



A Glivenko–Cantelli theorem for almost additive functions on lattices

Christoph Schumacher*, Fabian Schwarzenberger¹, Ivan Veselić

Fakultät für Mathematik, Technische Universität Chemnitz, 09107 Chemnitz, Germany

Received 11 September 2015; received in revised form 3 May 2016; accepted 5 June 2016

Abstract

We develop a Glivenko–Cantelli theory for monotone, almost additive functions of i.i.d. sequences of random variables indexed by \mathbb{Z}^d . Under certain conditions on the random sequence, short range correlations are allowed as well. We have an explicit error estimate, consisting of a probabilistic and a geometric part. We apply the results to yield uniform convergence for several quantities arising naturally in statistical physics. © 2016 Elsevier B.V. All rights reserved.

MSC: 60F99; 60B12; 62E20; 60K35

Keywords: Glivenko–Cantelli theory; Uniform convergence; Empirical measures; Large deviations; Statistical mechanics

1. Introduction

The classical Glivenko–Cantelli theorem states that the empirical cumulative distribution functions of an increasing set of independent and identically distributed random variables converge *uniformly* to the cumulative population distribution function almost surely. Due to its importance to applications, e.g. statistical learning theory, the Glivenko–Cantelli theorem

* Corresponding author.

E-mail addresses: Christoph.Schumacher@mathematik.tu-chemnitz.de (C. Schumacher), fabian.schwarzenberger@imise.uni-leipzig.de (F. Schwarzenberger), ivan.veselic@mathematik.tu-chemnitz.de (I. Veselić).

¹ Current address: Fakultät Informatik/Mathematik, Hochschule für Technik und Wirtschaft Dresden, 01069 Dresden, Germany.

is also called the “fundamental theorem of statistics”. The theorem has initiated the study of so-called Glivenko–Cantelli classes as they feature, for instance, in the Vapnik–Chervonenkis theory [24]. Generalizations of the fundamental theorem rewrite the uniform convergence with respect to the real variable as convergence of a *supremum over a family (of sets or functions)* and widen the family over which the supremum is taken, making the statement “more uniform”. However, there are limits to this uniformization: For instance, if the original distribution is continuous, there is no convergence if the supremum is taken w.r.t. the family of finite subsets of the reals. Thus, a balance has to be found between the class over which the supremum is taken and the distribution of the random variables, the details of which are often dictated by the application in mind. Another important extension are multivariate Glivenko–Cantelli theorems, where the i.i.d. random variables are generalized to i.i.d. random vectors with possibly dependent coordinates. Such results have been obtained e.g. in [18,22,2,30]. In contrast to the classical one-dimensional Glivenko–Cantelli theorem, where no assumptions on the underlying distribution is necessary, in the higher dimensional case, one has to exclude certain singular continuous measures, cf. Theorem 5.3. The multidimensional version of the Portmanteau theorem provides a hint why such conditions are necessary. We apply these results in Section 5.

To avoid confusion, let us stress that uniform convergence in the classical Glivenko–Cantelli Theorem and in our result involves discontinuous functions, so it is quite different to uniform convergence of differentiable functions, as it is encountered e.g. with power series.

In many models of statistical physics one shows that certain random quantities are self-averaging, i.e. possess a well defined non-random thermodynamic limit. This is not only true for random operators of Schrödinger type, cf. e.g. [21,16,27], but also for spin systems, cf. e.g. [5,6,28,29,1]. Note however that the latter papers, studying the free energy (and derived quantities), heavily use specific properties of the exponential function (entering the free energy) like convexity and smoothness. We lack these properties in the Glivenko–Cantelli setting and are thus dealing with a completely different situation. The geometric ingredients of the proof of the thermodynamic limit can be traced back to papers by Van Hove [23] and Følner [4]. This is why the exhaustion sets used in the thermodynamic limit are associated with their names.

While standard statistical problems concern i.i.d. samples, an independence assumption quickly appears unnatural in statistical physics. Neighboring entities in solid state models (such as atoms or spins) are unlikely to not influence each other. In order to treat physically relevant scenarios one introduces a geometry to encode location and adjacency relations between the random variables, which in turn are used to allow dependencies between close random variables. In the present paper we choose \mathbb{Z}^d as our model of physical space, although our methods should apply to amenable groups as well, at least with an additional monotile condition. The focus on \mathbb{Z}^d allows us to avoid technicalities of amenable groups with monotiles and can thus present our results in a simpler, more transparent manner. Furthermore, we can achieve more explicit error bounds due to the simple geometry of \mathbb{Z}^d .

Our main result is Theorem 2.6, which is a Glivenko–Cantelli type theorem for a class of monotone, almost additive functions and suitable distributions of the random variables, allowing spatial dependencies. Our precise hypotheses are spelled out in Assumption 2.1 and Definition 2.3. The theorem can be interpreted as a multi-dimensional ergodic theorem with values in the Banach space of right continuous and bounded functions with sup-norm, i.e. a uniform convergence result. Under slightly strengthened assumptions we obtain an explicit error term for the convergence, which is a sum of a geometric and a probabilistic part, cf. Theorem 2.8. While earlier Banach space valued ergodic theorems, e.g. [10,11], have been restricted to a finite set of colors, we are able to treat the real-valued case. To do this, we have

to assume a monotonicity property, which is satisfied in most cases of interest. We obtain a more explicit convergence estimate than [10], as well. This is due to the fact that we assume a short range correlation condition, while [10] assumes the existence of limiting frequencies. The Glivenko–Cantelli result is applied to several examples from statistical physics in Sections 7 and 8. The flexibility and generality of our probabilistic model is displayed in the Appendix.

For the proof we use two sets of ideas. The first one concerns geometric approximation and tiling arguments for almost additive functions based on the amenability of the group \mathbb{Z}^d going back to the mentioned seminal papers of Van Hove [23] and Følner [4]. In the context of Banach space valued ergodic theorems they have been used for instance in [9,12,13,10,11,17]. The second ingredient of the proof is multivariate Glivenko–Cantelli theory, as developed in [18,22,2,30]. Our Theorem 5.5 shows that in our setting a large deviations type estimate derived by Wright can be applied. The latter is a modification of the Dvoretzky–Kiefer–Wolfowitz inequality [3,14].

The structure of the paper is as follows: In Section 2 we present our notation and the two main theorems. Section 3 contains an intuitive sketch of the proof in the case $\mathbb{Z}^d = \mathbb{Z}$, Section 4 geometric tiling and approximation arguments, Section 5 multivariate Glivenko–Cantelli theory, Section 6 the proof of the main theorem, and Sections 7 and 8 examples.

2. Notation and main results

The geometric setting of this paper is given via \mathbb{Z}^d , which gives in a natural way rise to a graph $(\mathbb{Z}^d, \mathcal{E})$. Here, the set of edges \mathcal{E} is the subset of the power set of \mathbb{Z}^d , consisting exactly of those $\{x, y\} \subseteq \mathbb{Z}^d$ which satisfy $\|y - x\|_1 = 1$. As usual $\|x\|_1 = \sum_{i=1}^d |x_i|$ denotes the ℓ^1 -norm in \mathbb{Z}^d . By \mathcal{F} we denote the (countable) set which consists of all finite subsets of \mathbb{Z}^d . For $\Lambda \in \mathcal{F}$, we write $|\Lambda|$ for the number of elements in Λ . The metric on the set of vertices $\mathfrak{d}: \mathbb{Z}^d \times \mathbb{Z}^d \rightarrow \mathbb{N}_0$ is defined via the ℓ^1 -norm, i.e. for $x, y \in \mathbb{Z}^d$ we set $\mathfrak{d}(x, y) := \|y - x\|_1$. For two sets $\Lambda_1, \Lambda_2 \subseteq \mathbb{Z}^d$ we write $\mathfrak{d}(\Lambda_1, \Lambda_2) := \min\{\mathfrak{d}(x, y) \mid x \in \Lambda_1, y \in \Lambda_2\}$. In the case that $\Lambda_1 = \{x\}$ contains only one element we write $\mathfrak{d}(x, \Lambda_2)$ for $\mathfrak{d}(\{x\}, \Lambda_2)$.

For $\Lambda \subseteq \mathbb{Z}^d$ we write $\Lambda + z := \{x + z \mid x \in \Lambda\}$. A cube of side length $n \in \mathbb{N}$ is a set which is given by $([0, n)^d \cap \mathbb{Z}^d) + z$ for some $z \in \mathbb{Z}^d$.

Using the metric \mathfrak{d} , we define for $r \in \mathbb{N}_0$ the r -boundary of a set $\Lambda \subseteq \mathbb{Z}^d$ by

$$\partial^r(\Lambda) := \{x \in \Lambda \mid \mathfrak{d}(x, \mathbb{Z}^d \setminus \Lambda) \leq r\} \cup \{x \in \mathbb{Z}^d \setminus \Lambda \mid \mathfrak{d}(x, \Lambda) \leq r\}.$$

Moreover, we set

$$\Lambda^r := \Lambda \setminus \partial^r(\Lambda) = \{x \in \Lambda \mid \mathfrak{d}(x, \mathbb{Z}^d \setminus \Lambda) > r\}. \quad (2.1)$$

If $(\Lambda_n)_{n \in \mathbb{N}}$ (or short (Λ_n)) is a sequence of subsets of \mathbb{Z}^d , we write $(\Lambda_n^r)_{n \in \mathbb{N}}$ or (Λ_n^r) instead of $((\Lambda_n)^r)_{n \in \mathbb{N}}$.

Note that for a cube Λ_n of side length n and $r \leq n/2$ we have

$$|\Lambda_n| = n^d, \quad |\Lambda_n^r| = (n - 2r)^d \quad \text{and} \quad |\partial^r(\Lambda_n)| = (n + 2r)^d - (n - 2r)^d.$$

In the following we introduce colorings of the elements of \mathbb{Z}^d . To this end, let $\mathcal{A} \subseteq \mathbb{R}$ be the set of possible colors. The sample set, which describes the set of all possible colorings of \mathbb{Z}^d is given by

$$\Omega := \mathcal{A}^{\mathbb{Z}^d} := \{\omega = (\omega_z)_{z \in \mathbb{Z}^d} \mid \omega_z \in \mathcal{A}\} \subseteq \mathbb{R}^{\mathbb{Z}^d}.$$

For each $z \in \mathbb{Z}^d$ we define the translation

$$\tau_z: \Omega \rightarrow \Omega, \quad (\tau_z \omega)_x = \omega_{x+z}, \quad (z \in \mathbb{Z}^d), \quad (2.2)$$

i.e. \mathbb{Z}^d acts on Ω via translations. For $\Lambda \in \mathcal{F}$ we set $\Omega_\Lambda := \mathcal{A}^\Lambda := \{\omega = (\omega_z)_{z \in \Lambda} \mid \omega_z \in \mathcal{A}\}$ and define $\Pi_\Lambda: \Omega \rightarrow \Omega_\Lambda$ by

$$\Pi_\Lambda(\omega) := \omega_\Lambda := (\omega_z)_{z \in \Lambda} \quad \text{for } \omega = (\omega_z)_{z \in \mathbb{Z}^d} \in \Omega.$$

We simplify $\Pi_z := \Pi_{\{z\}}$ for $z \in \mathbb{Z}^d$. As usual, \mathcal{A} is equipped with the Borel σ -algebra $\mathcal{B}(\mathcal{A})$ inherited from \mathbb{R} . Let $\mathcal{B}(\Omega)$ be the product σ -algebra on Ω . Let \mathbb{P} be a probability measure on $(\Omega, \mathcal{B}(\Omega))$ satisfying:

- Assumption 2.1.** (M1) *Translation invariance:* For each $z \in \mathbb{Z}^d$ we have $\mathbb{P} \circ \tau_z^{-1} = \mathbb{P}$.
 (M2) *Existence of densities:* There are σ -finite measures μ_z , $z \in \mathbb{Z}^d$, on $(\mathcal{A}, \mathcal{B}(\mathcal{A}))$ such that for each $\Lambda \in \mathcal{F}$ the measure $\mathbb{P}_\Lambda := \mathbb{P} \circ \Pi_\Lambda^{-1}$ is absolutely continuous with respect to $\mu_\Lambda := \bigotimes_{z \in \Lambda} \mu_z$ on $(\Omega_\Lambda, \mathcal{B}(\Omega_\Lambda))$. We denote the density function by $\rho_\Lambda := \frac{d\mathbb{P}_\Lambda}{d\mu_\Lambda}$. The measure \mathbb{P}_Λ is called a *marginal measure* of \mathbb{P} . It is defined on $(\Omega_\Lambda, \mathcal{B}(\Omega_\Lambda))$, where $\mathcal{B}(\Omega_\Lambda)$ is again the product σ -algebra.
 (M3) *Independence at a distance:* There exists $r \geq 0$ such that for all $n \in \mathbb{N}$ and non-empty $\Lambda_1, \dots, \Lambda_n \subseteq \mathbb{Z}^d$ with $\min\{\mathfrak{d}(\Lambda_i, \Lambda_j) \mid i \neq j\} > r$ we have $\rho_\Lambda = \prod_{j=1}^n \rho_{\Lambda_j}$, where $\Lambda = \bigcup_{j=1}^n \Lambda_j$.

Remark 2.2. • The constant $r \geq 0$ in (M3) can be interpreted as *correlation length*. In particular, if $r = 0$ this property implies that the colors of the vertices are independent.

- Conditions (M2) and (M3) are trivially satisfied, if \mathbb{P} is a product measure.
- For examples of measures \mathbb{P} satisfying (M1), (M2) and (M3) we refer to [Appendix](#).

In the following we deal with partial orderings on Ω and Ω_Λ , $\Lambda \in \mathcal{F}$. We write $\omega \leq \omega'$ for $\omega, \omega' \in \Omega$ if we have $\omega_z \leq \omega'_z$ for all $z \in \mathbb{Z}^d$, and analogously for Ω_Λ .

We consider the Banach space

$$\mathbb{B} := \{F: \mathbb{R} \rightarrow \mathbb{R} \mid F \text{ right-continuous and bounded}\},$$

which is equipped with supremum norm $\|\cdot\| := \|\cdot\|_\infty$.

We now introduce a certain class of \mathbb{B} -valued functions.

Definition 2.3. A function $f: \mathcal{F} \times \Omega \rightarrow \mathbb{B}$ is called *admissible* if the following conditions are satisfied

- (i) *translation invariance:* For $\Lambda \in \mathcal{F}$, $z \in \mathbb{Z}^d$ and $\omega \in \Omega$ we have
- $$f(\Lambda + z, \omega) = f(\Lambda, \tau_z \omega).$$
- (ii) *locality:* For all $\Lambda \in \mathcal{F}$ and $\omega, \omega' \in \Omega$ satisfying $\Pi_\Lambda(\omega) = \Pi_\Lambda(\omega')$ we have
- $$f(\Lambda, \omega) = f(\Lambda, \omega').$$
- (iii) *almost additivity:* There exists a function $b = b_f: \mathcal{F} \rightarrow [0, \infty)$ such that for arbitrary $\omega \in \Omega$, pairwise disjoint $\Lambda_1, \dots, \Lambda_n \in \mathcal{F}$ and $\Lambda := \bigcup_{i=1}^n \Lambda_i$ we have

$$\left\| f(\Lambda, \omega) - \sum_{i=1}^n f(\Lambda_i, \omega) \right\| \leq \sum_{i=1}^n b(\Lambda_i),$$

and b satisfies

- Please cite this article in press as: C. Schumacher, et al., A Glivenko–Cantelli theorem for almost additive functions on lattices, *Stochastic Processes and their Applications* (2016), <http://dx.doi.org/10.1016/j.spa.2016.06.005>

To see this, we choose $\Lambda \in \mathcal{F}$ and $\omega \in \Omega$ arbitrarily and calculate as follows:

$$\begin{aligned} \|f(\Lambda, \omega)\| &\leq \left\| f(\Lambda, \omega) - \sum_{z \in \Lambda} f(\{z\}, \omega) \right\| + \left\| \sum_{z \in \Lambda} f(\{z\}, \omega) \right\| \\ &\leq \sum_{z \in \Lambda} b(\{z\}) + \sum_{z \in \Lambda} \|f(\{z\}, \omega)\| \\ &\leq D|\Lambda| + \sum_{z \in \Lambda} \|f(\{0\}, \tau_{-z}\omega)\| \leq (D + \sup_{\omega \in \Omega} \|f(\{0\}, \omega)\|)|\Lambda|. \end{aligned}$$

Thus, $K_f \leq D + \sup_{\omega \in \Omega} \|f(\{0\}, \omega)\| < \infty$.

