ALEA, Lat. Am. J. Probab. Math. Stat. **23**, 105–141 (2026)
DOI: 10.30757/ALEA.v23-06

On the tradeoff between almost sure error tolerance and its mean deviation frequency in martingale convergence

Luisa F. Estrada, Michael A. Högele and Alexander Steinicke

University of Warwick, Department of Mathematics, CV4 7AL, Coventry, UK.
E-mail address: luisa-fernanda.estrada-plata@warwick.ac.uk
URL: https://orcid.org/0009-0003-2126-331X

Universidad de los Andes, Departamento de Matemáticas, Cra. 1 #18a-12, 111711 Bogotá, Colombia.
E-mail address: ma.hoegele@uniandes.edu.co
URL: https://orcid.org/0000-0001-5744-0494

Technical University of Leoben, Chair of Applied Mathematics, Peter-Tunner-Str. 25/I, 8700 Leoben, Austria.
E-mail address: alexander.steinicke@unileoben.ac.at
URL: https://orcid.org/0000-0001-6330-0295

Abstract. In this article we quantify almost sure martingale convergence theorems in terms of the tradeoff between asymptotic almost sure rates of convergence (error tolerance) and the respective modulus of convergence. For this purpose we generalize an elementary quantitative version of the first Borel-Cantelli lemma on the statistics of the deviation frequencies, which was recently established by the authors. First we study martingale convergence in L2, and in the setting of the Azuma-Hoeffding inequality. In a second step we study the strong law of large numbers for martingale differences. Applications are the tradeoff for the multicolor generalized Pólya urn processes, the generalized Chinese restaurant process, statistical M-estimators, as well as excursion frequencies of the Galton-Watson branching process.

1. Introduction

The notion of almost sure (a.s.) convergence of a sequence of random variables $(X_n)_{n\in\mathbb{N}}$ to a random variable X as $n \to \infty$, is certainly one of the most natural concepts in probability and statistics in the assessment of the evolution of observed data. This type of convergence is intuitive to grasp due to its similarity to the pointwise converence of deterministic functions. However, we are not aware of a satisfactory quantification in the literature, since the modulus of convergence $\mathbb{m}_\varepsilon$, that is the *last index* $\mathbb{m}_\varepsilon \in \mathbb{N}$, when a given error threshold $\varepsilon > 0$ is broken in the sense of $|X_{\mathbb{m}_\varepsilon} - X| > \varepsilon$, is inherently random and seemingly not easily accessible.

Received by the editors January 11th, 2024; accepted November 12th, 2025; published under license CC BY 4.0.
2010 *Mathematics Subject Classification.* 60E15; 60F10; 60F15; 60G42; 60J80; 62F05.
Key words and phrases. Vanilla Azuma inequality; Azuma-Hoeffding inequality; Martingales inequality; almost sure martingale converence; Chinese Restaurant process; Pólya's urn; M-estimators; SLLN for martingales; excursion dynamics of the Galton-Watson branching process; Baum-Katz-Nagaev weak laws of large numbers.

In many situations, almost sure convergence is established by an application of the first Borel-Cantelli lemma (Billingsley, 1999; Borel, 1909; Cantelli, 1917; Chandra, 2012; Chung and Erdös, 1952; Hill, 1983; Shiryaev, 1996) to the sequence of the error events $A_n(\varepsilon) = \{|X_n - X| > \varepsilon\}$ for any $\varepsilon > 0$, $n \in \mathbb{N}$, and $n \geqslant n_0$ for some fixed $n_0 \in \mathbb{N}$. For an overview of the literature we refer to the introduction of Estrada and Högele (2022). Classical examples of this proof technique are Etemadi's strong law of large numbers (Etemadi, 1981), Lévy's construction of Brownian motion, the Kolmogorov-Chentsov theorem, and the law of the iterated logarithm. See Högele and Steinicke (2023) for more examples in the context of Brownian path property approximation. This particular notion of a.s. convergence stemming from the first Borel-Cantelli lemma is well-established in the literature as *complete convergence* (Hsu and Robbins, 1947; Ma and Sun, 2018; Yukich, 1999): A sequence of random variables $(X_n)_{n\geq 0}$ converges completely to a random variable X, if for all $\varepsilon > 0$ we have $\sum_{n=0}^{\infty} \mathbb{P}(A_n(\varepsilon)) = \sum_{n=0}^{\infty} \mathbb{P}(\{|X_n - X| > \varepsilon\}) < \infty$. We generalize this notion in the spirit of Estrada and Högele (2022) with the help of the following refined first Borel-Cantelli lemma: Recall that the classical first Borel-Cantelli lemma can be formulated as follows: On a given probability space $(\Omega, \mathcal{A}, \mathbb{P})$, the summability of the sequence of the probabilities of the events $(A_n)_{n\geqslant n_0}$ implies that the overlap statistic $\mathcal{O} := \sum_{n=n_0}^{\infty} \mathbf{1}(A_n)$ is finite with probability 1. The result $\mathcal{O} < \infty$ a.s. with its elegant one-line proof, however, is suboptimal since by monotone convergence we even know the average size of $\mathcal{O}$

