{t \in (0,\delta)} \left(t^{1-k} g^{\frac{9t}{4}}\right).$$

For the rest of the proof we assume that t is real and is contained in the interval $[0, \delta)$. The minima of $t^{1-k} g^t$ and $t^{1-k} g^{\frac{9t}{4}}$ on $[0, +\infty)$ are reached at $t = \frac{k-1}{\log g}$ and at $t = \frac{k-1}{\frac{9}{4} \log g}$ respectively. Hence, for $k - 1 \leq \delta \cdot \log g$, we obtain the following bounds:

$$\Upsilon_g^{(2;k)} \leq B_2 \cdot \left(\frac{8}{3}\right)^{4g} \cdot (\log g)^{14} \cdot g^{-3/2} \cdot (\log g)^{k-1} \cdot \left(\frac{e}{k-1}\right)^{k-1},$$

$$\sum_{V \geq 3} \Upsilon_g^{(V;k)} \leq B_3 \cdot \left(\frac{8}{3}\right)^{4g} \cdot (\log g)^{24} \cdot g^{-5/2} \times \left(\frac{9}{4}\right)^{k-1} (\log g)^{k-1} \left(\frac{e}{k-1}\right)^{k-1}.$$

Now, for g large enough we have $\sqrt{2\pi\delta\log g} \le \log g$. Hence, by Stirling formula, for g large enough and for all $k-1 \le \delta \cdot \log g$ we have

$$\Upsilon_g^{(2;k)} \le B_2 \cdot \left(\frac{8}{3}\right)^{4g} \cdot (\log g)^{15} \cdot g^{-3/2} \cdot \frac{(\log g)^{k-1}}{(k-1)!},$$
$$\sum_{V\ge 3} \Upsilon_g^{(V;k)} \le B_3 \cdot \left(\frac{8}{3}\right)^{4g} \cdot (\log g)^{25} \cdot g^{-5/2} \cdot \left(\frac{9}{4}\right)^{k-1} \cdot \frac{(\log g)^{k-1}}{(k-1)!}.$$

Combining expression (3.4) (in which we set $m = +\infty$) for $\Upsilon_g^{(1;k)} = \operatorname{Vol}\Gamma_k(g)$ from Theorem 3.4 with expression (3.40) for $\widetilde{H}_{3g-3,\infty,1/2}(k)$ from Corollary 3.18 we get the following asymptotics for $\Upsilon^{(1;k)}$ as $g \to +\infty$:

$$\Upsilon_g^{(1;k)} = \sqrt{\frac{2}{\pi}}\frac{1}{\sqrt{3g-3}}\left(\frac{8}{3}\right)^{4g-4}\left(\frac{1}{2}\right)^{k-1}\frac{\big(\log(6g-6)\big)^{k-1}}{(k-1)!}$$
$$\times\left(\frac{1}{\Gamma\left(1+\frac{k-1}{\log(6g-6)}\right)} + O\left(\frac{k-1}{(\log g)^2}\right)\right).$$

For all $k-1 < \delta\log g$ the rightmost factor in the above expression is greater than or equal to $1/\Gamma(1+44/19) + O((\log g)^{-1})$. We have $1/\Gamma(1+44/19) > 1/3$. Hence, for all $k-1 < \delta\log g$ and as $g \to +\infty$ we have

$$\frac{\Upsilon_g^{(2;k)} + \sum_{V\ge 3}\Upsilon_g^{(V;k)}}{\Upsilon_g^{(1;k)}} = O\left((\log g)^{25} \cdot \max\left(g^{-1}\cdot 2^k, g^{-2}\cdot\left(\frac{9}{2}\right)^k\right)\right). \tag{5.7}$$

Rewriting

$$2^k = g^{\frac{k}{\log g}\log 2} \qquad \text{and} \qquad \left(\frac{9}{2}\right)^k = g^{\frac{k}{\log g}(\log 9 - \log 2)}$$

we obtain

$$\max\left(g^{-1}\cdot 2^k, g^{-2}\cdot\left(\frac{9}{2}\right)^k\right) = g^{\max\left(-1+\frac{k\log 2}{\log g}, -2+\frac{k}{\log g}(\log 9-\log 2)\right)}$$

Solving the linear equation we find that

$$\begin{aligned} -2 + x\log\tfrac{9}{2} \le -1 + x\log 2 \quad &\text{for} \quad x \le \frac{1}{\log\frac{9}{4}} \approx 1.23315,\\ -1 + x\log 2 \le -2 + x\log\tfrac{9}{2} < 0 \quad &\text{for} \quad \frac{1}{\log\frac{9}{4}} \le x < \frac{2}{\log\frac{9}{2}} \approx 1.32972. \end{aligned} \tag{5.8}$$

Note that $\delta = \frac{44}{19} \approx 2.31$, so $\delta > \frac{2}{\log\frac{9}{2}}$. Note also that $\Upsilon_g^{(1;k)} = \operatorname{Vol}\Gamma_k(g)$. Using (5.7) and (5.8) we conclude that for any x satisfying $x < \frac{2}{\log\frac{9}{2}}$ we have

$$\begin{aligned} \operatorname{Vol}_k \mathcal{Q}_g &= \Upsilon_g^{(1;k)} \left(1 + \frac{\Upsilon_g^{(2;k)} + \sum\limits_{V\ge 3} \Upsilon_g^{(V;k)}}{\Upsilon_g^{(1;k)}} \right)\\ &= \begin{cases} \operatorname{Vol}\Gamma_k(g)\left(1 + O\left((\log g)^{25}\cdot g^{-1+\frac{k\log 2}{\log g}}\right)\right) & \text{for}\, k \le \frac{\log g}{\log\frac{9}{4}};\\ \operatorname{Vol}\Gamma_k(g)\left(1+O\left((\log g)^{25}\cdot g^{-2+\frac{k\log\frac{9}{2}}{\log g}}\right)\right) & \text{for}\, \frac{\log g}{\log\frac{9}{4}} \le k \le x\log g, \end{cases} \end{aligned}$$

uniformly in the corresponding range of k. This completes the proof of the first assertion of Theorem 5.2.

By (3.71) we have

$$\frac{\Upsilon_g^{(1;k)}}{\Upsilon_g^{(1)}} = p_{g,\infty}^{(1)}(k) = q_{3g-3,\infty,\frac{1}{2}}(k)\cdot\left(1 + O\left(\frac{(k+2\log g)^2}{g}\right)\right)$$

uniformly for all $k \le \frac{2\log g}{\log\frac{9}{2}}$. Combining Eq. (5.2) from Theorem 5.1 and Eq. (3.2) from Theorem 3.1 we get

$$\operatorname{Vol}\mathcal{Q}_g = \Upsilon_g^{(1)}\left(1 + O\left((\log g)^{24} g^{-1/4}\right)\right).$$

(Alternatively, we could use Eq. (5.5) for the particular value $t = 1$ to obtain the latter relation.)