Definition 2.5. For $K, D, D' > 0$ and $r' \in \mathbb{N}$, we form the set

$$\mathcal{U}_{K,D,D',r'} = \{f : \mathcal{F} \times \Omega \rightarrow \mathbb{B} \mid f \text{ admissible with } K_f \leq K, D_f \leq D, D'_f \leq D' \text{ and } r'_f \leq r'\}$$

where K_f, D_f, D'_f and r'_f are the constants from [Definition 2.3](#) associated to f .

Let us state the main theorem of this paper.

Theorem 2.6. Let $\mathcal{A} \subseteq \mathbb{R}$, $\Omega := \mathcal{A}^{\mathbb{Z}^d}$ and $(\Omega, \mathcal{B}(\Omega), \mathbb{P})$ a probability space such that \mathbb{P} satisfies (M1), (M2) and (M3) with correlation length $r \in \mathbb{N}_0$, and let $f : \mathcal{F} \times \Omega \rightarrow \mathbb{B}$ be an admissible function. Let further $\Lambda_n := [0, n) \cap \mathbb{Z}^d$ for $n \in \mathbb{N}$. Then there exists a set $\tilde{\Omega} \in \mathcal{B}(\Omega)$ of full measure and a function $f^* \in \mathbb{B}$ such that for every $\omega \in \tilde{\Omega}$ we have

$$\lim_{n \rightarrow \infty} \left\| \frac{f(\Lambda_n, \omega)}{|\Lambda_n|} - f^* \right\| = 0. \quad (2.4)$$

Remark 2.7. • The following special case illustrates the relation to the Glivenko–Cantelli theorem. Let $\mathbb{P} := \bigotimes_{z \in \mathbb{Z}} \mu$ be a product measure on $\times_{\mathbb{Z}} \mathbb{R}$, where μ is a probability measure on \mathbb{R} , and let $f(\Lambda, \omega)(E) := \sum_{z \in \Lambda} \chi_{(-\infty, E]}(\omega_z)$ for $\Lambda \in \mathcal{F}$, $\omega \in \Omega$ and $E \in \mathbb{R}$. Then f is an admissible function. The quantities $f(\Lambda_n, \omega)(E)/|\Lambda_n| = |\Lambda_n|^{-1} \sum_{z \in \Lambda_n} \delta_{\omega_z}((-\infty, E])$ turn out to be the distribution functions of empirical measures. [Theorem 2.6](#) now states that the empirical distribution functions converge uniformly. The limit f^* is of course the distribution function of μ : $f^*(E) = \mu((-\infty, E])$ for all $E \in \mathbb{R}$.

- We emphasize that the statement of [Theorem 2.6](#) does not contain the measurability of the set

$$\left\{ \omega \in \Omega \mid \lim_{n \rightarrow \infty} \left\| \frac{f(\Lambda_n, \omega)}{|\Lambda_n|} - f^* \right\| = 0 \right\}.$$

Instead, the claim is that this set contains a measurable subset $\tilde{\Omega}$ of full measure. If the probability space was complete, the above set would be measurable, too. We write all almost sure statements in explicit fashion, in order to avoid a completeness assumption and measurability issues.

- The limit function f^* inherits the boundedness from f , since there exists $\omega \in \Omega$ such that

$$\|f^*\| \leq \limsup_{n \rightarrow \infty} \left(\left\| \frac{f(\Lambda_n, \omega)}{|\Lambda_n|} - f^* \right\| + \left\| \frac{f(\Lambda_n, \omega)}{|\Lambda_n|} \right\| \right) \leq K_f.$$

- Note that [Theorem 2.6](#) readily generalizes to absolutely convergent linear combinations of admissible functions in the following sense. Let $K, \alpha_j \in \mathbb{R}$, $j \in \mathbb{N}$ such that $\sum_{j \in \mathbb{N}} |\alpha_j| < \infty$

$$\begin{aligned} \limsup_{n \rightarrow \infty} \left\| \frac{f(\Lambda_n, \omega)}{|\Lambda_n|} - f^* \right\| &\leq \limsup_{n \rightarrow \infty} \left(\sum_{j=1}^{J-1} |\alpha_j| \left\| \frac{f_j(\Lambda_n, \omega)}{|\Lambda_n|} - f_j^* \right\| \right. \\ &\quad \left. + \sum_{j=J}^{\infty} |\alpha_j| \left(\left\| \frac{f_j(\Lambda_n, \omega)}{|\Lambda_n|} \right\| + \|f_j^*\| \right) \right) \\ &< \varepsilon. \end{aligned}$$

Please cite this article in press as: C. Schumacher, et al., A Glivenko–Cantelli theorem for almost additive functions on lattices, Stochastic Processes and their Applications (2016), <http://dx.doi.org/10.1016/j.spa.2016.06.005>

- As before, the monotonicity can be weakened, see [Remark 2.7](#). Note in particular, that a convex combination of functions in $\mathcal{U}_{K,D,D',r'}$ still obeys the quantitative estimates given by K, D, D' and r' . The statement of [Theorem 2.8](#) remains valid for the convex combination since the geometric part is derived without the use of monotonicity and the argument from [Remark 2.7](#) applies to the probabilistic part.

Corollary 2.10. *Under the conditions of [Theorem 2.8](#), there exists a set $\tilde{\Omega} \in \mathcal{B}(\Omega)$ with $\mathbb{P}(\tilde{\Omega}) = 1$ such that for all $\omega \in \tilde{\Omega}$, we have*

$$\sup_{f \in \mathcal{U}_{K,D,D',r'}} \sup_{E \in \mathbb{R}} \left| \frac{f(A_n, \omega)(E)}{|A_n|} - f^*(E) \right| \xrightarrow{n \rightarrow \infty} 0. \quad (2.5)$$

If furthermore for an admissible f , $f(A_n, \omega): \mathbb{R} \rightarrow \mathbb{R}$ is an isotone function for all A_n and $\omega \in \tilde{\Omega}$, then the limit function $f^* \in \mathbb{B}$ is isotone, too. In particular, cumulative distribution functions are preserved.

Proof. By [Theorem 2.8](#), we have

$$0 \leq \lim_{n \rightarrow \infty} \sup_{f \in \mathcal{U}_{K,D,D',r'}} \left\| \frac{f(A_n, \omega)}{|A_n|} - f^* \right\| \leq 2^{2d+1} \frac{2(K+D)r^d + 3Dr'^d}{m-2r} \xrightarrow{m \rightarrow \infty} 0.$$

Recall that the norm in \mathbb{B} is the sup norm $\|\cdot\| = \sup_{E \in \mathbb{R}} |\cdot(E)|$ to see [\(2.5\)](#).

If the functions $f_{n,\omega} := f(A_n, \omega)/|A_n|: \mathbb{R} \rightarrow \mathbb{R}$ are increasing, then for all $E, E' \in \mathbb{R}$ with $E < E'$ and $\varepsilon > 0$ we find $n \in \mathbb{N}$ such that $\|f_{n,\omega} - f^*\| < \varepsilon/2$ and

$$f^*(E) \leq f_n(E) + \varepsilon/2 \leq f_n(E') + \varepsilon/2 \leq f^*(E') + \varepsilon.$$

Since $\varepsilon > 0$ was arbitrary, f^* is increasing, too. \square

3. Illustration of the idea of proof

Let us consider the exemplary situation of dimension $d = 1$ and independently chosen colors, i.e., the constant r from (M3) equals 0. In this case, the idea of the proof of [Theorem 2.6](#) is illustrated in the following line:

$$\frac{1}{mk} f([0, mk), \omega) \stackrel{(1)}{\approx} \frac{1}{m} \langle f_m, L_{m,mk}^\omega \rangle \stackrel{(2)}{\approx} \frac{1}{m} \langle f_m, \mathbb{P}_m \rangle \stackrel{(3)}{\approx} f^*, \quad (3.1)$$

where $0 \ll m \ll k$. Assume that $n = mk$ and $A_n = [0, n)$. Then the left hand side in [\(3.1\)](#) equals the approximant in [Theorem 2.6](#). The function $f_m: \Omega_{[0,m)} \rightarrow \mathbb{B}$ is defined by $f_m(\omega) := f([0, m), \omega')$ for $\omega' \in \Pi_{[0,m)}^{-1}(\{\omega\})$, cf. [Remark 2.4](#). $L_{m,n}^\omega(B)$ is the empirical probability measure counting the number of occurrences of elements of $B \in \mathcal{B}(\Omega_{[0,m)})$ at the positions $[jm, (j+1)m)$, $j = 0, 1, \dots, k-1$ in ω , i.e.

$$L_{m,n}^\omega: \mathcal{B}(\Omega_{[0,m)}) \rightarrow [0, 1], \quad L_{m,n}^\omega := \frac{1}{k} \sum_{j=0}^{k-1} \delta_{(\tau_{jm}\omega)_{[0,m)}} = \frac{1}{k} \sum_{j=0}^{k-1} \delta_{\Pi_{[0,m)}(\tau_{jm}\omega)}. \quad (3.2)$$

We use the shortcut notation

$$\langle f_m, L_{m,n}^\omega \rangle := \int_{\Omega_{[0,m)}} f_m(\omega') dL_{m,n}^\omega(\omega') = \frac{1}{k} \sum_{j=0}^{k-1} f([jm, (j+1)m), \omega).$$

Let us discuss the three approximation steps separately.

- (1) The “ $\stackrel{(1)}{\approx}$ ” means that the two expressions are getting close to each other if m increases. To see this we use almost additivity of an admissible function, cf. (iii) of Definition 2.3. The detailed calculations will be presented in Section 4.
- (2) In the second step we compare the empirical measure $L_{m,mk}^\omega$ with the marginal measure $\mathbb{P}_m := \mathbb{P}_{[0,m]}$. The method of choice is a multivariate Glivenko–Cantelli theorem, which we apply in a version of DeHardt and Wright. In this particular situation it shows that for increasing k the expression $\langle f_m, L_{m,mk}^\omega \rangle$ converges to $\langle f_m, \mathbb{P}_m \rangle$ almost surely. This approximation step is explicitly discussed in Section 5.
- (3) In the last step we make intensive use of almost additivity of f in order to obtain that $(\langle f_m, \mathbb{P}_m \rangle / m)_m$ is Cauchy sequence in \mathbb{B} . As \mathbb{B} is a Banach space, we can identify the limit with an element $f^* \in \mathbb{B}$. The details are found in Section 6.

Remark 3.1 (*Frequencies*). From the discussion of step (1) above it is clear that the empirical measure counts occurrences of patterns at the positions $[jm, (j+1)m)$ for $j = 0, 1, \dots, k-1$. Thus, the corresponding sets are disjoint and their union covers the whole interval $[0, n)$, $n = mk$. In this sense, the present technique of counting occurrences differs from the counting in certain papers. For instance in [10,11], the authors count occurrences of patterns at each possible consecutive position, ignoring the fact that they may overlap. In our setting, this would correspond to the situation where the empirical measure is defined to count occurrences at all positions $[j, j+m)$, $j = 0, 1, \dots, m(k-1)$, i.e.,

$$\bar{L}_{m,n}^\omega: \mathcal{B}(\Omega_{[0,m)}) \rightarrow [0, 1], \quad \bar{L}_{m,n}^\omega := \frac{1}{m(k-1)+1} \sum_{j=0}^{m(k-1)} \delta_{(\tau_j \omega)_{[0,m)}}. \quad (3.3)$$

However, both ways of counting can be related to each other. The link can be seen best by comparing with the average

$$\frac{1}{m(k-1)} \sum_{j=1}^{m(k-1)} \delta_{(\tau_j \omega)_{[0,m)}} = \frac{1}{m} \sum_{i=1}^m \frac{1}{k-1} \sum_{j=0}^{k-2} \delta_{(\tau_{jm+i} \omega)_{[0,m)}}, \quad (3.4)$$

where the first observation $\delta_{\omega_{[0,m)}}$ is discarded. Indeed, for large $n = mk$, the difference between $\bar{L}_{m,n}^\omega$ and (3.4) vanishes. The right hand side of (3.4) shows that $\bar{L}_{m,n}^\omega$ is essentially a convex combination of empirical measures of the type (3.2). As $k \rightarrow \infty$, all the empirical measures of type (3.2) in (3.4) converge to the same limit \mathbb{P}_m , rendering the convex combination harmless. Recall that in the approximation first the limit $k \rightarrow \infty$ and afterwards the limit $m \rightarrow \infty$ is performed. This shows that the empirical measure defined in (3.3) converges to the same limit as the empirical measures in (3.2).

4. Approximation via the empirical measure

In the following we show how to estimate an admissible function f in terms of the empirical measure. As in Theorem 2.6, let $\Lambda_n = ([0, n) \cap \mathbb{Z})^d$ for each $n \in \mathbb{N}$.

Our aim is to approximate for $m \ll n$ the set Λ_n using translates of the set Λ_m . To this end, we define the grid

$$T_{m,n} := \{t \in m\mathbb{Z}^d \mid \Lambda_m + t \subseteq \Lambda_n\}. \quad (4.1)$$

Thus, we have $|T_{m,n}| = \lfloor n/m \rfloor^d$, $\Lambda_{\lfloor n/m \rfloor m} = \dot{\bigcup}_{t \in T_{m,n}} (\Lambda_m + t) = \Lambda_m + T_{m,n}$, and

$$\Lambda_n \setminus \Lambda_{\lfloor n/m \rfloor m} \subseteq \partial^m(\Lambda_n) \quad \text{or equivalently} \quad \Lambda_n^m \subseteq \Lambda_{\lfloor n/m \rfloor m}. \quad (4.2)$$

As in Remark 2.4, define for an admissible f and $\Lambda \in \mathcal{F}$ the function

$$f_\Lambda: \Omega_\Lambda \rightarrow \mathbb{B}, \quad f_\Lambda(\omega) := f(\Lambda, \omega') \quad \text{where } \omega' \in \Pi_\Lambda^{-1}(\{\omega\}). \quad (4.3)$$

By locality (ii) of Definition 2.3, f_Λ is well-defined. In the case $\Lambda = \Lambda_n$, we write

$$f_n := f_{\Lambda_n} \quad \text{and} \quad f_n^m := f_{\Lambda_n^m} \quad (4.4)$$

for $m \in \mathbb{N}_0$. Next, we introduce the empirical measure $L_{m,n}^\omega$ by setting for $\omega \in \Omega$ and $m, n \in \mathbb{N}$:

$$L_{m,n}^\omega: \mathcal{B}(\Omega_{\Lambda_m}) \rightarrow [0, 1], \quad L_{m,n}^\omega = \frac{1}{|T_{m,n}|} \sum_{t \in T_{m,n}} \delta_{(\tau_t \omega)_{\Lambda_m}}. \quad (4.5)$$

Here, $\delta_\omega: \mathcal{B}(\Omega_{\Lambda_m}) \rightarrow [0, 1]$ is the point measure on $\omega \in \Omega_{\Lambda_m}$. In the same manner, we define $L_{m,n}^{r,\omega}$ as an adaption of $L_{m,n}^\omega$ which ignores the r -boundary of Λ_m . The precise definition is the following: for $r \in \mathbb{N}_0$ we set

$$L_{m,n}^{r,\omega}: \mathcal{B}(\Omega_{\Lambda_m^r}) \rightarrow [0, 1], \quad L_{m,n}^{r,\omega} = \frac{1}{|T_{m,n}|} \sum_{t \in T_{m,n}} \delta_{(\tau_t \omega)_{\Lambda_m^r}}. \quad (4.6)$$

The variable r we used here will eventually be the constant from (M3), but here in Section 4 we do not need that specific value.