$$\mathbb{E}[\mathcal{O}] = \sum_{n=n_0}^{\infty} \mathbb{P}(A_n), \tag{1.1}$$

which is finite by hypothesis. Moreover, the law of the random variable $\mathcal{O}$ has been known for a long time by the Schuette-Nesbitt formula (Gerber, 1979). Not surprisingly, the value $\mathbb{P}(\mathcal{O} = k)$ is given by means of an inclusion-exclusion principle as the sum of the probabilities of all the intersections of exactly k events of the sequence $(A_n)_{n\geqslant n_0}$. Unfortunately, the complete sequence of all such probabilities of event intersections is hardly ever available in applications (for the case of independent events we refer to Estrada and Högele (2022, Subsection 2.2, Theorem 3)). On the other hand, the (top level) null sequence $(\mathbb{P}(A_n))_{n\geqslant n_0}$ is often well-known and turns out to tend to 0 faster than just strictly necessary to be summable. In many situations, for instance in the presence of a large deviations principle, it is of exponential order of decay. It is natural to translate this structural surplus into the finiteness of higher moments of $\mathcal{O}$ and the tail asymptotics $\mathbb{P}(\mathcal{O} \geqslant k)$ as $k \to \infty$. In Estrada and Högele (2022, Theorem 1) it is shown for $n_0 = 1$ that for a sequence of positive, nondecreasing weights $(a_n)_{n\geqslant n_0}$ certain nonlinear higher moments of $\mathcal{O}$ (depending on the sequence $a = (a_n)_{n\geqslant n_0}$) can be bounded by the weighted sum

$$C_a := \sum_{n=n_0}^{\infty} a_n \sum_{m=n}^{\infty} \mathbb{P}(A_m), \tag{1.2}$$

whenever the preceding series converges. We show a slight generalization of this result, which turns out to be useful in applications. We illustrate the novelty of our results by the following example. Think of a Cramérs type estimate

$$\mathbb{P}(|\bar{X}_n - \mathbb{E}[X_1]| > \varepsilon_n) \leqslant 2\exp\big(-\tfrac{1}{2}n\varepsilon_n^2\big), \qquad n \in \mathbb{N},$$

for the law of large numbers with i.i.d. summands X_i with some finite exponential moment. While Estrada and Högele (2022) treats the case of constant ε we observe the following. The essentially optimal rates $\varepsilon_n = \sqrt{\alpha \ln(n)/n}$, $\alpha > 2$, yield for α close to 2 barely summable probabilities. This implies by (1.1) that $\mathbb{E}[\mathcal{O}] < \infty$ and therefore, by Markov's inequality, $\mathbb{P}(\mathcal{O} \geqslant \ell) \leqslant \mathbb{E}[\mathcal{O}]/\ell$. However, if we consider the slightly suboptimal rate $\tilde{\varepsilon}_n = n^{-1/3} > \varepsilon_n$, we obtain the by far better rate $\mathbb{P}(|\bar{X}_n - \mathbb{E}[X_1]| > \tilde{\varepsilon}_n) \leqslant 2\exp\big(-\frac{1}{2}n^{1/3}\big)$. Further, we get $\mathbb{E}[\exp(p\mathcal{O}^{1/3})] < \infty$ for any $p \in (0,1)$ (see Example 3) and hence the much faster observation $\mathbb{P}(\mathcal{O} \geqslant \ell) \leqslant \mathbb{E}[\exp(p\mathcal{O}^{1/3})]/\exp\big(p\ell^{1/3}\big)$ which then can still be minimized over all $p \in (0,1)$. More useful still, our results including all upper bounds are valid not only for $\mathcal{O}$, that is the *number* of error event indices, but also for the *last index* $\mathbb{m}$ (defined in (2.2)) where an error event occurs. In a word, there is often a tradeoff in the sense that relaxing the optimal a.s. rate of convergence to a slightly worse one, we often "speed up" its emergence substantially.

Our quantitative Borel-Cantelli result allows for the solution of problem (a) for the special sequence of events $(A_n(\varepsilon_n))_{n\in\mathbb{N}}$ defined above. More precisely, we study the relation between a given positive null sequence $\epsilon := (\varepsilon_n)_{n\in\mathbb{N}}$, called *error tolerance*, and the higher order integrability of $\mathcal{O}_\epsilon := \sum_{n=n_0}^\infty \mathbf{1}(A_n(\varepsilon_n))$, called the *error incidence* or *deviation frequency* or *overlap count*, which generalizes formula (1.1). That is to say, $\mathcal{O}_{\epsilon,n_0} = |\{n \geqslant n_0 \ : \ |X_n - X| > \varepsilon_n\}|$ and $\mathbb{m}_{\epsilon,n_o} = \max\{n - n_0 \geqslant 0 \ : \ |X_n - X| > \varepsilon_n\}$. The quantification of the a.s. convergence $X_n \to X$ relies in the finiteness of higher moments of $\mathcal{O}_{\epsilon,n_0}$ ("how many errors occur before dying out") and $\mathbb{m}_{\epsilon,n_0}$ ("at which position happens the last error"). The type of moments that we consider is specified in Lemma 1 in Section 2, a key result for the rest of the article. It states the following. Given events $A_n = A_n(\varepsilon_n)$ and a chosen sequence a such that C_a in (1.2) is finite, then