Note that

$$-1 + x\log 2 \le -\frac{1}{4} \quad \text{for} \quad x \le \frac{3}{4\log 2} \approx 1.08202,$$
$$-\frac{1}{4} \le -1 + x\log 2 \quad \text{for} \quad x \ge \frac{3}{4\log 2}.$$

Combining these considerations with (5.8) we conclude that for any $x < \frac{2}{\log \frac{9}{2}}$ we have

$$p_g(k) = \begin{cases} q_{3g-3,\infty,\frac{1}{2}}(k)\left(1 + O\left((\log g)^{25} \cdot g^{-\frac{1}{4}}\right)\right) & \text{for} \quad k \le \frac{3\log g}{4\log 2}; \\ q_{3g-3,\infty,\frac{1}{2}}(k)\left(1 + O\left((\log g)^{25} \cdot g^{-1+\frac{k\log 2}{\log g}}\right)\right) & \text{for} \quad \frac{3\log g}{4\log 2} \le k \le \frac{\log g}{\log \frac{9}{4}}; \\ q_{3g-3,\infty,\frac{1}{2}}(k)\left(1 + O\left((\log g)^{25} \cdot g^{-2+\frac{k\log\frac{9}{2}}{\log g}}\right)\right) & \text{for} \quad \frac{\log g}{\log \frac{9}{4}} \le k \le x\log g, \end{cases}$$

uniformly for all k in the corresponding ranges. Theorem 5.2 is proved. □

5.2 Remaining proofs

Next we deduce the statements stated in Sect. 1 from Theorems 5.1 and 5.2.

Proof of Theorem 1.12 Let $K_g(\gamma)$ be the number of components of a random multicurve γ on a surface of genus g. Let

$$F_g(t) = \sum_{k=1}^{+\infty} \mathrm{Vol}_k\, \mathcal{Q}_g\, t^k.$$

By definition $F_g(1) = \mathrm{Vol}\, \mathcal{Q}_g$ and we have

$$\mathbb{E}_g(t^{K_g(\gamma)}) = \frac{F_g(t)}{F_g(1)}.$$

Applying Theorem 5.1 we obtain the result. □

We say that a multicurve $\gamma = m_1\gamma_1 + \cdots + m_k\gamma_k$ is *non-separating* if primitive components $\gamma_1, \ldots, \gamma_k$ of γ represent linearly independent homology cycles. Otherwise we say that a multicurve is *separating*. Clearly, $S \setminus \{\gamma_1 \cup \cdots \cup \gamma_k\}$ is connected if and only if γ is non-separating, so non-separating multicurves correspond to stable graphs with a single vertex, while separating multicurves correspond to stable graphs with two and more vertices.

The Corollary below is a quantitative version of an analogous statement [3, Proposition 10.7] due to A. Aggarwal. We originally conjectured a weaker form of this assertion in [26, Conjecture 1.33].

Corollary 5.3 *For any $k \geq 1$ we have*

$$\mathbb{P}(\gamma \textit{ is separating} \mid K_g(\gamma) = k) = O\left((\log g)^{25} \cdot g^{-1}\right).$$

Proof By definition,

$$\begin{aligned}\mathbb{P}(\gamma \text{ is separating} \mid K_g(\gamma) = k) &= \frac{\Upsilon_g^{(2;k)} + \sum_{V\geq 3} \Upsilon_g^{(V;k)}}{\operatorname{Vol}_k \mathcal{Q}_g} \\ &\leq \frac{\Upsilon_g^{(2;k)} + \sum_{V\geq 3} \Upsilon_g^{(V;k)}}{\Upsilon_g^{(1;k)}}.\end{aligned}$$

Equation (5.7) in the proof of Theorem 5.2 and analysis of the error term following it provides an upper bound for the right hand-side of the expression above from which the corollary follows. □

Remark 5.4 We proved in [29, Theorem 1.18], that for $k = 1$ we, actually, have the following exponential decay

$$\mathbb{P}(\gamma \text{ is separating} \mid K_g(\gamma) = 1) \sim \sqrt{\frac{2}{3\pi g}} \cdot 4^{-g}.$$

Proof of Theorems 1.1 and 1.4 It follows from combination of Propositions 8.3, 8.4, 8.5 in [3] proved by A. Aggarwal that the relative contribution to the Masur–Veech volume $\operatorname{Vol} \mathcal{Q}_g$ coming from all stable graphs in $\mathcal{G}_g$ which have more than one vertex, tends to zero as $g \to +\infty$. Translated to the language of multicurves or to the language of square-tiled surfaces, this statement corresponds to assertion (a) of Theorems 1.1 and 1.4.

In terms of the results of the present paper, the same statement can be justified comparing Theorems 3.1 and 5.1 and observing that the asymptotics of $\sum_{k\geq 1} \operatorname{Vol}(\Gamma_k(g))$ and of $\operatorname{Vol}(\mathcal{Q}_g)$ are the same up to a factor which tends to 1 as $g \to +\infty$.

Assertion (b) is a particular case, corresponding to the value $m = 1$ of the parameter m, of more general Theorems 1.2 and 1.5. These Theorems are proved independently below.

Let us prove assertion (c). By Theorem 5.2, in the range $k = o(\log g)$ the volume contributions $\operatorname{Vol}_k \mathcal{Q}_g$ and $\Upsilon_g^{(1;k)}$ are asymptotically equivalent. We have

$$\Upsilon_g^{(1;k)} = \sum_{\substack{m_1,\dots,m_k \\ m_i \ge 1 \text{ for } i=1,\dots,k}} \operatorname{Vol}\big(\Gamma_k(g), (m_1, \dots m_k)\big),$$

where the contribution of primitive multicurves is equal to $\operatorname{Vol}\big(\Gamma_g(k), (1, \dots 1)\big)$. By Theorem 3.4, the contribution to $\operatorname{Vol} \mathcal{Q}_g$ coming from all non-separating multicurves and from all primitive non-separating multicurves are respectively proportional to $\widetilde{H}_{3g-3,\infty,1/2}(k)$ and to $\widetilde{H}_{3g-3,1,1/2}(k)$ with the same coefficient of proportionality $\frac{2\sqrt{2}}{\sqrt{\pi}} \cdot \sqrt{3g-3} \cdot \left(\frac{8}{3}\right)^{4g-4}$. By Corollary 3.18, the quantities $\widetilde{H}_{3g-3,\infty,1/2}$ and $\widetilde{H}_{3g-3,1,1/2}(k)$ are asymptotically equivalent in the range $k = o(\log g)$, which completes the proof. □