As illustrated before in Section 3 we use for $\Lambda \in \mathcal{F}$, a bounded and measurable $f: \Omega_\Lambda \rightarrow \mathbb{B}$, and a measure ν on $(\Omega_\Lambda, \mathcal{B}(\Omega_\Lambda))$ the notation

$$\langle f, \nu \rangle := \int_{\Omega_\Lambda} f(\omega) \, d\nu(\omega). \quad (4.7)$$

Lemma 4.1. Recall $\Lambda_n := ([0, n] \cap \mathbb{Z})^d$. For any admissible function $f: \mathcal{F} \times \Omega \rightarrow \mathbb{B}$ we have, for all $\omega \in \Omega$ and all $n, m, r \in \mathbb{N}$ with $n > 2m$,

$$\begin{aligned} \left\| \frac{f(\Lambda_n, \omega)}{n^d} - \frac{\langle f_m^r, L_{m,n}^{r,\omega} \rangle}{m^d} \right\| &\leq \frac{b(\Lambda_{\lfloor n/m \rfloor m})}{|\Lambda_{\lfloor n/m \rfloor m}|} + \frac{(2K_f + D)|\partial^m(\Lambda_n^m)|}{|\Lambda_n^m|} \\ &\quad + \frac{b(\Lambda_m) + b(\Lambda_m^r) + (K_f + D)|\partial^r(\Lambda_m)|}{|\Lambda_m|}. \end{aligned} \quad (4.8)$$

Moreover,

$$\lim_{m \rightarrow \infty} \lim_{n \rightarrow \infty} \left\| \frac{f(\Lambda_n, \omega)}{n^d} - \frac{\langle f_m^r, L_{m,n}^{r,\omega} \rangle}{m^d} \right\| = 0. \quad (4.9)$$

Proof. Let $\omega \in \Omega$ and $n, m, r \in \mathbb{N}$ be given such that $n > 2m$. This condition ensures that $\Lambda_n^m \neq \emptyset$. First we verify (4.8), and afterwards we show that this implies (4.9). By the triangle

inequality we obtain

$$\begin{aligned} \left\| \frac{f(\Lambda_n, \omega)}{n^d} - \frac{\langle f_m^r, L_{m,n}^{r,\omega} \rangle}{m^d} \right\| &\leq \left\| \frac{f(\Lambda_n, \omega)}{|\Lambda_n|} - \frac{f(\Lambda_n, \omega)}{|\Lambda_{[n/m]m}|} \right\| \\ &+ \left\| \frac{f(\Lambda_n, \omega) - f(\Lambda_{[n/m]m}, \omega)}{|\Lambda_{[n/m]m}|} \right\| + \left\| \frac{f(\Lambda_{[n/m]m}, \omega)}{|\Lambda_{[n/m]m}|} - \frac{\langle f_m, L_{m,n}^{\omega} \rangle}{m^d} \right\| \\ &+ \frac{\|\langle f_m, L_{m,n}^{\omega} \rangle - \langle f_m^r, L_{m,n}^{r,\omega} \rangle\|}{m^d}. \end{aligned} \quad (4.10)$$

We bound the four terms on the right hand side separately. To estimate the first term, we use $|\Lambda_{[n/m]m}| \geq |\Lambda_n^m|$, see (4.2), and obtain

$$0 \leq \frac{1}{|\Lambda_{[n/m]m}|} - \frac{1}{|\Lambda_n|} \leq \frac{1}{|\Lambda_n^m|} - \frac{1}{|\Lambda_n|} = \frac{|\Lambda_n| - |\Lambda_n^m|}{|\Lambda_n||\Lambda_n^m|} \leq \frac{|\partial^m(\Lambda_n^m)|}{|\Lambda_n||\Lambda_n^m|}.$$

Applying the bound given by (2.3) in Remark 2.4, we get

$$\left\| \frac{f(\Lambda_n, \omega)}{|\Lambda_n|} - \frac{f(\Lambda_n, \omega)}{|\Lambda_{[n/m]m}|} \right\| \leq K_f \frac{|\partial^m(\Lambda_n^m)|}{|\Lambda_n^m|}. \quad (4.11)$$

In order to find an appropriate upper bound for the second term in (4.10) we use almost additivity (iii), the inclusion (4.2) and $\hat{\Lambda}_{m,n} := \Lambda_n \setminus \Lambda_{[n/m]m}$ to obtain

$$\begin{aligned} \left\| \frac{f(\Lambda_n, \omega) - f(\Lambda_{[n/m]m}, \omega)}{|\Lambda_{[n/m]m}|} \right\| &\leq \frac{b(\Lambda_{[n/m]m})}{|\Lambda_{[n/m]m}|} + \frac{b(\hat{\Lambda}_{m,n})}{|\Lambda_{[n/m]m}|} + \frac{\|f(\hat{\Lambda}_{m,n}, \omega)\|}{|\Lambda_{[n/m]m}|} \\ &\leq \frac{b(\Lambda_{[n/m]m})}{|\Lambda_{[n/m]m}|} + \frac{D|\hat{\Lambda}_{m,n}|}{|\Lambda_{[n/m]m}|} + \frac{K_f|\hat{\Lambda}_{m,n}|}{|\Lambda_{[n/m]m}|} \\ &\leq \frac{b(\Lambda_{[n/m]m})}{|\Lambda_{[n/m]m}|} + \frac{(K_f + D)|\partial^m(\Lambda_n^m)|}{|\Lambda_n^m|}. \end{aligned} \quad (4.12)$$

To approximate the third term in (4.10), we calculate using translation invariance (i) of admissible functions

$$\begin{aligned} \langle f_m, L_{m,n}^{\omega} \rangle &= \int_{\Omega_{\Lambda_m}} f_m(\omega') dL_{m,n}^{\omega}(\omega') = \frac{1}{|T_{m,n}|} \sum_{t \in T_{m,n}} \int_{\Omega_{\Lambda_m}} f_m(\omega') d\delta_{(\tau_t \omega)_{\Lambda_m}}(\omega') \\ &= \frac{1}{|T_{m,n}|} \sum_{t \in T_{m,n}} f_m((\tau_t \omega)_{\Lambda_m}) = \frac{1}{|T_{m,n}|} \sum_{t \in T_{m,n}} f(\Lambda_m + t, \omega). \end{aligned} \quad (4.13)$$

This and (iii) of Definition 2.3 give

$$\begin{aligned} \left\| \frac{f(\Lambda_{[n/m]m}, \omega)}{|\Lambda_{[n/m]m}|} - \frac{\langle f_m, L_{m,n}^{\omega} \rangle}{|\Lambda_m|} \right\| &= \frac{1}{|T_{m,n}||\Lambda_m|} \left\| f(\Lambda_{[n/m]m}, \omega) - \sum_{t \in T_{m,n}} f(\Lambda_m + t, \omega) \right\| \\ &\leq \frac{1}{|T_{m,n}||\Lambda_m|} \sum_{t \in T_{m,n}} b(\Lambda_m + t) = \frac{b(\Lambda_m)}{|\Lambda_m|}. \end{aligned} \quad (4.14)$$

Finally, we estimate the fourth term. In the same way as in (4.13) we obtain

$$\langle f_m^r, L_{m,n}^{r,\omega} \rangle = \frac{1}{|T_{m,n}|} \sum_{t \in T_{m,n}} f(\Lambda_m^r + t, \omega).$$

Application of the triangle inequality, $\Lambda_m \setminus \Lambda_m^r = \Lambda_m \cap \partial^r(\Lambda_m) \subseteq \partial^r(\Lambda_m)$ and (iii) of Definition 2.3 as well as (2.3) lead to

$$\begin{aligned} \|\langle f_m, L_{m,n}^\omega \rangle - \langle f_m^r, L_{m,n}^{r,\omega} \rangle\| &\leq \frac{1}{|T_{m,n}|} \sum_{t \in T_{m,n}} \|f(\Lambda_m + t, \omega) - f(\Lambda_m^r + t, \omega)\| \\ &\leq \frac{1}{|T_{m,n}|} \sum_{t \in T_{m,n}} (b(\Lambda_m^r) + b(\Lambda_m \setminus \Lambda_m^r) + \|f((\Lambda_m \setminus \Lambda_m^r) + t, \omega)\|) \\ &\leq b(\Lambda_m^r) + (K_f + D)|\partial^r(\Lambda_m)|. \end{aligned} \quad (4.15)$$

It remains to combine (4.10) with the bounds (4.11), (4.12), (4.14) and (4.15) to obtain (4.8).

Let us turn to (4.9). As required, we first perform the limit $n \rightarrow \infty$. In (4.8), the bounding terms affected by this limit vanish, due to property (iii) and the fact that \mathbb{Z}^d is amenable:

$$\lim_{n \rightarrow \infty} \left(\frac{b(\Lambda_{\lfloor n/m \rfloor m})}{|\Lambda_{\lfloor n/m \rfloor m}|} + \frac{(2K_f + D)|\partial^m(\Lambda_n^m)|}{|\Lambda_n^m|} \right) = 0.$$

Secondly, we let $m \rightarrow \infty$. Since $b(\Lambda_m^r)/|\Lambda_m| \leq b(\Lambda_m^r)/|\Lambda_m^r|$ for $m > 2r$, this takes care of the remaining terms of the upper bound in (4.8).

$$\lim_{m \rightarrow \infty} \frac{b(\Lambda_m) + b(\Lambda_m^r) + (K_f + D)|\partial^r(\Lambda_m)|}{|\Lambda_m|} = 0.$$

Thus, (4.9) follows. \square

Remark 4.2. Let us emphasize that the statement of the lemma is not an “almost sure”-statement, but rather holds for all $\omega \in \Omega$.

5. Application of multivariate Glivenko–Cantelli theory

We briefly restate multivariate Glivenko–Cantelli results in Theorem 5.3 and apply this result to our setting in Theorem 5.6. To do so, we need some notions concerning monotonicity in \mathbb{R}^k .

Definition 5.1. Let $k \in \mathbb{N}$.

- Let $s \in \{-1, 1\}^k$. The closed cone \mathcal{C}_s with sign vector s is the set

$$\mathcal{C}_s := \{x = (x_j)_{j=1,\dots,k} \in \mathbb{R}^k \mid \forall j \in \{1, \dots, k\}: x_j s_j \geq 0\}.$$

The closed cone with sign vector s and apex $x \in \mathbb{R}^k$ is $\mathcal{C}_s(x) := x + \mathcal{C}_s$.

- A function $f: \mathbb{R}^k \rightarrow \mathbb{R}$ is *monotone*, if it is monotone in each coordinate, i.e. there exists $s \in \{-1, 1\}^k$ such that, for all $x, y \in \mathbb{R}^k$,

$$y \in \mathcal{C}_s(x) \implies f(y) \geq f(x).$$

- A set $\mathcal{T} \subseteq \mathbb{R}^k$ is a *monotone graph*, if there exists a sign vector $s \in \{-1, 1\}^k$ such that, for all $x \in \mathcal{T}$,

$$\mathcal{T} \cap \mathcal{C}_s(x) \subseteq \partial \mathcal{C}_s(x),$$

where ∂C denotes the boundary of C in \mathbb{R}^k .

- A set $\mathcal{T} \subseteq \mathbb{R}^k$ is a *strictly monotone graph*, if there exists a sign vector $s \in \{-1, 1\}^k$ such that, for all $x \in \mathcal{T}$,

$$\mathcal{T} \cap \mathcal{C}_s(x) = \{x\}.$$

Remark 5.2. • This notion of monotonicity is compatible with (iv) in Definition 2.3.

- We want to emphasize that in the above definition a second meaning of the notion of a *graph* was used. In Section 2 a graph was introduced as a pair consisting of a set of vertices and a set of edges. In contrast to that, Definition 5.1 states that a *monotone graph* is a certain subset of \mathbb{R}^k . In order to distinguish both meanings we will always insert the term *monotone* when speaking about subsets of \mathbb{R}^k .

The following theorem is proven in [30, Theorems 1 and 2]. Recall that the continuous part μ_c of a measure μ on \mathbb{R}^k is given by $\mu_c(A) := \mu(A) - \sum_{x \in A} \mu\{x\}$ for all Borel sets $A \in \mathcal{B}(\mathbb{R}^k)$.

Theorem 5.3 (DeHardt, Wright). *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $X_t: \Omega \rightarrow \mathbb{R}^k$, $t \in \mathbb{N}$, independent and identically distributed random variables with distribution μ , i.e., $\mu(A) := \mathbb{P}(X_1 \in A)$ for all $A \in \mathcal{B}(\mathbb{R}^k)$. For each $J \subseteq \{1, \dots, k\}$, $J \neq \emptyset$, let μ^J be the distribution of the vector $(X_1^j)_{j \in J}$ consisting of the coordinates $j \in J$ of the vector $X_1 = (X_1^j)_{j \in \{1, \dots, k\}}$, i.e. a marginal of μ . We denote by*

$$L_n: \Omega \times \mathcal{B}(\mathbb{R}^k) \rightarrow \mathbb{R}, \quad L_n^{(\omega)}(A) := \frac{1}{n} \sum_{t=1}^n \delta_{X_t(\omega)}$$

the empirical distribution corresponding to the sample (X_1, \dots, X_n) , $n \in \mathbb{N}$. Fix further $M > 0$ and let

$$\mathcal{M} := \{f: \mathbb{R}^k \rightarrow \mathbb{R} \mid f \text{ monotone and } \sup |f(\mathbb{R}^k)| \leq M\}.$$

Then the following assertions are equivalent:

- For all $J \subseteq \{1, \dots, k\}$, $J \neq \emptyset$, the continuous part μ_c^J of the marginal μ^J of μ vanishes on every strictly monotone graph $\Upsilon \subseteq \mathbb{R}^J$:

$$\mu_c^J(\Upsilon) = 0.$$

- There exists a set $\Omega' \in \mathcal{A}$ of full probability $\mathbb{P}(\Omega') = 1$ such that, for all $\omega \in \Omega'$,

$$\sup_{f \in \mathcal{M}} |\langle f, L_n^{(\omega)} - \mu \rangle| \xrightarrow{n \rightarrow \infty} 0.$$

- For all $\varepsilon > 0$, there are $a = a(\varepsilon) > 0$ and $b = b(\varepsilon) > 0$ such that for all $n \in \mathbb{N}$ there exists an $\Omega_{\varepsilon, n} \in \mathcal{A}$, such that for all $\omega \in \Omega_{\varepsilon, n}$, we have

$$\sup_{f \in \mathcal{M}} |\langle f, L_n^{(\omega)} - \mu \rangle| \leq \varepsilon \quad \text{and} \quad \mathbb{P}(\Omega_{\varepsilon, n}) \geq 1 - b \exp(-an).$$

Remark 5.4. Note that if we knew that the set $\{\omega \in \Omega \mid \sup_{f \in \mathcal{M}} |\langle f, L_n^{(\omega)} - \mu \rangle| \geq \varepsilon\}$ was measurable, we could rephrase (iii) as follows. For all $\varepsilon > 0$, the probabilities $\mathbb{P}(\sup_{f \in \mathcal{M}} |\langle f, L_n^{(\omega)} - \mu \rangle| \geq \varepsilon)$ converge exponentially fast to zero as $n \rightarrow \infty$.

