Hereditary Hsu-Robbins-Erdős Law of Large Numbers *

István Berkes † Ioannis Karatzas ‡ Walter Schachermayer §

March 19, 2025

To the memory of Herbert Ellis Robbins (1915–2001)

Abstract

We show that every sequence $f_1, f_2, \cdots$ of real-valued random variables with $\sup_{n \in \mathbb{N}} \mathbb{E}(f_n^2) < \infty$ contains a subsequence $f_{k_1}, f_{k_2}, \cdots$ converging in Cesàro mean to some $f_\infty \in \mathbb{L}^2$ *completely,* to wit,

$$\sum_{N \in \mathbb{N}} \mathbb{P}\left(\left| \frac{1}{N} \sum_{n=1}^{N} f_{k_n} - f_\infty \right| > \varepsilon \right) < \infty \,, \quad \forall\, \varepsilon > 0\,;$$

and *hereditarily,* i.e., along all further subsequences as well. We also identify a condition, slightly weaker than boundedness in $\mathbb{L}^2$, which turns out to be not only sufficient for the above hereditary complete convergence in Cesàro mean, but necessary as well.

AMS 2020 Subject Classification: Primary 60A10, 60F15; Secondary 60G42, 60G46.
Keywords: Strong law of large numbers, Hsu-Robbins-Erdős theorem, complete convergence, hereditary convergence, uniform integrability, KPR Lemma, strong exchangeability at infinity, Wasserstein distance

1 Introduction

The strong law of large numbers (SLLN; Kolmogorov [28], [29]; also [17]) is one of the pillars of the theory of probability. For a sequence of real-valued and integrable functions $f_1, f_2, \cdots$, defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and independent, with common distribution μ, it states that the sample, or "Cesàro", averages

$$\frac{1}{N} \sum_{n=1}^{N} f_n \,, \quad N \in \mathbb{N}\,, \tag{1.1}$$

converge $\mathbb{P}$−a.e. to the ensemble average $\mathbb{E}(f_1) = \int_{\mathbb{R}} x\, \mu(\mathrm{d}x)$, as $N \to \infty$.

---

* Work on this project started in June 2022 during a brief visit at the "Centro di Ricerca Matematica Ennio de Giorgi" of the Scuola Normale Superiore, Pisa. The authors are greatly indebted to Tomoyuki Ichiba for bringing important results to their attention, and to Amol Aggarwal for his suggestions and interest in the work.

† Institute of Statistics, Graz University of Technology, Kopernikusgasse 24, 8010 Graz, Austria (*berkes@tugraz.at*) and Alfred Rényi Institute of Mathematics, Reáltanoda utca 13-15, 1053 Budapest, Hungary (*berkes.istvan@renyi.hu*).

‡ Departments of Mathematics and Statistics, Columbia University, New York, NY 10027 (*ik1@columbia.edu*).
Support from the National Science Foundation under Grant DMS-20-04977 is gratefully acknowledged.

§ Faculty of Mathematics, University of Vienna, Oskar-Morgenstern-Platz 1, 1090 Vienna, Austria (*walter.schachermayer@univie.ac.at*). Support from the Austrian Science Fund (FWF) under grants P-35519 and P-35197 is gratefully acknowledged.

In 1947, Hsu & Robbins [23] obtained a remarkable strengthening of this result. In the same setting as above, but now under the square-integrability condition

$$\mathbb{E}\big(f_1^2\big) = \int_{\mathbb{R}} x^2 \, \mu(\mathrm{d}x) < \infty \tag{1.2}$$

and with $\mathbb{E}\big(f_1\big) = 0$, they established the stronger (so-called "complete") convergence

$$\sum_{N \in \mathbb{N}} \mathbb{P} \left(\left| \frac{1}{N} \sum_{n=1}^{N} f_n \right| > \varepsilon \right) < \infty \,, \qquad \forall \, \varepsilon > 0 \,. \tag{1.3}$$

Then in 1949/50, Erdős ([18], [19]) gave a very concise proof of the Hsu-Robbins theorem; and showed additionally that the square-integrability condition (1.2) is not only sufficient for the validity of (1.3) but also *necessary,* as indeed had been conjectured in [23].

A useful juxtaposition of these results comes about, when one considers the sojourn times

$$T_\varepsilon := \sum_{N \in \mathbb{N}} \mathbf{1}_{\left\{ \left| \sum_{n=1}^{N} f_n \right| > \varepsilon N \right\}} \,, \quad \varepsilon > 0 \tag{1.4}$$

the sequence of averages in (1.1) spends outside $\varepsilon-$neighborhoods of the ensemble average $\mathbb{E}(f_1) = 0$. For a sequence of independent and equi-distributed $f_1, f_2, \cdots$, the SLLN amounts to the statement

$$\mathbb{E}\big(|f_1|\big) < \infty \implies \mathbb{P}\big(T_\varepsilon < \infty\big) = 1 \,, \quad \forall \, \varepsilon > 0 \,; \tag{1.5}$$

the Hsu-Robbins theorem to the statement

$$\mathbb{E}\big(f_1^2\big) < \infty \implies \mathbb{E}\big(T_\varepsilon\big) < \infty \,, \quad \forall \, \varepsilon > 0 \,; \tag{1.6}$$

and Erdős's result to the validity of the reverse implication in (1.6). The connection between the condition $\mathbb{E}\big(f_1^2\big) < \infty$ of (1.2), and complete convergence, was further sharpened by Heyde [22] who showed that, under (1.2), the variance $\sigma^2 := \mathbb{E}\big(f_1^2\big)$ admits the representation

$$\sigma^2 = \lim_{\varepsilon \downarrow 0} \left(\varepsilon^2 \sum_{N \in \mathbb{N}} \mathbb{P}\Big(\Big| \sum_{n=1}^{N} f_n \Big| > \varepsilon \, N \Big) \right) = \lim_{\varepsilon \downarrow 0} \Big(\varepsilon^2 \cdot \mathbb{E}\big(T_\varepsilon\big) \Big) \,. \tag{1.7}$$

• Another classical result along a similar, yet somewhat distinct, strand of inquiry, is that of Komlós [30]. Also known for a long time (58 years already) but always very striking, it asserts that such "ergodicity" (stabilization via averaging) as manifest in the SLLN, occurs within *any* sequence $f_1, f_2, \cdots$ of measurable, real-valued functions that satisfy only the boundedness-in-$\mathbb{L}^1$ condition

$$\sup_{n \in \mathbb{N}} \mathbb{E}\big(|f_n|\big) < \infty \,. \tag{1.8}$$

More precisely, under (1.8) *there exist an integrable function f_∞ and a subsequence $f_{k_1}, f_{k_2}, \cdots$ with*

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} f_{k_n} = f_\infty \,, \qquad \mathbb{P}-a.e.; \tag{1.9}$$

and the same holds "hereditarily", i.e., along every further subsequence of $f_{k_1}, f_{k_2}, \cdots$ (this hereditary aspect is immediate, when the $f_1, f_2, \cdots$ are independent and equi-distributed). Whereas, it is shown in [31] that every convex subset of $\mathbb{L}^1$ which satisfies the conclusion of the Komlós theorem, i.e., the statement in italics right above, must necessarily be $\mathbb{L}^1-$norm bounded as in (1.8).

The proof in [30] is one of the early applications in probability limit theory of martingale techniques; such techniques were used in [12]–[14], [21] to derive similar hereditary results for the central limit theorem and for the law of the iterated logarithm. They will be deployed liberally here as well.

1.1 Preview

Our result, a special case of which is stated succinctly as the first sentence of the abstract, and as Corollary 2.3, appears in Theorem 2.2. In the spirit of Komlós [30], it provides a hereditary version for the Hsu-Robbins-Erdős theorem. It consists of a *sufficiency part* (i), corresponding to [23] and proved in sections 4, 5; and of a *necessity part* (ii), corresponding to [18] and proved in section 6.

We establish first an auxiliary sufficiency result, Proposition 2.4, under an additional requirement ($\star$) which posits the existence of a subsequence whose squares converge weakly in $\mathbb{L}^1$ to some bounded function. This auxiliary result is proved in section 3 via uniform integrability arguments based on the KPR Lemma 7.1 ([24], [36]; [11]), perturbation methodologies, and martingale-theoretic arguments as in [30], [12] (pp. 137-141). These latter are reviewed in an Appendix (section 8) and are used to effect reductions to simple martingale differences. A uniform version of the result (1.7), quite important in the present context, is established in another Appendix (section 9).

1.2 Additional Aspects

The statement of Corollary 2.3 formulates, in the context of the Hsu-Robbins result, the heuristic "general principle of subsequences" which appears on the first pages of Chatterji [14] and of Berkes-Péter [7] (cf. [13], [2] also). Its proof proceeds by approximating appropriate subsequences of $f_1, f_2, \cdots$ by sequences *strongly exchangeable at infinity;* pioneered by Aldous ([1], [2]), this approach was refined by Berkes-Péter [7] and is adapted to our $\mathbb{L}^2$ setting here in somewhat simplified form (section 4). Exchangeability methods have been deployed (cf. [1], [2], [7]) to establish instances of the subsequence principle for "almost-sure" and for distributional results; but not for complete convergence as done here, or for convergence in probability as done in [26]. An additional hereditary feature, of the Komlós and of related results, was established by Berkes and Tichy in [6], [10]: not only does every subsequence of $f_{k_1}, f_{k_2}, \cdots$ converge a.e. in Cesàro mean to f_∞, but so do all further *permutations;* subsection 3.1 here discusses this aspect in our setting.

Another noteworthy phenomenon occurs here. Benoîst & Quint [5] construct *a bounded-in-$\mathbb{L}^2$ martingale-difference sequence with* $\sum_{N\in\mathbb{N}} \mathbb{P}(\big|\sum_{n=1}^N f_n\big| > \varepsilon N) = \infty$ *for every* $\varepsilon > 0$ (but containing, in accordance with Corollary 2.3 below, a *subsequence* $f_{k_1}, f_{k_2}, \cdots$ along which, and along whose every further subsequence, $\sum_{N\in\mathbb{N}} \mathbb{P}(\big|\sum_{n=1}^N f_{k_n}\big| > \varepsilon N) < \infty$ holds for every $\varepsilon > 0$). Thus, for proving Theorem 2.2 (i) or even its Corollary 2.3, approximation via martingale differences alone leads to a dead end; rather, one needs to approximate a suitable subsequence by an *exchangeable* sequence. That such a situation should arise, had been predicted in [1], [2]; but the argument was supported there only by (very) artificial examples. Our Theorem 2.2 seems to provide the first "concrete" instance, of an important limiting result in Probability Theory valid for independent, equi-distributed functions, whose "hereditary extension" *requires* methods based on exchangeability.

2 A Hereditary Hsu-Robbins-Erdős Law of Large Numbers

Definition 2.1. HRE Property. *We say that a sequence* $\big(f_n\big)_{n\in\mathbb{N}} \subset \mathbb{L}^0$ satisfies the HRE property for some given $f_\infty \in \mathbb{L}^0$, *if it contains a subsequence* $\big(f_{k_n}\big)_{n\in\mathbb{N}}$ *converging in* Cesàro *mean to* f_∞ *completely, i.e.,*

$$\sum_{N\in\mathbb{N}} \mathbb{P}\Big(\Big|\sum_{n=1}^{N} f_{k_n} - N\,f_\infty\Big| > \varepsilon\,N\Big) < \infty\,, \qquad \forall\;\varepsilon > 0\,, \tag{2.1}$$

and "hereditarily", i.e., also along all subsequences of $\big(f_{k_n}\big)_{n\in\mathbb{N}}$.

We establish in this paper the following hereditary version of the HSU-ROBBINS-ERDŐS Law of Large Numbers ([23], [18], [19]).

Theorem 2.2 (HEREDITARY HSU-ROBBINS-ERDŐS LLN)**.** *On a probability space* $(\Omega, \mathcal{F}, \mathbb{P})$*, consider a sequence of real-valued, measurable functions* $f_1, f_2, \cdots$ *.*

(i) *Suppose that for some sequence of sets* A_1 , A_2 , $\cdots$ *in* $\mathcal{F}$ *with* $\lim_{n\to\infty} \mathbb{P}(A_n) = 1$,

. *the sequence* $\big(f_n \mathbf{1}_{A_n}\big)_{n\in\mathbb{N}}$ *is bounded in* $\mathbb{L}^2$; *while*

. *the sequence* $\big(f_n \mathbf{1}_{A_n^c}\big)_{n\in\mathbb{N}}$ *converges to zero in* $\mathbb{L}^1$.

There exists then a subsequence $f_{k_1}, f_{k_2}, \cdots$ *of* $f_1, f_2, \cdots$ *with the HRE property for some* $f_\infty \in \mathbb{L}^2$.

(ii) *Conversely, suppose* $f_1, f_2, \cdots$ *has the HRE property for* $f_\infty \equiv 0$. *There exist then a sequence* $\big\{A_n\big\}_{n\in\mathbb{N}} \subset \mathcal{F}$ *with* $\lim_{n\to\infty} \mathbb{P}(A_n) = 1$, *and a subsequence* $f_{k_1}, f_{k_2}, \cdots$ *of* $f_1, f_2, \cdots$, *such that*

. *the sequence* $\big(f_{k_n} \mathbf{1}_{A_{k_n}}\big)_{n\in\mathbb{N}}$ *is bounded in* $\mathbb{L}^2$, *and*

. *the sequence* $\big(f_{k_n} \mathbf{1}_{A_{k_n}^c}\big)_{n\in\mathbb{N}}$ *converges to zero in* $\mathbb{L}^1$.

If all its sets A_1 , A_2 , $\cdots$ satisfy $\mathbb{P}(A_n) = 1$, then Theorem 2.2 (i) amounts to the statement that follows. This corresponds to the first sentence of the paper's abstract, and vindicates in the present context the heuristic "general principle of subsequences" enunciated in [14].

Corollary 2.3. *Every sequence of functions* $f_1, f_2, \cdots$ *which is bounded in* $\mathbb{L}^2$ *, i.e., satisfies*

$$\sup_{n\in\mathbb{N}} \mathbb{E}\big(f_n^2\big) < \infty \, , \tag{2.2}$$

contains a subsequence $f_{k_1}, f_{k_2}, \cdots$ *with the HRE property for some* $f_\infty \in \mathbb{L}^2$*.*

In section 3 we shall establish, using martingale methods, the preliminary result of Proposition 2.4 below with its additional assumption $(\star)$; then deploy this result in sections 5, 6 as a crucial stepping stone that allows us eventually to ascend to the generality of Theorem 2.2.

Proposition 2.4. *Suppose that the functions* $f_1, f_2, \cdots$ *satisfy* (2.2)*, as well as the following:*

$(\star)$ *The sequence* $f_1, f_2, \cdots$ *contains a subsequence* $f_{k_1}, f_{k_2}, \cdots$ *whose squares* $f_{k_1}^2, f_{k_2}^2, \cdots$ *converge weakly in* $\mathbb{L}^1$ *to a function* $\boldsymbol{\eta} \in \mathbb{L}^\infty$, *namely,*

$$\lim_{n\to\infty} \mathbb{E}\big(f_{k_n}^2 \cdot \xi\big) = \mathbb{E}\big(\boldsymbol{\eta} \cdot \xi\big) \, , \quad \forall \;\; \xi \in \mathbb{L}^\infty \, . \tag{2.3}$$

There exist then a real-valued function $f_\infty \in \mathbb{L}^2$ *and a suitable (further, relabelled) subsequence* $f_{k_1}, f_{k_2}, \cdots$ *of* $f_1, f_2, \cdots$ *with the HRE property for this* f_∞ *.*

2.1 Ramifications

Remark 2.5. As shall see in the start of section 4 (leading to Remark 4.1), for establishing the HRE property (2.1) in the context of Theorem 2.2 (i), we may assume that

$$\text{the sequence } \big(f_n^2\big)_{n\in\mathbb{N}} \text{ is uniformly integrable.} \tag{2.4}$$

Then the DUNFORD-PETTIS Theorem (T25 in Chapter II of [16]) gives a subsequence $f_{k_1}, f_{k_2}, \cdots$ satisfying (2.3) for some $\boldsymbol{\eta}$ in $\mathbb{L}^1$; *though not necessarily in* $\mathbb{L}^\infty$, as posited in Proposition 2.4.