- for the a.s. asymptotic upper error rate, we have

$$\limsup_{n\to\infty} |X_n - X| \cdot \varepsilon_n^{-1} \leqslant 1 \qquad \mathbb{P}\text{-a.s.} \tag{1.3}$$

- Further, for the respective mean deviation frequency (MDF) quantification we have

$$\mathbb{E}[\mathcal{S}_{a,n_0}(\mathcal{O}_\epsilon)] \leqslant \mathbb{E}[\mathcal{S}_{a,n_0}(\mathbb{m}_\epsilon)] \leqslant C_a, \quad \text{where} \quad \mathcal{S}_{a,n_0}(N) := \sum_{n=0}^{N-1} a_{n_0+n}, \quad N \in \mathbb{N}, \tag{1.4}$$

  with the convention $\mathcal{S}_{a,n_0}(0) = 0$.

A choice for the sequence $(a_n)_{n\geq n_0}$ that will appear often is a power sequence $a_n = n^p$ for some $p > 0$. Then $\mathcal{S}_{a,n_0}(N)$ grows polynomially in N with degree $p+1$. It will be used to estimate moments such as $\mathbb{E}[\mathbb{m}_\epsilon^{p+1}]$. Another choice are exponential sequences $a_n = e^{\alpha n}$ for some $\alpha > 0$. Then also $\mathcal{S}_{a,n_0}(N)$ grows exponentially in N. We use it to bound exponential moments of $\mathbb{m}_\epsilon$. Note further that (1.4) implies that for any $k \geqslant 1$

$$\mathbb{P}(\mathcal{O}_\epsilon \geqslant k) \leqslant \mathbb{P}(\mathbb{m}_\epsilon \geqslant k) \leqslant \inf_a C_a \cdot (\mathcal{S}_{a,n_0}(k))^{-1},$$

where the infimum is taken over some meaningful subset of positive sequences of weights $(a_n)_{n\in\mathbb{N}}$ such that $C_a < \infty$. Particular cases of such quantifications can be found in Högele and Steinicke (2023) in the context of Brownian sample path approximations. This result has two main benefits:

(a) The tradeoff relation between $\epsilon = (\varepsilon_n)_{n\in\mathbb{N}}$ and $\mathbb{P}(\mathbb{m}_\epsilon \geqslant k)$ for $|X_n - X| \to 0$ a.s. is completely intuitive and analogous to the convergence in any metric space. It can be described informally as follows: The <u>faster</u> $\varepsilon_n \searrow 0$, as $n \to \infty$, the <u>higher</u> the last index at which $|X_n - X| > \varepsilon_n$. Consequently, we have larger values of $\mathbb{m}_\epsilon$ and less integrability and a <u>slower</u> decay of $\mathbb{P}(\mathbb{m}_\epsilon \geqslant k)$ as $k \to \infty$. Conversely, the slower $\varepsilon_n \searrow 0$, as $n \to \infty$, the lower the number of deviations and the smaller $\mathbb{m}_\epsilon$. The same mechanism is valid for $\mathcal{O}_\epsilon$.

(b) The relation (1.4) bounds *nonlinear* higher order moments of $\mathcal{O}_\epsilon = \sum_{n=n_0}^\infty \mathbf{1}(A_n(\varepsilon_n))$ and $\mathbb{m}_\epsilon$ by the constant C_a, whose finiteness is an elementary, weighted, *linear* condition (1.2) on $(\mathbb{P}(A_n(\varepsilon_n)))_{n\in\mathbb{N}}$. We refer to Example 1, 2 and 3. Condition (1.2) is easy to verify and therefore allows for the retroactive and meaningful quantification of many known results of complete convergence (or even only a sufficiently strong convergence in probability) in the literature. A sample of applications (still for fixed $\varepsilon > 0$) is given in Estrada and Högele (2022). This article shows the utility of such a concept for almost sure martingale convergence and strong laws for martingale differences more generally for nonincreasing sequences $\epsilon = (\varepsilon_n)_{n\in\mathbb{N}}$.

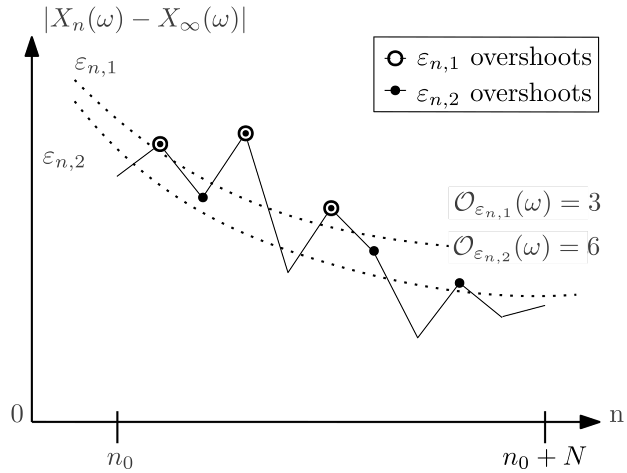

FIGURE 1.1. Schematic of the error $|X_n(\omega) - X(\omega)|$ along the time index n. A larger error tolerance $\varepsilon_{n,1}$ yields a smaller number of overshoots $\mathcal{O}_{\varepsilon_{n,1}}(\omega)$, vice versa, a smaller error tolerance $\varepsilon_{n,2}$ yields a larger number of overshoots $\mathcal{O}_{\varepsilon_{n,2}}(\omega)$.