We provide a sufficient condition for (i) in Theorem 5.3 and apply the theorem to our setting. The idea to use product measures in the context of Glivenko–Cantelli type theorems appears already in [22].

Theorem 5.5. *Let μ be a measure on \mathbb{R}^k which is absolutely continuous with respect to a product measure $\bigotimes_{j=1}^k \mu_j$ on \mathbb{R}^k , where μ_j , $j \in \{1, \dots, k\}$ are measures on \mathbb{R} . Then, for each strictly monotone graph $\Upsilon \subseteq \mathbb{R}^k$ we have $\mu_c(\Upsilon) = 0$, where μ_c is the continuous part of μ . Moreover, (i) from Theorem 5.3 is satisfied.*

Proof. Let ρ be the density of μ with respect to $\bigotimes_{j=1}^k \mu_j$. We define the set of atoms of μ ,

$$S := \{x \in \mathbb{R}^k \mid \mu\{x\} > 0\}, \quad \text{and} \quad S_j := \{x_j \in \mathbb{R} \mid \mu_j\{x_j\} > 0\} \quad (j \in \{1, \dots, k\}).$$

Then we have $S \subseteq S_1 \times \dots \times S_k$, and for each $x = (x_1, \dots, x_k) \in S_1 \times \dots \times S_k$, we have

$$\mu\{x\} = \rho(x) \prod_{j=1}^k \mu_j\{x_j\}. \quad (5.1)$$

This implies in particular that for all $x \in S_1 \times \dots \times S_k \setminus S$, we have $\rho(x) = 0$.

In order to prove $\mu_c(\mathcal{T}) = 0$ it is sufficient to show

$$\mu(\mathcal{T}) = \sum_{x \in S \cap \mathcal{T}} \mu\{x\}. \quad (5.2)$$

We will prove this by induction over k . If $k = 1$ then a strictly monotone graph is a singleton, i.e. $\mathcal{T} = \{x\}$ for some $x \in \mathbb{R}$. Thus, (5.2) holds true. In the case $k > 1$ we assume that (5.2) holds for $k - 1$ and proceed by disintegration. Note that, for $z \in \mathbb{R}$, the cross section $\mathcal{T}_z := \{y \in \mathbb{R}^{k-1} \mid (y, z) \in \mathcal{T}\}$ is itself a strictly monotone graph in \mathbb{R}^{k-1} . Using the cross section $\rho_z: \mathbb{R}^{k-1} \rightarrow \mathbb{R}$, $\rho_z(y) := \rho(y, z)$, $z \in \mathbb{R}$, of the density, we define the cross section $\mu_z := \rho_z \bigotimes_{j=1}^{k-1} \mu_j$ of the measure μ . By Fubini's Theorem, the disintegration of μ is

$$\mu(d(y, z)) = \rho_z(y) \bigotimes_{j=1}^{k-1} \mu_j(dy_j) \otimes \mu_k(dz).$$

By the induction hypothesis we obtain

$$\begin{aligned} \mu(\mathcal{T}) &= \int_{\mathbb{R}} \left(\int_{\mathbb{R}^{k-1}} \chi_{\mathcal{T}_z}(y) \mu_z(dy) \right) \mu_k(dz) \\ &= \int_{\mathbb{R}} \mu_z(\mathcal{T}_z) \mu_k(dz) = \int_{\mathbb{R}} \sum_{y \in S \cap \mathcal{T}_z} \mu_z\{y\} \mu_k(dz), \end{aligned} \quad (5.3)$$

where $\bar{S} := S_1 \times \dots \times S_{k-1}$. The next aim is to show that the set $\mathcal{Z} := \{z \in \mathbb{R} \mid \bar{S} \cap \mathcal{T}_z \neq \emptyset\}$ is countable. To this end, we will use that \bar{S} is countable, define two mappings

$$\varphi: \bar{S} \rightarrow (\bar{S} \times \mathbb{R}) \cap \mathcal{T} \quad \text{and} \quad \psi: (\bar{S} \times \mathbb{R}) \cap \mathcal{T} \rightarrow \mathcal{Z}$$

and show that they are surjective. We first define φ . Let $(y, z), (y, z') \in (\bar{S} \times \mathbb{R}) \cap \mathcal{T}$ be given and assume without loss of generality that $z \leq z'$. Let $s \in \{-1, 1\}^k$ be the sign vector of \mathcal{T} from Definition 5.1, and, again without loss of generality, consider the case $s(k) = 1$. Then we have

$$\mathcal{C}_s(y, z) \cap \mathcal{T} = \{(y, z)\} \quad \text{and} \quad \mathcal{C}_s(y, z') \cap \mathcal{T} = \{(y, z')\}.$$

As $z \leq z'$ and $s(k) = 1$, we have $\mathcal{C}_s(y, z) \supseteq \mathcal{C}_s(y, z')$, such that we obtain

$$\{(y, z)\} = \mathcal{C}_s(y, z) \cap \mathcal{T} \supseteq \mathcal{C}_s(y, z') \cap \mathcal{T} = \{(y, z')\}.$$

This shows that if $y \in \bar{S}$ is such that there exists an element $z \in \mathbb{R}$ with $(y, z) \in \mathcal{T}$, then this z is unique. Let $h \in (\bar{S} \times \mathbb{R}) \cap \mathcal{T}$ be arbitrary but fixed and set

$$\varphi: \bar{S} \rightarrow (\bar{S} \times \mathbb{R}) \cap \mathcal{T}, \quad \varphi(y) := \begin{cases} (y, z) & \text{if } (y, z) \in \mathcal{T}, \text{ and} \\ h & \text{if } (\{y\} \times \mathbb{R}) \cap \mathcal{T} = \emptyset. \end{cases}$$

This φ is well-defined and surjective. The mapping ψ is defined by

$$\psi: (\bar{S} \times \mathbb{R}) \cap \mathcal{T} \rightarrow \mathcal{Z}, \quad \psi(y, z) := z.$$

To check that ψ is surjective let $z \in \mathcal{Z}$ be given. Then there exists $y \in \bar{S} \cap \mathcal{T}_z$. Thus, by definition of \mathcal{T}_z we have $(y, z) \in \mathcal{T}$ and $(y, z) \in \bar{S} \times \mathbb{R}$. This shows that (y, z) is in the domain of ψ and $\psi(y, z) = z$.

The surjectivity of φ and ψ and the fact that \bar{S} is countable show that \mathcal{Z} is countable. Therefore the last integral in (5.3) is actually a sum:

$$\mu(\mathcal{T}) = \int_{\mathbb{R}} \sum_{y \in \bar{S} \cap \mathcal{T}_z} \mu_z\{y\} \mu_k(dz) = \sum_{z \in S_k} \sum_{y \in \bar{S} \cap \mathcal{T}_z} \mu_z\{y\} \mu_k\{z\} = \sum_{x \in S \cap \mathcal{T}} \mu\{x\}.$$

Here, the last equality follows from (5.1), $\bigcup_{z \in S_k} (\bar{S} \cap \mathcal{T}_z) \times \{z\} \supseteq S \cap \mathcal{T}$, and the fact that ρ vanishes on

$$\bigcup_{z \in S_k} (\bar{S} \cap \mathcal{T}_z) \times \{z\} \setminus (S \cap \mathcal{T}) \subseteq S_1 \times \cdots \times S_k \setminus S.$$

This finishes the induction and we obtained (5.2) and $\mu_c(\mathcal{T}) = 0$.

Let $J \subseteq \{1, \dots, k\}$ such that $J \neq \emptyset$ and $J^c := \{1, \dots, k\} \setminus J \neq \emptyset$. Define $\rho^J: \mathbb{R}^J \rightarrow \mathbb{R}$ via

$$\rho^J(x^J) := \int_{\mathbb{R}^{J^c}} \rho(x) d \bigotimes_{j \in J^c} \mu_j(x^{J^c}),$$

where $x = (x^J, x^{J^c}) \in \mathbb{R}^J \times \mathbb{R}^{J^c}$. The function ρ^J is the density of the marginal μ^J of μ with respect to $\bigotimes_{j \in J} \mu_j$, since by Fubini for all $A \in \mathcal{B}(\mathbb{R}^J)$

$$\mu^J(A) = \int_{\mathbb{R}^k} \chi_A(x^J) \rho(x) d \bigotimes_{j=1}^k \mu_j(x) = \int_A \rho^J(x^J) d \bigotimes_{j \in J} \mu_j(x^J).$$

Thus, the above calculation applies for all marginals of μ , too. This shows (i) from Theorem 5.3. \square

Now we approximate the empirical measure $L_{m,n}^{r,\omega}$ using the measure \mathbb{P}_m^r , see step (2) in Section 3. The connection to Assumption 2.1 is established by Theorem 5.5. As announced before we apply the multivariate Glivenko–Cantelli Theorem 5.3 for the proof of Theorem 5.6.

Theorem 5.6. Let $\Lambda_n := [0, n) \cap \mathbb{Z}^d$, $n \in \mathbb{N}$, a set $\mathcal{A} \subseteq \mathbb{R}$, $\Omega := \mathcal{A}^{\mathbb{Z}^d}$, a probability space $(\Omega, \mathcal{B}(\Omega), \mathbb{P})$ such that \mathbb{P} satisfies (M1), (M2) and (M3) and an admissible function f be given. Besides this let for $m, n \in \mathbb{N}$ and $\omega \in \Omega$ the empirical measure $L_{m,n}^{r,\omega}$ be given as in (4.6) and let $\mathbb{P}_m^r := \mathbb{P}_{\Lambda_m^r}$ be the marginal measure, where r is the constant given by (M3). Then there exists a set $\tilde{\Omega} \in \mathcal{B}(\Omega)$ of full measure, such that for all $\omega \in \tilde{\Omega}$ and all $m \in \mathbb{N}$:

$$\lim_{n \rightarrow \infty} \|\langle f_m^r, L_{m,n}^{r,\omega} - \mathbb{P}_m^r \rangle\| = 0. \quad (5.4)$$

Furthermore, for $K, D, D' > 0$ and $r' \in \mathbb{N}$, we have for all $\omega \in \tilde{\Omega}$ and $m \in \mathbb{N}$

$$\lim_{n \rightarrow \infty} \sup_{f \in \mathcal{U}_{K,D,D',r'}} \|\langle f_m^r, L_{m,n}^{r,\omega} - \mathbb{P}_m^r \rangle\| = 0. \quad (5.5)$$

Additionally, for each $\varepsilon > 0$ there exist $a = a(\varepsilon, m, K) > 0$ and $b = b(\varepsilon, m, K)$ such that for all $n \in \mathbb{N}$ there is a measurable set $\Omega(\varepsilon, n)$ with $\mathbb{P}(\Omega(\varepsilon, n)) \geq 1 - b \exp(-a \lfloor n/m \rfloor^d)$ and

$$\sup_{f \in \mathcal{U}_{K,D,D',r'}} \|\langle f_m^r, L_{m,n}^{r,\omega} - \mathbb{P}_m^r \rangle\| \leq \varepsilon \quad \text{for all } \omega \in \Omega(\varepsilon, n). \quad (5.6)$$

Proof. Let $m \in \mathbb{N}$ be given. We set $k := |\Lambda_m^r|$ and embed $\Omega_{\Lambda_m^r} \subseteq \mathbb{R}^k$. Fix an admissible function f . For each $E \in \mathbb{R}$, there exists a monotone and bounded function $g_{m,E}^r: \mathbb{R}^k \rightarrow \mathbb{R}$ which extends $f_m^r(\cdot)(E): \Omega_{\Lambda_m^r} \rightarrow \mathbb{R}$, i.e. $f_m^r(\omega)(E) = g_{m,E}^r(\omega)$ for all $\omega \in \Omega_{\Lambda_m^r}$. In fact, $g_{m,E}^r$ can be bounded by kK_f , where K_f is the constant introduced in (2.3). Thus, the set $\mathcal{M}_f := \{g_{m,E}^r \mid E \in \mathbb{R}\}$ is monotone and bounded by kK_f , see Remark 2.4.

In order to apply the Glivenko–Cantelli Theorem 5.3, we enumerate $[0, \infty)^d \cap m\mathbb{Z}^d$ with a sequence $(t_\ell)_{\ell \in \mathbb{N}}$ such that, for all $q \in \mathbb{N}$,

$$\{t_1, \dots, t_{q^d}\} = [0, mq^d)^d \cap m\mathbb{Z}^d.$$

Consider further for each $\ell \in \mathbb{N}$ the mapping

$$X_\ell^r: \Omega \rightarrow \Omega_{\Lambda_m^r} \subseteq \mathbb{R}^k, \quad X_\ell^r(\omega) := \Pi_{\Lambda_m^r}(\tau_{t_\ell}^{-1}\omega).$$

By (M3) the random variables X_ℓ^r , $\ell \in \mathbb{N}$ are independent with respect to the measure \mathbb{P} on $(\Omega, \mathcal{B}(\Omega))$. Moreover, applying (M1) shows that X_ℓ^r , $\ell \in \mathbb{N}$, are identically distributed. By definition, the empirical measure of X_ℓ^r , $\ell \in \{1, \dots, |T_{m,n}|\}$, where $|T_{m,n}| = \lfloor n/m \rfloor^d$, is exactly the empirical measure $L_{m,n}^{r,\omega}$ given in (4.6). According to (M2), the measure \mathbb{P}_m^r is absolutely continuous with respect to a product measure on $\Omega_{\Lambda_m^r}$. We trivially extend \mathbb{P}_m^r and $L_{m,n}^{r,\omega}$ to measures on \mathbb{R}^k (and use the same names for the extensions). This allows to apply Theorem 5.5, which gives (i) of Theorem 5.3. Thus, the Glivenko–Cantelli theorem implies that (for the $m \in \mathbb{N}$ chosen above) there is a set $\Omega_m \in \mathcal{B}(\Omega)$ of probability one such that for each $\omega \in \Omega_m$ we have

$$\|\langle f_m^r, L_{m,n}^{r,\omega} - \mathbb{P}_m^r \rangle\| = \sup_{g \in \mathcal{M}_f} |\langle g, L_{m,n}^{r,\omega} - \mathbb{P}_m^r \rangle| \xrightarrow{n \rightarrow \infty} 0,$$

since the supremum is bounded by the supremum in (ii) from Theorem 5.3. By the same token,

$$\sup_{f \in \mathcal{U}_{K,D,D',r'}} \|\langle f_m^r, L_{m,n}^{r,\omega} - \mathbb{P}_m^r \rangle\| = \sup_{f \in \mathcal{U}_{K,D,D',r'}} \sup_{g \in \mathcal{M}_f} |\langle g, L_{m,n}^{r,\omega} - \mathbb{P}_m^r \rangle| \xrightarrow{n \rightarrow \infty} 0.$$

In the light of that, the claimed convergences in (5.4) and (5.5) hold independently from m for all $\omega \in \Omega := \bigcap_{m \in \mathbb{N}} \Omega_m$. To obtain (5.6) we apply Theorem 5.3, (iii). \square

6. Almost additivity and limits, Proof of Theorem 2.6

Next we investigate the expression $\langle f_m^r, \mathbb{P}_m^r \rangle$ for large m . This is the third and last step in our approximation scheme. Thus, this step brings us in the position to prove our main results, namely Theorems 2.6 and 2.8.