Remark 2.6. Stable Convergence. The functions f_∞ , $\boldsymbol{\eta}$ play the rôles of randomized limiting first and second moments for the subsequence $f_{k_1}, f_{k_2}, \cdots$; they are measurable with respect to the tail $\sigma-$algebra

$$\mathcal{T} := \bigcap_{n\in\mathbb{N}} \mathcal{T}_n\,, \qquad \mathcal{T}_n := \sigma\big(f_{k_n}, f_{k_{n+1}}, \cdots\big). \tag{2.5}$$

Let us elaborate. As shown in § 6.1.1 here, every sequence $f_1, f_2, \cdots$ with the HRE property contains a subsequence $f_{k_1}, f_{k_2}, \cdots$ *bounded in* $\mathbb{L}^0$ ("bounded in probability", "tight"): namely, with $\sup_{n\in\mathbb{N}} \mathbb{P}\big(|f_{k_n}| > \lambda\big) \to 0$ as $\lambda \to \infty$. Thus, for Theorem 2.2 we may assume the sequence $f_1, f_2, \cdots$ to contain a subsequence bounded in $\mathbb{L}^0$. Now, from a long line of inquiry initiated by Rényi (cf. [34]; [4]; [3]; [9], Theorem 2.2), such a sequence contains also a (relabelled) "determining" subsequence $f_{k_1}, f_{k_2}, \cdots$, along which the *stable convergence* (extended Helly-Bray lemma)

$$\lim_{n\to\infty} \mathbb{P}\big(f_{k_n} \le x, B\big) =: \mathbb{Q}(x,B) = \int_B \boldsymbol{H}(x,\omega)\,\mathbb{P}(\mathrm{d}\omega)\,, \qquad \forall\;\; B \in \mathcal{F} \tag{2.6}$$

holds at every point x of a countable, dense set $\boldsymbol{D} \subset \mathbb{R}$. For each $x \in \boldsymbol{D}$, the set-function $B \mapsto \mathbb{Q}(x,B)$ is a measure on $\mathcal{F}$, absolutely continuous with respect to $\mathbb{P}$; and $\omega \longmapsto \boldsymbol{H}(x,\omega)$ a version of the Radon-Nikodým derivative $\mathrm{d}\mathbb{Q}(x,\cdot)/\mathrm{d}\mathbb{P}$, measurable with respect to the tail $\sigma-$algebra $\mathcal{T}$ in (2.5). Whereas, for $\mathbb{P}-$a.e. $\omega \in \Omega$, the mapping $x \longmapsto \boldsymbol{H}(x,\omega)$ is a probability distribution function on the real line (cf. [1], Lemma 2), called *limit random distribution function of the determining subsequence* $f_{k_1}, f_{k_2}, \cdots$. We denote its first (when it exists) and second moments, respectively, by

$$f_\infty(\omega) = \int_{\mathbb{R}} x\,\mathrm{d}\boldsymbol{H}(x,\omega)\,, \quad \boldsymbol{\eta}(\omega) = \int_{\mathbb{R}} x^2\,\mathrm{d}\boldsymbol{H}(x,\omega)\,, \qquad \omega \in \Omega\,. \tag{2.7}$$

The mapping $f_\infty : \Omega \to \mathbb{R}$ is well-defined and integrable under the condition (1.8) of the Komlós theorem, and then (1.9) holds; whereas, $\boldsymbol{\eta} : \Omega \to [0,\infty]$ is the mapping in Proposition 2.4 and Remark 2.5. These same $f_\infty(\omega)$, $\boldsymbol{\eta}(\omega)$, as well as the distribution function $\boldsymbol{H}(\cdot,\omega)$ and the measure $\boldsymbol{\mu}(\omega)$ it induces on the Borel sets of the real line, are generated also by the sequence $\big(f_n \mathbf{1}_{A_n}\big)_{n\in\mathbb{N}}$.

When the distribution function $x \longmapsto \boldsymbol{H}(x)$ in (2.6) happens not to depend on ω, the stable convergence $\lim_{n\to\infty} \mathbb{P}\big(f_{k_n} \le x, B\big) = \boldsymbol{H}(x)\,\mathbb{P}(B)$, $\forall\; B \in \mathcal{F}$ in (2.6) is called *mixing* (e.g., [34], [3]). The quantities of (2.7) are then real constants, and the condition $\boldsymbol{\eta} \in \mathbb{L}^\infty$ is satisfied trivially.

Remark 2.7. The conditions of Proposition 2.4 are of course satisfied in the context of the original Hsu-Robbins-Erdős Law of Large Numbers when the $f_1, f_2, \cdots$ are independent, equi-distributed, and square-integrable with $\mathbb{E}(f_1) = 0$; for then we can take $\mathbb{P}(A_n) = 1$, $\forall\; n \in \mathbb{N}$ in part **(i)** of Theorem 2.2 and have (2.2) trivially, (2.3) with $\boldsymbol{\eta} = \mathbb{E}(f_1^2)$, and (2.1) with $f_\infty = 0$.

2.2 Martingale-Difference Sequences

Let us suppose that $f_1, f_2, \cdots$ is a martingale-difference sequence with respect to its own filtration. If it is also bounded in $\mathbb{L}^p$ for some $p > 2$ (this requirement is stronger than (2.4)), Theorem 3.6 in Lesigne - Volný [32] shows that the $f_1, f_2, \cdots$ converge completely in Cesàro mean to zero: i.e., (2.1) holds with $f_\infty \equiv 0$. In a significant recent development, Benoîst & Quint construct (cf. Remark 2.3 in [5]) a martingale-difference sequence $f_1, f_2, \cdots$, bounded in $\mathbb{L}^2$, *for which the complete convergence of* (2.1) *with* $f_\infty \equiv 0$ *fails;* they point out that the functions $f_1, f_2, \cdots$ may even be independent. It is important to stress that, in both references [32], [5], the results refer *to the martingale difference sequences themselves, without any passage to subsequences.*

2.3 Exchangeable Sequences

Suppose the $f_1, f_2, \cdots$ are *exchangeable,* i.e., their finite-dimensional marginal distributions are invariant under permutations of indices, with $\mathbb{E}(|f_1|) < \infty$. The celebrated DE FINETTI theorem (e.g., [27]) provides then a random probability distribution function $x \mapsto \boldsymbol{H}(x, \omega)$, measurable with respect to the tail $\sigma-$algebra $\mathcal{T}$ and satisfying $\int_{\mathbb{R}} |x| \, \mathrm{d}\boldsymbol{H}(x, \omega) < \infty$ for $\mathbb{P}-$a.e. $\omega \in \Omega$, as well as the following property: *"Conditional on* $\mathcal{T}$, *the* $f_1, f_2, \cdots$ *are independent with common distribution function* $\mathbb{P}\big[f_1 \le x \,\big|\, \mathcal{T} \big](\omega) = \boldsymbol{H}\big(x; \omega\big), \; x \in \mathbb{R}$, *for* $\mathbb{P}-$*a.e.* $\omega \in \Omega$." The first and second moments

$$f_\infty(\omega) \;=\; \mathbb{E}\big[\, f_1 \,\big|\, \mathcal{T}\,\big](\omega) \;=\; \int_{\mathbb{R}} x \, \mathrm{d}\boldsymbol{H}\big(x; \omega\big)\,, \qquad \boldsymbol{\eta}(\omega) = \mathbb{E}\big[\, f_1^2 \,\big|\, \mathcal{T}\,\big](\omega) = \int_{\mathbb{R}} x^2 \, \mathrm{d}\boldsymbol{H}\big(x; \omega\big) \tag{2.8}$$

of this random distribution provide here the functions in (2.7), and we have $\mathbb{E}(\boldsymbol{\eta}) = \mathbb{E}(f_1^2)$.

For *exchangeable* $f_1, f_2, \cdots$, it follows from Proposition 9.1 that $\mathbb{E}(f_1^2) < \infty$ is equivalent to

$$\sum_{N \in \mathbb{N}} \mathbb{P}\left(\left| \frac{1}{N} \sum_{n=1}^{N} f_n - f_\infty \right| > \varepsilon \right) < \infty\,, \qquad \forall\, \varepsilon > 0\,, \tag{2.9}$$

and to the following strengthening, inspired by HEYDE's identity (1.7), of (2.9):

$$\limsup_{\varepsilon \downarrow 0} \left[\varepsilon^2 \sum_{N \in \mathbb{N}} \mathbb{P}\left(\left| \frac{1}{N} \sum_{n=1}^{N} f_n - f_\infty \right| > \varepsilon \right) \right] \;<\; \infty\,. \tag{2.10}$$

Proposition 2.8. *Let* $f_1, f_2, \cdots$ *be an exchangeable sequence of real-valued, integrable functions, and* f_∞, $\boldsymbol{\eta}$ *be as in* (2.8). *Then* (2.9) $\iff$ (2.10) $\iff$ $f_1 \in \mathbb{L}^2$ $\iff$ $\mathbb{E}(\boldsymbol{\eta}) < \infty$.

3 The Proof of Proposition 2.4

The proof of Proposition 2.4 will involve several steps, which we have tried to outline clearly.

First, we need to prepare the ground. On the strength of the boundedness–in–$\mathbb{L}^2$ assumption (2.2), the sequence of functions $f_1, f_2, \cdots$ contains a (relabelled) subsequence which

- is *determining,* to wit, satisfies (2.6) for some limit random probability distribution function $\boldsymbol{H}(\cdot\,, \omega)$ with first and second moments $f_\infty \in \mathbb{L}^2$ and $\boldsymbol{\eta} \in \mathbb{L}^\infty$, as in (2.7), and
- *converges weakly in* $\mathbb{L}^2$ *to this* $f_\infty \in \mathbb{L}^2$, i.e.,

$$\lim_{n \to \infty} \mathbb{E}\big(f_n \cdot \xi\big) = \mathbb{E}\big(f_\infty \cdot \xi\big)\,, \quad \forall \;\; \xi \in \mathbb{L}^2\,; \tag{3.1}$$

- whereas, the sequence $f_1^2, f_2^2, \cdots$ *converges weakly in* $\mathbb{L}^1$ *to* $\boldsymbol{\eta} \in \mathbb{L}^\infty$ in the manner of (2.3).

Without sacrificing generality (cf. (4.1) and the paragraph following it), we take $f_\infty \equiv 0\,,\; \big\|\boldsymbol{\eta}\big\|_\infty \le 1$.

Finally we may assume, in accordance with the perturbation arguments of CHATTERJI ([12], pp. 137-141; also discussed in section 8 here) and passing inductively to a subsequence if necessary, that the $f_1, f_2, \cdots$ are *simple,* $\mathbb{L}^2-$*bounded martingale differences* ("strongly orthogonal" in the terminology of [14]), i.e., satisfy

$$\mathbb{E}\big(f_{n+1} \,\big|\, \mathcal{F}_n\big) = 0\,, \quad \mathbb{E}\big(f_{n+1}^2 \,\big|\, \mathcal{F}_n\big) \le 1\,, \quad \mathbb{P} - \text{a.e.}, \qquad \text{with} \qquad \mathcal{F}_n := \sigma\big(f_1, \cdots, f_n\big) \tag{3.2}$$

for every $n \in \mathbb{N}$. We set $\mathcal{F}_0 := \{\emptyset, \Omega\}$. Consequently, the sequence $X_N := \sum_{n=1}^{N} f_n\,,\; N \in \mathbb{N}$ is a *square-integrable martingale* with $\mathbb{E}\big(X_N^2\big) \le N$, of the resulting filtration

$$\mathbb{F} \;:=\; \big\{ \mathcal{F}_n \big\}_{n \in \mathbb{N}_0}\,. \tag{3.3}$$

Since $f_\infty \equiv 0$, it is enough to establish (2.1) with $\varepsilon = 1$ (because we can then replace the $f_1, f_2, \cdots$ by $f_1/\varepsilon, f_2/\varepsilon, \cdots$); i.e., to show that for

$$B_N := \Big\{ \big| X_N \big| > N \Big\} = \Big\{ \Big| \sum_{n=1}^{N} f_n \Big| > N \Big\} \qquad \text{we have} \qquad \sum_{N \in \mathbb{N}} \mathbb{P}\big(B_N\big) < \infty . \tag{3.4}$$

• We start by introducing the stopping times

$$\boldsymbol{\tau}_N := \min \Big\{ n = 1, \cdots, N : \big| X_n \big| > N/3 \Big\} , \qquad N \in \mathbb{N} \tag{3.5}$$

of the filtration $\mathbb{F}$ in (3.3) (with the understanding $\min \emptyset \equiv \infty$), so that the Čebyšev inequality and the Doob maximal inequality (e.g., [17], pp. 249-250) give

$$\mathbb{P}\big(\boldsymbol{\tau}_N < \infty\big) = \mathbb{P}\Big(\max_{1 \le n \le N} \big| X_n \big| > \frac{N}{3} \Big) \le \frac{9}{N^2}\, \mathbb{E}\Big[\max_{1 \le n \le N} X_n^2 \Big] \le \frac{36}{N^2}\, \mathbb{E}\Big(X_N^2 \Big) \le \frac{36}{N} . \tag{3.6}$$

Summation over N yields a divergent series, so we need to come up with an improved estimate. In particular, we shall try to show that, with high probability, $\big| X_{\boldsymbol{\tau}_N} \big|$ will be bounded by $2N/3$.

To this effect, we introduce the event

$$A_N := \Big\{ \max_{n=1,\cdots,N} \big| f_n \big| \le \frac{N}{3} \Big\} \tag{3.7}$$

and claim

$$\sum_{N \in \mathbb{N}} \mathbb{P}\big(A_N^c\big) \;=\; \sum_{N \in \mathbb{N}} \mathbb{P} \left(\max_{1 \le n \le N} \big| f_n \big| > \frac{N}{3} \right) < \infty . \tag{3.8}$$

• We postpone the proof of (3.8) but note already its implication that, in order to prove the claim in (3.4), we need only show

$$\sum_{N \in \mathbb{N}} \mathbb{P}\Big(B_N \cap A_N \Big) < \infty . \tag{3.9}$$

We introduce for this purpose the $\mathbb{L}^2$−bounded martingale

$$Y_n \;:=\; X_n - X_{n \wedge \boldsymbol{\tau}_N} \;=\; \Big(X_n - X_{\boldsymbol{\tau}_N} \Big) \cdot \mathbf{1}_{\{\boldsymbol{\tau}_N < n\}} , \qquad n = 1, \cdots, N \tag{3.10}$$

with $\mathbb{E}\big[\big(Y_n - Y_{n-1} \big)^2 \,\big|\, \mathcal{F}_{\boldsymbol{\tau}_N} \big] \le 1$, $\mathbb{P}\big(\big| Y_N \big| > N/3 \,\big|\, \mathcal{F}_{\boldsymbol{\tau}_N} \big) \le 9/N$. Recalling (3.6), and noting that on the event $B_N \cap A_N$ both $\boldsymbol{\tau}_N < \infty$ and $\big| X_{\boldsymbol{\tau}_N} \big| \le 2N/3$ hold, we obtain $\big| Y_N \big| > N/3$. These estimates provide the bounds

$$\mathbb{P}\Big(B_N \cap A_N \Big) = \mathbb{P}\Big(\big| X_N \big| > N,\, \max_{1 \le n \le N} \big| f_n \big| \le \frac{N}{3} \Big) = \mathbb{P}\Big(\big| X_N \big| > N,\, \boldsymbol{\tau}_N < \infty,\, \max_{1 \le n \le N} \big| f_n \big| \le \frac{N}{3} \Big)$$
$$\le \mathbb{P}\Big(\boldsymbol{\tau}_N < \infty,\, \big| X_{\boldsymbol{\tau}_N} \big| \le \frac{2N}{3},\, \big| Y_N \big| > \frac{N}{3} \Big) \le \frac{36}{N} \cdot \frac{9}{N} ,$$

and summing over $N \in \mathbb{N}$ we obtain a convergent series, as posited in (3.9).