Note that the estimates obtained through Lemma 1 are necessarily suboptimal, however, not by much. The reason is, that our quantitative version of the first Borel-Cantelli lemma uses in a crucial step a suboptimal union bound. This union bound encodes the possible sparseness of error indices before finally dying out. However, we show that this effect only affects the integrability of $\mathcal{O}_\epsilon$ and $\mathbb{m}_\epsilon$ for rates of $\mathbb{P}(A_n(\varepsilon_n))$ given by inverse monomials with small exponents, see Example 1. For high order polynomially, exponentially or Weibull-type fast rates $\mathbb{P}(A_n(\varepsilon_n)) \searrow 0$, as $n \to \infty$, this effect is essentially negligible (see Example 2 and 3).

We highlight the utility of the previously mentioned tradeoff between error tolerance and deviation frequency in the context of martingale convergence theorems and the strong laws for martingale differences. There is a large literature on discrete martingales, which we cannot review here. The concept of martingale differences first emerged in Lévy's monography (Lévy, 1937) as a technical device to relax the independence in the central limit theorem even before the term martingale was coined and conceptualized by Ville (1939) in the context of fair games and still formulated in the controversial language of von Mises' collectives (Mazliak, 2009, Section 1.3). We refer to the classical monographs (Doob, 1953; Föllmer and Schied, 2025; Protter, 1990; Williams, 1991) for an introduction to discrete martingales. Nowadays, martingales are at the core of many applications.

First we study martingales which are uniformly bounded in L^p, $p \geqslant 2$, and with a.s. uniformly bounded increments with the help of the Azuma-Hoeffding inequality. Next we establish the strong law for martingale differences, for the cases where: they are are not necessarily bounded in L^p ; they are uniformly bounded in L^p; and when they have uniformly bounded exponential moments. Nowadays, there are many very fine martingale estimates in probability well-established, for an overview see Fan et al. (2015). Many of them are suitable for a run-off between the almost sure error tolerance (1.3) and the mean deviation frequency in (1.4). The preceding tradeoff is applied in four major applications: 1) (multicolor) Pólya's urns with applications, 2) the Generalized Chinese Restaurant Process with applications in machine learning, 3) a quantification of the a.s. convergence of statistical M-estimators and 4) the number of outliers for the Galton branching processes.

Organization of the article

We start in Section 2 with the proof of a quantitative version of the Borel-Cantelli lemma in Lemma 1 and the tradeoff between (1.3) and (1.4) in Lemma 2. In Section 3 we study martingale convergence theorems. First we quantify the Pythagorean theorem of martingale convergence in L^2 in Subsection 3.1, in Subsection 3.2 we quantify the Azuma-Hoeffding exponential closure and its MDF consequences. Section 4 starts with an a.s. MDF convergence result with the strong law of large numbers for not necessarily bounded data in L^p. For bounded data in L^p we use the optimal Baum-Katz-Nagaev type results in Subsection 4.2. Finally, Subsection 4.3 treats the strong law for martingale differences which have uniformly bounded exponential moments. In Section 5 we present several applications. Subsection 5.1 is dedicated to the assessment of the a.s. convergence of multicolor Pólya urn models. Subsection 5.2 illustrates the convergence of a Generalized Chinese Restaurant Process. In Subsection 5.3 we establish the statistical convergence results on M-estimators. Finally, Subsection 5.4 is dedicated to the MDF quantification of the convergence of the martingales associated to the Galton-Watson branching process.

Preliminaries and notation

In this article the natural numbers $\mathbb{N} = \{1, 2, \ldots, \}$ do not contain 0, while $\mathbb{N}_0 = \{0, 1, 2, \ldots\}$. Throughout this article all random vectors are defined over a common given probability space $(\Omega, \mathcal{A}, \mathbb{P})$. A filtered probability space is a probability space $(\Omega, \mathcal{A}, \mathbb{P}, \mathbb{F})$ equipped with a filtration $\mathbb{F} = (\mathcal{F}_n)_{n\in\mathbb{N}_0}$ that is a sequence of sub σ-algebras $\mathcal{F}_n \subseteq \mathcal{A}$ which satisfy $\mathcal{F}_n \subseteq \mathcal{F}_{n+1}$ for all $n \in \mathbb{N}_0$. We use the convention that for sums $\sum_{n=n_0}^{n_0+N-1} a_n$ for some $n_0, N \in \mathbb{N}_0$ and a real sequence $(a_n)_{n\in\mathbb{N}_0}$, the value $\sum_{n=n_0}^{n_0-1} a_n$ is 0.