Proof of Theorems 1.2 and 1.5 Taking the ratio of expression (3.1) from Theorem 3.1 evaluated at $t = 1$ over expression (3.2) from the same Theorem we get

$$\lim_{g\to+\infty} \frac{\displaystyle\sum_{k=1}^{3g-3} \sum_{\substack{m_1,\dots,m_k \\ m_i \ge 1 \text{ for } i=1,\dots,k}} \operatorname{Vol}(\Gamma_k(g), (m_1, \dots, m_k))}{\displaystyle\sum_{k=1}^{3g-3} \operatorname{Vol}\big(\Gamma_k(g)\big)} = \sqrt{\frac{m}{m+1}}.$$

Since the contribution from stable graphs with $V \ge 2$ vertices is negligible, it is a fortiori negligible when we consider bounded multiplicities $m_i \le m$. Hence, we have as $g \to +\infty$ the asymptotics

$$\lim_{g\to+\infty} \frac{\displaystyle\sum_{k=1}^{+\infty} \sum_{\substack{m_1,\dots,m_k \\ m_i \ge 1 \text{ for } i=1,\dots,k}} \operatorname{Vol}(\Gamma_k(g), (m_1, \dots, m_k))}{\displaystyle\sum_{k=1}^{+\infty} \operatorname{Vol}\big(\Gamma_k(g)\big)}$$

$$= \lim_{g\to+\infty} \frac{\displaystyle\sum_{\Gamma \in \mathcal{G}_g} \sum_{\substack{m_1,\dots,m_k \\ m_i \ge 1 \text{ for } i=1,\dots,k}} \operatorname{Vol}(\Gamma, (m_1, \dots, m_k))}{\operatorname{Vol}(\mathcal{Q}_g)},$$

which concludes the proof. □

Proof of Theorem 1.3 and of Theorem 1.6 The central limit theorem for $K_g(\gamma)$ follows from the general Theorem 3.19 that holds under mod-Poisson convergence. The mod-Poisson convergence was stated in Theorem 1.12. It only remains to justify the normalization used in Theorem 1.3 and in Theorem 1.6.

It follows from (2.1) that

$$\operatorname{card}(\mathcal{ST}_g(N)) \sim m(g) \cdot N^{6g-6},$$

where

$$m(g) = \frac{\operatorname{Vol} \mathcal{Q}_g}{(12g-12) \cdot 2^{6g-6}}.$$

By the central limit theorem (Theorem 3.19) we obtain

$$\lim_{g\to+\infty} \lim_{N\to+\infty} \frac{1}{m(g) \cdot N^{6g-6}} \cdot \operatorname{card}\left\{ S \in \mathcal{ST}_g(N) \,\middle|\, \frac{K_g(S) - \lambda_{3g-3}}{\sqrt{\lambda_{3g-3}}} \le x \right\}$$
$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt.$$

It remains to use (2.15) from Theorem 2.3 of A. Aggarwal (see Theorem 1.7 in the original paper [3]) to compute

$$\frac{1}{m(g)} = \frac{(12g-12) \cdot 2^{6g-6}}{\operatorname{Vol} \mathcal{Q}_g} \sim 3\pi g \cdot \left(\frac{9}{8}\right)^{2g-2},$$

which proves Theorem 1.6.

The proof of Theorem 1.3 analogous.

M. Mirzakhani proved in [61] that for any integral multicurve $\eta \in \mathcal{ML}(\mathbb{Z})$ one has

$$\operatorname{card}\left(\left\{\gamma \in \mathcal{ML}_g(\mathbb{Z}) \,\middle|\, \iota(\eta, \gamma) \le N\right\} / \operatorname{Stab}(\eta)\right) \sim \tilde{c}(\eta) \cdot N^{6g-6}.$$

Now let $\eta = \rho_g$, where ρ_g is a simple closed non-separating curve on a surface of genus g. Note that the stable graph associated to ρ_g is $\Gamma_1(g)$ and that the associated weight m_1 is equal to 1. By [33, Proposition 8.8] the asymptotic frequency $\tilde{c}(\rho_g)$ in the expression above is proportional to the asymptotic frequency $c(\rho_g)$ defined in (2.16) with the following factor:

$$c(\rho_g) = 2^{2g-3} \cdot \tilde{c}(\rho_g).$$

Combining this relation with (2.21) and (2.20) (where we let $n = 0$) we get

$$\operatorname{Vol}(\Gamma_1(g), 1) = 2(6g-6) \cdot (4g-4)! \cdot 2^{6g-6} \cdot \tilde{c}(\rho_g),$$

Since $\operatorname{Vol}(\Gamma_1(g)) = \zeta(6g-6) \cdot \operatorname{Vol}(\Gamma_1(g), 1)$ and since $\zeta(6g-6) \sim 1$ as $g \to +\infty$ we conclude that

$$\frac{1}{\tilde{c}(g)} \sim \frac{12g \cdot (4g-4)! \cdot 2^{6g-6}}{\operatorname{Vol}(\Gamma_1(g))} \sim \sqrt{\frac{3\pi g}{2}} \cdot 12g \cdot (4g-4)! \cdot \left(\frac{9}{8}\right)^{2g-2},$$

where we used

$$\operatorname{Vol}\Gamma_1(g) = \sqrt{\frac{2}{3\pi g}} \cdot \left(\frac{8}{3}\right)^{4g-4} \cdot (1 + o(1)) \quad \text{as } g \to +\infty.$$

that is obtained by a combination of Theorem 3.4 and Corollary 3.18. Actually, the latter asymptotic equivalence was originally proved in Equation (4.5) from Theorem 4.2 in [29]. □

Proof of Theorem 1.13 At the current stage, we can prove mod-Poisson convergence of p_g only for a relatively small radius $R = 8/7 \approx 1.14286$. Thus, a straightforward application of Corollary 3.17 to p_g does not provide sufficiently strong estimates for the distribution p_g. This is why we proceed differently.

Relation (1.17) follows from combination of relations (1.12) for $q_{3g-3,\infty,1/2}(k)$ with relations expressing $p_g(k)$ through $q_{3g-3,\infty,1/2}(k)$ proved in Theorem 5.2.

To estimate the left and right tails of the distribution p_g we use relation (5.3)

$$p_g(k) = q_{3g-3,\infty,1/2}(k) \cdot \left(1 + O\left(g^{-1/4} \cdot (\log g)^{24}\right)\right),$$

proved in Theorem 5.2 for k satisfying $k \le \frac{3}{4\log 2}\log g$.

Estimate (1.19) for the left tails follows directly from Eq. (1.14) of Theorem 1.10.