• We still need to argue the claim (3.8), and do this by reprising Erdős's original argument from p. 287 in [18]. For a function $f_1 \in \mathbb{L}^2$, we let $a_i := \mathbb{P}\big(|f_1| > 2^i \big)$, $i \in \mathbb{N}_0$ and check, in a straightforward manner,

$$\sum_{i \in \mathbb{N}_0} 2^{\,2i-1} a_i \;\le\; \sum_{i \in \mathbb{N}_0} 2^{\,2i} \big(a_i - a_{i+1} \big) \;\le\; \mathbb{E}(f_1^2) \;\le\; \sum_{i \in \mathbb{N}_0} 2^{\,2(i+1)} \big(a_i - a_{i+1} \big) \;\le\; \sum_{i \in \mathbb{N}_0} 2^{\,2i+2} a_i .$$

These inequalities show that the second-moment condition $\mathbb{E}(f_1^2) < \infty$ is equivalent to

$$\sum_{i \in \mathbb{N}_0} 2^{\,2i}\, a_i < \infty . \tag{3.11}$$

Now, consider $f_1, f_2, \cdots$ with common distribution and $\mathbb{E}(f_1^2) < \infty$ (note that Erdős imposes also independence on these $f_1, f_2, \cdots$ as a "blanket" assumption, but this part of his argument does not need independence). With integers $2^i \le N < 2^{i+1}$, $i \ge 2$ we let

$$F_N := \bigcup_{n=1}^{N} \big\{ |f_n| > 2^{\,i-2} \big\} = \Big\{ \max_{n=1,\cdots,N} |f_n| > 2^{\,i-2} \Big\}$$

and observe $\mathbb{P}(F_N) \le \sum_{n=1}^N \mathbb{P}\big(|f_n| > 2^{\,i-2}\big) = N\, a_{i-2} \le 2^{\,i+1}\, a_{i-2}$. Thus, on account of (3.11), we obtain (3.8) on account of

$$\sum_{N\in\mathbb{N}} \mathbb{P}(F_N) = \sum_{i\in\mathbb{N}\,,\, i\ge 2} \; \sum_{N=2^i}^{2^{\,i+1}-1} \mathbb{P}(F_N) \;\le\; \sum_{i\in\mathbb{N}\,,\, i\ge 2} 2^i \cdot 2^{\,i+1}\, a_{i-2} \;\le\; \sum_{i\in\mathbb{N}_0} 2^{\,2i+5}\, a_i \;<\infty\,.$$

• Summing up, we have proved (2.1) with $f_\infty \equiv 0$ for the (relabelled) subsequence $f_1, f_2, \cdots$ and for every further subsequence. The proof of Proposition 2.4 is complete. □

3.1 Permutations

Let us show now that (3.8), thus (3.9) and (3.4) as well, hold (after passing once more to a subsequence) also *along any permutation* $f_{k_1}, f_{k_2}, \cdots$ *of the (relabelled) subsequence* $f_1, f_2, \cdots$; namely, that we have

$$\sum_{N\in\mathbb{N}} \mathbb{P}\left(\max_{n=1,\ldots,N} |f_{k_n}| > \frac{N}{3} \right) \;<\; \infty\,. \tag{3.12}$$

To this end, we adapt the previous argument to the general setting of a sequence $f_1, f_2, \cdots$ bounded in $\mathbb{L}^2$, in fact satisfying (in the spirit of Remark 2.5) the uniform integrability (2.4) of its squares; we may assume $f_\infty \equiv 0$. By passing to a subsequence, we may assume that

$$a_i^n := \mathbb{P}\big(|f_n| > 2^i\big)\,, \quad n\in\mathbb{N} \tag{3.13}$$

converges to some limit a_i, as $n\to\infty$, for each $i\in\mathbb{N}$; and by $\mathbb{L}^2-$boundedness, we have again the property (3.11) for these limits. Whereas, by passing to a (relabelled) subsequence again, we obtain for each $i\in\mathbb{N}$ the bound

$$a_i^n \;<\; a_i + 2^{\,-3i}\,, \qquad \forall\;\; n\ge i\,. \tag{3.14}$$

Pretending for a moment that (3.14) holds for all $(n,i)\in\mathbb{N}^2$, allows us to establish the estimate

$$\sum_{N=2^{\,i}}^{2^{\,i+1}-1} \mathbb{P}\left(\max_{n=1,\cdots,N} |f_{k_n}| > N \right) \;\le\; C\cdot 2^{\,2i+3}\,\big(a_i + 2^{\,-3i}\big) \tag{3.15}$$

for every $i\in\mathbb{N}$; and we note that the right-hand-side is summable over $i\in\mathbb{N}$ on account of (3.11).

Of course, for given, fixed $i\in\mathbb{N}$, we cannot guarantee the validity of (3.14) for all $n\in\mathbb{N}$; but we *can* obtain a (relabelled) subsequence of $f_1, f_2, \cdots$ which satisfies, for some $i_0\in\mathbb{N}$, the bound

$$a_i^n = \mathbb{P}\big(|f_n| > 2^i\big) \;<\; 2^{\,-2i}\,, \qquad \forall\;\; n\in\mathbb{N} \tag{3.16}$$

for all integers $i\ge i_0$; for otherwise the presumed uniform integrability of the $f_1^2, f_2^2, \cdots$ would fail. And combining the two estimates (3.14), (3.16) we deduce that a further subsequence, again denoted $f_1, f_2, \cdots$, can be selected, along which the inequality (3.16) holds now for *every* $i\in\mathbb{N}$.

On the other hand, for $2^{\,i} \le N < 2^{\,i+1}$, there are at most i terms of the "permuted" subsequence from (3.12) among the $f_{k_1}, \cdots, f_{k_N}$, for which (3.14) can fail, and the corresponding terms do not affect the convergence in (3.15); whereas (3.16) holds for each one of the terms $f_{k_1}, \cdots, f_{k_N}$.

Putting everything together, we conclude that (3.12) holds.

• Finally, we consider any *permutation* $f_{k_1}, f_{k_2}, \cdots$ of the (relabelled) subsequence $f_1, f_2, \cdots$ constructed above, and try to prove that $\sum_{N\in\mathbb{N}} \mathbb{P}\big(\big|\sum_{n=1}^N f_{k_n}\big| > N\big) < \infty$, i.e., (2.1) with $f_\infty \equiv 0$, holds also along it. On the strength of (3.12), it suffices to show the convergence

$$\sum_{N\in\mathbb{N}} \mathbb{P}\left(\Big|\sum_{n=1}^{N} f_{k_n}\Big| > N\,;\ \max_{n=1,\ldots,N} \big|f_{k_n}\big| \le \frac{N}{3}\right) < \infty\,. \tag{3.17}$$

We focus on the N^{th} term of this sum; re-arrange the $\{k_1, \cdots, k_N\}$ in increasing order, i.e., as $\{n_1, \cdots, n_N\}$ with $n_1 < \cdots < n_N$; and apply the argument right above to the thus re-arranged finite collection. This leads eventually to (3.17) in a straightforward manner.

4 Strong Exchangeability at Infinity; Independence

We prepare now the ground for the proof of Theorem 2.2 (i), following the trail blazed in the seminal works Aldous [1], [2], Aldous-Eagleson [3] and, very significantly, Berkes-Péter [7]. Our path here will be straighter, as we can work in an $\mathbb{L}^2-$setting and thus measure distances between probability measures using the quadratic Wasserstein, rather than the more delicate Prokhorov, metric. This allows simpler arguments which we have endeavored, nevertheless, to spell out in detail.

We impose as in [7] the blanket assumption, that the underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is *separable* and rich enough to accommodate all the objects we need to place on it; in particular, a sequence $f_1, f_2, \cdots$ of functions as in Theorem 2.2 (i), for which *we may assume* $f_1^2\, \mathbf{1}_{A_1}, f_2^2\, \mathbf{1}_{A_2}, \cdots$ *to be uniformly integrable* in the manner of (2.4); see the justification right below. In accordance with section 3, we may take the $f_1\, \mathbf{1}_{A_1}, f_2\, \mathbf{1}_{A_2}, \cdots$ to be *simple, martingale-differences* as in (3.2); and observe from (3.1) that, after passing to a (re-labelled) subsequence, we may assume also that

$$\lim_{n\to\infty} \mathbb{E}\big(f_n\, \mathbf{1}_{A_n} \cdot \xi\big) = \mathbb{E}\big(f_\infty \cdot \xi\big)\,, \quad \forall\ \ \xi \in \mathbb{L}^2 \tag{4.1}$$

holds for some $f_\infty \in \mathbb{L}^2$. We simplify typography by taking $f_\infty \equiv 0$.

• *To justify the just posited uniform integrability of* $f_1^2\, \mathbf{1}_{A_1}, f_2^2\, \mathbf{1}_{A_2}, \cdots$, let us recall the setting of Theorem 2.2 (i). We observe first, from Proposition 7.2 (i), that a (relabelled) subsequence of $\big(f_n \cdot \mathbf{1}_{A_n^c}\big)_{n\in\mathbb{N}}$ satisfies the HRE property with $f_\infty \equiv 0$. Consequently, in order to prove Theorem 2.2 (i), **it suffices to establish this HRE property for the sequence** $\big(f_n \cdot \mathbf{1}_{A_n}\big)_{n\in\mathbb{N}}$.

Now, the sequence $\big(f_n^2 \cdot \mathbf{1}_{A_n}\big)_{n\in\mathbb{N}}$ is bounded in $\mathbb{L}^1$, so Lemma 7.1 provides a sequence $B_1, B_2, \cdots$ of disjoint sets in $\mathcal{F}$ such that, after passing to a subsequence, $\big(f_n^2 \cdot \mathbf{1}_{A_n \setminus B_n}\big)_{n\in\mathbb{N}}$ is uniformly integrable. On the other hand, the functions $h_n := f_n \cdot \mathbf{1}_{A_n \cap B_n}$, $n \in \mathbb{N}$ are bounded in $\mathbb{L}^2$ and supported on disjoint sets, so $\sum_{n\in\mathbb{N}} \mathbb{E}(h_n^2) = \mathbb{E}\left(\sum_{n\in\mathbb{N}} h_n\right)^2 < \infty$. Consequently, $\big(h_n\big)_{n\in\mathbb{N}}$ converges to zero in $\mathbb{L}^2$, thus also in $\mathbb{L}^1$; and invoking Proposition 7.2 (i) once again, we deduce that this sequence satisfies the HRE property with $f_\infty \equiv 0$. It suffices, therefore, to establish this HRE property for the sequence of functions $\big(f_n \cdot \mathbf{1}_{A_n \setminus B_n}\big)_{n\in\mathbb{N}}$, *whose squares are uniformly integrable.*

Remark 4.1. As stressed in the preceding paragraphs, for the purposes of proving Theorem 2.2 (i) it suffices to establish the HRE property for the sequence of functions $\big(f_n \cdot \mathbf{1}_{A_n}\big)_{n\in\mathbb{N}}$, whose squares can be assumed uniformly integrable. *To simplify typography we shall denote this sequence, in the remainder of this section and in section 5, simply by* $\big(f_n\big)_{n\in\mathbb{N}}$, *and assume without loss of generality that it has the martingale-difference property.*

• We recall then the filtration $\mathbb{F} = \{\mathcal{F}_n\}_{n\in\mathbb{N}_0}$ from (3.3); consider for each $n \in \mathbb{N}$ a collection $A_j^{(n)}$, $j = 1, \cdots, J_n$ of atoms of positive $\mathbb{P}-$measure, which generate the $\sigma-$algebra $\mathcal{F}_n$; and denote the conditional distribution of f_n, given a "generic" one of these atoms A with $\mathbb{P}(A) > 0$, by

$$\mu_{(n)}^A(\cdot) := \mathbb{P}\big(f_n \in \cdot \,\big|\, A\big) . \tag{4.2}$$

This distribution is an element of the space $\mathcal{P}_2(\mathbb{R})$ of probability measures on the BOREL sets of the real line, with finite second moment. We endow this space with the quadratic WASSERSTEIN distance right below, which renders it Polish (i.e., a complete, separable, metric space; cf. [37]):

$$\mathcal{W}_2\big(\mu, \nu\big) := \left(\inf_{X\sim\mu,\, Y\sim\nu} \mathbb{E}\Big(X - Y\Big)^2 \right)^{1/2} .$$

Now, the uniform integrability of the $\big(f_n^2\big)_{n\in\mathbb{N}}$ from Remark 4.1 implies that, for each fixed atom A of the collection $\big(A_j^{(n)}\, ;\, j = 1, \cdots, J_n\, ,\, n \in \mathbb{N}\big)$, the sequence of probability measures

$$\mathcal{M}_A := \Big\{\mu_{(n)}^A\Big\}_{n\in\mathbb{N}} \tag{4.3}$$

in (4.2) is *tight:* given any $\varepsilon > 0$, there exists a compact set $K_\varepsilon \subset \mathbb{R}$ with $\int_{\mathbb{R}\setminus K_\varepsilon} x^2\, \nu(\mathrm{d}x) < \varepsilon\, ,\ \forall\ \nu \in \mathcal{M}_A$. Thus $\mathcal{M}_A$ is also a *relatively compact* subset of $\mathcal{P}_2(\mathbb{R})$: given any fixed such atom A, we can find a probability measure $\mu^A \in \mathcal{P}_2(\mathbb{R})$ and a subsequence $f_{k_1}^A, f_{k_2}^A, \cdots$ of $f_1, f_2, \cdots$, such that

$$\lim_{n\to\infty} \mathcal{W}_2\Big(\mu^A,\, \mu_{(k_n)}^A\Big) = 0 \tag{4.4}$$

(Theorem 7.12 in [37]); and eventually, by diagonalization, also an "omnibus" subsequence $f_{k_1}, f_{k_2}, \cdots$ of $f_1, f_2, \cdots$ with (4.4) valid for *each* such atom A. This $f_{k_1}, f_{k_2}, \cdots$ is a martingale-difference sequence of the "thinned" sub-filtration $\{\mathcal{F}_{k_n}\}_{n\in\mathbb{N}}$ of $\{\mathcal{F}_n\}_{n\in\mathbb{N}}$ in (3.3). To ease (the already heavy) notation, we denote these thinned subsequences as $\{f_n\}_{n\in\mathbb{N}}$, $\{\mu_{(n)}^A\}_{n\in\mathbb{N}}$ and $\{\mathcal{F}_n\}_{n\in\mathbb{N}}$, respectively.

• We follow again the trail from RÉNYI [34] (as well as [3], [9]) and obtain the existence of a measurable mapping $\boldsymbol{\mu} : \Omega \to \mathcal{P}_2(\mathbb{R})$ which "aggregates" the limiting probability measures in (4.4) in the sense that, for *each* atom A as above, we have for every bounded and continuous $\varphi : \mathbb{R} \to \mathbb{R}$ the property

$$\int_{\mathbb{R}} \varphi(x)\, \mu^A(\mathrm{d}x) = \int_A \left(\int_{\mathbb{R}} \varphi(x)\, \boldsymbol{\mu}(\mathrm{d}x, \omega) \right) \mathbb{P}(\mathrm{d}\omega)\, ; \quad \text{equivalently,} \quad \mu^A(\cdot) = \int_A \boldsymbol{\mu}(\cdot\,, \omega)\, \mathbb{P}(\mathrm{d}\omega)\, . \tag{4.5}$$

This "aggregating probability measure" $\boldsymbol{\mu}(\cdot\,, \omega)$ has $x \longmapsto \boldsymbol{H}(x, \omega)$ of (2.6) as its distribution function. We define also, for each fixed $n \in \mathbb{N}$, an $\mathcal{F}_n-$measurable mapping $\boldsymbol{\mu}_{(n)} : \Omega \to \mathcal{P}_2(\mathbb{R})$ by setting

$$\boldsymbol{\mu}_{(n)}(\omega, \cdot) := \mu_{(n)}^A(\cdot)\, , \qquad \omega \in \Omega \tag{4.6}$$

as in (4.2), with A the unique atom among the $A_j^{(n)}$, $j = 1, \cdots, J_n$ for which $\omega \in A$. In this manner we obtain the aggregated (and relabelled) version on (4.4), namely,

$$\lim_{n\to\infty} \mathcal{W}_2\big(\boldsymbol{\mu}\, , \boldsymbol{\mu}_{(n)}\big) = 0\, ,\ \mathbb{P} - \text{a.e.}; \quad \text{as well as} \quad \lim_{n\to\infty} \mathbb{E}\left[\mathcal{W}_2\big(\boldsymbol{\mu}\, , \boldsymbol{\mu}_{(n)}\big)\right] = 0\, . \tag{4.7}$$

Furthermore, appealing to EGOROV's theorem, we obtain sets $E_1 \subseteq E_2 \subseteq \cdots$ in $\mathcal{F}$ with $\mathbb{P}(E_m) \geq 1 - \varepsilon_m$ and $\lim_{m\to\infty} \downarrow \varepsilon_m = 0$, such that the $\mathbb{P}-$a.e. convergence in (4.7) is uniform on each E_m. We may (and will) assume additionally that, for each of these sets E_m, the restriction $\boldsymbol{\mu}\big|_{E_m}$ of the aggregator $\boldsymbol{\mu}$ in (4.5) is supported on a relatively compact subset of $\mathcal{P}_2(\mathbb{R})$.