Lemma 6.1. *Let $\mathcal{A} \subseteq \mathbb{R}$, $\Omega := \mathcal{A}^{\mathbb{Z}^d}$, a probability space $(\Omega, \mathcal{B}(\Omega), \mathbb{P})$ such that \mathbb{P} satisfies (M1), (M2) and (M3), an admissible function f and the sequence (Λ_n) with $\Lambda_n = ([0, n) \cap \mathbb{Z})^d$, $n \in \mathbb{N}$ be given. Besides this, let r be the constant from (M3) and let for $m \in \mathbb{N}$ the*

marginal measure $\mathbb{P}_m^r := \mathbb{P}_{\Lambda_m^r}$ and the function f_m^r be given as in (4.4). Then there exists a function $f^* \in \mathbb{B}$ with

$$\lim_{m \rightarrow \infty} \left\| \frac{\langle f_m^r, \mathbb{P}_m^r \rangle}{m^d} - f^* \right\| = 0. \quad (6.1)$$

Furthermore, we have, with b and D from Definition 2.3 and K_f from Remark 2.4, for all $m \in \mathbb{N}$

$$\left\| \frac{\langle f_m^r, \mathbb{P}_m^r \rangle}{m^d} - f^* \right\| \leq \frac{b(\Lambda_m^r)}{m^d} + (K_f + D) \frac{|\partial^r(\Lambda_m)|}{m^d}.$$

Proof. Let us define $F: \mathcal{F} \rightarrow \mathbb{B}$ by setting for each $\Lambda \in \mathcal{F}$:

$$F(\Lambda) := \langle f_\Lambda, \mathbb{P}_\Lambda \rangle = \int_{\Omega_\Lambda} f_\Lambda(\omega) d\mathbb{P}_\Lambda(\omega) = \int_{\Omega} f(\Lambda, \omega) d\mathbb{P}(\omega).$$

With this notation, it is sufficient to show that $(F(\Lambda_m^r)/m^d)_{m \in \mathbb{N}}$ is a Cauchy sequence.

First, we note that F is translation invariant, i.e. $F(\Lambda + t) = F(\Lambda)$. To see this, use (M1) and (i) of Definition 2.3. Note also, that F is almost additive with the same b and D as f , see (iii) of the same definition. Furthermore, it follows from Remark 2.4 that F is bounded in the following sense: For all $\Lambda \in \mathcal{F}$, we have $F(\Lambda) \leq K_f |\Lambda|$ with the same constant K_f as in (2.3).

Next, assume that two integers m, M with $m \leq M$ are given. As in (4.1), set

$$T_{m,M} := \{t \in (m\mathbb{Z})^d \mid \Lambda_m + t \subseteq \Lambda_M\}.$$

We are interested in an estimate of the difference

$$\delta(m, M) := \left\| \frac{F(\Lambda_M^r)}{M^d} - \frac{F(\Lambda_m^r)}{m^d} \right\|. \quad (6.2)$$

To study this we use the triangle inequality and get

$$\delta(m, M) \leq \frac{\alpha(m, M)}{M^d} + \beta(m, M) \quad (6.3)$$

with

$$\begin{aligned} \alpha(m, M) &:= \left\| F(\Lambda_M^r) - \sum_{t \in T_{m,M}} F(\Lambda_m^r + t) \right\|, \\ \beta(m, M) &:= \left\| \frac{F(\Lambda_m^r)}{m^d} - \sum_{t \in T_{m,M}} \frac{F(\Lambda_m^r + t)}{M^d} \right\|. \end{aligned}$$

In order to estimate $\alpha(m, M)$, note that

$$\Lambda_M^r = \bigcup_{t \in T_{m,M}} (\Lambda_m^r + t) \dot{\cup} \left(\Lambda_M^r \cap \bigcup_{t \in T_{m,M}} ((\Lambda_m \cap \partial^r(\Lambda_m)) + t) \right) \dot{\cup} \Lambda_M^r \setminus (\Lambda_{\lfloor M/m \rfloor m}).$$

This and (iii) of Definition 2.3 yield

$$\begin{aligned} \alpha(m, M) &\leq \sum_{t \in T_{m,M}} \left(b(\Lambda_m^r) + b(\Lambda_M^r \cap ((\Lambda_m \cap \partial^r(\Lambda_m)) + t)) \right) \\ &\quad + \left\| F(\Lambda_M^r \cap ((\Lambda_m \cap \partial^r(\Lambda_m)) + t)) \right\| \\ &\quad + b(\Lambda_M^r \setminus \Lambda_{\lfloor M/m \rfloor m}) + \left\| F(\Lambda_M^r \setminus \Lambda_{\lfloor M/m \rfloor m}) \right\| \\ &\leq |T_{m,M}| b(\Lambda_m^r) + (K_f + D) |T_{m,M}| |\partial^r(\Lambda_m)| + (K_f + D) |\Lambda_M^r \setminus \Lambda_{\lfloor M/m \rfloor m}|. \end{aligned}$$

Here, we also used translation invariance of F and property (v) of [Definition 2.3](#). Dividing this term by M^d and using $|T_{m,M}|m^d \leq M^d$ as well as [\(4.2\)](#) leads to

$$\frac{\alpha(m, M)}{M^d} \leq \frac{b(\Lambda_m^r)}{m^d} + (K_f + D) \frac{|\partial^r(\Lambda_m)|}{m^d} + (K_f + D) \frac{|\partial^m(\Lambda_M)|}{M^d}.$$

To estimate $\beta(m, M)$ we apply again translation invariance of F and obtain

$$\beta(m, M) = \left\| \frac{F(\Lambda_m^r)}{m^d} - \frac{|T_{m,M}|F(\Lambda_m^r)}{M^d} \right\| = \left(\frac{1}{m^d} - \frac{\lfloor M/m \rfloor^d}{M^d} \right) \|F(\Lambda_m^r)\|.$$

Using the properties of the boundary term b , the above bounds on $\alpha(m, M)$ and $\beta(m, M)$ yield

$$\lim_{m \rightarrow \infty} \lim_{M \rightarrow \infty} \delta(m, M) = 0. \quad (6.4)$$

This is equivalent to $(F(\Lambda_m^r)/m^d)_{m \in \mathbb{N}}$ being a Cauchy sequence. To see this in detail, choose $\varepsilon > 0$ arbitrarily. Then, by [\(6.4\)](#), there exists $m_0 \in \mathbb{N}$ such that $\lim_{M \rightarrow \infty} \delta(m_0, M) \leq \varepsilon/4$. Therefore, we find $M_0 \in \mathbb{N}$ satisfying $\delta(m_0, M) \leq \varepsilon/2$ for all $M \geq M_0$. Now, let $j, k \geq M_0$ be arbitrary. Then we obtain using the triangle inequality

$$\begin{aligned} \left\| \frac{F(\Lambda_j^r)}{j^d} - \frac{F(\Lambda_k^r)}{k^d} \right\| &\leq \left\| \frac{F(\Lambda_j^r)}{j^d} - \frac{F(\Lambda_{m_0}^r)}{m_0^d} \right\| + \left\| \frac{F(\Lambda_{m_0}^r)}{m_0^d} - \frac{F(\Lambda_k^r)}{k^d} \right\| \\ &= \delta(m_0, j) + \delta(m_0, k) \leq \varepsilon. \end{aligned}$$

This shows that $(F(\Lambda_m^r)/|\Lambda_m|)_{m \in \mathbb{N}}$ is a Cauchy sequence and hence convergent in the Banach space \mathbb{B} .

Now, that we know that the limit f^* exists, we can study the speed of convergence.

$$\begin{aligned} \left\| \frac{\langle f_m^r, \mathbb{P}_m^r \rangle}{m^d} - f^* \right\| &= \lim_{M \rightarrow \infty} \left\| \frac{\langle f_m^r, \mathbb{P}_m^r \rangle}{m^d} - \frac{\langle f_M^r, \mathbb{P}_M^r \rangle}{M^d} \right\| = \lim_{M \rightarrow \infty} \delta(m, M) \\ &\leq \lim_{M \rightarrow \infty} \left(\frac{\alpha(m, M)}{M^d} + \beta(m, M) \right) \leq \frac{b(\Lambda_m^r)}{m^d} + (K_f + D) \frac{|\partial^r(\Lambda_m)|}{m^d}. \quad \square \end{aligned}$$

Now we are in the position to prove the main theorem of this paper.

Proof of Theorems 2.6 and 2.8. The proof is basically a combination of [Lemmas 4.1](#) and [6.1](#) and [Theorem 5.6](#). We choose $\tilde{\Omega}$ as in [Theorem 5.6](#), r as the constant from (M3) and $f^* \in \mathbb{B}$ according to [Lemma 6.1](#). Then we have for arbitrary $m \in \mathbb{N}$ and $\omega \in \tilde{\Omega}$:

$$\begin{aligned} \left\| \frac{f(\Lambda_n, \omega)}{n^d} - f^* \right\| &\leq \left\| \frac{f(\Lambda_n, \omega)}{n^d} - \frac{\langle f_m^r, L_{m,n}^{r,\omega} \rangle}{m^d} \right\| + \left\| \frac{\langle f_m^r, L_{m,n}^{r,\omega} \rangle}{m^d} - \frac{\langle f_m^r, \mathbb{P}_m^r \rangle}{m^d} \right\| \\ &\quad + \left\| \frac{\langle f_m^r, \mathbb{P}_m^r \rangle}{m^d} - f^* \right\|. \end{aligned}$$

Each of the above mentioned results controls one of the error terms on the right hand side, which leads to

$$\left\| \frac{f(\Lambda_n, \omega)}{|\Lambda_n|} - f^* \right\| \leq G(m, n) + m^{-d} \left\| \langle f_m^r, L_{m,n}^{r,\omega} \rangle - \langle f_m^r, \mathbb{P}_m^r \rangle \right\|$$

with

$$G(m, n) := \frac{b(A_{\lfloor n/m \rfloor m})}{|A_{\lfloor n/m \rfloor m}|} + \frac{(2K + D)|\partial^m(A_n^m)|}{|A_n^m|} + \frac{2b(A_m^r) + b(A_m) + 2(K + D)|\partial^r(A_m)|}{|A_m|}.$$

Taking first the limit $n \rightarrow \infty$ and afterwards the limit $m \rightarrow \infty$ on both sides proves [Theorem 2.6](#).

To establish [Theorem 2.8](#), we use the additional hypotheses on the boundary term and estimate $G(m, n)$. First we note for $n \geq 2r$

$$|\partial^r(A_n)| = (n + 2r)^d - (n - 2r)^d = \sum_{j=0}^d \binom{d}{j} (1 - (-1)^j) (2r)^j n^{d-j} \leq 2^{2d+1} r^d n^{d-1}.$$

Therefore,

$$\frac{b(A_n)}{|A_n|} \leq \frac{|\partial^{r'}(A_n)| D'}{|A_n|} \leq \frac{2^{2d+1} r'^d D'}{n}$$

holds for all $n \geq 2r'$. With $A_n^m = A_{n-2m} + (m, m, \dots, m)$, it is now straightforward to verify

$$G(m, n) \leq 2^{2d+1} \left(\frac{(2K + D)m^d + D' r'^d}{n - 2m} + \frac{2(K + D)r^d + 3D' r'^d}{m - 2r} \right).$$

The two claims about $\sup_{f \in \mathcal{U}_{K,D,D',r'}} \|\langle f_m^r, L_{m,n}^{r,\omega} \rangle - \langle f_m^r, \mathbb{P}_m^r \rangle\|$ follow from [Theorem 5.6](#). \square

7. Eigenvalue counting functions for the Anderson model

In the following, we introduce the Anderson model on \mathbb{Z}^d or, more precisely, on the graph with nodes \mathbb{Z}^d and nearest neighbor bonds. For the corresponding Schrödinger operators we show that the associated eigenvalue counting functions almost surely converge uniformly.

The Laplace operator $\Delta: \ell^2(\mathbb{Z}^d) \rightarrow \ell^2(\mathbb{Z}^d)$ is given by

$$(\Delta\varphi)(z) = \sum_{x: \mathfrak{D}(x,z)=1} (\varphi(x) - \varphi(z)) \quad (z \in \mathbb{Z}^d).$$

In order to define a random potential, we introduce the corresponding probability space. We fix the canonical space $\Omega := \mathcal{A}^{\mathbb{Z}^d}$, where $\mathcal{A} \subseteq \mathbb{R}$ is an arbitrary subset of \mathbb{R} . As before we equip Ω with $\mathcal{B}(\Omega)$, the σ -algebra on Ω generated by the cylinder sets. Moreover, we chose a probability measure $\mathbb{P}: \mathcal{B}(\Omega) \rightarrow [0, 1]$ satisfying (M1), (M2) and (M3). In particular, a product measure $\mathbb{P} = \prod_{z \in \mathbb{Z}} \mu$ is allowed, where $\mu: \mathcal{B}(\mathcal{A}) \rightarrow [0, 1]$ is a measure on $(\mathcal{A}, \mathcal{B}(\mathcal{A}))$. An alternative way to specify such a product measure is to say that the projections $\Omega \ni (\omega_x)_{x \in \mathbb{Z}} \rightarrow \omega_z, z \in \mathbb{Z}$, are \mathcal{A} -valued i.i.d. random variables.

The random potential $V = (V_\omega)_{\omega \in \Omega}$ is now defined by setting for each $\omega = (\omega_z)_{z \in \mathbb{Z}^d} \in \Omega$:

$$V_\omega: \ell^2(\mathbb{Z}^d) \rightarrow \ell^2(\mathbb{Z}^d), \quad (V_\omega\varphi)(z) = \omega_z\varphi(z) \quad (\varphi \in \ell^2(\mathbb{Z}^d), z \in \mathbb{Z}^d). \quad (7.1)$$

Together, the Laplace operator and the random potential form the random Schrödinger operator $H = (H_\omega)_{\omega \in \Omega}$:

$$H_\omega: \ell^2(\mathbb{Z}^d) \rightarrow \ell^2(\mathbb{Z}^d), \quad H_\omega := -\Delta + V_\omega. \quad (7.2)$$

This operator is almost surely self-adjoint and ergodic by (M1) and (M3). Thus, the spectrum $\sigma(H_\omega)$ of H_ω is a non-random subset of \mathbb{R} , cf. [16]. In the following we are interested in the

distribution of $\sigma(H_\omega)$ on \mathbb{R} . The function which describes this distribution is called *integrated density of states*.