For the right tail, the equivalence between $p_g(k)$ and $q_{3g-3,\infty,1/2}(k)$ is not known beyond $\frac{3}{4\log 2}$ and we need to pass to estimates on the complementary event. Let $\lambda_{3g-3} = \frac{1}{2}\log(6g-6)$. Relation (5.3) implies, that for x in the range $0 \le x \le \frac{3}{2\log 2} \approx 2.16$ we have

$$\sum_{k=1}^{\lceil x\lambda_{3g-3}\rceil} p_g(k) = \left(\sum_{k=1}^{\lceil x\lambda_{3g-3}\rceil} q_{3g-3,\infty,1/2}(k)\right) \cdot \left(1 + O\left(g^{-1/4} \cdot (\log g)^{24}\right)\right).$$

In particular, this relation is applicable to $x_1 = 1.236$ and to $x_2 = 1.24$. Passing to complementary probabilities, we get for $0 \le x \le \frac{3}{2\log 2}$:

$$\sum_{k=\lceil x\lambda_{3g-3}\rceil+1}^{3g-3} p_g(k) = \sum_{k=\lceil x\lambda_{3g-3}\rceil+1}^{3g-3} q_{3g-3,\infty,1/2}(k) + O\left(g^{-1/4}\cdot(\log g)^{24}\right). \tag{5.9}$$

We use now relation (1.13) from Theorem 1.10 for the bound for the tail of distribution $q_{3g-3,\infty,1/2}$. Applying (1.13) with $n = 3g-3$, $\lambda = \frac{1}{2}\log(6g-6)$, we get

$$\begin{aligned}\sum_{k=\lceil x\lambda_{3g-3}\rceil+1}^{3g-3} q_{3g-3,\infty,1/2}(k) &= O\Big(\exp\big(-\lambda_{3g-3}(-x\log x - x + 1)\big)\Big).\\ &= O\left(g^{(-x\log x - x+1)/2}\right).\end{aligned}$$

For $x_1 = 1.236$ we have

$$(-x_1\log x_1 - x_1 + 1)/2 > -0.249.$$

Since the function $-x\log x - x + 1$ takes value 0 at $x = 1$ and is monotonously decreasing on $[1, +\infty]$ we conclude that for $x \in (1, x_1]$ we have

$$g^{-1/4}\cdot(\log g)^{24} = o\left(g^{(-x\log x - x+1)/2}\cdot\frac{1}{\log g}\right).$$

This implies that for this range of x the error term in the right-hand side of (5.9) is negligible with respect to the error term in (1.13) evaluated with parameters $n = 3g-3$, $\lambda = \frac{1}{2}\log(6g-6)$, and (1.18) follows.

For $x = x_2 = 1.24$ we have

$$(-x_2\log x_2 - x_2 + 1)/2 < -0.253,$$

so

$$O\left(g^{(-x\log x - x+1)/2}\right) = o\left(g^{-1/4}\right).$$

Taking into consideration monotonicity of $-x\log x - x + 1$ for $x \ge 1$, this implies that for $x \ge x_2$ the first summand in the right-hand side of (5.9) becomes negligible with respect to the second summand. Note also that

$$x_2\lambda_{3g-3} = 0.62\log(6g-6) < 0.62\log g + 0.62\log 6 < 0.62\log g + 1.12.$$

Thus, the sum of $q_{3g-3,\infty,1/2}(k)$ starting from $k = \lfloor 0.62 \log g \rfloor + 1$ might contain at most three extra terms with respect to the sum starting from $\lceil x_2 \lambda_{6g-6} \rceil + 1$. Clearly, each of these three terms has order $o\left(g^{-1/4}\right)$. We have proved that

$$\sum_{k=\lceil x_0 \log g \rceil + 1}^{3g-3} q_{3g-3,\infty,1/2}(k) = o\left(g^{-1/4}\right).$$

Plugging the above estimates in (5.9) we obtain our estimate for the right part. To estimate the left part we rely (1.19) that we already proved. It is sufficient to notice that for $x_0 = 0.18$ we have

$$-(x_0 \log x_0 - x_0 + 1)/2 < -0.255$$

and (1.20) follows. □

Proof of Theorem 1.14 The mod-Poisson convergence of $p_g(k)$ proved in Theorem 1.12 together with the general asymptotics of cumulants in Theorem 3.22 implies Theorem 1.14. □

6 Numerical and experimental data and further conjectures

6.1 Numerical and experimental data

In this section we compare the distribution $p_g(k)$ of the number of components of a random multicurve in genus g (see Theorems 1.1 and 1.13 from Sect. 1) with the approximation given by the mod-Poisson convergence.

Recall that for any $\lambda > 0$, we defined in (1.10) the real numbers $u_{\lambda,1/2}(k)$, for $k \in \mathbb{N}$, as the coefficients of the Taylor expansion of

$$e^{\lambda(t-1)} \cdot \frac{t \cdot \Gamma\left(\frac{3}{2}\right)}{\Gamma\left(1 + \frac{t}{2}\right)} = \sum_{k \geq 1} u_{\lambda,1/2}(k) \cdot t^k$$

We have the formula

$$u_{\lambda,1/2}(k) = \sqrt{\pi} \cdot e^{-\lambda} \cdot \frac{1}{k!} \cdot \sum_{i=1}^{k} \binom{k}{i} \cdot \phi_i \cdot \left(\frac{1}{2}\right)^i \cdot \lambda^{k-i},$$

where ϕ_k are the coefficients of the Taylor series of $1/\Gamma(t)$. Even though the sequence $\{u_{\lambda,1/2}(k)\}_{k\geq 1}$ is not a probability distribution we refer to this collection of numbers as the $(\mathrm{Poi}_\lambda, \Gamma(\frac{1}{2}))$-distribution

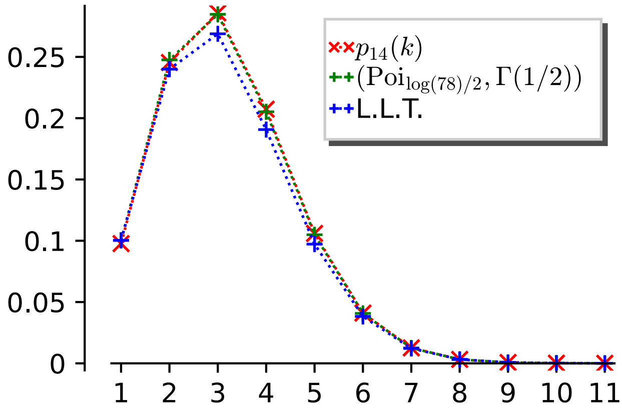

Fig. 4 The exact distribution $p_g(k)$ of the number of components k of a random multicurves (red), the $(\mathrm{Poi}_{\lambda_{3g-3}}, \Gamma(\frac{1}{2}))$-distribution (green) and the distribution given by the local limit theorem (blue) for $g = 14$