• Next, we adapt the notion of "strong exchangeability at infinity" from [7] to our $\mathbb{L}^2-$ setting.

Definition 4.2. Strong $\mathbb{L}^2-$Exchangeability at Infinity. *Fix a sequence of square-integrable functions* $\big(g_n\big)_{n\in\mathbb{N}}$, *and a sequence* $\boldsymbol{\varepsilon} := \{\varepsilon_k\}_{k\in\mathbb{N}} \subset (0, 1)$ *with* $\lim_{k\to\infty} \downarrow \varepsilon_k = 0$.

We call $\big(g_n\big)_{n\in\mathbb{N}}$ **strongly $\mathbb{L}^2-$exchangeable at infinity with speed $\boldsymbol{\varepsilon}$** *if,* for each $k \in \mathbb{N}$, *there exists a partition* $\big\{A_0^{(k)}, A_1^{(k)}, \cdots, A_{J_k}^{(k)}\big\} \subset \mathcal{F}$ *of* Ω *by disjoint sets of positive* $\mathbb{P}-$*measure with the following properties:*

(i) $A_0^{(k+1)} \subseteq A_0^{(k)}$, *and the partition* $\big(A_j^{(k+1)},\ j = 0, 1, \cdots, J_{k+1}\big)$ *refines* $\big(A_j^{(k)},\ j = 0, 1, \cdots, J_k\big)$;

(ii) $\mathbb{P}\big(A_0^{(k)}\big) \le \varepsilon_k$, $\sup_{n\in\mathbb{N}} \mathbb{E}\Big(g_n^2 \cdot \mathbf{1}_{A_0^{(k)}}\Big) \le \varepsilon_k$;

(iii) *For each set* A *among the* $A_1^{(k)}, \cdots, A_{J_k}^{(k)}$, *there exist independent, square-integrable functions* $h_{k+1}^A, h_{k+2}^A, \cdots$, *with common distribution* μ^A *as in* (4.4) *and the sentence preceding it, as well as the property* $\mathbb{E}^{\mathbb{P}^A}\big[\,\big(g_n - h_n^A\big)^2\,\big] \le \varepsilon_k$, $\forall\ n \ge k+1$ *under the conditional probability measure*

$$\mathbb{P}^A\big(\,\cdot\,\big) := \mathbb{P}\big(\,\cdot \cap A\,\big) \,/\, \mathbb{P}\big(A\big)\,. \tag{4.8}$$

Here is an analogue of Theorem 1 in [7], tailored to our $\mathbb{L}^2-$ setting. It approximates (subsequences of) martingale-difference sequences by exchangeable ones, in a "progressively improving" manner.

Theorem 4.3. *Fix a sequence of functions* $f_1, f_2, \cdots$ *as in this section; and a decreasing sequence* $\boldsymbol{\varepsilon} := \big\{\varepsilon_k\big\}_{k\in\mathbb{N}} \subset (0,1)$ *with* $\lim_{k\to\infty} \downarrow \varepsilon_k = 0$. *There exist then a subsequence* $f_{\ell_1}, f_{\ell_2}, \cdots$ *and,* for each $k \in \mathbb{N}$, *an* exchangeable *sequence* $\big(\widehat{h}_n^{(k)}\big)_{n\ge k+1}$ *of square-integrable functions, so that*

$$\mathbb{E}\big[\,\big(f_{\ell_n} - \widehat{h}_n^{(k)}\big)^2\,\big|\,\mathcal{F}_k\,\big] \le \varepsilon_k\,, \quad \forall\ n \ge k+1 \quad \textit{holds with} \quad \mathcal{F}_k := \sigma\big(f_{\ell_1}, f_{\ell_2}, \cdots, f_{\ell_k}\big)\,. \tag{4.9}$$

In a similar spirit, we formulate an analogue of Theorem 2 in [7], dealing with strong exchangeability and tailored once again to our $\mathbb{L}^2-$setting. Just as in Berkes - Péter [7] (whose Theorem 1 follows from Theorem 2 there), Theorem 4.3 right above is a direct consequence of our next result, Theorem 4.4; this is proved in subsection 4.2 below.

Theorem 4.4. *Fix a sequence* $f_1, f_2, \cdots$ *as in this section; and a decreasing sequence* $\boldsymbol{\varepsilon} := \big\{\varepsilon_k\big\}_{k\in\mathbb{N}} \subset (0,1)$ *with* $\lim_{k\to\infty} \downarrow \varepsilon_k = 0$. *There exist then a subsequence* $f_{\ell_1}, f_{\ell_2}, \cdots$ *of* $f_1, f_2, \cdots$, *and a sequence* $\big(g_n\big)_{n\in\mathbb{N}} \subseteq \mathbb{L}^2$ *strongly* $\mathbb{L}^2-$*exchangeable at infinity with speed* $\boldsymbol{\varepsilon}$, *so that the analogue* $\mathbb{E}\big[\big(f_{\ell_n} - g_n\big)^2\,\big|\,\mathcal{F}_k\big] \le \varepsilon_k$ *of* (4.9) *holds for* $(n,k) \in \mathbb{N}^2$ *with* $n \ge k+1$.

We deduce in the next subsection some important consequences of this result. We start by casting it in an equivalent but more detailed and operational form, recalling Definition 4.2.

Corollary 4.5. *In the setting of Theorem 4.4, there exist a subsequence* $f_{\ell_1}, f_{\ell_2}, \cdots$ *of the given sequence* $f_1, f_2, \cdots$, *and a double array of row-wise disjoint sets* $\big(A_j^{(k)},\ j = 0, 1, \cdots, J_k\big)_{k\in\mathbb{N}}$ *with* $\mathbb{P}\big(A_j^{(k)}\big) > 0$, *such that,* for each given $k \in \mathbb{N}$, *we have:*

(i) $A_0^{(k+1)} := \Omega \backslash \bigcup_{j=1}^{J_{k+1}} A_j^{(k+1)} \subseteq \Omega \backslash \bigcup_{j=1}^{J_k} A_j^{(k)} =: A_0^{(k)}$, *and the partition* $\big(A_j^{(k+1)},\ j = 0, 1, \cdots, J_{k+1}\big)$ *is a refinement of the preceding partition* $\big(A_j^{(k)},\ j = 0, 1, \cdots, J_k\big)$;

(ii) $\mathbb{P}\big(A_0^{(k)}\big) \le \varepsilon_k$; $\sup_{n\in\mathbb{N}} \mathbb{E}\Big(f_{\ell_n}^2 \cdot \mathbf{1}_{A_0^{(k)}}\Big) \le \varepsilon_k$;

(iii) *the* $\sigma-$*algebra* $\mathcal{F}_k$ *of* (4.9) *is included in* $\mathcal{G}_k = \sigma\big(A_j^{(k)}\ ,\ j = 0, 1, \cdots, J_k\big)$; *and*

(iv) *for each set* A *among the* $A_1^{(k)}, \cdots, A_{J_k}^{(k)}$, *there exist square-integrable functions* $h_{k+1}^A, h_{k+2}^A, \cdots$, *independent and with common distribution* μ^A *as in* (4.4) *under the probability measure* $\mathbb{P}^A$ *of* (4.8), *which satisfy* $\mathbb{E}^{\mathbb{P}^A}\big(f_{\ell_n} - h_n^A\big)^2 \le \varepsilon_k$, $\forall\ n \ge k+1$; *or equivalently,*

$$\mathbb{E}\,\bigg[\,\Big(f_{\ell_n} - h_n^{A_j^{(k)}}\Big)^2 \cdot \mathbf{1}_{A_j^{(k)}}\,\bigg] \le \varepsilon_k \cdot \mathbb{P}\big(A_j^{(k)}\big)\,, \qquad \forall\ j = 1, \cdots, J_k\,, \quad n = k+1,\, k+2, \cdots. \tag{4.10}$$

4.1 Approximation by an Omnibus Sequence

Corollary 4.5 casts Theorem 4.4 in terms of a "progressively improving" (i.e., with diminishing error $\varepsilon_k \downarrow 0$) $\mathbb{L}^2-$approximation of an appropriate subsequence $f_{\ell_1}, f_{\ell_2}, \cdots$ of $f_1, f_2, \cdots$, using double arrays $\big(\widetilde{h}_n^{(k)},\, n \geq k+1\big)_{k\in\mathbb{N}}$ of row-wise independent functions in $\mathbb{L}^2$ with common distribution.

This setup is almost exactly that of Berkes-Péter [7], who impose only 'tightness' on the $f_1, f_2, \cdots$ and use Prokhorov distances. But it comes at a price: *at each level* $k \in \mathbb{N}$, it has to start a *new* register $k+1,\, k+2, \cdots$, and discard an exceptional set $A_0^{(k)}$ of small $\mathbb{P}-$measure.

For our purposes we shall need only $\mathbb{L}^2-$*bounded*, as apposed to $\mathbb{L}^2-$small-and-diminishing, approximations of the terms in the subsequence $f_{\ell_1}, f_{\ell_2}, \cdots$. The fact that we have already proved Proposition 2.4 will afford us this small luxury; and will enable us to put together, instead of a double array $\big(\widetilde{h}_n^{(k)},\, n \geq k+1\big)_{k\in\mathbb{N}}$ as in Corollary 4.5, *a single, "omnibus" sequence* of independent, centered functions $h_1, h_2, \cdots$, with the following properties:
(a) their conditional distributions, given each set in a partition of the space, will be the same ; and
(b) the omnibus sequence $h_1, h_2, \cdots$ will approximate, in a good $\mathbb{L}^2-$sense, a sequence $f_1^*, f_2^*, \cdots$ suitably close to the subsequence $f_{\ell_1}, f_{\ell_2}, \cdots$.
. In this manner *we will not have to restart a new exchangeable sequence* $\big(\widetilde{h}_n^{(k)},\, n \geq k+1\big)$ *at each level* $k \in \mathbb{N}$ *of approximation.*

We put together now the omnibus sequence $h_1, h_2, \cdots$. We construct the subsequence $f_{\ell_1}, f_{\ell_2}, \cdots$ and the partitions $\mathcal{A}^{(k)} = \big(A_j^{(k)},\; j = 0, 1, \cdots, J_k\big)$, $k \in \mathbb{N}$ inductively, as follows.
(i) Suppose that the indices $\ell_1, \cdots, \ell_{k-1}$ and the atoms $\mathcal{A}^{(\kappa)} = \big(A_j^{(\kappa)},\; j = 0, 1, \cdots, J_\kappa\big)$, $\kappa = 1, \cdots, k-1$, have been selected.
(ii) We define ℓ_k and $\mathcal{A}^{(k)}$ by splitting the "exceptional atom" $A_0^{(k-1)}$ into disjoint sets $\big(A_j^{(k)},\; j = 0, 1, \cdots, M_k\big)$ in such a way that, for each given $j = 1, \cdots, M_k$ and with $A \equiv A_j^{(k)}$, there are independent functions $h_{\ell_k}^A, h_{\ell_k+1}^A, \cdots$ supported on A, with the same distribution, and satisfying

$$\big\| h_n^A - f_n \big\|_{\mathbb{L}^2(\mathbb{P}^A)} < 1\,, \quad \forall\; n \geq \ell_k\,. \tag{4.11}$$

(iii) As for $A_0^{(k)}$, we require $\mathbb{P}\big(A_0^{(k)}\big) < 2^{-(k+1)}$ and $\big\| f_n \cdot \mathbf{1}_{A_0^{(k)}} \big\|_{\mathbb{L}^2(\mathbb{P})} < 2^{-(k+2)}$, $\forall\; n \geq \ell_k$; whereas, on the strength of this relation for the preceding step $k-1$, we may (and do) assume also

$$\Big\| \sum_{j=1}^{M_k} h_n^{A_j^{(k)}} \Big\|_{\mathbb{L}^2(\mathbb{P})} < 2^{-(k+1)}\,. \tag{4.12}$$

(iv) Regarding the remaining atoms $\big(A_j^{(k-1)},\; j = 1, \cdots, J_{k-1}\big)$ of the partition $\mathcal{A}^{(k-1)}$, we *do not split them any further;* rather, we keep them in the partition $\mathcal{A}^{(k)}$, *but relabelled,* namely as $\big(A_j^{(k)},\; j = M_k + 1, \cdots, J_k\big)$, so that $M_k + J_{k-1} = J_k$.

We continue in an obvious manner, and obtain a subsequence $f_{\ell_1}, f_{\ell_2}, \cdots$, as well as a countable partition $\mathcal{B}$ consisting of those atoms which appear as $A_j^{(k)} \in \mathcal{A}^{(k)}$ for some $(j,k) \in \mathbb{N}^2$ (and therefore also as elements of $\mathcal{A}^{(m)}$ for $m > k$). Fixing an atom B of this countable partition $\mathcal{B}$, we choose a sequence $h_1^B, h_2^B, \cdots$ of independent and equi-distributed functions, supported by B and satisfying, with $\boldsymbol{\kappa}(B)$ the smallest integer k for which $B \in \mathcal{A}^{(k)}$, the bound

$$\big\| h_n^B - f_n \big\|_{\mathbb{L}^2(\mathbb{P}^B)} < 1\,, \quad \forall\; n \geq \boldsymbol{\kappa}(B)\,. \tag{4.13}$$

This bound provides the desired estimate for integers $n \geq \boldsymbol{\kappa}(B)$.

But what about integers $n < \boldsymbol{\kappa}(B)$? To take care of these, we *modify* the sequence $f_{\ell_1}, f_{\ell_2}, \cdots$, and obtain a new approximating sequence $f_1^*, f_2^*, \cdots$ in the following manner. For $B \in \mathcal{B}$ we define $f_n^* \mathbf{1}_B = h_n^B$ for $n < \boldsymbol{\kappa}(B)$, so that $\|h_n^B - f_n^*\|_{\mathbb{L}^2(\mathbb{P}^B)} = 0 < 1$ holds trivially. Whereas, using (4.11)–(4.13), we have the estimate $\mathbb{E}\big[\,\big(f_{\ell_k} - f_k^*\big)^2\,\big] < 2^{-k}$.

In this manner, the sequence $f_{\ell_1}, f_{\ell_2}, \cdots$ satisfies (2.1) with $f_\infty = 0$ (in hereditary fashion) if, and only if, the so-constructed "approximating sequence" $f_1^*, f_2^*, \cdots$ does.

We reprise all the above by formulating the central result of the present section, an "omnibus version" of Theorem 4.4 and of Corollary 4.5.

Theorem 4.6. An Omnibus $\mathbb{L}^2-$Approximation. *For a sequence of functions $f_1, f_2, \cdots$ satisfying* (4.1) *and the conditions in the paragraph preceding it, there exist*

- *a subsequence $\big(f_{\ell_n}\big)_{n\in\mathbb{N}}$ of $\big(f_n\big)_{n\in\mathbb{N}}$;*
- *an approximating sequence $\big(f_n^*\big)_{n\in\mathbb{N}} \subset \mathbb{L}^2$ satisfying $\mathbb{E}\big[\big(f_{\ell_n} - f_n^*\big)^2\,\big] \le 2^{-n}$ for each $n \in \mathbb{N}$;*
- *a countable partition $\mathcal{B} = \{B_1, B_2, \cdots\}$ of Ω by sets in $\mathcal{F}$ with positive $\mathbb{P}-$measure; and*
- *a sequence $\big(h_n\big)_{n\in\mathbb{N}} \subset \mathbb{L}^2$, such that,* for each set $B \in \mathcal{B}$*, the functions $h_1, h_2, \cdots$ are independent and identically distributed under the conditional probability measure $\mathbb{P}^B$ as in* (4.8)*, with*

$$\mathbb{E}^{\mathbb{P}^B}\big[\,\big(f_n^* - h_n\big)^2\,\big] \le 1\,, \quad \textit{that is}\,, \quad \mathbb{E}\Big[\,\big(f_n^* - h_n\big)^2 \cdot \mathbf{1}_B\,\Big] \le \mathbb{P}\,(B)\;, \quad \forall\; n \in \mathbb{N}\,. \tag{4.14}$$

The "omnibus" sequence of square-integrable functions $\big(h_n\big)_{n\in\mathbb{N}}$ in this Theorem has properties particularly well-suited to our context, as the following result demonstrates.