Let us define finite dimensional restrictions of H . To this end, consider for a given set $\Lambda \in \mathcal{F}$ the projection

$$p_\Lambda: \ell^2(\mathbb{Z}^d) \rightarrow \ell^2(\Lambda), \quad (p_\Lambda \varphi)(z) := \varphi(z), \quad (\varphi \in \ell^2(\mathbb{Z}^d), z \in \Lambda) \quad (7.3)$$

and the inclusion

$$i_\Lambda: \ell^2(\Lambda) \rightarrow \ell^2(\mathbb{Z}^d), \quad (i_\Lambda \varphi)(z) := \begin{cases} \varphi(z) & \text{if } z \in \Lambda, \\ 0 & \text{otherwise,} \end{cases} \quad (\varphi \in \ell^2(\Lambda), z \in \mathbb{Z}^d). \quad (7.4)$$

Now, for any $\omega \in \Omega$ and $\Lambda \in \mathcal{F}$ we set

$$H_\omega^\Lambda: \ell^2(\Lambda) \rightarrow \ell^2(\Lambda), \quad H_\omega^\Lambda := p_\Lambda H_\omega i_\Lambda.$$

The corresponding eigenvalue counting function is given by

$$f(\Lambda, \omega) := \left(\mathbb{R} \ni x \mapsto \text{Tr}(\chi_{(-\infty, x]}(H_\omega^\Lambda)) \right). \quad (7.5)$$

Here, $\chi_{(-\infty, x]}(H_\omega^\Lambda)$ denotes the spectral projection of H_ω^Λ on the interval $(-\infty, x]$. Thus, $f(\Lambda, \omega)(x)$ equals the number of eigenvalues (counted with multiplicities) of H_ω^Λ which do not exceed x .

Lemma 7.1. *The eigenvalue counting function $f: \mathcal{F} \times \Omega \rightarrow \mathbb{B}$ given by (7.5) is admissible in the sense of Definition 2.3. It admits a proper boundary term, and possible constants for f are $D = D' = 8$, $K = 9$ and $r' = 1$.*

Proof. We verify (i)–(v) of Definition 2.3.

- (i) Let $\Lambda \in \mathcal{F}$ and $z \in \mathbb{Z}^d$ be given. Using the definitions of the potential V_ω in (7.1), the translation τ_z in (2.2), the projection p_Λ in (7.3) and the inclusion i_Λ in (7.4) we obtain

$$p_\Lambda V_{\tau_z \omega} i_\Lambda = p_{\Lambda+z} V_\omega i_{\Lambda+z}.$$

This generalizes to $H_{\tau_z \omega}^\Lambda = H_\omega^{\Lambda+z}$ and hence implies for each $x \in \mathbb{R}$

$$f(\Lambda, \tau_z \omega)(x) = \text{Tr}(\chi_{(-\infty, x]}(H_{\tau_z \omega}^\Lambda)) = \text{Tr}(\chi_{(-\infty, x]}(H_\omega^{\Lambda+z})) = f(\Lambda+z, \omega)(x).$$

- (ii) Let $\Lambda \in \mathcal{F}$ be given. Obviously, for all $\omega, \omega' \in \Omega$ with $\Pi_\Lambda(\omega) = \Pi_\Lambda(\omega')$ we have $H_\omega^\Lambda = H_{\omega'}^\Lambda$. Thus, we obtain $f(\Lambda, \omega) = f(\Lambda, \omega')$.
- (iii) In order to show almost additivity, we make use of the following estimate, which holds for $\Lambda' \subseteq \Lambda \in \mathcal{F}$ and arbitrary $\omega \in \Omega$:

$$\|f(\Lambda, \omega) - f(\Lambda', \omega)\| \leq 4|\Lambda \setminus \Lambda'|. \quad (7.6)$$

This bound can be verified using the min–max–principle, cf. appendix of [11]. Now let $n \in \mathbb{N}$, disjoint sets $\Lambda_i \in \mathcal{F}$, $i = 1, \dots, n$ and $\Lambda := \bigcup_{i=1}^n \Lambda_i \in \mathcal{F}$ be given. With triangle

inequality and (7.6) we obtain for each $\omega \in \Omega$:

$$\begin{aligned} & \left\| f(\Lambda, \omega) - \sum_{i=1}^n f(\Lambda_i, \omega) \right\| \\ & \leq \left\| f(\Lambda, \omega) - f\left(\bigcup_{i=1}^n \Lambda_i^1, \omega\right) \right\| + \left\| f\left(\bigcup_{i=1}^n \Lambda_i^1, \omega\right) - \sum_{i=1}^n f(\Lambda_i, \omega) \right\| \\ & \leq 4 \sum_{i=1}^n |\partial^1(\Lambda_i)| + \left\| f\left(\bigcup_{i=1}^n \Lambda_i^1, \omega\right) - \sum_{i=1}^n f(\Lambda_i^1, \omega) \right\| \\ & \quad + \sum_{i=1}^n \left\| f(\Lambda_i^1, \omega) - f(\Lambda_i, \omega) \right\| \\ & \leq 8 \sum_{i=1}^n |\partial^1(\Lambda_i)| + \left\| f\left(\bigcup_{i=1}^n \Lambda_i^1, \omega\right) - \sum_{i=1}^n f(\Lambda_i^1, \omega) \right\|. \end{aligned}$$

In order to deal with the last difference, we use that the operator in consideration has hopping range 1, which gives for $\tilde{\Lambda} := \bigcup_{i=1}^n \Lambda_i^1$:

$$H_{\omega}^{\tilde{\Lambda}} = \bigoplus_{i=1}^n H_{\omega}^{\Lambda_i^1}.$$

Thus, the eigenvalues of $H_{\omega}^{\tilde{\Lambda}}$ are exactly the union of the eigenvalues of the operators $H_{\omega}^{\Lambda_i^1}$, $i = 1, \dots, n$. This implies

$$f(\tilde{\Lambda}, \omega) = \sum_{i=1}^n f(\Lambda_i^1, \omega)$$

and hence

$$\left\| f(\Lambda, \omega) - \sum_{i=1}^n f(\Lambda_i, \omega) \right\| \leq 8 \sum_{i=1}^n |\partial^1(\Lambda_i)|.$$

We set $b: \mathcal{F} \rightarrow [0, \infty)$ and $b(\Lambda) := 8|\partial^1(\Lambda)|$. Let $\Lambda \in \mathcal{F}$ and $z \in \mathbb{Z}^d$. Then obviously $b(\Lambda + z) = b(\Lambda)$ and $b(\Lambda) \leq 8|\Lambda|$, and for any sequence of cubes (Λ_n) with increasing side length, we have $b(\Lambda_n)/|\Lambda_n| \rightarrow 0$ as $n \rightarrow \infty$.

- (iv) For $\Lambda \in \mathcal{F}$ and $\omega \in \Omega$ we denote the $|\Lambda|$ eigenvalues of H_{ω}^{Λ} (counted with multiplicities) by $E_1(H_{\omega}^{\Lambda}) \leq \dots \leq E_{|\Lambda|}(H_{\omega}^{\Lambda})$. Choose $n \in \{1, \dots, |\Lambda|\}$ and $\omega \leq \omega'$, i.e. for each $z \in \mathbb{Z}^d$ we have $\omega_z \leq \omega'_z$. By the min-max-principle we get for the n th eigenvalue:

$$\begin{aligned} E_n(H_{\omega}^{\Lambda}) &= \min_{\substack{U \subseteq \mathbb{R}^{\Lambda}, \\ \dim(U)=n}} \max_{\substack{\varphi \in U, \\ \|\varphi\|=1}} \langle H_{\omega}^{\Lambda} \varphi, \varphi \rangle \\ &= \min_{\substack{U \subseteq \mathbb{R}^{\Lambda}, \\ \dim(U)=n}} \max_{\substack{\varphi \in U, \\ \|\varphi\|=1}} (\langle H_{\omega'}^{\Lambda} \varphi, \varphi \rangle - \langle (V_{\omega'} - V_{\omega}) \varphi, \varphi \rangle) \leq E_n(H_{\omega'}^{\Lambda}). \end{aligned}$$

Therefore, we have for each $x \in \mathbb{R}$ the inequality $f(\Lambda, \omega)(x) \geq f(\Lambda, \omega')(x)$.

- (v) Let arbitrary $\omega \in \Omega$ be given. Since the operator $H_\omega^{(0)}$ has exactly one eigenvalue, we have $\|f(\{0\}, \omega)\| = 1$. \square

Let us state the main result of this section.

Theorem 7.2. Let $\Lambda_n := [0, n)^d \cap \mathbb{Z}^d$. Moreover, let $\mathcal{A} \subseteq \mathbb{R}$, $\Omega := \mathcal{A}^{\mathbb{Z}^d}$ and $(\Omega, \mathcal{B}(\Omega), \mathbb{P})$ be a probability space satisfying (M1)–(M3). Consider the random Schrödinger operator H defined in (7.2) and the associated f given in (7.5). Then there exists a set $\tilde{\Omega} \in \mathcal{B}(\Omega)$ of full measure, such that for all $\omega \in \tilde{\Omega}$:

$$\lim_{n \rightarrow \infty} \left\| \frac{f(\Lambda_n, \omega)}{|\Lambda_n|} - f^* \right\| = 0, \quad (7.7)$$

where $f^* \in \mathbb{B}$ is given by

$$f^*(x) := \mathbb{E}(\langle \delta_0, \chi_{(-\infty, x]}(H_\omega) \delta_0 \rangle). \quad (7.8)$$

Here, $\delta_0 \in \ell^2(\mathbb{Z}^d)$ is given by $\delta_0(0) = 1$ and $\delta_0(x) = 0$ for $x \neq 0$. Moreover, $\chi_{(-\infty, x]}(H_\omega)$ is the spectral projection of H_ω on the interval $(-\infty, x]$. The convergence is quantified by

$$\begin{aligned} \left\| \frac{f(\Lambda_n, \omega)}{|\Lambda_n|} - f^* \right\| &\leq 2^{d+1} \left(\frac{26m^d + 8}{n - m} + \frac{34r^d + 24}{m - r} \right) \\ &+ \sup_{f \in \mathcal{U}_{K,D,D',r'}} \frac{\| \langle f \Lambda_m^r, L_{m,n}^{r,\omega} - \mathbb{P} \Lambda_m^r \rangle \|}{|\Lambda_m|} \end{aligned} \quad (7.9)$$

for $n, m \in \mathbb{N}$, $m < n$.

Proof. By Lemma 7.1 we know that the eigenvalue counting function $f: \mathcal{F} \times \Omega \rightarrow \mathbb{B}$ is admissible. Thus we can apply Theorem 2.6 and obtain that there exists a function $\tilde{f} \in \mathbb{B}$ and a set $\Omega_1 \in \mathcal{B}(\Omega)$ of full measure such that for each $\omega \in \Omega_1$ we have

$$\lim_{n \rightarrow \infty} \left\| \frac{f(\Lambda_n, \omega)}{|\Lambda_n|} - \tilde{f} \right\| = 0. \quad (7.10)$$

Thus, it remains to show that \tilde{f} equals the function

$$f^*: \mathbb{R} \rightarrow [0, 1], \quad f^*(x) := \mathbb{E}(\langle \delta_0, \chi_{(-\infty, x]}^\omega \delta_0 \rangle).$$

Here we use ergodicity of H and infer from [16, Theorem 4.8] that there is a set $\Omega_2 \in \mathcal{B}(\Omega)$ of full measure such that for each $\omega \in \Omega_2$

$$\lim_{n \rightarrow \infty} \frac{f(\Lambda_n, \omega)(x)}{|\Lambda_n|} = f^*(x) \quad (7.11)$$

for all $x \in \mathbb{R}$ which are continuity points of f^* . By definition, this is weak convergence of distribution functions. Thus, as for all $\omega \in \Omega_1 \cap \Omega_2$ we have that $f(\Lambda_n, \omega)/|\Lambda_n|$ converges weakly to f^* and uniformly to \tilde{f} , which implies $\tilde{f} = f^*$. \square

Remark 7.3. • The limit f^* of the normalized eigenvalue counting functions is called the *integrated density of states* or *spectral distribution function* of the operator H . The fact that f^* can be expressed as the function given in (7.8) is often referred to as the *Pastur–Shubin trace*

formula, named after the pioneering works [15,20]. For more recent results in the specific context we are treating here, c.f. [27,19,13] and the references therein.

- Let us also emphasize that the f^* is a deterministic function. On the one hand this is interesting as this implies that the normalized eigenvalue counting function converges for almost all realizations to same limit function. On the other hand this is not surprising as we mentioned that H is ergodic, and in this setting it is well-known that the spectrum (as a set) is deterministic, see for instance [16].
- The result is easily generalized to sequences of cubes $(\Lambda_n)_n$ of diverging side length with $\Lambda_n \subsetneq \Lambda_{n+1}$. The validity of the Pastur–Shubin formula shows that the limit f^* is independent of the specific choice sequence of cubes $(\Lambda_n)_n$.
- The statement of Theorem 7.2 has been obtained before in a different setting. In [10,13] ergodic random operators have been considered. The assumption of ergodicity concerns the measure \mathbb{P} (in our notation) and is weaker than the assumptions (M1)–(M3) which we use here. With this regard the result of [13] is more general than the one obtained here. However, under the mere assumption of ergodicity it is not possible to obtain explicit error estimates as in (7.9). The paper [10] obtains an error estimate, similar to, but weaker than (7.9). There the setting is also different from ours here: \mathcal{A} needs to be countable and instead of a probability measure properties of frequencies are used.
- Similar, but weaker results have been proven for Anderson-percolation Hamiltonians in [25, 26,13]. These models are particularly interesting since their integrated density of states exhibits typically an infinite set of discontinuities, which lie dense in the spectrum. The random variables entering the Hamiltonian may take uncountably many different values.

8. Cluster counting functions in percolation theory

We introduce briefly percolation on \mathbb{Z}^d . Percolation comes in two flavors, site and bond percolation. We focus on site percolation here. Part of the results have already been obtained in [17]. However, we go far beyond since we not only obtain convergence of densities, but are even able to identify the limit objects.

As before, we let $\Omega := \mathbb{R}^{\mathbb{Z}^d}$. We fix the alphabet $\mathcal{A} := \{0, 1\}$ and a probability measure $\mathbb{P}: \mathcal{B}(\Omega) \rightarrow [0, 1]$ which is supported on $\mathcal{A}^{\mathbb{Z}^d} \in \mathcal{B}(\Omega)$, i.e. $\mathbb{P}(\mathcal{A}^{\mathbb{Z}^d}) = 1$. A configuration $\omega \in \mathcal{A}^{\mathbb{Z}^d} \subseteq \Omega$ determines a *percolation graph* $\Gamma_\omega = (\mathbb{Z}^d, \mathcal{E}_\omega)$ as follows. The set of vertices of Γ_ω is \mathbb{Z}^d , and an edge connects two vertices if and only if they have distance 1 and are both “switched on” in the configuration $\omega = (\omega_z)_{z \in \mathbb{Z}^d}$:

$$\mathcal{E}_\omega := \{ \{x, y\} \subseteq \mathbb{Z}^d \mid \mathfrak{d}(x, y) = 1, \omega_x = \omega_y = 1 \}.$$

By this, the percolation graph Γ_ω is well-defined for \mathbb{P} -almost all $\omega \in \Omega$, and Γ_ω is a random graph. For our purposes, we want \mathbb{P} to satisfy (M1), (M2) and (M3). This setting includes but is not limited to the product measure $\mathbb{P} = \prod_{z \in \mathbb{Z}^d} \mu$, where $\mu: \mathcal{B}(\mathbb{R}) \rightarrow [0, 1]$ is any probability measure supported on \mathcal{A} .