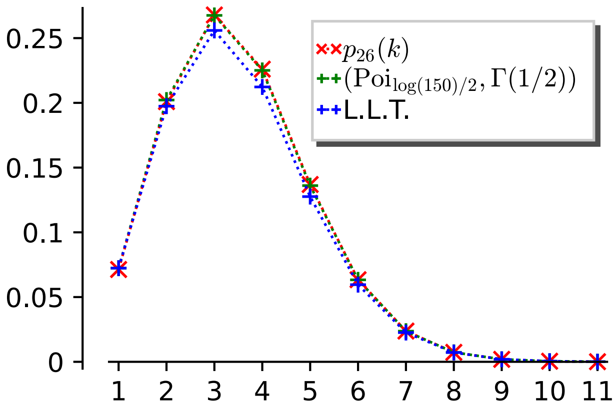

Fig. 5 (Experimental) distribution $p_g(k)$ of the number of components k of a random multi-cuvrves (red), the $(\mathrm{Poi}_{\lambda_{3g-3}}, \Gamma(\frac{1}{2}))$-distribution (green) and the distribution given by the local limit theorem (blue) for $g = 26$

Corollary 1.9 shows that $q_g(k)$ (and hence also $p_g(k)$) is well-approximated by $u_{\lambda_{3g-3},1/2}(k)$. Theorem 1.13 also shows that $u_{\lambda_{3g-3},1/2}(k)$ can be approximated by a much simpler formula, namely

$$u_{\lambda_{3g-3},1/2}(k+1) \sim p_g(k+1) \sim e^{-\lambda_{3g-3}} \frac{(\lambda_{3g-3})^k}{k!} \frac{\Gamma(1/2)}{2 \cdot \Gamma\left(1 + \frac{k}{2} \cdot \lambda_{3g-3}\right)}.$$

In the tables below and in Figs. 4, 5 and 6 we refer to the approximation given by the function $u_{\lambda_{3g-3},1/2}$ as the $(\mathrm{Poi}_{\lambda_{3g-3}}, \Gamma(\frac{1}{2}))$-approximation, and to the approximation in the right-hand side of the above expression as "LLT"-approximation (for "Local Limit Theorem"). We provide numerical

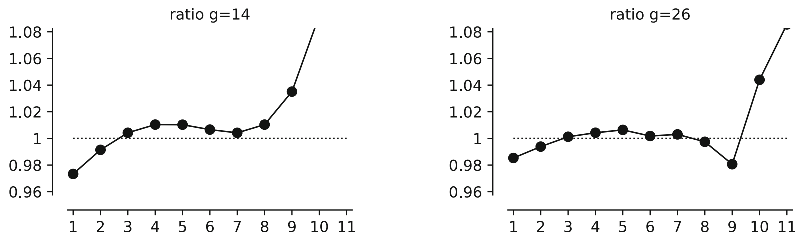

Fig. 6 Ratios $p_g(k)/u_{\lambda_{3g-3},1/2}(k)$ for $g = 14$ (exact) and for $g = 26$ (experimental)

data comparing the three distributions in the tables below. For $g = 14$ the distribution p_{14} was rigorously computed as a sequence of explicit rational numbers. For $g = 26$ the distribution p_{26} was computed experimentally, collecting statistics of random integral generalized interval exchange transformations (linear involutions). The graphic comparison of this data is presented in Figs. 4, 5 and 6. We observe a particularly rapid convergence of p_g to $(\mathrm{Poi}_{\lambda_{3g-3}}, \Gamma(\frac{1}{2}))$. In particular, already for $g = 14$ the graphs of these two distributions are visually indistinguishable.

k	p_{26}	$(\mathrm{Poi}_{\lambda_{3\cdot 26-3}}, \Gamma(\frac{1}{2}))$	LLT
1	0.0713	0.0724	0.0724
2	0.2009	0.2022	0.1974
3	0.2679	0.2675	0.2559
4	0.2260	0.2251	0.2123
5	0.1369	0.1361	0.1276
6	0.0634	0.0633	0.0596
7	0.0237	0.0237	0.0226
8	0.0073	0.0073	0.0072
9	0.0019	0.0019	0.0020
10	0.0005	0.0004	0.0005
11	0.0001	0.0001	0.0001

6.2 Further conjectures

Recall that the square-tiled surfaces which we study in this paper are integer points in the total space of the bundle of quadratic differentials $\mathcal{Q}_g$ over $\mathcal{M}_g$. In this section, we also consider square-tiled *translation* surfaces that correspond to integer points in the total space of the Hodge bundle $\mathcal{H}_g$ of holomorphic Abelian differentials over $\mathcal{M}_g$. Note that $\mathcal{H}_g$ can be considered as a subspace of $\mathcal{Q}_g$ of codimension $2g - 2$.

We first conjecture an analogue of our local limit theorem (Theorem 1.13) in the context of square-tiled *translation* surfaces.

Conjecture 1 *Let $p_g^{Ab}(k)$ be the probability that a random square-tiled translation surface in $\mathcal{H}_g$ has k cylinders. Then for all $x > 0$, we have uniformly in $k \in \{0, 1, \ldots, \lfloor x \log(g) \rfloor\}$*

$$p_g^{Ab}(k+1) = e^{-\mu_g} \cdot \frac{(\mu_g)^k}{k!} \cdot \left(\frac{1}{\Gamma\left(1 + \frac{k}{\mu_g}\right)} + o(1) \right) \textit{ as } g \to \infty,$$

where $\mu_g = \log(4g - 3)$.

In plain words, Conjecture 1 implies that the statistics $p_g^{Ab}(k)$ becomes practically indistinguishable from the statistics of the number of disjoint cycles in the cycle decomposition of a random permutation in S_{4g-3}, with respect to the uniform probability measure on the symmetric group of $4g - 3$ elements. The latter was denoted by $\mathbb{P}_{4g-3,\infty,1}$ in Sect. 3.

Both $\mathcal{Q}_g$ and $\mathcal{H}_g$ are stratified by the partition of the order of the zeros of respectively the quadratic and Abelian differentials. The parameters $6g - 6$ and $4g - 3$ that appear in the mod-Poisson convergence of random square-tiled surfaces from respectively Theorem 1.13 and Conjecture 1 coincide with the dimensions $\dim_{\mathbb{C}} \mathcal{Q}_g = 6g - 6$ and $\dim_{\mathbb{C}} \mathcal{H}_g = 4g - 3$. We conjecture the following strong form of mod-Poisson convergence uniform for all non-hyperelliptic connected components of all strata.