Proposition 4.7. *Suppose the functions $\big(f_n\big)_{n\in\mathbb{N}}$ are square-integrable, and that $\mathcal{B} = \{B_1, B_2, \cdots\}$ is a partition of Ω by sets in $\mathcal{F}$ of positive measure and the property that, conditioned on each B_m, the functions $\big(f_n\big)_{n\in\mathbb{N}}$ are independent and have common distribution with zero mean.*

Then the functions $\big(f_n\big)_{n\in\mathbb{N}}$ satisfy the property (2.9) *with $f_\infty \equiv 0$, i.e., converge to zero completely in* Cesàro *mean.*

Proof: As already observed, it suffices to establish (2.9) for $\varepsilon = 1$. The uniform version of the representation (1.7) in Proposition 9.1, provides a universal constant $C > 0$ with

$$\sum_{N\in\mathbb{N}} \mathbb{P}^{B_m}\bigg(\Big|\sum_{n=1}^N f_n\Big| > N\bigg) \le C\,\big(\sigma_m^2 \vee 1\big)\,, \quad \forall\;\; m \in \mathbb{N}\,, \qquad \text{where} \quad \sigma_m^2 := \mathbb{E}^{\mathbb{P}^{B_m}}\big(f_1^2\big)\,.$$

Multiplying by $\mathbb{P}(B_m)$, then summing up over $m \in \mathbb{N}$, we obtain from the law of total probability

$$\sum_{N\in\mathbb{N}} \mathbb{P}\left(\Big|\sum_{n=1}^N f_n\Big| > N\right) \le\; C \sum_{m\in\mathbb{N}} \mathbb{P}(B_m)\,\big(\sigma_m^2 \vee 1\big) \;\le\; C\,\Big(1 + \mathbb{E}^{\mathbb{P}}\big(f_1^2\big)\Big) < \infty\,. \qquad \square$$

4.2 Proof of Theorem 4.4 and of Corollary 4.5

Following the trail of [7], we fix the subsequence $\big(f_{k_n}\big)_{n\in\mathbb{N}}$ of simple functions from the construction leading to (4.4), and relabel it $\big(f_n\big)_{n\in\mathbb{N}}$ for simplicity. This sequence is adapted to the filtration $\mathbb{F}$ of (3.3), and a martingale difference with respect to it. We recall also the "Egorov" sets $E_1 \subseteq E_2 \subseteq \cdots$ in $\mathcal{F}$ from below (4.7), and assume: $\mathbb{P}\big(E_k\big) \ge 1 - \varepsilon_k$ as well as $\sup_{n\in\mathbb{N}} \mathbb{E}\Big(f_n^2 \cdot \mathbf{1}_{\Omega\setminus E_k}\Big) \le \varepsilon_k\,;\;\; \forall\; k \in \mathbb{N}$.

In the next two subsections we shall construct a subsequence $\big(f_{\ell_k}\big)_{k\in\mathbb{N}}$ of the relabelled $\big(f_n\big)_{n\in\mathbb{N}}$, which is strongly $\mathbb{L}^2-$exchangeable at infinity with speed $\boldsymbol{\varepsilon} = \big(\varepsilon_k\big)_{k\in\mathbb{N}}$.

4.2.1 The Induction Step

We establish here Corollary 4.5, which is a rephrasing of Theorem 4.4.

We proceed by induction on k. Starting with $A_0^{(0)} = \Omega$ for $k = 0$, suppose that the partition $\big(A_j^{(k-1)},\ j = 0, 1, \cdots, J_{k-1}\big)$ has been constructed, along with functions $f_{\ell_1}, f_{\ell_2}, \cdots, f_{\ell_{k-1}}$ with the desired properties. We construct the next partition level $\big(A_j^{(k)},\ j = 0, 1, \cdots, J_k\big)$ as follows:

Let $A_0^{(k)} := A_0^{(k-1)} \cap \big(\Omega \setminus E_k\big)$. Using the uniform $\mathcal{W}_2-$convergence of the sequence $\big(\boldsymbol{\mu}_{(n)}\big)_{n\in\mathbb{N}}$ as in (4.5)-(4.7) to the "aggregator" $\boldsymbol{\mu} : \Omega \to \mathcal{P}_2(\mathbb{R})$, whose restriction to E_k has relatively compact range, we find an integer $J_k > J_{k-1}$, and a partition $\big(A_j^{(k)},\ j = 0, 1, \cdots, J_k\big)$ of Ω which

- has $A_0^{(k)}$ as first element;
- refines the previous-level partition $\big(A_j^{(k-1)},\ j = 0, 1, \cdots, J_{k-1}\big)$ and the $\sigma-$algebra $\mathcal{F}_{k-1}$ of (4.9);
- and is such that, *for every* $j = 1, \cdots, J_k$, *the restrictions to* $A_j^{(k)}$ *of the measure-valued mappings* $\big(\boldsymbol{\mu}_{(n)}\big)_{n\geq\ell_k}$ *and* $\boldsymbol{\mu}$, *all lie in a set* $\mathcal{M}_j^{(k)} \subset \mathcal{P}_2(\mathbb{R})$ *of* $\mathcal{W}_2-$*diameter less than* $\sqrt{\varepsilon_k}$.

Furthermore, we note that we may (and do) assume the functions $f_{\ell_1}, f_{\ell_2}, \cdots, f_{\ell_{k-1}}$ to be measurable with respect to the $\sigma-$algebra generated by the partition $\big(A_j^{(k)},\ j = 0, 1, \cdots, J_k\big)$. As a consequence, and in the notation of (4.2), (4.4), we may find an integer $\ell_k > \ell_{k-1}$ with the property

$$\mathcal{W}_2\Big(\mu_{(n)}^{A_j^{(k)}}, \mu^{A_j^{(k)}}\Big) \leq \sqrt{\varepsilon_k}\,, \qquad \forall\ n \geq \ell_k\,,\ j = 1, \cdots, J_k\,.$$

Results of Berkes-Philipp ([8], Theorems 1, 2) along with the assumed richness of the $\sigma-$algebra $\mathcal{F}$ and properties of the quadratic Wasserstein distance (the "joining step" in § 4.2.2), provide now a sequence of independent functions $\big(h_n^{A_j^{(k)}}\big)_{n\geq k+1}$, with common distribution $\mu^{A_j^{(k)}}$ as in (4.4) under the probability measure $\mathbb{P}^{A_j^{(k)}}$ of (4.8), and $\mathbb{E}^{\mathbb{P}^{A_j^{(k)}}}\Big(h_n^{A_j^{(k)}} - f_{\ell_n}\Big)^2 \leq \varepsilon_k\,,\ \forall\ n \geq \ell_k\,,\ j = 1, \cdots, J_k\,.$

4.2.2 The Joining Step

The above arguments need the following, simple property of the Wasserstein distance.

Suppose two measures μ, ν *in* $\mathcal{P}_2(\mathbb{R})$ *satisfy* $\mathcal{W}_2^2\big(\mu, \nu\big) < \varepsilon$, *and that a given function* $f \in \mathbb{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ *has distribution* μ. *With* $\big(\overline{\Omega}, \overline{\mathcal{F}}, \overline{\mathbb{P}}\big) := \big([0, 1] \times \Omega,\ \mathcal{B}([0, 1]) \otimes \mathcal{F},\ \mathrm{Leb} \otimes \mathbb{P}\big)$, *there exists a function* $g \in \mathbb{L}^2\big(\overline{\Omega}, \overline{\mathcal{F}}, \overline{\mathbb{P}}\big)$ *with distribution* ν, *and such that* $\overline{\mathbb{E}}\,\big(f - g\big)^2 < \varepsilon$.

The verification of this property is particularly straightforward, when the function f is simple. As we only need this case, we take $f = \sum_{j=1}^N \alpha_j\, \mathbf{1}_{A_j}$ for some real numbers $\alpha_1, \alpha_2, \cdots, \alpha_N$ and a finite partition $A_1, A_2, \cdots, A_N$ of Ω. By definition of the Wasserstein distance, there are probability measures $\kappa_1, \kappa_2, \cdots, \kappa_N$ in $\mathcal{P}_2(\mathbb{R})$ with $\nu = \sum_{j=1}^N \mathbb{P}\big(A_j\big)\, \kappa_j$, $\sum_{j=1}^N \mathbb{P}\big(A_j\big)\, \mathcal{W}_2^2\big(\boldsymbol{\delta}_{\alpha_j}, \kappa_j\big) < \varepsilon$.

Now, for every $j = 1, \cdots, N$, there exists a function $g_j \in \mathbb{L}^2\big([0, 1], \mathcal{B}([0, 1]), \mathrm{Leb}\big)$ with distribution κ_j; so we use these functions to define $g(t, \omega) := \sum_{j=1}^N g_j(t)\, \mathbf{1}_{A_j}(\omega)$, $(t, \omega) \in [0, 1] \times \Omega$. This new function g has the desired property $\overline{\mathbb{E}}\,\big(f - g\big)^2 < \varepsilon$. □

5 The Proof of Theorem 2.2 (i)

We recall the first few paragraphs of section 4, up to and including Remark 4.1. On their strength, it is enough to consider functions $f_1, f_2, \cdots$ bounded in $\mathbb{L}^2$; and reasoning as in the preamble of section 3 (cf. section 8), assume these are simple, centered martingale differences with $f_\infty \equiv 0$.

We appeal now to Theorem 4.6, recalling its bounded-in-$\mathbb{L}^2$ "approximating" and "omnibus" sequences $\big(f^*_n\big)_{n\in\mathbb{N}}$ and $\big(h_n\big)_{n\in\mathbb{N}}$, respectively, as well as to Remark 2.5; and introduce the functions

$$\xi_n \,:=\, f^*_n - h_n\,, \quad n\in\mathbb{N}\,. \tag{5.1}$$

Denoting by $\boldsymbol{\zeta}\in\mathbb{L}^1$ the weak-$\mathbb{L}^1$ limit of $\xi_1^2, \xi_2^2, \cdots$ in the manner of (2.3), we observe from (4.14) the bound $\mathbb{E}\big[\,\xi_n^2\,\big|\,\mathcal{B}\big]\le 1$, valid for every $n\in\mathbb{N}$ and leading to $\mathbb{P}\big(0\le\boldsymbol{\zeta}\le 1\big)=1$.

At this point, Proposition 2.4 takes over: after passing once again to a (relabelled) subsequence, Proposition 2.4 applies to the sequence $\big(\xi_n\big)_{n\in\mathbb{N}}$ and gives

$$\sum_{N\in\mathbb{N}}\mathbb{P}\left(\left|\frac{1}{N}\sum_{n=1}^{N}\xi_n\right|>\frac{\varepsilon}{4}\right)<\infty\,,\quad \forall\,\varepsilon>0\,. \tag{5.2}$$

The "omnibus" sequence $\big(h_n\big)_{n\in\mathbb{N}}$ satisfies the tenets of Proposition 4.7, thus also its conclusion

$$\sum_{N\in\mathbb{N}}\mathbb{P}\left(\left|\frac{1}{N}\sum_{n=1}^{N}h_n\right|>\frac{\varepsilon}{4}\right)<\infty\,,\quad\text{which leads to}\quad \sum_{N\in\mathbb{N}}\mathbb{P}\left(\left|\frac{1}{N}\sum_{n=1}^{N}f^*_n\right|>\frac{\varepsilon}{2}\right)<\infty \tag{5.3}$$

for every $\varepsilon>0$, on account of (5.1)–(5.2). Therefore, and in the context of Theorem 4.6 again, in order to establish the HRE property $\sum_{N\in\mathbb{N}}\mathbb{P}\left(\big|\sum_{n=1}^{N}f_{\ell_n}\big|>\varepsilon\,N\right)<\infty\,,\ \varepsilon>0$ it suffices to show $\sum_{N\in\mathbb{N}}\mathbb{P}\left(\sum_{n\in\mathbb{N}}\big|f_{\ell_n}-f^*_n\big|>\varepsilon N/2\right)<\infty$. But the elementary observation

$$\sum_{N\in\mathbb{N}}\mathbb{P}(Z>N)\,\le\,\mathbb{E}(Z)\,=\,\int_0^\infty\mathbb{P}\big(Z>t\big)\,\mathrm{d}t\,\le\,\sum_{N\in\mathbb{N}_0}\mathbb{P}(Z>N)\,,\qquad\forall\quad Z\in\mathbb{L}^0_+\,, \tag{5.4}$$

leads to the bounds

$$\frac{\varepsilon}{2}\cdot\sum_{N\in\mathbb{N}}\mathbb{P}\left(\sum_{n\in\mathbb{N}}\Big|f_{\ell_n}-f^*_n\Big|>\frac{\varepsilon}{2}\,N\right)\le\mathbb{E}\sum_{n\in\mathbb{N}}\Big|f_{\ell_n}-f^*_n\Big|\,\le\,\sum_{n\in\mathbb{N}}\left(\mathbb{E}\,\Big|f_{\ell_n}-f^*_n\Big|^2\right)^{1/2}<\infty \tag{5.5}$$

on account of $\mathbb{E}\left(f_{\ell_n}-f^*_n\right)^2\le 2^{-n}$ from Theorem 4.6, and the argument is now complete. □

6 The Proof of Theorem 2.2 (ii)

We turn now to Part (ii) of Theorem 2.2. Namely, we consider a sequence $f_1, f_2, \cdots$ of measurable, real-valued functions satisfying the HRE property with $f_\infty\equiv 0$, that is,

$$\sum_{N\in\mathbb{N}}\mathbb{P}\left(\left|\sum_{n=1}^{N}f_{k_n}\right|>\varepsilon\,N\right)\,<\,\infty\,,\quad\forall\,\varepsilon>0 \tag{6.1}$$

along some subsequence $f_{k_1}, f_{k_2},\cdots$ and along all its subsequences. Because complete convergence implies convergence a.e., this leads to the a.e. convergence of the Cesàro averages in (1.9) to $f_\infty\equiv 0$, also hereditarily. Whereas, passing to $f_{k_1}/\varepsilon, f_{k_2}/\varepsilon,\cdots$, it is enough to require (6.1) only for $\varepsilon=1$:

$$\sum_{N\in\mathbb{N}}\mathbb{P}\left(\left|\sum_{n=1}^{N}f_{k_n}\right|>N\right)\,<\,\infty\,. \tag{6.2}$$

We shall show presently that *there exist sets* $A_1\,,\,A_2\,,\cdots$ *in* $\mathcal{F}$ *with* $\lim_{n\to\infty}\mathbb{P}(A_n)=1$ (cf. (6.10) below) *as well as a further, relabelled subsequence* $f_{k_1}, f_{k_2},\cdots$, *such that* $\big(f_{k_n}\mathbf{1}_{A_{k_n}}\big)_{n\in\mathbb{N}}$ *is bounded in* $\mathbb{L}^2$ *while* $\big(f_{k_n}\mathbf{1}_{A^c_{k_n}}\big)_{n\in\mathbb{N}}$ *converges to zero in* $\mathbb{L}^1$.

6.1 The Plan

This program will be carried out in four distinct steps.

Step 1: *The* $f_{k_1}, f_{k_2}, \cdots$ *may be assumed bounded in* $\mathbb{L}^0$, *i.e.,* $\sup_{n\in\mathbb{N}} \mathbb{P}(|f_n| > \lambda) \to 0$ *as* $\lambda \to \infty$.

Step 2: *The* $f_{k_1}, f_{k_2}, \cdots$ *may be assumed integrable.*

Step 3: *The function* $\boldsymbol{\eta} : \Omega \to [0,\infty]$ *of* (2.7) *satisfies* $\mathbb{E}(\boldsymbol{\eta}) < \infty$.