In this article, all appearing Polish spaces $\mathcal{X}$ are considered to be equipped with their respective Borel σ-algebra, that is, the σ-algebra generated by the open sets. In case of a separable Banach space $(B, \|\cdot\|)$ equipped with its Borel-sigma-algebra $\mathcal{B}$, we recall the definition of a martingale (and the one of a martingale difference sequence) with values in B:

(a) A stochastic process $(X_n)_{n\in\mathbb{N}_0}$ on a given filtered probability space $(\Omega, \mathcal{A}, \mathbb{P}, \mathbb{F})$ with values in B is called martingale with respect to $\mathbb{F}$ if it satisfies the following three conditions:
  - (i) $\mathbb{E}[\|X_n\|] < \infty$ for all $n \in \mathbb{N}_0$.
  - (ii) $(X_n)_{n\in\mathbb{N}_0}$ is $\mathbb{F}$-adapted, that is, X_n is $(\mathcal{F}_n, \mathcal{B})$-measurable for all $n \in \mathbb{N}_0$.
  - (iii) $\mathbb{E}[X_n \mid \mathcal{F}_{n-1}] = X_{n-1}$ $\mathbb{P}$-a.s. for all $n \in \mathbb{N}$.

(b) A stochastic process $(X_n)_{n\in\mathbb{N}_0}$ with values in B is called a sequence of martingale differences (MDs) with respect to $\mathbb{F}$ if it satisfies the following three conditions: items (i) and (ii) of (a) and

$$\mathbb{E}[X_n \mid \mathcal{F}_{n-1}] = 0 \qquad \mathbb{P}\text{-a.s. for all } n \in \mathbb{N}.$$

In Section 3, 4 and 5 we apply the results of Section 2 to several examples of martingales. Results for martingales with values in infinite dimensional spaces require the notion of p-smooth Banach spaces (following e.g. Luo (2022) or Pisier (1975)) which we state here in brevity: A Banach space is called *p-uniformly smooth* for a fixed $p \in (1, 2]$ if there is a constant $s \geq 0$ such that for all $\tau > 0$,

$$\sup\left\{\tfrac{\|x+\tau y\|+\|x-\tau y\|}{2} - 1 : \|x\| = \|y\| = 1\right\} \leq s\tau^p.$$

Note that all Hilbert spaces are 2-uniformly smooth (by the parallelogram identity). Most of our results for martingales in infinite dimensions rely on concentration equalities for Banach spaces. Our choices of such inequalities (a variety of the Azuma inequalities from Luo (2022) and Baum-Katz type-estimates (Giraudo, 2019)) can of course be extended, e.g. using the findings in Naor (2012); Pinelis (1994) or Pisier (1975).

2. A quantitative version of the first Borel-Cantelli lemma

We start by extending the result given in Estrada and Högele (2022, Theorem 1).

Definition 1. Let $(A_n)_{n\in\mathbb{N}_0}$ be a sequence of events in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. For $n_0 \in \mathbb{N}_0$ we call

$$\mathcal{O}_{n_0}(\omega) := \sum_{n=n_0}^{\infty} \mathbf{1}(A_n)(\omega), \qquad \omega \in \Omega, \tag{2.1}$$

the **overlap count** of $(A_n)_{n\in\mathbb{N}_0}$ and

$$\mathbb{m}_{n_0}(\omega) := \max\{i - n_0 \geqslant 0 \mid \omega \in A_i\} \tag{2.2}$$

the **last occurrence index** of $(A_n)_{n\in\mathbb{N}_0}$.
For a nonnegative, nondecreasing sequence $a = (a_n)_{n\in\mathbb{N}_0}$ we define

$$\mathcal{S}_{a,n_0}(N) := \sum_{n=0}^{N-1} a_{n_0+n} \qquad \text{for } N \in \mathbb{N} \qquad \text{and} \qquad \mathcal{S}_{a,n_0}(0) := 0. \quad [1] \tag{2.3}$$

The function $\mathcal{S}_{a,n_0}$ represents the order of the moments of $\mathcal{O}_{\epsilon,n_0}$ and $\mathbb{m}_{\epsilon,n_0}$. It is (due to summation by parts) the "antiderivate" of the sequence of "weights" $(a_n)_{n\in\mathbb{N}}$. The following lemma gives sufficient conditions on upper bounds of $\mathbb{E}[\mathcal{S}_{a,n_0}(\mathcal{O}_{\epsilon,n_0})$, and $\mathbb{E}[\mathcal{S}_{a,n_0}(\mathbb{m}_{\epsilon,n_0})$, respectively. The examples afterwards illustrate how these moments are upper bounds of polynomial, exponential or Weibull type moments in concrete situations.