Conjecture 2 *There exist a constant $R_2 > 1$ such that the mod-Poisson convergence as in Theorem 1.12 but with radius R_2 holds uniformly for all non-hyperelliptic components of strata of holomorphic quadratic differentials.*

More precisely, let $\mathcal{C}$ be a non-hyperelliptic component of a stratum of holomorphic quadratic differentials. Let $p_{\mathcal{C}}(k)$ denote the probability that a random square-tiled surface in $\mathcal{C}$ has k cylinders. Then

$$\sum_{k \geq 1} p_{\mathcal{C}}(k) t^k = (\dim_{\mathbb{C}} \mathcal{C})^{\frac{t-1}{2}} \cdot \frac{\sqrt{\pi}}{\Gamma(t/2)} \left(1 + O\left(\frac{1}{g}\right) \right),$$

where the error term is uniform over all non-hyperelliptic components of all strata of holomorphic quadratic differentials and uniform in t on compact subsets of the complex disk $|t| < R_2$.

Similarly, a uniform local limit theorem should be valid for strata in the moduli space of Abelian differentials.

Conjecture 3 *Conjecture 1 holds uniformly for all non-hyperelliptic connected components of all strata of Abelian differentials.*

More precisely, let $x > 0$. *Let* $\mathcal{C}$ *be a non-hyperelliptic connected component of a stratum of Abelian differentials. Let* $p_{\mathcal{C}}(k)$ *denote the probability that a random Abelian square-tiled surface in* $\mathcal{C}$ *has* k *cylinders. Then uniformly for* k *in* $\{0, 1, \ldots, \lfloor x \log(\dim_{\mathbb{C}} \mathcal{C}) \rfloor\}$ *and uniformly in* $\mathcal{C}$ *such that* $\dim \mathcal{C} \to \infty$ *we have*

$$p_{\mathcal{C}}^{Ab}(k+1) = \frac{1}{\dim_{\mathbb{C}} \mathcal{C}} \cdot \frac{(\log \dim_{\mathbb{C}} \mathcal{C})^k}{k!} \cdot \left(\frac{1}{\Gamma\left(1 + \frac{k}{\log \dim_{\mathbb{C}} \mathcal{C}}\right)} + o(1) \right).$$

Conjecture 3 is based on analyzing huge experimental data. We experimentally collected statistics of the number $K_{\mathcal{C}}(S)$ of maximal horizontal cylinders in cylinder decompositions of random square-tiled surfaces in about 30 connected components $\mathcal{C}$ of strata in genera from 40 to 10,000. In particular, the least squares linear approximation for components $\mathcal{C}$ of dimension $\dim_{\mathbb{C}} \mathcal{C}$ between 400 and 20,000 gives:

$$\begin{aligned}
\mathbb{E}(K_{\mathcal{C}}) &\sim 0.999 \log \dim_{\mathbb{C}} \mathcal{C} + 0.581 \approx 0.999 \log \dim_{\mathbb{C}} \mathcal{C} + \gamma + 0.004 \\
\mathbb{V}(K_{\mathcal{C}}) &\sim 0.996 \log \dim_{\mathbb{C}} \mathcal{C} - 1.043 \approx 0.996 \log \dim_{\mathbb{C}} \mathcal{C} + \gamma - \zeta(2) + 0.02
\end{aligned}$$

(compare to (1.2)). Visually the graphs of distributions $p_{\mathcal{C}}^{Ab}(k)$ and $\frac{s(\dim \mathcal{C}, k)}{(\dim \mathcal{C})!}$ are, basically, indistinguishable for large genera.

Acknowledgements We are deeply indebted to A. Aggarwal for transforming our dreams into reality by proving all our conjectures from [26]. We also very much appreciate his suggestions, including the indication on how to compute multi-variate harmonic sums, which was crucial for making correct predictions in [26]. His numerous precious comments on the preliminary versions of this paper allowed us to correct a technical mistake and many typos, and improve the presentation. Results of this paper were directly or indirectly influenced by beautiful and deep ideas of Maryam Mirzakhani. We are grateful to the anonymous referee for careful reading of the manuscript and for helpful comments which allowed to clean some typos and to improve the presentation. We thank S. Schleimer, who was the first person to notice that our experimental data on statistics of cylinder decompositions of random Abelian square-tiled surfaces seems to have resemblance with statistics of cycle decomposition of random permutations. We thank F. Petrov for the reference to the paper [37] in the context of cycle decomposition of random permutations. We thank M. Bertola, A. Borodin, G. Borot, D. Chen, A. Eskin, V. Feray, M. Kazarian, S. Lando, M. Liu, H. Masur, M. Möller, B. Petri, K. Rafi, A. Sauvaget, J. Souto, D. Zagier and D. Zvonkine for useful discussions. We thank B. Green for the talk at the conference "CMI at 20" and T. Tao for his blog both of which were very inspiring for us. We thank D. Calegari for kind permission to use a picture from his book [19] in Fig. 1. Figure 2 originally appeared in our paper [29] (copyright Duke University Press). We are grateful to MPIM (Bonn) and EIMI (St. Petersburg), where part of the work was done, and to MSRI (Berkeley) and MFO (Oberwolfach) for providing us with friendly and stimulating environments.