We need some standard terminology of graph theory. Let $\Gamma = (V, \mathcal{E})$ be a graph. For each subset $A \subseteq V$ of the set of nodes, Γ induces a graph $\Gamma^A := (A, \mathcal{E}^A)$ by

$$\mathcal{E}^A := \{e \in \mathcal{E} \mid e \subseteq A\}.$$

A *walk* of length $n \in \mathbb{N} \cup \{0, \infty\}$ in the graph Γ is a sequence of nodes $(z_j)_{j=0}^n \in (\mathbb{Z}^d)^{n+1}$ such that $\{z_j, z_{j+1}\}$ is an edge of Γ , i.e. $\{z_j, z_{j+1}\} \in \mathcal{E}$, for all $j \in \mathbb{N} \cup \{0\}$, $j < n$. Note that a finite walk of length n contains n edges but $n + 1$ nodes.

If the walk $(z_j)_{j=0}^n$ has finite length $n < \infty$, we say that it *connects* the points z_0 and z_n . Being connected by a walk is an equivalence relation on the nodes. We denote the fact that two points $x, y \in \mathbb{Z}^d$ are connected in the graph Γ as $x \overset{\Gamma}{\rightsquigarrow} y$.

The equivalence classes of $\overset{\Gamma}{\rightsquigarrow}$ are called *clusters*. Let $\Lambda \subseteq \mathbb{Z}^d$ and $x \in \Lambda$. The cluster of x in the percolation graph Γ_ω^Λ restricted to Λ consists of all nodes which are connected to x by a walk in Γ_ω^Λ :

$$C_x^\Lambda(\omega) := \{y \in \mathbb{Z}^d \mid x \overset{\Gamma_\omega^\Lambda}{\rightsquigarrow} y\},$$

again for $\omega \in \mathcal{A}^{\mathbb{Z}^d}$, $\Lambda \subseteq \mathbb{Z}^d$ and $x, y \in \Lambda$.

8.1. Convergence of cluster counting functions

We now define a cumulative counting function for clusters. As before, let \mathcal{F} be the set of finite subsets of \mathbb{Z}^d and \mathbb{B} the set of bounded functions from \mathbb{R} to \mathbb{R} which are continuous from the right. The function $f: \mathcal{F} \times \Omega \rightarrow \mathbb{B}$ counts the number of clusters in Γ_ω^Λ which are smaller then the given threshold:

$$f(\Lambda, \omega)(\lambda) := |\{C_z^\Lambda(\omega) \mid z \in \Lambda, |C_z^\Lambda(\omega)| \leq \lambda\}|. \quad (8.1)$$

Note that f counts clusters and not vertices in clusters.

Lemma 8.1. *The cluster counting function $f: \mathcal{F} \times \Omega \rightarrow \mathbb{B}$ given by (8.1) is admissible in the sense of Definition 2.3 and permits a proper boundary term. Possible constants are $D = D' = 2$, $r' = 1$ and $K = 3$.*

Proof. We verify (i)–(v) of Definition 2.3.

- (i) Let $\Lambda \in \mathcal{F}$ and $x, z \in \mathbb{Z}^d$ be given. The percolation graph Γ_ω is determined by the configuration $\omega \in \Omega$, for almost all $\omega \in \Omega$. The shifted configuration gives shifted clusters, i.e.

$$|C_x^\Lambda(\tau_z \omega)| = |C_{x+z}^{\Lambda+z}(\omega)|$$

for all $x, z \in \mathbb{Z}^d$. Accordingly, for all $\lambda \in \mathbb{R}$,

$$\begin{aligned} f(\Lambda, \tau_z \omega)(\lambda) &= |\{C_x^\Lambda(\tau_z \omega) \mid x \in \Lambda, |C_x^\Lambda(\tau_z \omega)| \leq \lambda\}| \\ &= |\{C_{x+z}^{\Lambda+z}(\omega) \mid x \in \Lambda, |C_{x+z}^{\Lambda+z}(\omega)| \leq \lambda\}| = f(\Lambda + z, \omega)(\lambda). \end{aligned}$$

- (ii) Fix $\Lambda \in \mathcal{F}$ and $\omega, \omega' \in \Omega$ with $\omega_\Lambda = \omega'_\Lambda$, where $\omega_\Lambda := (\omega_x)_{x \in \Lambda}$ as before. The edges of Γ_ω^Λ are determined by ω_Λ . Hence, $\Gamma_\omega^\Lambda = \Gamma_{\omega'}^\Lambda$, thus $C_z^\Lambda(\omega) = C_z^\Lambda(\omega')$ for all $z \in \Lambda$, and we obtain $f(\Lambda, \omega) = f(\Lambda, \omega')$.
- (iii) In order to show almost additivity, fix $\omega \in \Omega$, $n \in \mathbb{N}$ and disjoint sets $\Lambda_j \in \mathcal{F}$, $j \in \{1, \dots, n\}$. We name the union $\Lambda := \bigcup_{j=1}^n \Lambda_j$. For $x \in \mathbb{R}$,

$$\sum_{j=1}^n f(\Lambda_j, \omega)(x)$$

is the total number of clusters of size not larger than x in the graphs $\Gamma_\omega^{A_j}$. Whenever $\partial(A_j, A_k) = 1$, the graph Γ_ω^A could contain edges connecting a point in A_j with a point in A_k , depending on ω . Each of these edges join two possibly different clusters, so for each edge, there are two less small clusters and one more large one. By this mechanism, the number of clusters below the threshold x changes at most by twice the number of added edges. We note

$$\left| \mathcal{E}_\omega^A \setminus \bigcup_{j=1}^n \mathcal{E}_\omega^{A_j} \right| \leq \sum_{j=1}^n |\partial^1 A_j|$$

and conclude

$$\left| f(A, \omega)(x) - \sum_{j=1}^n f(A_j, \omega)(x) \right| \leq 2 \sum_{j=1}^n |\partial^1 A_j|$$

for all $x \in \mathbb{R}$. The choice $b(A) := 2|\partial^1 A|$ for $A \in \mathcal{F}$ gives a proper boundary term for f , cf. Lemma 7.1.

- (iv) Let $A \in \mathcal{F}$ and $\omega, \omega' \in \Omega$, $\omega \leq \omega'$. Then each edge of Γ_ω is also an edge in $\Gamma_{\omega'}$: $\mathcal{E}_\omega \subseteq \mathcal{E}_{\omega'}$. As reasoned in (iii), a new edge never increases the number of clusters below a threshold $x \in \mathbb{R}$, so

$$f(A, \omega)(x) \geq f(A, \omega')(x).$$

- (v) For all $\omega \in \Omega$, $f(\{0\}, \omega)(x) = 0$ for $x < 1$ and $f(\{0\}, \omega)(x) = 1$ for $x \geq 1$. \square

Theorem 2.6 and Lemma 8.1 immediately give the following.

Corollary 8.2. Let $A_n := [0, n)^d \cap \mathbb{Z}^d$ for $n \in \mathbb{N}$ and $f: \mathcal{F} \times \Omega \rightarrow \mathbb{B}$ be the cumulative cluster counting function given in (8.1). There exists a set $\tilde{\Omega} \subseteq \Omega$ of full measure and a function $f^* \in \mathbb{B}$ such that, for each $\omega \in \tilde{\Omega}$,

$$\lim_{n \rightarrow \infty} \left\| \frac{f(A_n, \omega)}{|A_n|} - f^* \right\| = 0.$$

For all $m, n \in \mathbb{N}$, $m < n$, we have

$$\left\| \frac{f(A_n, \omega)}{|A_n|} - f^* \right\| \leq 2^{d+1} \left(\frac{8m^d + 2}{n - m} + \frac{10r^d + 6}{m - r} \right) + \sup_{f \in \mathcal{U}_{K,D,D',r'}} \frac{\| \langle f_{A_m^r}, L_{m,n}^{r,\omega} - \mathbb{P}_{A_m^r} \rangle \|}{|A_m|}.$$

8.2. Identification of the limit

In the previous section we studied the convergence of the counting function in (8.1) normalized with $|A_n|$. Next, we give a brief overview on closely related convergence results. We sketch the proofs only briefly since these results are not in the main focus of this paper. The heart of the section is that we do not just give statements about convergence, but even present closed expressions of the limits.

We start with defining

$$K_\omega(A) := |\{C_x^A(\omega) \mid x \in A\}|, \quad (8.2)$$

which counts the number of all clusters in Γ_ω^A . Using this quantity we set:

$$\begin{aligned} a_n^{(m)}(\omega) &:= |\Lambda_n|^{-1} |\{C_x^{\Lambda_n}(\omega) \mid x \in \Lambda_n, |C_x^{\Lambda_n}(\omega)| = m\}|, \\ b_n^{(m)}(\omega) &:= K_\omega(\Lambda_n)^{-1} |\{C_x^{\Lambda_n}(\omega) \mid x \in \Lambda_n, |C_x^{\Lambda_n}(\omega)| = m\}|, \quad \text{and} \\ c_n^{(m)}(\omega) &:= |\Lambda_n|^{-1} |\{x \in \Lambda_n \mid |C_x^{\Lambda_n}(\omega)| = m\}|, \end{aligned}$$

where again $\Lambda_n := [0, n)^d \cap \mathbb{Z}^d$ for $n \in \mathbb{N}$.

Lemma 8.3. *In the above setting, we have almost surely*

$$\begin{aligned} \lim_{n \rightarrow \infty} a_n^{(m)}(\omega) &= \frac{1}{m} \mathbb{P}(|C_0| = m), \\ \lim_{n \rightarrow \infty} b_n^{(m)}(\omega) &= \frac{1}{\kappa m} \mathbb{P}(|C_0| = m), \quad \text{and} \\ \lim_{n \rightarrow \infty} c_n^{(m)}(\omega) &= \mathbb{P}(|C_0| = m), \end{aligned}$$

where $\kappa := \mathbb{E}(|C_0|^{-1})$.

Note that the existence of the limit in the case corresponding to $a_n^{(m)}$ was treated in Section 8.1. The existence of the limits in Lemma 8.3 has already been proved in [17] in the setting of bond percolation. However, the authors did not give explicit expressions for the limit objects. For the proof of Lemma 8.3 one may use Theorem 2.6 in combination with the d -dimensional version of Birkhoff's ergodic theorem, see [8], and the fact [7] that for almost all ω :

$$\lim_{n \rightarrow \infty} \frac{K_\omega(\Lambda_n)}{|\Lambda_n|} = \kappa.$$

The above convergence results can again be extended to the associated distribution functions. To formulate the corresponding result, we introduce for $n \in \mathbb{N}$ and $\omega \in \Omega$ the maps $\Theta_\omega^n, \Phi_\omega^n, \Psi_\omega^n : \mathbb{R} \rightarrow \mathbb{R}$ by setting for each $m \in \mathbb{N}$

$$\begin{aligned} \Theta_\omega^n(m) &:= \sum_{j=1}^{\lfloor m \rfloor} a_n^{(j)}(\omega) = \frac{|\{C_x^{\Lambda_n}(\omega) \mid x \in \Lambda_n, |C_x^{\Lambda_n}(\omega)| \leq m\}|}{|\Lambda_n|} \\ \Phi_\omega^n(m) &:= \sum_{j=1}^{\lfloor m \rfloor} b_n^{(j)}(\omega) = \frac{|\{C_x^{\Lambda_n}(\omega) \mid x \in \Lambda_n, |C_x^{\Lambda_n}(\omega)| \leq m\}|}{K_\omega(\Lambda_n)}, \quad \text{and} \\ \Psi_\omega^n(m) &:= \sum_{j=1}^{\lfloor m \rfloor} c_n^{(j)}(\omega) = \frac{|\{x \in \Lambda_n \mid |C_x^{\Lambda_n}(\omega)| \leq m\}|}{|\Lambda_n|}. \end{aligned}$$

Moreover, we define the deterministic functions $\Theta, \Phi, \Psi : \mathbb{R} \rightarrow \mathbb{R}$ by

$$\Theta(m) := \sum_{j=1}^{\lfloor m \rfloor} \frac{1}{j} \mathbb{P}(|C_0| = j), \quad \Phi(m) := \frac{1}{\kappa} \Theta(m), \quad \text{and} \quad \Psi(m) := \mathbb{P}(|C_0| \leq m) \quad (8.3)$$

for $m \in \mathbb{N}$.

Theorem 8.4. *In the above setting, we can find a set $\tilde{\Omega} \subseteq \Omega$ of full measure such that for all $\omega \in \tilde{\Omega}$ we have*

$$\lim_{n \rightarrow \infty} \|\Theta_\omega^n - \Theta\| = 0, \quad \lim_{n \rightarrow \infty} \|\Phi_\omega^n - \Phi\| = 0 \quad \text{and} \quad \lim_{n \rightarrow \infty} \|\Psi_\omega^n - \Psi\| = 0.$$

Here $\|\cdot\|$ denotes the supremum norm in $\mathcal{B}(\mathbb{R})$.

Let us give a brief sketch of the proof. The convergence of Θ_ω^n and Φ_ω^n follows rather direct from Theorem 2.6 and Lemma 8.3. However, in order to obtain the convergence of Ψ_ω^n one has to apply a different scheme, which was used in the context of the eigenvalue counting function in [13, Section 6]. The strategy consist of the following steps: One first verifies weak convergence of the distribution functions and second, shows that $\nu_\omega^n(\{\lambda\}) \rightarrow \nu(\{\lambda\})$ for each $\lambda \in \mathbb{R}$. Here ν and ν_ω^n are the measures corresponding to Ψ and Ψ_ω^n , respectively. Both steps together imply uniform convergence. To verify these convergences one applies again Lemma 8.3 as well as Birkhoff's ergodic theorem.

Remark 8.5. The first statement of Theorem 8.4 identifies the limit f^* from Corollary 8.2, namely it shows $f^* = \Theta$, where Θ is given in (8.3).

Appendix. Examples of measures

Let us discuss three classes of examples of measures \mathbb{P} satisfying (M1), (M2) and (M3).