Lemma 1 (Quantitative version of the first Borel-Cantelli lemma).
Given a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, $n_0 \in \mathbb{N}_0$, and a sequence of events $(A_n)_{n\geqslant n_0}$, such that

$$\sum_{n=n_0}^{\infty} \mathbb{P}(A_n) < \infty.$$

Then for any positive, nondecreasing sequence $(a_n)_{n\geqslant n_0}$, the following statements are true:

(a) *If the sequence $(A_n)_{n\geqslant n_0}$ is nested, that is, $A_{n+1} \subseteq A_n$, $n \geqslant n_0$, it follows that*

$$\mathbb{E}[\mathcal{S}_{a,n_0}(\mathcal{O}_{n_0})] = \sum_{n=n_0}^{\infty} a_n \mathbb{P}(A_n).$$

(b) *Consider a sequence $(A_n)_{n\geqslant n_0}$, which is not necessarily nested. Then the following relations are valid:*

i) *For all $\omega \in \Omega$ we have*

$$\mathbb{m}_{n_0}(\omega) = \sum_{n=n_0}^{\infty} \mathbf{1}\Big(\bigcup_{m=n}^{\infty} A_m\Big)(\omega). \tag{2.4}$$

ii) *We have the moment estimate*

$$\mathbb{E}[\mathcal{S}_{a,n_0}(\mathcal{O}_{n_0})] \leqslant \mathbb{E}[\mathcal{S}_{a,n_0}(\mathbb{m}_{n_0})] = \sum_{n=n_0}^{\infty} a_n \mathbb{P}\Big(\bigcup_{m=n}^{\infty} A_m\Big) \leqslant \sum_{n=n_0}^{\infty} a_n \sum_{m=n}^{\infty} \mathbb{P}(A_m) = K_a. \tag{2.5}$$

Remark 1.

(a) The nestedness hypothesis in item (a) in Lemma 1 only applies directly under particular circumstances, see for instance Corollary 2 item (b). However, we obtain an exact formula, whereas in the general case of item (b) we only obtain an upper bound. For $\mathcal{O}_{n_0}$ the difference between the nested case (a) and (b) lies in the replacement of the sequence $\mathbb{P}(A_n)$ by the sequence $\sum_{m=n}^{\infty} \mathbb{P}(A_n)$, which is clearly suboptimal, as can be seen in Example 1. However, in Example 2 and 3 below, we see that this gap in the order is often negligible.

[1] Note that this definition of $\mathcal{S}_a$ corrects an off-by-one error in Estrada and Högele (2022, Thm 1). Compare with Example 1 and Example 2 below.

(b) For a positive sequence of real numbers $a = (a_n)_{n\geqslant n_0}$ and $N \geqslant 0$, $n_0 \in \mathbb{N}$, we note that $\mathcal{S}_{a,n_0}(N) := \sum_{n=0}^{N-1} a_{n_0+n}$ is a 'discrete antiderivate' of a w.r.t. the counting measure. This function itself might seem a bit involved, however, it is often estimated from below without much effort. In order to obtain a lower bound of $\mathbb{E}[\mathcal{S}_{a,n_0}(\mathcal{O}_\epsilon)]$ in many cases the comparison principle for sums and (Riemann-)integrals are used.

(c) Since a is nondecreasing, the relation (2.5) implies that $\sum_{m=n}^{\infty} \mathbb{P}(A_m) < \infty$ such that the classical first Borel-Cantelli lemma applies. Note that Lemma 1 can only quantify the excess of summability in $(\mathbb{P}(A_n))_{n\geqslant n_0}$, it cannot turn non-summable sequences into summable ones.

(d) Note that the finiteness on the right-hand side in estimate (2.5) is a linear condition in a for a nonlinear higher moment of $\mathcal{O}_{n_0}$.

Proof of Lemma 1: We start with the proof of (a). Fix some $n_0, N \in \mathbb{N}$ and define $\mathcal{O}_{n_0,N} := \sum_{m=n_0}^{N+n_0} \mathbf{1}(A_m)$. Note that by construction

$$\mathcal{O}_{n_0,N} \in \{0, \ldots, N+1\}.$$

By the nestedness we have for each $k = 1, \ldots, N$ that

$$\mathbb{P}(\mathcal{O}_{n_0,N} = k) = \mathbb{P}(A_{n_0+k-1} \setminus A_{n_0+k}) = \mathbb{P}(A_{n_0+k-1}) - \mathbb{P}(A_{n_0+k}).$$

In addition, $\mathbb{P}(\mathcal{O}_{n_0,N} = 0) = \mathbb{P}(\Omega \setminus A_{n_0})$ and $\mathbb{P}(\mathcal{O}_{n_0,N} = N+1) = \mathbb{P}(A_{N+n_0})$, compare with Figure 2.2.

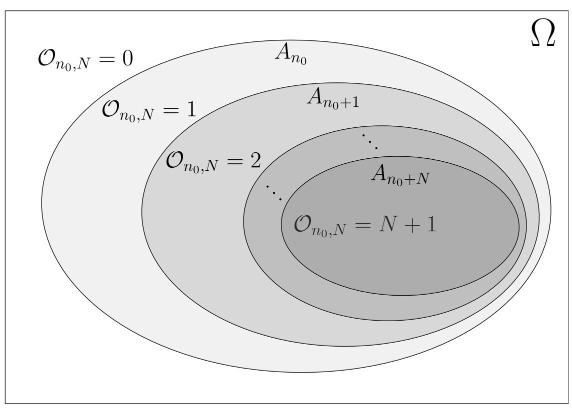

FIGURE 2.2. Overlap statistic $\mathcal{O}_{n_0,N}$ of the nested events $A_{n_0} \supseteq A_{n_0+1} \supseteq \cdots \supseteq A_{n_0+N}$