References

1. Abramowitz, M., Stegun, I.A. (eds.): Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 10th edn., pp. 258–259. Dover, New York (1972)
2. Aggarwal, A.: Large genus asymptotics for volumes of strata of abelian differentials. J. Am. Math. Soc. **33**(4), 941–989 (2020)
3. Aggarwal, A.: Large genus asymptotics for intersection numbers and principal strata volumes of quadratic differentials. Invent. Math. **226**(3), 897–1010 (2021)
4. Anantharaman, N., Monk, L.: A high-genus asymptotic expansion of Weil–Petersson volume polynomials (2020). arXiv:2011.14889 [math.GT]
5. Angel, O., Schramm, O.: Uniform infinite planar triangulations. Commun. Math. Phys. **241**(2–3), 191–213 (2003)
6. Aggarwal, A., Delecroix, V., Goujard, E., Zograf, P., Zorich, A.: Conjectural large genus asymptotics of Masur–Veech volumes and of area Siegel–Veech constants of strata of quadratic differentials. Arnold Math. J. **6**(2), 149–161 (2020)
7. Alexeev, A., Zograf, P.: Random matrix approach to the distribution of genomic distance. J. Comput. Biol. **21**(8), 622–631 (2014)
8. Arratia, R., Barbour, A.D., Tavaré, S.: Logarithmic Combinatorial Structures: A Probabilistic Approach. EMS Monographs in Mathematics. European Mathematical Society, Zürich (2003)
9. Arana-Herrera, F.: Equidistribution of families of expanding horospheres on moduli spaces of hyperbolic surfaces. Geom. Dedicata **210**, 65–102 (2021)
10. Arana-Herrera, F.: Counting square-tiled surfaces with prescribed real and imaginary foliations and connections to Mirzakhani's asymptotics for simple closed hyperbolic geodesics. J. Mod. Dyn. **16**, 81–107 (2020)
11. Arana-Herrera, F.: Counting hyperbolic multi-geodesics with respect to the lengths of individual components and asymptotics of Weil–Petersson volumes, geometry and topology (to appear)
12. Athreya, J., Eskin, A., Zorich, A.: Right-angled billiards and volumes of moduli spaces of quadratic differentials on $\mathbb{C}P^1$. Ann. Scient. ENS, 4ème série, **49**, 1307–1381 (2016)
13. Bettinelli, J.: Scaling limits for random quadrangulations of positive genus. Electron. J. Probab. **15**, 1594–1644 (2010)
14. Bettinelli, J.: The topology of scaling limits of positive genus random quadrangulations. Ann. Probab. **40**, 1897–1944 (2012)
15. Brooks, R., Makover, E.: Random construction of Riemann surfaces. J. Differ. Geom. **68**(1), 121–157 (2004)
16. Budzinski, T., Curien, N., Petri, B.: Universality for random surfaces in unconstrained genus. Electron. J. Combin. **26**(4), Paper 4.2 (2019)
17. Budzinski, T., Curien, N., Petri, B.: The diameter of random Belyĭ surfaces Algebr. Geom. Topol. **21**(6), 2929–2957 (2021)
18. Budzinski, T., Louf, B.: Local limits of uniform triangulations in high genus. Invent. Math. **223**(1), 1–47 (2021)
19. Calegari, D.: Foliations and the Geometry Of 3-Manifolds. Oxford Mathematical Monographs. Oxford University Press, Oxford (2007)
20. Chassaing, P., Durhuus, B.: Local limit of labeled trees and expected volume growth in a random quadrangulation. Ann. Probab. **34**(3), 879–917 (2006)
21. Chen, D., Möller, M., Sauvaget, A.: With an appendix by Borot, G., Giacchetto, A., Lewanski, D., Masur–Veech volumes and intersection theory: the principal strata of quadratic differentials. arXiv:1912.02267 [math.AG]
22. Choi, J., Cvijović, D.: Values of the polygamma functions at rational arguments. J. Phys. A **40**(50), 15019–15028 (2007)

23. Chen, D., Möller, M., Sauvaget, A., Zagier, D.: Masur–Veech volumes and intersection theory on moduli spaces of Abelian differentials. Invent. Math. **222**(1), 283–373 (2020)
24. Chmutov, S., Pittel, B.: The genus of a random chord diagram is asymptotically normal. J. Comb. Theory Ser. A **120**(1), 102–110 (2013)
25. Curien, N.: Planar stochastic hyperbolic triangulations. Probab. Theory Relat. Fields **165**(3–4), 509–540 (2016)
26. Delecroix, V., Goujard, E., Zograf, P., Zorich, A.: Masur–Veech volumes, frequencies of simple closed geodesics and intersection numbers of moduli spaces of curves. arXiv:1908.08611 [math.GT] (2019)
27. Delecroix, V., Goujard, E., Zograf, P., Zorich, A.: Enumeration of meanders and Masur-Veech volumes. In: Forum of Mathematics Π, vol. 8:e4 (2020)
28. Delecroix, V., Goujard, E., Zograf, P., Zorich, A.: With appendix of Ph. Engel, Contribution of one-cylinder square-tiled surfaces to Masur–Veech volumes. In: Crovisier, S., et al. (ed.) Some Aspects of the Theory of Dynamical Systems: A Tribute to Jean-Christophe Yoccoz, vol I, , vol. 415, pp. 223–274. Astérisque, SMF, Paris (2020)
29. Delecroix, V., Goujard, E., Zograf, P., Zorich, A.: Masur–Veech volumes, frequencies of simple closed geodesics and intersection numbers of moduli spaces of curves. Duke Math. J. **170**(12), 2633–2718 (2021)
30. Delecroix, V., Liu, M.: Length partition of random multicurves on large genus hyperbolic surfaces. arXiv:2202.10255 [math.GT] (2022)
31. Erdős, P., Kac, M.: On the Gaussian law of errors in the theory of additive functions. Proc. Natl. Acad. Sci. USA **25**, 206–207 (1939)
32. Erlandsson, V., Parlier, H., Souto, J.: Counting curves, and the stable length of currents. J. EMS **22**(6), 1675–1702 (2020)
33. Erlandsson, V., Souto, J.: Geodesic currents and Mirzakhani's curve counting (to appear)
34. Féray, V., Méliot, P.-L., Nikeghbali, A.: Mod-ϕ convergence. Normality zones and precise deviations. Springer Briefs in Probability and Mathematical Statistics (2016)
35. Flajolet, Ph., Sedgewick, R.: Analytic Combinatorics. Cambridge University Press, Cambridge (2009)
36. Gamburd, A.: Poisson–Dirichlet distribution for random Belyĭ surfaces. Ann. Probab. **34**(5), 1827–1848 (2006)
37. Goncharov, V.L.: Some facts from combinatorics. Izvestia Akad. Nauk USSR Ser. Mat. **8**, 3–48 (1944)
38. Goujard, E.: Volumes of strata of moduli spaces of quadratic differentials: getting explicit values. Ann. Inst. Fourier **66**(6), 2203–2251 (2016)
39. Guo, J., Yang, D.: On the large genus asymptotics of psi-class intersection numbers (2021). arXiv:2110.06774 [math.AG]
40. Guth, L., Parlier, H., Young, R.: Pants decomposition of random surfaces. Geom. Funct. Anal. **21**(5), 1069–1090 (2011)
41. Harer, J., Zagier, D.: The Euler characteristic of the moduli space of curves. Invent. Math. **85**(3), 457–485 (1986)
42. Hide, W., Magee, M.: Near optimal spectral gaps for hyperbolic surfaces (2021). arXiv:2107.05292 [math.SP]
43. Hwang, H.K.: Théorèmes limites pour les structures combinatoires et les fonctions arithmétiques. Ph.D. thesis, École Polytechnique (1994)
44. Hwang, H.K.: Asymptotics of Poisson approximation to random discrete distrubtions: an analytic approach. Adv. Appl. Probab. **31**(2), 448–491 (1999)
45. Hwang, H.K.: Asymptotic expansions for the Stirling numbers of the first kind. J. Comb. Theory Ser. A **71**, 343–351 (1995)
46. Janson, S., Louf, B.: Short cycles in high genus unicellular maps. arXiv:2103.02549 [math.PR]