Step 4: *Steps 1–3 lead to the claims in Part (ii) of Theorem 2.2.*

6.1.1 Step 1

Assume the $f_{k_1}, f_{k_2}, \cdots$ were not bounded in $\mathbb{L}^0$. Then for some constant $\alpha > 0$ and (relabelled) subsequence we would have $\mathbb{P}\big(|f_{k_N}| > 2N\big) \geq \alpha$, $\forall\, N \in \mathbb{N}$; and reasoning as in Erdős [18], also

$$\Big\{ |f_{k_N}| > 2N \Big\} \subseteq \left\{ \Big| \sum_{n=1}^{N} f_{k_n} \Big| > N \right\} \cup \left\{ \Big| \sum_{n=1}^{N-1} f_{k_n} \Big| > N-1 \right\}, \qquad \forall\, N \geq 2\,. \tag{6.3}$$

The assumption would then imply $\sum_{N\in\mathbb{N}} \mathbb{P}\big(|\sum_{n=1}^{N} f_{k_n}| > N\big) = \infty$, contradicting the HRE property.

6.1.2 Step 2

The HRE property (6.2) for $f_{k_1}, f_{k_2}, \cdots$, gives

$$\sum_{N\in\mathbb{N}} \mathbb{P}\left(\sum_{n=1}^{N} f_{k_n} > N \right) < \infty\,, \qquad \sum_{N\in\mathbb{N},\, N\geq 2} \mathbb{P}\left(\sum_{n=2}^{N} f_{k_n} > N \right) < \infty\,,$$

thus also $\sum_{N\in\mathbb{N},\,N\geq 2} \mathbb{P}\big(f_{k_1} > 2N\big) < \infty$. Applying the same reasoning to $-f_{k_1}, -f_{k_2}, \cdots$, and adding, we are led to $\sum_{N\in\mathbb{N}} \mathbb{P}\big(|f_{k_1}| > 4N\big) < \infty$, thus also via (5.4) to $\mathbb{E}\big(|f_{k_1}|\big) < \infty$. Similar arguments lead to $\mathbb{E}\big(|f_{k_n}|\big) < \infty$ for all $n \in \mathbb{N}$.

Remark 6.1. The I.I.D. Case. In particular, if the $f_1, f_2, \cdots$ are independent, have the same distribution, and satisfy $\sum_{N\in\mathbb{N}} \mathbb{P}\big(|\sum_{n=1}^{N} f_n| > \varepsilon N\big) < \infty$ for all $\varepsilon > 0$, this complete convergence holds along every subsequence as well — and leads to $\mathbb{E}\big(|f_1|\big) < \infty$, thus also to $\mathbb{E}\big(f_1\big) = 0$ by the strong law of large numbers and the fact that complete convergence implies convergence a.e. (cf. [19]).

6.1.3 Step 3

On the strength of Step 1, and after passing to an appropriate subsequence, we may assume that the functions $f_1, f_2, \cdots$ are "determining", in the sense of satisfying for $\mathbb{P}$−a.e. $\omega \in \Omega$ the stable convergence (2.6) for some limiting probability distribution function $\boldsymbol{H}(\cdot\,,\omega)$ and corresponding random measure $\boldsymbol{\mu}(\omega)$. As in (2.7), we let $\boldsymbol{\eta}(\omega) = \int_{\mathbb{R}} x^2 \, \mathrm{d}\boldsymbol{H}(x,\omega) \leq \infty$.

We consider now a sequence $g_1, g_2, \cdots$ of random variables, conditionally independent and with common distribution $\boldsymbol{\mu}$, given the sigma-algebra $\sigma(\boldsymbol{\mu})$ (cf. [1], p. 72 for the requisite construction). We recall the following generalization of (2.6), due to Dacunha-Castelle [15] (cf. [2], p. 122; [1], Corollary 8): for every $M \in \mathbb{N}$, and passing again to a suitable subsequence of $f_1, f_2, \cdots$, the random vector

$$\big(f_{k+1}, \cdots, f_{k+M}\big) \quad \text{converges in distribution as } k \to \infty \text{ to} \quad \big(g_1, \cdots, g_M\big). \tag{6.4}$$

• For expository reasons, we consider a special case first. Assume the following sharpened version of the HRE property (6.2): namely, that there exists a constant $K \in (0, \infty)$ with

$$\sum_{N \in \mathbb{N}} \mathbb{P}\Big(\Big| \sum_{n=1}^{N} f_{k_n} \Big| > N \Big) \,\leq\, K \tag{6.5}$$

valid for every subsequence $f_{k_1}, f_{k_2}, \cdots$ of $f_1, f_2, \cdots$. With $R \in \mathbb{N}_0$ fixed, we obtain then

$$\sum_{N \in \mathbb{N}} \mathbb{P}\left(\left| \sum_{n=1}^{N} f_{R+n} \right| > N \right) \leq K\,; \qquad \text{thus also} \qquad \sum_{N=1}^{M} \mathbb{P}\left(\left| \sum_{n=1}^{N} f_{R+n} \right| > N \right) \leq K\,,$$

for every given $M \in \mathbb{N}$. Letting $R \to \infty$ and recalling (6.4), we obtain

$$\sum_{N=1}^{M} \mathbb{P}\left(\left| \sum_{n=1}^{N} g_n \right| > N \right) \leq K\,, \quad \forall\, M \in \mathbb{N}\,, \qquad \text{thus} \qquad \sum_{N \in \mathbb{N}} \mathbb{P}\left(\left| \sum_{n=1}^{N} g_n \right| > N \right) \leq K\,.$$

But when conditioned on the sigma-algebra $\sigma(\boldsymbol{\mu})$, the functions $g_1,\, g_2, \cdots$ are independent with common random distribution function $\boldsymbol{H}(\cdot\,, \omega)$. For this sequence, applying (9.2) of Proposition 9.1 conditionally on $\sigma(\boldsymbol{\mu})$ and taking expectations, leads to $K \geq \sum_{N \in \mathbb{N}} \mathbb{P}\big(\big| \sum_{n=1}^{N} g_n \big| > N \big) \geq c \cdot \mathbb{E}\big(g_1^2\big)$ with a suitable universal constant $c > 0$. We have thus shown that, under (6.5), the "randomized second moment" $\omega \longmapsto \boldsymbol{\eta}(\omega) = \int_{\mathbb{R}} x^2 \, \mathrm{d}\boldsymbol{H}(x, \omega) \in [0, \infty]$ in (2.7) is integrable, namely, satisfies $\mathbb{E}(\boldsymbol{\eta}) = \mathbb{E}(g_1^2) < \infty$. (Similar arguments establish Proposition 2.8.)

• We drop now the condition (6.5); assume only (6.2), that is, $\sum_{N \in \mathbb{N}} \mathbb{P}\big(\big| \sum_{n=1}^{N} f_{k_n} \big| > N \big) \,<\, \infty\,$, for every subsequence $f_{k_1}, f_{k_2}, \cdots$ of $f_1, f_2, \cdots$; and try to show that this leads again to $\mathbb{E}(\boldsymbol{\eta}) \,<\, \infty$.

We argue by contradiction: namely, assume $\mathbb{E}(\boldsymbol{\eta}) = \infty$, and work toward finding a subsequence $f_{\ell_1}, f_{\ell_2}, \cdots$ of $f_1, f_2, \cdots$ and an increasing sequence of integers $1 < M_1 < M_2 < \cdots$ with

$$\sum_{N=1}^{M_j} \mathbb{P}\left(\left| \sum_{n=1}^{N} f_{\ell_n} \right| > N \right) \,\geq\, j - 1\,, \qquad \forall\, j \in \mathbb{N}\,; \tag{6.6}$$

this will then contradict the assumed HRE property (6.2) for the sequence $\big\{ f_{\ell_n} \big\}_{n \in \mathbb{N}}$.

As a first reduction step, we may *assume that each f_n is bounded,* i.e., $f_n \in \mathbb{L}^\infty$. Indeed, as each f_n can be assumed integrable on the strength of Step 2, we may find $f_n^* \in \mathbb{L}^\infty$ with $\big\| f_n - f_n^* \big\|_1 < 2^{-n}$. Now $\big\{ f_n^* \big\}_{n \in \mathbb{N}}$ inherits the HRE property from $\big\{ f_n \big\}_{n \in \mathbb{N}}$, as $\big\{ f_n - f_n^* \big\}_{n \in \mathbb{N}}$ satisfies this property on the strength of Proposition 7.2 (i); and $\big\{ f_n^* \big\}_{n \in \mathbb{N}}$ is still determining, with the same exchangeable sequence $\big\{ g_n \big\}_{n \in \mathbb{N}}$ as the original $\big\{ f_n \big\}_{n \in \mathbb{N}}$. In conclusion, we shall assume $\big\{ f_n \big\}_{n \in \mathbb{N}} \subset \mathbb{L}^\infty$.

We return now to the task of finding a subsequence $f_{\ell_1}, f_{\ell_2}, \cdots$ of $f_1, f_2, \cdots$ with the properties spelled out in the previous two paragraphs; and proceed to prove (6.6) via an induction (on $j \in \mathbb{N}$), whose first step $j = 1$ is clear. Next, we assume that (6.6) holds for some $j \in \mathbb{N}$ and natural numbers $M_1 < M_2 < \cdots < M_j\,,\ \ell_1 < \ell_2 < \cdots < \ell_{M_j}$. For the function $F_j := \sum_{n=1}^{M_j} f_{\ell_n} \in \mathbb{L}^\infty$, we can find then a real constant $C_j > 0$ so large, that $\big| F_j \big| \leq C_j$ holds a.e. Summoning the sequence $\big(g_n \big)_{n \in \mathbb{N}}$ of random variables which are conditionally independent and with common distribution $\boldsymbol{\mu}$ given the sigma-algebra $\sigma(\boldsymbol{\mu})$, applying Lemma 9.2 conditionally on $\sigma(\boldsymbol{\mu})$, taking expectations, and using monotone convergence as well as the standing assumption $\mathbb{E}(\boldsymbol{\eta}) = \infty$, we obtain

$$\sum_{\substack{N \in \mathbb{N} \\ N \geq M_j + 1}} \mathbb{P}\Big(\Big| \sum_{n = M_j + 1}^{N} g_n \Big| > N + C_j \Big) \,\geq\, c \cdot \mathbb{E}(\boldsymbol{\eta}) - 1 = \infty\,.$$

We extend now the string $\big(\ell_1, \cdots, \ell_{M_j}\big)$ to the increasing string $\big(\ell_1, \cdots, \ell_{M_j}, \ell^R_{M_j+1}, \cdots, \ell^R_{M_j+K}\big) := \big(\ell_1, \cdots, \ell_{M_j},\ R+1, \cdots, R+K\big)$, with $\ell_{M_j} \le R$ and $K \in \mathbb{N}$ fixed, so that the random vector $\big(f_{\ell^R_{M_j+1}}, \cdots, f_{\ell^R_{M_j+K}}\big)$ converges in distribution, as $R \to \infty$, to the random vector $\big(g_{M_j+1}, \cdots, g_{M_j+K}\big)$. This yields

$$\lim_{K\to\infty} \lim_{R\to\infty} \sum_{N=M_j+1}^{M_j+K} \mathbb{P}\left(\Big| \sum_{n=1}^{M_j} f_{\ell_n} + \sum_{n=M_j+1}^{M_j+K} f_{\ell^R_n} \Big| > N \right)$$

$$\ge \lim_{K\to\infty} \lim_{R\to\infty} \sum_{N=M_j+1}^{M_j+K} \mathbb{P}\left(\Big| \sum_{n=M_j+1}^{M_j+K} f_{\ell^R_n} \Big| > N + C_j \right) = \lim_{K\to\infty} \sum_{N=M_j+1}^{M_j+K} \mathbb{P}\left(\Big| \sum_{n=M_j+1}^{M_j+K} g_n \Big| > N + C_j \right) = \infty \,.$$

Taking R and M_{j+1} large enough, we define $\big(\ell_{M_j+1}, \cdots, \ell_{M_{j+1}}\big) := \big(\ell^R_{M_j+1}, \cdots, \ell^R_{M_{j+1}}\big)$ to obtain $\sum_{N=M_j+1}^{M_{j+1}} \mathbb{P}\big(\big| \sum_{n=1}^{N} f_{\ell_n} \big| > N \big) \ge 1$. This is the inductive step $j \mapsto j+1$ needed for establishing the claim (6.6), and completes Step 3.

6.1.4 Step 4

From Step 1, we may assume the $f_1, f_2, \cdots$ to be determining; in this manner, $\lim_{n\to\infty} \mathbb{P}\big(f_n \le x\big) = H(x) := \int_\Omega \boldsymbol{H}(x,\omega)\, \mathbb{P}(\mathrm{d}\omega)$, $\forall\ x \in \boldsymbol{D}$ holds from (2.6) with $B = \mathbb{R}$. Whereas, Step 3 gives

$$\lim_{n\to\infty} \mathbb{E}\Big(f_n^2 \cdot \mathbf{1}_{\{|f_n| \le K\}} \Big) \;=\; \int_{[-K,K]} x^2\, \mathrm{d}H(x) \;=\; \mathbb{E} \int_{[-K,K]} x^2\, \mathrm{d}\boldsymbol{H}(x,\omega) \;\le\; \mathbb{E}\big(\boldsymbol{\eta}\big) \;=:\; M < \infty \tag{6.7}$$

provided $\pm K \in \boldsymbol{D}$. We select now a sequence $0 < K_1 < K_2 < \cdots$ with $\pm K_n \in \boldsymbol{D}$ as well as

$$\mathbb{P}\big(|f_n| > K_n\big) \le 2^{-n}\,, \qquad \mathbb{E}\big(\big|f_n\big| \cdot \mathbf{1}_{\{|f_n| > K_n\}} \big) \le 2^{-n}\,, \qquad \forall\ n \in \mathbb{N} \tag{6.8}$$

(the latter because the $f_1, f_2, \cdots$ can be assumed integrable, on the strength of Step 2). Thus, after passing again to a subsequence, (6.7) gives

$$\mathbb{E}\big[f_n^2 \cdot \mathbf{1}_{\{|f_n| \le K_n\}} \big] \;<\; 2\, M \qquad \text{for all } n \in \mathbb{N}\,. \tag{6.9}$$

- *We introduce at this point the sets*

$$A_n := \big\{ |f_n| \le K_n \big\} \quad \text{for} \quad n \in \mathbb{N}\,, \quad \text{which satisfy } \ \mathbb{P}(A_n) > 1 - 2^{-n} \tag{6.10}$$

(and passing to the sets $\widetilde{A}_n = \bigcap_{m \ge n} A_m$ if necessary, we may also assume that these $\big(A_n\big)_{n\in\mathbb{N}}$ are increasing). From (6.9), the sequence $\big(f_n\, \mathbf{1}_{A_n}\big)_{n\in\mathbb{N}}$ is bounded in $\mathbb{L}^2$, as posited in Theorem 2.2 (ii); whereas $\big(f_n\, \mathbf{1}_{A_n^c}\big)_{n\in\mathbb{N}}$ converges to zero in $\mathbb{L}^1$, on the strength of the second display in (6.8).

Step 4 is thus established, and the Proof of Theorem 2.2 (ii) is now complete. □

7 Appendix: Functions Supported on Disjoint Sets

We recall from [11], Lemma 2.1.3 (cf. [24]; [36]; [25], Lemma A.44) the important result known colloquially as "KPR Lemma".

Lemma 7.1 (Kadeć-Pełzyński-Rosenthal)**.** *Every bounded–in–$\mathbb{L}^1$ collection of real-valued functions, contains a subsequence $\big(f_n\big)_{n\in\mathbb{N}}$ of the form $f_n = h_n + g_n$, $n \in \mathbb{N}$, where $\big(h_n\big)_{n\in\mathbb{N}}$ are integrable functions supported on disjoint sets $B_1, B_2, \cdots$, and $\big(g_n\big)_{n\in\mathbb{N}}$ are uniformly integrable functions with each g_n supported on the set $\Gamma_n := \bigcup_{m>n} B_m \in \mathcal{F}$.*

The following result shows that fast convergence to zero in $\mathbb{L}^1$, is sufficient for the HRE property; as well as necessary, after passage to an appropriate subsequence, for functions with disjoint or "essentially disjoint" (in a manner reminiscent of Lemma 7.1) supports.