Note that by the definition of $\mathcal{O}_{n_0,N}$ we have the following representation

$$\begin{aligned}
&\mathbb{E}[\mathcal{S}_{a,n_0}(\mathcal{O}_{n_0,N})]\\
&\quad= \mathcal{S}_{a,n_0}(0)\mathbb{P}(\mathcal{O}_{n_0,N} = 0) + \sum_{k=1}^{N} \mathcal{S}_{a,n_0}(k)\mathbb{P}(\mathcal{O}_{n_0,N} = k) + \mathcal{S}_{a,n_0}(N+1)\mathbb{P}(\mathcal{O}_{n_0,N} = N+1)\\
&\quad= \mathcal{S}_{a,n_0}(0)\mathbb{P}(\Omega \setminus A_{n_0}) + \sum_{k=1}^{N} \mathcal{S}_{a,n_0}(k)\mathbb{P}(\mathcal{O}_{n_0,N} = k) + \mathcal{S}_{a,n_0}(N+1)\mathbb{P}(A_{N+n_0}).
\end{aligned}$$

Integration by parts yields for any sequences $(f_k)_{k\in\mathbb{N}_0}$ and $(g_k)_{k\in\mathbb{N}_0}$ that

$$\sum_{k=0}^{N} f_k g_k = f_N \sum_{k=0}^{N} g_k - \sum_{j=0}^{N-1} (f_{j+1} - f_j) \sum_{\ell=0}^{j} g_k.$$

For notational convenience we set $p_{n_0+k} = \mathbb{P}(A_{n_0+k})$. Hence for $f_k = p_{n_0+k}$ and $g_k = a_{n_0+k}$ we obtain

$$\sum_{k=0}^{N} a_{n_0+k} p_{n_0+k} = p_{n_0+N} \sum_{k=0}^{N} a_{n_0+k} + \sum_{j=0}^{N-1} (p_{n_0+j} - p_{n_0+j+1}) \sum_{\ell=0}^{j} a_{n_0+k}.$$

In other words, for all $N \in \mathbb{N}_0$ we have the formula

$$\begin{aligned}\sum_{k=0}^{N} a_{n_0+k}\mathbb{P}(A_{n_0+k}) &= \mathbb{P}(A_{n_0+N}) \sum_{k=0}^{N} a_{n_0+k} + \sum_{j=0}^{N-1} (\mathbb{P}(A_{n_0+j}) - \mathbb{P}(A_{n_0+j+1})) \sum_{\ell=0}^{j} a_{n_0+\ell} \\ &= \sum_{j=1}^{N+1} \Big(\sum_{\ell=0}^{j-1} a_{n_0+\ell}\Big)\mathbb{P}(\mathcal{O}_{n_0,N} = j) = \sum_{j=0}^{N+1} \mathcal{S}_{a,n_0}(j)\mathbb{P}(\mathcal{O}_{n_0,N} = j) \\ &= \mathbb{E}[\mathcal{S}_{a,n_0}(\mathcal{O}_{n_0,N})] - \mathcal{S}_{a,n_0}(N+1)\mathbb{P}(A_{N+n_0}),\end{aligned}$$

if and only if $\mathcal{S}_{a,n_0}(N) = \sum_{\ell=0}^{N-1} a_{n_0+\ell}$ with the convention that $\mathcal{S}_{a,n_0}(0) = 0$. Sending $N \to \infty$, the monotone convergence theorem implies

$$\mathbb{E}[\mathcal{S}_{a,n_0}(\mathcal{O}_{n_0})] = \sum_{k=0}^{\infty} a_{n_0+k}\mathbb{P}(A_{n_0+k}) = \sum_{\ell=n_0}^{\infty} a_\ell \cdot \mathbb{P}(A_\ell).$$

Here, Cesàro's lemma (Williams, 1991, 12.6) implies the error $\mathcal{S}_{a,n_0}(N+1)\mathbb{P}(A_{N+n_0})$ tends to 0 as $\mathbb{P}(A_{N+n_0}) \to 0$ for $N \to \infty$. This shows item (a).

We continue with item (b)(ii). Define $\tilde{A}_n := \bigcup_{m=n}^{\infty} A_m$ and $\tilde{O}_{n_0} = \sum_{n=n_0}^{\infty} \mathbf{1}(\tilde{A}_n)$. Note that the sequence $(\tilde{A}_n)_{n \geqslant n_0}$ is a nested sequence of events by construction. At the same time we have by construction the monotonicity $\mathcal{O}_{n_0} \leqslant \tilde{\mathcal{O}}_{n_0}$ a.s. and by the nonnegativity of the sequence $(a_n)_{n \geqslant n_0}$ that $\mathcal{S}_{a,n_0}$ is nondecreasing and

$$\mathbb{E}[\mathcal{S}_{a,n_0}(\mathcal{O}_{n_0})] \leqslant \mathbb{E}[\mathcal{S}_{a,n_0}(\tilde{\mathcal{O}}_{n_0})] = \sum_{n=n_0}^{\infty} a_n \mathbb{P}(\tilde{A}_n),$$

while by a union bound we have

$$\mathbb{E}[\mathcal{S}_{a,n_0}(\mathcal{O}_{n_0})] \leqslant \mathbb{E}[\mathcal{S}_{a,n_0}(\tilde{\mathcal{O}}_{n_0})] \leqslant \sum_{n=n_0}^{\infty} a_n \sum_{m=n}^{\infty} \mathbb{P}(A_m).$$

This shows item (b)(ii).