47. Kazakov, V.A., Kostov, I.K., Migdal, A.A.: Critical properties of randomly triangulated planar random surfaces. Phys. Lett. B **157**(4), 295–300 (1985)
48. Kazarian, M.: Recursion for Masur-Veech volumes of moduli spaces of quadratic differentials. J. Inst. Math. Jussieu 1–6 (2021) https://doi.org/10.1017/S1474748020000638
49. Krikun, M.: Local structure of random quadrangulations. arXiv:math/0512304 [math.PR]
50. Kontsevich, M.: Intersection theory on the moduli space of curves and the matrix Airy function. Commun. Math. Phys. **147**, 1–23 (1992)
51. Kowalski, E., Nikeghbali, A.: Mod-Poisson convergence in probability and number theory. Int. Math. Res. Not. **18**, 3549–3587 (2010)
52. Le Gall, J.-F.: Brownian geometry (Takagi lectures). Jpn. J. Math. **14**(2), 135–174 (2019)
53. Lipnowski, M., Wright, A.: Towards optimal spectral gaps in large genus. arXiv:2103.07496 [math.GT]
54. Liu, M.: Length statistics of random multicurves on closed hyperbolic surfaces. arXiv:1912.11155 [math.GT]; to appear in Groups, Geometry, Dynamics
55. Louf, B.: Planarity and non-separating cycles in uniform high genus quadrangulations. arXiv:2012.06512 [math.PR]
56. Magee, M., Naud, F., Puder, D.: A random cover of a compact hyperbolic surface has relative spectral gap $\frac{3}{16} - \varepsilon$ Geom. Funct. Anal. **32**(3), 595–661 (2022)
57. Masur, H.: Interval exchange transformations and measured foliations. Ann. Math. **115**, 169–200 (1982)
58. Masur, H., Rafi, K., Randecker, A.: Expected covering radius of a translation surface. Int. Math. Res. Not. arXiv:1809.10769
59. Ménard, L.: The two uniform infinite quadrangulations of the plane have the same law. Ann. Inst. Henri Poincaré Probab. Stat. **46**(1), 190–208 (2010)
60. Miermont, G.: Aspects of random maps. Lecture notes of Saint-Flour school 2014. http://perso.ens-lyon.fr/gregory.miermont/coursSaint-Flour.pdf
61. Mirzakhani, M.: Simple geodesics on hyperbolic surfaces and the volume of the moduli space of curves. Ph.D. Thesis, Harvard University (2004)
62. Mirzakhani, M.: Growth of the number of simple closed geodesics on hyperbolic surfaces. Ann. Math. (2) **168**(1), 97–125 (2008)
63. Mirzakhani, M.: Ergodic theory of the earthquake flow. IMRN **3**, 1–39 (2008)
64. Mirzakhani, M.: Growth of Weil–Petersson volumes and random hyperbolic surfaces of large genus. J. Differ. Geom. **94**(2), 267–300 (2013)
65. Mirzakhani, M.: Counting mapping class group orbits on hyperbolic surfaces. arXiv:1601.03342 [math.GT]
66. Mirzakhani, M., Petri, B.: Lengths of closed geodesics on random surfaces of large genus. Comment. Math. Helv. **94**(4), 869–889 (2019)
67. Mirzakhani, M., Zograf, P.: Towards large genus asymtotics of intersection numbers on moduli spaces of curves. GAFA **25**, 1258–1289 (2015)
68. Monk, L.: Benjamini–Schramm convergence and spectrum of random hyperbolic surfaces of high genus. Analysis & PDE (2021)
69. Monk, L., Thomas, J.: The tangle-free hypothesis on random hyperbolic surfaces. Int. Math. Res. Not. rnab160 (2021)
70. Monin, L., Telpukhovskiy, V.: On normalizations of Thurston measure on the space of measured laminations. Topol. Appl. **267**, 1–12 (2019)
71. Nikeghbali, A., Zeindler, D.: The generalized weighted probability measure on the symmetric group and the asymptotic behavior of the cycles. Ann. Inst. Henri Poincaré Probab. Stat. **49**(4), 961–981 (2013)
72. Pitman, J.: Combinatorial stochastic processes. Ecole d'Eté de Probabilités de Saint-Flour XXXII—2002. Lecture Notes in Mathematics. Springer (2006)
73. Rafi, K., Souto, J.: Geodesic currents and counting problems. GAFA **29**(3), 871–889 (2019)
74. Rényi, A., Turán, P.: On a theorem of Erdős-Kac. Acta Arith. **4**, 71–84 (1958)

75. Riordan, J.: Moment recurrence relations for binomial, Poisson and hypergeometric frequency distributions. Ann. Math. Stat. **8**(2), 103–111 (1937)
76. Selberg, A.: Note on a paper by L. G. Sathe. J. Indian Math. Soc. B **18**, 83–87 (1954)
77. Shrestha, S.: The topology and geometry of random square-tiled surfaces. arXiv:2005.00099 [math.GT]
78. Tao, T.: The Poisson–Dirichlet process, and large prime factors of a random number. https://terrytao.wordpress.com/2013/09/21/the-poisson-dirichlet-process-and-large-prime-factors-of-a-random-number
79. Thurston, W.P.: Geometry and Topology of Three-Manifolds. Lecture Notes. Princeton University, Princeton (1979)
80. Tutte, W.T.: A census of of planar maps. Can. J. Math. **15**, 249–271 (1963)
81. Veech, W.: Gauss measures for transformations on the space of interval exchange maps. Ann. Math. **115**, 201–242 (1982)
82. Vershik, A.M.: Statistical mechanics of combinatorial partitions, and their limit shapes. Funct. Anal. Appl. **30**, 90–105 (1996)
83. Walsh, T.R.S., Lehman, A.B.: Counting rooted maps by genus. I. J. Comb. Theory B **13**, 192–218 (1972)
84. Witten, E., Two-dimensional gravity and intersection theory on moduli space. In: Surveys in Differential Geometry (Cambridge, MA, 1991), pp. 243–310. Lehigh Univ., Bethlehem (1990)
85. Wu, Y., Xue, Y.: Random hyperbolic surfaces of large genus have first eigenvalues greater than $\frac{3}{16} - \varepsilon$. Geom. Funct. Anal. **32**(2), 340–410 (2022)
86. Yang, D., Zagier, D., Zhang, Y.: Masur–Veech volumes of quadratic differentials and their asymptotics. J. Geom. Phys. **158**, 1–12 (2020)