Proposition 7.2. Conditions Sufficient, and Conditions Necessary, for the HRE Property: *Consider a sequence of real-valued, measurable functions* $h_1, h_2, \cdots$ *.*

(i) *The condition*

$$\sum_{n\in\mathbb{N}} \mathbb{E}\big(|h_n|\big) < \infty \tag{7.1}$$

implies, for the sequence in question and for all its subsequences, the complete convergence

$$\sum_{N\in\mathbb{N}} \mathbb{P}\Big(\Big|\sum_{n=1}^{N} h_n\Big| > \varepsilon N\Big) < \infty\,, \qquad \forall\ \varepsilon > 0\,. \tag{7.2}$$

(ii) *Conversely, if the* $h_1, h_2, \cdots$ *are supported on disjoint sets* $B_1, B_2, \cdots$ *in* $\mathcal{F}$*, the hereditary validity of* (7.2) *implies*

$$\underline{\lim}_{n\to\infty}\, \mathbb{E}\big(|h_n|\big) = 0\,, \tag{7.3}$$

therefore also (7.1) *along an appropriate subsequence and along all its subsequences.*

(iii) *More generally, with* $h_1, h_2, \cdots$ *as in (ii), consider a sequence of measurable functions* $g_1, g_2, \cdots$ *with each* g_n *supported on the set* $\Gamma_n := \bigcup_{m>n} B_m \in \mathcal{F}$ *. Then a necessary condition for the sequence* $h_1+g_1,\ h_2+g_2\,, \cdots$ *to contain a subsequence converging completely in* Cesàro *mean to zero, is again* (7.3); *which leads to* (7.1) *along a (relabelled) subsequence and all its subsequences.*

7.1 The Proof of Proposition 7.2

Proof of Part (i): This follows directly from (5.4), which gives $\sum_{N\in\mathbb{N}} \mathbb{P}\big(\big|\sum_{n=1}^{N} h_n\big| > N\big) \le \sum_{N\in\mathbb{N}} \mathbb{P}\big(\sum_{n\in\mathbb{N}} |h_n| > N\big) \le \sum_{n\in\mathbb{N}} \mathbb{E}\big(|h_n|\big) < \infty$ .This applies also to every subsequence. ☐

Proof of Part (ii): We argue by contradiction. Suppose that (7.3) fails; to wit, that after passing to a subsequence, we have $\mathbb{E}\big(|h_n|\big) > 2\,\beta\,,\ \forall\ n \in \mathbb{N}$ for some $\beta > 0$. Then, passing to a subsequence once again and recalling that the $h_1, h_2, \cdots$ are supported on disjoint sets, we may also assume

$$\mathbb{E}\Big(|h_n| \cdot \mathbf{1}_{\{|h_n|>n\}}\Big)\ >\ \beta\,, \qquad \forall\ \ n \in \mathbb{N}\,. \tag{7.4}$$

Now, for fixed $m \in \mathbb{N}$, we note

$$\sum_{N\ge m} \mathbb{P}\big(|h_m| > N\big)\ \ge\ \mathbb{E}\Big(|h_m| \cdot \mathbf{1}_{\{|h_m|>m\}}\Big) \ge \sum_{N>m} \mathbb{P}\big(|h_m| > N\big) \tag{7.5}$$

$$= \sum_{N\ge m} \mathbb{P}\big(|h_m| > N\big) - \mathbb{P}\big(|h_m| > m\big) \ge \sum_{N\ge m} \mathbb{P}\big(|h_m| > N\big) - \mathbb{P}\big(B_m\big)$$

(for the first two inequalities, recall (5.4)) as well as

$$\sum_{N\in\mathbb{N}} \mathbb{P}\Big(\Big|\sum_{n=1}^{N} h_n\Big| \cdot \mathbf{1}_{B_m} > N\Big) = \sum_{N\ge m} \mathbb{P}\big(|h_m| > N\big) \ge\ \mathbb{E}\Big(|h_m| \cdot \mathbf{1}_{\{|h_m|>m\}}\Big) > \beta \tag{7.6}$$

from (7.4), (7.5). We sum now in (7.6) over the disjoint sets B_m , $m \in \mathbb{N}$, and deduce

$$\sum_{N\in\mathbb{N}} \mathbb{P}\left(\Big|\sum_{n=1}^{N} h_n\Big| > N\right) = \infty\,; \tag{7.7}$$

i.e., that (7.2) fails for $\varepsilon = 1$, thus for all $\varepsilon > 0$ by scaling. This yields the contradiction. ☐

Proof of Part (iii): We argue again by contradiction: assume $\underline{\lim}_{n\to\infty} \mathbb{E}(|h_n|) > 0$, and show that the following analogue of (7.7) has then to hold along some subsequence:

$$\sum_{N\in\mathbb{N}} \mathbb{P}\left(\Big| \sum_{\ell=1}^{N} \Big(h_{k_\ell} + g_{k_\ell} \Big) \Big| > N \right) = \infty . \tag{7.8}$$

More precisely, we suppose again that $\mathbb{E}(|h_n|) > 2\beta$, $\forall\, n \in \mathbb{N}$ holds for some $\beta > 0$; also, without loss of generality, that $\mathbb{P}(B_n) < n^{-3}$ holds for all $n \in \mathbb{N}$. We construct now inductively a sequence of integers $1 = k_1 < k_2 < \cdots$ with each k_n, $n \geq 2$ equal to either $2n-1$ or $2n$, as follows.

For $n = 1$, we take $k_1 = 1$ as already mentioned, so that, with the help of (5.4), we obtain

$$\sum_{N\in\mathbb{N}} \mathbb{P}\Big(\Big| \sum_{n=1}^{N} \Big(h_\ell + g_\ell \Big) \mathbf{1}_{B_1} \Big| > N \Big) = \sum_{N\in\mathbb{N}} \mathbb{P}\big(|h_1| > N \big) \geq \mathbb{E}\big(|h_1|\big) - \mathbb{P}(B_1) > 2\beta - \mathbb{P}(B_1)\,;$$

and note that this holds also along any subsequence $(k_\ell)_{\ell\in\mathbb{N}}$ with $k_1 = 1$. Proceeding inductively, suppose that $1 = k_1 < k_2 < \cdots < k_n$ have been chosen so that, for each $j = 1, \cdots, n$, the expression

$$\sum_{N\in\mathbb{N}} \mathbb{P}\left(\Big| \sum_{\ell=1}^{N} \Big(h_{k_\ell} + g_{k_\ell} \Big) \mathbf{1}_{B_{2j-1}} \Big| > N \right) = \sum_{N\in\mathbb{N}} \mathbb{P}\left(\Big| \sum_{\ell=1}^{N\wedge n} \Big(h_{k_\ell} + g_{k_\ell} \Big) \mathbf{1}_{B_{2j-1}} \Big| > N \right)$$

dominates $2\beta - \big(2j-1\big) \cdot \mathbb{P}\big(B_{2j-1}\big)$.

We pair now the function h_{2n+1} with the sum $\sum_{\ell=1}^{n} g_{k_\ell} \cdot \mathbf{1}_{B_{2n+1}}$, and distinguish two possibilities:

- If

$$\mathbb{E}\left(\Big| h_{2n+1} + \sum_{\ell=1}^{n} g_{k_\ell} \cdot \mathbf{1}_{B_{2n+1}} \Big| \right) > \beta \tag{7.9}$$

holds, we set $k_{n+1} := 2n+1$ and proceed to estimate

$$\sum_{N\in\mathbb{N}} \mathbb{P}\left(\Big| \sum_{\ell=1}^{N} \Big(h_{k_\ell} + g_{k_\ell} \Big) \mathbf{1}_{B_{2n+1}} \Big| > N \right) = \sum_{N\in\mathbb{N}} \mathbb{P}\left(\left| \sum_{\ell=1}^{N\wedge(n+1)} \Big(h_{k_\ell} + g_{k_\ell} \Big) \mathbf{1}_{B_{2n+1}} \right| > N \right)$$

$$\geq \sum_{N\in\mathbb{N}} \mathbb{P}\left(\Big| \Big(h_{2n+1} + \sum_{\ell=1}^{n} g_{k_\ell} \Big) \mathbf{1}_{B_{2n+1}} \Big| > N \right) - n \cdot \mathbb{P}\big(B_{2n+1}\big) \geq \beta - n \cdot \mathbb{P}\big(B_{2n+1}\big)\,.$$

- If, on the other hand, (7.9) fails, we have $\mathbb{E}\,\big| \sum_{\ell=1}^{n} g_{k_\ell} \cdot \mathbf{1}_{B_{2n+1}} \big| > \beta$ on account of the assumption $\mathbb{E}\big(|h_{2n+1}|\big) > 2\beta$; take $k_{n+1} := 2n+2$; and obtain

$$\sum_{N\in\mathbb{N}} \mathbb{P}\left(\Big| \sum_{\ell=1}^{N} \Big(h_{k_\ell} + g_{k_\ell} \Big) \mathbf{1}_{B_{2n+1}} \Big| > N \right) = \sum_{N\in\mathbb{N}} \mathbb{P}\left(\left| \sum_{\ell=1}^{N\wedge(n+1)} \Big(h_{k_\ell} + g_{k_\ell} \Big) \mathbf{1}_{B_{2n+1}} \right| > N \right)$$

$$\geq \sum_{N\in\mathbb{N}} \mathbb{P}\left(\Big| \sum_{\ell=1}^{n} g_{k_\ell} \mathbf{1}_{B_{2n+1}} \Big| > N \right) - n \cdot \mathbb{P}\big(B_{2n+1}\big) \geq \beta - n \cdot \mathbb{P}\big(B_{2n+1}\big)\,.$$

Since $\sum_{n\in\mathbb{N}} n \cdot \mathbb{P}\big(B_{2n+1}\big) < \infty$ by assumption, we complete the argument summing up in the above display over the sets B_m, $m \in \mathbb{N}$, and arrive at the desired conclusion (7.8). $\square$

8 Appendix: Approximation by Simple Martingale Differences

We illustrate here how to approximate bounded-in-$\mathbb{L}^2$ sequences of functions by simple, square-integrable martingale differences, in the manner of Komlós [30] and Chatterji ([12], [13]).

Let us consider then a bounded-in-$\mathbb{L}^2$ sequence $f_1, f_2, \cdots$ satisfying (2.2); this contains a (relabelled) subsequence converging weakly in $\mathbb{L}^2$ to some $f_\infty \in \mathbb{L}^2$, as in (3.1). We take $f_\infty = 0$ for concreteness; and approximate each f_n by a *simple* function $h_n \in \mathbb{L}^2$ with

$$\mathbb{E}\big|f_n - h_n\big|^2 \le 4^{-n}, \ \forall\, n \in \mathbb{N}; \quad \mathbb{E}\left(\sum_{n\in\mathbb{N}} \big|f_n - h_n\big|\right) \le \sum_{n\in\mathbb{N}} \sqrt{\mathbb{E}\big|f_n - h_n\big|^2} \le \sum_{n\in\mathbb{N}} 2^{-n} = 1; \quad (8.1)$$

thus also $\sum_{n\in\mathbb{N}} \big|f_n - h_n\big| < \infty$, $\mathbb{P}-$a.e. In addition, for every test function $\xi \in \mathbb{L}^2$ we have $\mathbb{E}\big(h_n \cdot \xi\big) = \mathbb{E}\big(f_n \cdot \xi\big) - \mathbb{E}\big(\big(f_n - h_n\big) \cdot \xi\big) \longrightarrow 0$, as $n \to \infty$ from (3.1), (8.1). It develops that the sequence of simple functions $h_1, h_2, \cdots$ is bounded in $\mathbb{L}^2$ (since the inequality $\|h_m\|_2 \le \|f_m\|_2 + \|f_m - h_m\|_2 \le \sup_{n\in\mathbb{N}} \|f_n\|_2 + (1/2) < \infty$ holds for all $m \in \mathbb{N}$), and as such contains a (relabelled) subsequence converging weakly in $\mathbb{L}^2$ to $f_\infty \equiv 0$.

We construct now inductively a sequence $1 = k_1 < k_2 < \cdots$ *of integers* $k_n \ge n$ *such that*

$$\boldsymbol{\vartheta}_n := \mathbb{E}\Big(h_{k_n} \Big| \mathcal{H}_{n-1}\Big) \qquad \textit{with} \qquad \mathcal{H}_{n-1} := \sigma\big(h_{k_1}, \cdots h_{k_{n-1}}\big), \quad n = 2, 3, \cdots, \qquad (8.2)$$

which are simple functions, satisfy the bound $\big|\, \boldsymbol{\vartheta}_n \,\big| \le 2^{-n}$, $\mathbb{P} - a.e.$

This is done as follows: the function $h_{k_1} \equiv h_1$ is simple, so the conditional expectation of a generic h_n, given h_1, is also simple: $\mathbb{E}\big(h_n \,\big|\, h_1\big) = \sum_{j=1}^J \gamma_j^{(n)} \cdot \mathbf{1}_{A_j}$. Here the disjoint sets $A_1, \cdots, A_J$ form a partition of Ω; each of them has positive measure; and for every $j = 1, \cdots, J$, we have $\gamma_j^{(n)} := \big(\mathbb{P}(A_j)\big)^{-1} \cdot \mathbb{E}\big(h_n \cdot \mathbf{1}_{A_j}\big) \longrightarrow 0$, as $n \to \infty$.

We can select, therefore, $k_2 > 1 = k_1$ so that $\big|\gamma_j^{(k_2)}\big| \le 2^{-2}$, $j = 1, \cdots, J$ holds; thus also $\big|\boldsymbol{\vartheta}_2\big| = \big|\,\mathbb{E}\big(h_{k_2} \big| h_{k_1}\big)\,\big| = \sum_{j=1}^J \big|\gamma_j^{(k_2)}\big| \cdot \mathbf{1}_{A_j} \le 2^{-2}$, $\mathbb{P}-$a.e. We keep repeating this procedure; at each of its stages, the vector of simple functions $\big(h_{k_1}, \cdots, h_{k_{n-1}}\big)$ generates a finite partition of Ω, and we arrive this way inductively at the claim $\big|\, \boldsymbol{\vartheta}_n \,\big| \le 2^{-n}$, $\mathbb{P}-$a.e. □

We have now the following result.