(a) *Countable colors:* Consider the case $d = 1$ and let $\mathcal{A} = \mathbb{N}_0$. Let $\Omega = \mathbb{R}^{\mathbb{Z}}$ and fix an arbitrary product measure $\tilde{\mathbb{P}}: \mathcal{B}(\Omega) \rightarrow [0, 1]$ with support

$$\text{supp } \tilde{\mathbb{P}} \subseteq \mathcal{A}^{\mathbb{Z}}.$$

We define a transformation of $\tilde{\mathbb{P}}$. To this end, let constants $c, \beta, \alpha_{-c}, \alpha_{-c+1}, \dots, \alpha_c \in \mathbb{N}_0$ be given and consider the function

$$\varphi: \Omega \rightarrow \Omega, \quad (\varphi(\omega))_z := \beta + \sum_{k=-c}^c \alpha_k \omega_{z-k}.$$

We define $\mathbb{P} := \tilde{\mathbb{P}} \circ \varphi^{-1}$. Let us check the conditions (M1), (M2) and (M3) for \mathbb{P} . In order to check (M1) let $z \in \mathbb{Z}$ be given. Then, using stationarity of the product measure $\tilde{\mathbb{P}}$,

$$\begin{aligned} \mathbb{P} \circ \tau_z^{-1} &= \tilde{\mathbb{P}} \circ \varphi^{-1} \circ \tau_z^{-1} = \tilde{\mathbb{P}} \circ (\tau_z \circ \varphi)^{-1} \\ &= \tilde{\mathbb{P}} \circ (\varphi \circ \tau_z)^{-1} = \tilde{\mathbb{P}} \circ \tau_z^{-1} \circ \varphi^{-1} = \tilde{\mathbb{P}} \circ \varphi^{-1} = \mathbb{P}. \end{aligned}$$

Let us verify condition (M2) for \mathbb{P} . We define for each $A \in \mathcal{F}$ the function

$$\rho_A: \mathcal{A}^A \rightarrow \mathbb{R}, \quad x = (x_z)_{z \in A} \mapsto \rho_A(x) = \mathbb{P}(\Pi_A^{-1}(\{x\})).$$

Then ρ_A is the density of the marginal measure \mathbb{P}_A with respect to the counting measure on \mathbb{N}_0 , since we have for each $A \in \mathcal{F}$ and $A \in \mathcal{B}(\mathcal{A}^A)$

$$\mathbb{P}_A(A) = \sum_{x \in A} \mathbb{P}(\Pi_A^{-1}(\{x\})) = \sum_{x \in A} \rho_A(x).$$

It remains to verify condition (M3). To this end, let $\Lambda_1, \dots, \Lambda_n \subseteq \mathbb{Z}$ with $\min\{\mathfrak{d}(\Lambda_i, \Lambda_j) \mid i \neq j\} > 2c$ be given. Then, using the definition of φ , we have for each $x = (x_z)_{z \in \Lambda} \in \Lambda := \bigcup_{i=1}^n \Lambda_i$

$$\mathbb{P}(\Pi_{\Lambda}^{-1}(\{x\})) = (\tilde{\mathbb{P}} \circ \varphi^{-1} \circ \Pi_{\Lambda}^{-1})(\{x\}) = \prod_{i=1}^n (\tilde{\mathbb{P}} \circ \varphi^{-1} \circ \Pi_{\Lambda_i}^{-1})(\{x\}),$$

which proves that $\rho_{\Lambda} = \prod_{i=1}^n \rho_{\Lambda_i}$.

- (b) *Normal distribution:* Here, we treat the case $d = 1$, $\mathcal{A} = \mathbb{R}$, $\Omega = \mathbb{R}^{\mathbb{Z}}$ and set $\tilde{\mathbb{P}} := \bigotimes_{z \in \mathbb{Z}} \mathcal{N}(0, 1) : \mathcal{B}(\Omega) \rightarrow [0, 1]$, where $\mathcal{N}(0, 1)$ is the standard normal distribution. For $c \in \mathbb{N}_0$ and $\beta, \alpha_{-c}, \alpha_{-c+1}, \dots, \alpha_c \in \mathbb{R}$ we use

$$\varphi : \Omega \rightarrow \Omega, \quad (\varphi(\omega))_z = \beta + \sum_{k=-c}^c \alpha_k \omega_{z-k}$$

to define $\mathbb{P} := \tilde{\mathbb{P}} \circ \varphi^{-1}$. As before, the conditions (M1) and (M3) are implied by the choice of φ and the product structure of $\tilde{\mathbb{P}}$. For (M2), let $\Lambda \subseteq \mathbb{Z}$ be finite and first assume that $\Lambda = [a, b] \cap \mathbb{Z}$, $a, b \in \mathbb{Z}$. We define the matrix

$$A_{\Lambda} \in \mathbb{R}^{\Lambda \times \{a-c, \dots, b+c\}}, \quad (A_{\Lambda})_{i,j} = \alpha_{i-j},$$

where $\alpha_k := 0$ if $k \notin \{-c, \dots, c\}$. Recall that $\mathbb{P}_{\Lambda} = \tilde{\mathbb{P}} \circ \varphi^{-1} \circ \Pi_{\Lambda}^{-1} = \tilde{\mathbb{P}} \circ (\Pi_{\Lambda} \circ \varphi)^{-1}$. For $\omega \in \Omega$ we get

$$\Pi_{\Lambda}(\varphi(\omega)) = A_{\Lambda} \Pi_{[a-c, b+c]}(\omega) + \beta e_{\Lambda},$$

where $e_{\Lambda} = (1, \dots, 1)^{\top} \in \mathbb{R}^{\Lambda}$. Now, it follows that \mathbb{P}_{Λ} is normal distributed with mean βe_{Λ} and covariance matrix $A_{\Lambda} A_{\Lambda}^{\top}$. Note that $A_{\Lambda} A_{\Lambda}^{\top}$ is invertible since the rows of A_{Λ} are linearly independent. Thus, the measure \mathbb{P}_{Λ} is absolutely continuous with respect to the multi-dimensional Lebesgue measure.

In the situation where Λ is not of the form $[a, b] \cap \mathbb{Z}$, consider the interval $I := [\min \Lambda, \max \Lambda] \cap \mathbb{Z}$. The measure \mathbb{P}_{Λ} is a marginal measure of \mathbb{P}_I and therefore has a density.

- (c) *Abstract densities and finite range:* In the following we develop a more general example with densities. Again, we consider for simplicity reasons the case $d = 1$, however this is easily generalized to higher dimensions. Choose $\mathcal{A}, \mathcal{B} \in \mathcal{B}(\mathbb{R})$ and independent \mathcal{B} -valued random variables X'_x , $x \in \mathbb{Z}$ with density $g : \mathcal{A} \rightarrow \mathbb{R}_+$. We use the abbreviation $X_{[m, \ell]} := (X_m, \dots, X_{\ell})$. We utilize a function $\varphi : \mathcal{B}^{k+1} \rightarrow \mathcal{A}$ to introduce the \mathcal{A} -valued random variables

$$X_x := \varphi(X'_{[x, x+k]}) \quad x \in \mathbb{Z}.$$

We require from φ , that there is a function $\psi : \mathcal{A} \times \mathcal{B}^k \rightarrow \mathcal{B}$ such that

$$\psi(\varphi(x_{[0, k]}), x_{[1, k]}) = x_0$$

for all $x_{[0, k]} \in \mathcal{B}^{k+1}$. Further, ψ shall be continuously differentiable w.r.t. its first argument: $\psi' := D_1 \psi$. An example of such a pair of functions is

$$\varphi(x_{[0, k]}) := \frac{1}{k+1} \sum_{j=0}^k x_j, \quad \psi(\xi_0, x_{[1, k]}) := (k+1)\xi_0 - \sum_{j=1}^k x_j,$$

where $\mathcal{A} := \mathcal{B} := [0, 1]$ and $\psi'(\xi_0, x_{[1, k]}) = k+1$. In this example, $(X_x)_x$ is a moving average process. By suitable modifications, all moving average processes are seen to be included in our setting.

Proposition A.1. Fix a finite set $\Lambda \subseteq \mathbb{Z}$. Under the specified circumstances, the joint distribution of $(X_x)_{x \in \Lambda}$, is absolutely continuous with respect to Lebesgue measure on \mathcal{A}^Λ .

Proof. Without loss of generality, we treat only the case $\Lambda = \{1, \dots, \ell\}$. By construction, for $A_1, \dots, A_\ell \subseteq \mathcal{A}$ measurable,

$$\begin{aligned} p &:= \mathbb{P}(X_1 \in A_1, \dots, X_\ell \in A_\ell) \\ &= \int_{\mathcal{B}^{\ell+k}} dx_{[1, \ell+k]} \prod_{m=1}^{\ell} \chi_{A_m}(\varphi(x_{[m, m+k]})) \cdot \prod_{m=1}^{\ell+k} g(x_m). \end{aligned}$$

By Fubini and induction on $j \in \{0, \dots, \ell\}$, we see

$$\begin{aligned} p &= \int_{A_1 \times \dots \times A_j} d\xi_{[1, j]} \int_{\mathcal{B}^{\ell-j+k}} dx_{[j+1, \ell+k]} \\ &\quad \times \prod_{m=1}^j (g(\psi(\tilde{x}_m^{(j)})) |\psi'(\tilde{x}_m^{(j)})|) \cdot \prod_{m=j+1}^{\ell} \chi_{A_m}(\varphi(x_{[m, m+k]})) \cdot \prod_{m=j+1}^{\ell+k} g(x_m), \end{aligned}$$

where $\tilde{x}_j^{(j)} := (\xi_j, x_{[j+1, j+k]})$ and, for $1 \leq m < j$, the term $\tilde{x}_m^{(j)} := \tilde{x}_m^{(j-1)}|_{x_j \mapsto \psi(\xi_j, x_{[j+1, j+k]})}$ is generated from $\tilde{x}_m^{(j-1)}$ by substituting x_j by $\psi(\xi_j, x_{[j+1, j+k]})$. For the induction step use the substitution $\xi_j := \varphi(x_{[j, j+k]})$ or $x_j = \psi(\xi_j, x_{[j+1, j+k]})$ in

$$\begin{aligned} &\int_{\mathcal{B}} dx_j \chi_{A_j}(\varphi(x_{[j, j+k]})) f_j(x_j) g(x_j) \\ &= \int_{A_j} d\xi_j f_j(\psi(\xi_j, x_{[j+1, j+k]})) g(\psi(\xi_j, x_{[j+1, j+k]})) |\psi'(\xi_j, x_{[j+1, j+k]})|, \end{aligned}$$

for any $x_{[j+1, j+k]} \in \mathcal{B}^k$ and suitable $f_j: \mathcal{B} \rightarrow \mathbb{R}_+$. For $j = \ell$, we conclude

$$p = \int_{A_1 \times \dots \times A_\ell} d\xi_{[1, \ell]} \int_{\mathcal{B}^k} dx_{[\ell+1, \ell+k]} \prod_{m=1}^{\ell} (g(\psi(\tilde{x}_m^{(\ell)})) |\psi'(\tilde{x}_m^{(\ell)})|) \cdot \prod_{m=\ell+1}^{\ell+k} g(x_m).$$

We hereby identified the density with respect to the product Lebesgue measure on \mathcal{A}^ℓ . \square

References

- [1] A. Bovier, Statistical Mechanics of Disordered Systems. A Mathematical Perspective, in: Cambridge Series in Statistical and Probabilistic Mathematics, vol. 18, Cambridge University Press, Cambridge, 2006.
- [2] J. DeHardt, Generalizations of the Glivenko–Cantelli theorem, Ann. Math. Statist. 42 (6) (1971) 2050–2055.
- [3] A. Dvoretzky, J. Kiefer, J. Wolfowitz, Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator, Ann. Math. Statist. 27 (3) (1956) 642–669.
- [4] E. Følner, On groups with full Banach mean value, Math. Scand. 3 (1955) 243–254.
- [5] R. Griffiths, A proof that the free energy of a spin system is extensive, J. Math. Phys. 5 (1964) 1215–1222.
- [6] R. Griffiths, J. Lebowitz, Random spin systems: some rigorous results, J. Math. Phys. 9 (8) (1968) 1284–1292.
- [7] G. Grimmett, Percolation, in: Grundlehren der Mathematischen Wissenschaften, vol. 321, Springer, Berlin, 1999.
- [8] G. Keller, Equilibrium States in Ergodic Theory, in: London Mathematical Society Student Texts, vol. 42, Cambridge University Press, Cambridge, 1998.
- [9] D. Lenz, Uniform ergodic theorems on subshifts over a finite alphabet, Ergodic Theory Dynam. Systems 22 (1) (2002) 245–255.
- [10] D. Lenz, P. Müller, I. Veselić, Uniform existence of the integrated density of states for models on \mathbb{Z}^d , Positivity 12 (4) (2008) 571–589. arxiv.org/abs/math-ph/0607063.

- [11] D. Lenz, F. Schwarzenberger, I. Veselić, A Banach space-valued ergodic theorem and the uniform approximation of the integrated density of states, *Geom. Dedicata* 150 (1) (2010) 1–34.
- [12] D. Lenz, P. Stollmann, An ergodic theorem for Delone dynamical systems and existence of the integrated density of states, *J. Anal. Math.* 97 (2005) 1–24.
- [13] D. Lenz, I. Veselić, Hamiltonians on discrete structures: Jumps of the integrated density of states and uniform convergence, *Math. Z.* 263 (2009) 813–835.
- [14] P. Massart, The tight constant in the Dvoretzky–Kiefer–Wolfowitz inequality, *Ann. Probab.* 18 (3) (1990) 1269–1283.
- [15] L.A. Pastur, Selfaverageability of the number of states of the Schrödinger equation with a random potential, *Mat. Fiz. i Funkcional. Anal. Vyp.* 2 (238) (1971) 111–116.
- [16] L.A. Pastur, A.L. Figotin, *Spectra of Random and Almost-Periodic Operators*, Springer Verlag, Berlin, 1992.
- [17] F. Pogorzelski, F. Schwarzenberger, A Banach space-valued ergodic theorem for amenable groups and applications, *J. Anal. Math.* (2012) in press. [arXiv:1205.3649 \[math.SP\]](https://arxiv.org/abs/1205.3649).
- [18] R. Ranga Rao, Relations between weak and uniform convergence of measures with applications, *Ann. Math. Statist.* 33 (2) (1962) 659–680.
- [19] F. Schwarzenberger, The integrated density of states for operators on groups (Ph.D. thesis), Technische Universität Chemnitz, 2013, <http://nbn-resolving.de/urn:nbn:de:bsz:ch1-qucosa-123241>.
- [20] M.A. Shubin, Spectral theory and the index of elliptic operators with almost-periodic coefficients, *Uspekhi Mat. Nauk* 34 (2 (206)) (1979) 95–135. [English translation: *Russ. Math. Surveys*, 34:109–157, 1979].
- [21] T. Spencer, The Schrödinger equation with a random potential—a mathematical review, in: K. Osterwalder, R. Stora (Eds.), *Phénomènes Critiques, Systèmes aléatoires, Théories de jauge, Part I, II*, in: *Les Houches*, vol. 1984, North-Holland, Amsterdam, 1986, pp. 895–944.
- [22] W. Stute, On a generalization of the Glivenko–Cantelli theorem, *Z. Wahrscheinlichkeitstheor. Verwandte Geb.* 35 (1976) 167–175.
- [23] L. Van Hove, Quelques propriétés générales de l'intégrale de configuration d'un système de particules avec interaction, *Physica* 15 (1949) 951–961.
- [24] V.N. Vapnik, A.Ya. Chervonenkis, On the uniform convergence of relative frequencies of events to their probabilities, *Theory Probab. Appl.* 16 (1971) 264–280.
- [25] I. Veselić, Spectral analysis of percolation Hamiltonians, *Math. Ann.* 331 (4) (2005) 841–865.
- [26] I. Veselić, Spectral properties of Anderson-percolation Hamiltonians, *Oberwolfach Rep.* 3 (1) (2006) 545–547.
- [27] I. Veselić, Existence and Regularity Properties of the Integrated Density of States of Random Schrödinger Operators, in: *Lecture Notes in Mathematics*, vol. 1917, Springer, Berlin, 2008.
- [28] P.A. Vuillermot, Thermodynamics of quenched random spin systems, and application to the problem of phase transitions in magnetic (spin) glasses, *J. Phys. A: Math. Gen.* 10 (8) (1977) 1319.
- [29] J. Wehr, M. Aizenman, Fluctuations of extensive functions of quenched random couplings, *J. Stat. Phys.* 60 (3–4) (1990) 287–306.
- [30] F.T. Wright, The empirical discrepancy over lower layers and a related law of large numbers, *Ann. Probab.* 9 (2) (1981) 323–329.