Proposition 8.1. *For the integers* $1 = k_1 < k_2 < \cdots$ *with* $k_n \ge n$ *selected above, the resulting subsequence* $f_{k_1}, f_{k_2}, \cdots$ *of the original bounded–in–*$\mathbb{L}^2$ *sequence* $f_1, f_2, \cdots$ *has the HRE property* (2.1) *with* $f_\infty \equiv 0$ *if, and only if, this is true for the simple martingale-differences*

$$\boldsymbol{\beta}_n = h_{k_n} - \boldsymbol{\vartheta}_n = h_{k_n} - \mathbb{E}\Big(h_{k_n} \Big| h_{k_1}, \cdots h_{k_{n-1}}\Big), \quad n \in \mathbb{N} \qquad (8.3)$$

which are generated by the simple functions $h_{k_1}, h_{k_2}, \cdots$ *in* (8.1)–(8.2) *and satisfy*

$$\Big\| f_{k_n} - \boldsymbol{\beta}_n \Big\|_2 \le 2^{1-n}, \quad \forall \ n \in \mathbb{N}. \qquad (8.4)$$

Proof: For the "if" part, we note the inequality

$$\sum_{N\in\mathbb{N}} \mathbb{P}\left(\left|\sum_{n=1}^N f_{k_n}\right| > N\right) \le \sum_{N\in\mathbb{N}} \mathbb{P}\left(\left|\sum_{n=1}^N \Big(h_{k_n} - \boldsymbol{\vartheta}_n\Big)\right| > \frac{N}{3}\right) + \sum_{N\in\mathbb{N}} \mathbb{P}\big(\Phi_N\big) + \sum_{N\in\mathbb{N}} \mathbb{P}\big(\Theta_N\big) \quad (8.5)$$

with

$$\Phi_N := \left\{ \Big| \sum_{n=1}^{N} \Big(f_{k_n} - h_{k_n} \Big) \Big| > \frac{N}{3} \right\}, \qquad \Theta_N := \left\{ \Big| \sum_{n=1}^{N} \boldsymbol{\vartheta}_n \Big| > \frac{N}{3} \right\}.$$

We recall also the bound (5.4), valid for measurable $Z \geq 0$; in conjunction with $\big| \boldsymbol{\vartheta}_n \big| \leq 2^{-n}$ and the resulting $\big| \sum_{n \in \mathbb{N}} \boldsymbol{\vartheta}_n \big| \leq 1$, as well as (8.1), this leads to $\sum_{N \in \mathbb{N}} \mathbb{P}(\Theta_N) \leq 4$ and $\sum_{N \in \mathbb{N}} \mathbb{P}(\Phi_N) \leq 4$. On account of (8.5), the "if" part of the claim is thus established; whereas, repetition of the same argument establishes the "only if" part, via the inequality

$$\sum_{N \in \mathbb{N}} \mathbb{P} \left(\Big| \sum_{n=1}^{N} \Big(h_{k_n} - \boldsymbol{\vartheta}_n \Big) \Big| > N \right) \leq \sum_{N \in \mathbb{N}} \mathbb{P} \left(\Big| \sum_{n=1}^{N} f_{k_n} \Big| > \frac{N}{3} \right) + \sum_{N \in \mathbb{N}} \mathbb{P}\big(\Phi_N\big) + \sum_{N \in \mathbb{N}} \mathbb{P}\big(\Theta_N\big).$$

As for (8.4), this follows from (8.1), the bound $\big| \boldsymbol{\vartheta}_n \big| \leq 2^{-n}$, and the triangle inequality. ☐

This martingale methodology establishes, under (2.2), the convergence $\sum_{n \in \mathbb{N}} b_n \big(f_{k_n} - f_\infty \big) < \infty$, $\mathbb{P}$−a.e., of Révész [35], and hereditarily, for any sequence of real numbers $b_1, b_2, \cdots$ that satisfy $\sum_{n \in \mathbb{N}} b_n^2 < \infty$ (cf. [12]); then the choice $b_n = 1/n$ leads to (1.9), via the Kronecker Lemma.

9 Appendix: Quantitative Results of Hsu-Robbins-Erdős Type

The proof of Theorem 2.2 uses the following uniform sharpening of the Hsu-Robbins-Erdős theorem. This extends a result of Heyde [22], on which the characterization (1.7) is based.

Proposition 9.1. *Let* $f_1, f_2, \cdots$ *be a sequence of independent copies of a random variable* $f \in \mathbb{L}^1$ *with* $\mathbb{E}(f) = 0$, $\sigma^2 := \mathbb{E}(f^2) \leq \infty$. *For some constants* $0 < c < C_1 < \infty$, $0 < C_2 < \infty$ *which are* universal (*i.e., do not depend on the distribution of* f), *we have then the double inequality*

$$c \cdot \sigma^2 \leq \sum_{N \in \mathbb{N}} \mathbb{P}\Big(\Big| \sum_{n=1}^{N} f_n \Big| > N \Big) + 1 \leq C_1 \cdot \sigma^2 + C_2, \tag{9.1}$$

as well as its scaled version

$$c \cdot \sigma^2 \leq \varepsilon^2 \Bigg(\sum_{N \in \mathbb{N}} \mathbb{P}\Big(\Big| \sum_{n=1}^{N} f_n \Big| > \varepsilon N \Big) + 1 \Bigg) \leq C_1 \cdot \sigma^2 + C_2 \cdot \varepsilon^2, \qquad \forall\, \varepsilon \in (0,1]. \tag{9.2}$$

Proof: The proof of the upper bound in (9.1) follows from the inequality (47) in Fuk-Nagaev [20] (see also [22], p. 175), which gives

$$\mathbb{P}\Big(\Big| \sum_{n=1}^{N} f_n \Big| > N \Big) \leq N\, \mathbb{P}\Big(|f| > \frac{N}{4} \Big) + \frac{128\,\big(1 + 2e^4\big)}{N^2}, \qquad \forall\;\; N \in \mathbb{N}; \tag{9.3}$$

adding over $N \in \mathbb{N}$, we obtain the upper bound in (9.1) for some universal constants $C_1 > 0$, $C_2 > 0$. The lower bound in (9.1) follows from the next result; the scaled version (9.2) is obvious. ☐

Lemma 9.2. *There is a universal constant* $c \in (0, \infty)$ *such that, for independent copies* $f_1, f_2, \cdots$ *of a random variable* f *with arbitrary distribution which is symmetric* (*i.e.,* $-f$ *has the same distribution as* f), *and with* $\sigma^2 := \mathbb{E}(f^2) \leq \infty$, *we have*

$$\sum_{N \in \mathbb{N}} \mathbb{P}\Big(\sum_{n=1}^{N} f_n > 2\,N \Big) + 1 \geq c \cdot \sigma^2. \tag{9.4}$$

Whereas, if we drop the symmetry assumption on f, we still have for a (possibly different) universal constant $c \in (0,\infty)$, the bound

$$\sum_{N\in\mathbb{N}} \mathbb{P}\Big(\Big| \sum_{n=1}^{N} f_n \Big| > N \Big) + 1 \,\geq\, c \cdot \sigma^2 \,. \tag{9.5}$$

Proof: **(i)** We argue first (9.4), which pertains to the symmetric case. By analogy with [18], p. 289, and using the symmetry assumption, we obtain for fixed $N \in \mathbb{N}$ the inequalities

$$\mathbb{P}\Big(\sum_{n=1}^{N} f_n > 2\,N \Big) \,\geq\, \mathbb{P}\Big[\bigcup_{n=1}^{N} \Big(\{ f_n > 2\,N \} \cap \Big\{ \sum_{\substack{k=1\\ k\neq n}}^{N} f_k \geq 0 \Big\} \Big) \Big]$$

$$\geq \sum_{n=1}^{N} \Big[\mathbb{P}\Big(f_n > 2\,N\,,\, \sum_{\substack{k=1\\ k\neq n}}^{N} f_k \geq 0 \Big) - \mathbb{P}\Big(\bigcup_{\substack{k=1\\ k<n}}^{N} \{ f_n > 2\,N \} \cap \{ f_k > 2\,N \} \Big) \Big]$$

$$\geq \sum_{n=1}^{N} \Big(\frac{1}{2}\, \mathbb{P}\big(f > 2\,N\big) - N \cdot \Big(\mathbb{P}\big(f > 2\,N\big) \Big)^2 \Big)^{+} . \tag{9.6}$$

We distinguish at this point two cases:
(a) If $\mathbb{E}(|f|) < \infty$, we have $\lim_{N\to\infty} \big(N \cdot \mathbb{P}(|f| > N) \big) = 0$ and thus $\mathbb{P}\big(f > 2\,N\big) \leq 1/(4N)$ for all $N \geq N_0$ with N_0 sufficiently large; then (9.6) gives $\mathbb{P}\big(\sum_{n=1}^{N} f_n > 2\,N \big) \,\geq\, (N/4) \cdot \mathbb{P}\big(f > 2\,N\big)$, thus also, for some universal real constant $K > 0$, the bound

$$\mathbb{E}\big(f^2\big) \,\leq\, K \cdot \sum_{N\in\mathbb{N}} \frac{N}{4} \cdot \mathbb{P}\big(f > 2\,N\big) \leq K \cdot \sum_{N\in\mathbb{N}} \mathbb{P}\Big(\sum_{n=1}^{N} f_n > 2\,N \Big) .$$

(b) On the other hand, if $\mathbb{E}(|f|) = \infty$, the HRE property fails for the sequence $f_1, f_2, \cdots$, i.e., we have $\mathbb{E}(f^2) = \infty = \mathbb{P}\big(\sum_{n=1}^{N} f_n > 2\,N \big)$; recall Remark 6.1.
. In either case, $\sigma^2 \,\leq\, K \cdot \sum_{N\in\mathbb{N}} \mathbb{P}\big(\big| \sum_{n=1}^{N} f_n \big| \,>\, 2\,N \big)$ holds for $\sigma^2 \,=\, \mathbb{E}\big(f^2\big) \,\geq\, 1$ and some universal real constant $K \,>\, 0$; so (9.4) holds with $c \,=\, 1/K$ after noting that, for $\sigma^2 \,<\, 1$, it is satisfied trivially. We observe also that this proof gives, possibly with a different universal constant $c > 0$, the inequality

$$\sum_{N\in\mathbb{N}} \mathbb{P}\Big(\sum_{n=1}^{N} f_n > 2\,N \Big) \,\geq\, c \cdot \sigma^2 \,. \tag{9.7}$$

(ii) For the general, non-symmetric case, we argue as follows. We let $\big(f_n^{\pm}\big)_{n\in\mathbb{N}}$ be independent copies of the sequence $\big(f_n\big)_{n\in\mathbb{N}}$ and set $g_n := f_n^{+} - f_n^{-}$, $n \in \mathbb{N}$; these $\big(g_n\big)_{n\in\mathbb{N}}$ are independent, identically distributed with variance $2\,\sigma^2 = 2\,\mathbb{E}\big(f^2\big)$, and *symmetric,* so we obtain from (9.7) the bound

$$\sum_{N\in\mathbb{N}} \mathbb{P}\Big(\sum_{n=1}^{N} g_n > 2\,N \Big) \,\geq\, c \cdot \mathbb{E}\big(g_1^2\big). \tag{9.8}$$

Arguing as in [18], we observe $\big\{ \sum_{n=1}^{N} g_n \,>\, 2\,N \big\} \,\subseteq\, \big\{ \big| \sum_{n=1}^{N} f_n^{+} \big| \,>\, N \big\} \cup \big\{ \big| \sum_{n=1}^{N} f_n^{-} \big| \,>\, N \big\}$, therefore also $\mathbb{P}\big(\sum_{n=1}^{N} g_n > 2\,N \big) \,\leq\, 2 \cdot \mathbb{P}\big(\big| \sum_{n=1}^{N} f_n \big| > N \big)$; consequently, for some new universal constant $c > 0$, we have in conjunction with (9.7) also the bound

$$\sum_{N\in\mathbb{N}} \mathbb{P}\Big(\Big| \sum_{n=1}^{N} f_n \Big| > N \Big) \,\geq\, (1/2) \,\cdot \sum_{N\in\mathbb{N}} \mathbb{P}\Big(\sum_{n=1}^{N} g_n > 2\,N \Big) \,\geq\, c \cdot \mathbb{E}\big(f^2\big) \,. \qquad \square$$

References

[1] Aldous, D. (1977) Limit theorems for subsequences of arbitrarily-dependent sequences of random variables. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* **40**, 59-82.

[2] Aldous, D. (1977) Subsequences of sequences of random variables. *Bulletin of the American Mathematical Society* **83**, 121-123.

[3] Aldous, D. and Eagleson, G.K. (1978) On mixing and stability of limit theorems. *Annals of Probability* **6**, 325-331.

[4] Baxter, J.R. and Chacon, R.V. (1977) Compactness of stopping times. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* **40**, 169-182.

[5] Benoîst, Y. and Quint, J.F. (2016) Central limit theorem for linear groups. *Annals of Probability* **44**, 1308-1340.

[6] Berkes, I. (1990) An extension of the Komlós subsequence theorem. *Acta Mathematica Hungarica* **55**, 103-110.

[7] Berkes, I. and Péter, E. (1986) Exchangeable random variables and the subsequence principle. *Probability Theory & Related Fields* **73**, 395-413.

[8] Berkes, I. and Philipp, W. (1979) Approximation theorems for independent and weakly dependent random vectors. *Annals of Probability* **7**, 29-54.

[9] Berkes, I. and Rosenthal, H.P. (1985) Almost exchangeable sequences of random variables. *Probability Theory & Related Fields* **70**, 473-507.

[10] Berkes, I. and Tichy, R. (2015) On permutation-invariance of limit theorems. *Journal of Complexity* **31**, 372-379.

[11] Cembranos, P. and Mendoza, J. (1997) Banach Spaces of Vector-Valued Functions. *Lecture Notes in Mathematics* **1676**. Springer-Verlag, Berlin.

[12] Chatterji, S.D. (1973) Les Martingales et Leurs Applications Analytiques. *Lecture Notes in Mathematics* **304**, 27-164. Springer-Verlag, Berlin.

[13] Chatterji, S.D. (1974) A principle of subsequences in probability. *Advances in Mathematics* **33**, 31-54.

[14] Chatterji, S.D. (1985) A subsequence principle in probability theory. *Jahresbericht der Deutschen Mathematiker-Vereiningung* **87**, 91-107.

[15] Dacunha-Castelle, D. (1975) Indiscernibility and exchangeability in $\mathbb{L}^p-$spaces. *Proceedings of the Seminar on Random Series, Convex Sets, and Geometry of Banach Spaces,* Aarhus Universitet, Matematisk Institut Various Publ. **25**, 50-56.

[16] Dellacherie, C. and Meyer, P.A. (1978) *Probabilities and Potential. Part B: Theory of Martingales.* North-Holland Mathematics Studies, Volume 29. North-Holland Publishing Co., Amsterdam and New York.

[17] Durrett, R. (2010) *Probability: Theory and Examples.* Cambridge University Press.

[18] Erdős, P. (1949) On a theorem of Hsu and Robbins. *Annals of Mathematical Statistics* **20**, 286-291.

[19] Erdős, P. (1950) Remark on my paper "On a theorem of Hsu and Robbins". *Annals of Mathematical Statistics* **21**, 138.

[20] Fuk, D.K. and Nagaev, S.V. (1971) Probability inequalities for sums of independent random variables. *Theory of Probability & Its Applications* **16**, 643-660.

[21] Gaposhkin, V.F. (1972) Convergence and limit theorems for sequences of random variables. *Theory of Probability & Its Applications* **17**, 379-400.

[22] Heyde, C.C. (1975) A supplement to the strong law of large numbers. *Journal of Applied Probability* **12**, 173-175.

[23] Hsu, P.L. and Robbins, H.E. (1947) Complete convergence and the law of large numbers. *Proceedings of the National Academy of Sciences, USA* **33**, 25-31.

[24] Kadeć, M.I. and Pełzyński, A. (1962) Bases, lacunary sequences, and complemented subspaces, in the spaces $\mathbb{L}^p$. *Studia Mathematica* **21**, 161-176.

[25] Karatzas, I. and Kardaras, C. (2021) *Portfolio Theory & Arbitrage.* Graduate Studies in Mathematics, Volume **214**. American Mathematical Society, Providence, RI.

[26] Karatzas, I. and Schachermayer, W. (2023) A weak law of large numbers for dependent random variables. *Theory of Probability & Its Applications* **48**, 501-509.

[27] Kingman, J.F.C. (1978) Uses of exchangeability. *Annals of Probability* **6**, 183-197.

[28] Kolmogorov, A.N. (1930) Sur la loi forte des grandes nombres. *Comptes Rendus de l' Académie des Sciences, Paris* **191**, 910-912.

[29] Kolmogorov, A.N. (1933) Grundbegriffe der Wahrscheinlichkeitsrechnung. *Ergebnisse der Mathematik* **2** (3), Springer-Verlag, Berlin. English Translation published as *"Foundations of Probability Theory",* Chelsea Publishing Co., New York, 1950.

[30] Komlós, J. (1967) A generalization of a problem of Steinhaus. *Acta Math. Acad. Sci. Hung.* **18**, 217-229.

[31] Lennard, C.J. (1993) A converse to a theorem of Komlós for convex subsets of $\mathbb{L}^1$. *Pacific Journal of Mathematics* **159**, 75–85.

[32] Lesigne, E. and Volný, D. (2001) Large deviations for martingales. *Stochastic Processes & Their Applications* **96**, 143-159.

[33] Meyer, P.A. (1966) *Probability and Potentials.* Blaisdell Publishing Company, Waltham, Mass.

[34] Rényi, A. (1965) On stable sequences of events. *Sankhyā, Ser. A* **16**, 310-318.

[35] Révész, P. (1965) On a problem of Steinhaus. *Acta Math. Acad. Sci. Hung.* **16**, 310-318.

[36] Rosenthal, H.P. (1973) On subspaces of $\mathbb{L}^p$. *Annals of Mathematics* **97**, 344-373.

[37] Villani, C. (2003) *Topics in Optimal Transportation.* Graduate Studies in Mathematics, Volume **58**. American Mathematical Society, Providence, RI.