It remains to show (b)(i). By definition, $\mathbb{m}_{n_0}$ being the last index of the sets A_j to which ω belongs, we have that $\mathbb{m}_{n_0}(\omega) = i$ implies that $\omega \in A_{n_0+i}$ and $\omega \notin A_{n_0+j}$ for all $j \geqslant i+1$. In particular, $\mathbb{m}_{n_0}(\omega) = i$ yields

$$\omega \in \tilde{A}_i = A_i \cup \bigcup_{j \geqslant i+1}^{\infty} A_j \qquad \text{and} \qquad \omega \notin \tilde{A}_j = \bigcup_{\ell \geqslant j} A_\ell \quad \text{for all} \quad j \geqslant i+1.$$

In addition, $\omega \in \tilde{A}_k$ for $k \leqslant i$, since $\tilde{A}_k \supseteq \tilde{A}_i$ for all $k \leqslant i$ by construction. This implies $\tilde{\mathcal{O}}_{n_0}(\omega) = i$. Conversely, if we assume that $\tilde{\mathcal{O}}_{n_0}(\omega) = i$, then the nestedness of the sequence $(\tilde{A}_i)_{i \geqslant n_0}$ yields that $\omega \in \tilde{A}_{n_0+i} \setminus \tilde{A}_{n_0+i+1}$. By definition, this yields

$$\omega \in \Big(\bigcup_{\ell=i}^{\infty} A_\ell\Big) \setminus \Big(\bigcup_{k=i+1}^{\infty} A_k\Big) = A_i \setminus \Big(\bigcup_{k=i+1}^{\infty} A_k\Big).$$

That is, $\omega \in A_i$ and $\omega \notin A_j$ for all $j \geqslant i+1$. That is, $\mathbb{m}_{n_0} = i$. This finishes the proof of (b)(i). □

Example 1 (**Polynomial probability decay**). Assume $\mathbb{P}(A_m) \leqslant cm^{-q}$ for all $m \geqslant n_0$ for some given constants $q, c > 0$ and $n_0 \geqslant 1$. Then it is shown below that for any $0 \leqslant p < q-2$ we have

$$\mathbb{E}[\mathcal{O}_{n_0}^{p+1}] \leqslant \mathbb{E}[\mathbb{m}_{n_0}^{p+1}] \leqslant cq\zeta(q-p-1; n_0). \tag{2.6}$$

For $n_0 = 1$ this result coincides with Estrada and Högele (2022, Example 1) except for the corrected prefactor cq here. In addition, for any $0 \leqslant p < q-2$ it follows by Markov's inequality and (2.6) that

$$\mathbb{P}(\mathcal{O}_{n_0} \geqslant k) \leqslant \mathbb{P}(\mathbb{m}_{n_0} \geqslant k) \leqslant cq \cdot k^{-(p+1)} \cdot \zeta(q-p-1; n_0) \qquad \text{for } k \geqslant 1, \tag{2.7}$$

where $\zeta(z; n_0) = \sum_{n=n_0} \frac{1}{n^z}$ is the classical Hurwitz zeta-function. The rate can be optimized and we obtain

$$\mathbb{P}(\mathcal{O}_{n_0} \geqslant k) \leqslant \mathbb{P}(\mathbb{m}_{n_0} \geqslant k) \leqslant c_1 \cdot k^{-(q-1)} \cdot \Big(\ln(k) + \frac{1}{n_0} - \psi(n_0)\Big) \qquad \text{for } k \geqslant e^{\frac{1}{q-2}+\psi(n_0)}, \tag{2.8}$$

where the constants c_1 and $\psi(n_0)$ in the case with optimal rate are given below. We note that the optimal rate $k^{-(q-1)}$ is only valid for sufficiently large values of k.
Statements 2.6 and 2.7 are seen as follows. For $a_n = n^p$, $p > 0$, we have the following estimate

$$\sum_{n=n_0}^{\infty} n^p \sum_{m=n}^{\infty} cm^{-q} \leqslant c \sum_{n=n_0}^{\infty} n^p \Big(n^{-q} + \int_n^{\infty} x^{-q} dx\Big) \leqslant \frac{cq}{q-1} \zeta(q-p-1; n_0). \tag{2.9}$$

where $\zeta(z; n_0) = \sum_{n=n_0}^{\infty} \frac{1}{n^z}$ is the Hurwitz zeta-function. The right-hand side of (2.9) is finite if and only if $p - q + 1 < -1$, i.e. $0 < p < q-2$. At the same time due to $n_0 \geqslant 1$ we have

$$\mathcal{S}_{a,n_0}(N) = \sum_{n=n_0}^{N+n_0-1} n^p \geqslant \sum_{n=1}^{N} n^p \geqslant \int_0^N x^p dx = \frac{N^{p+1}}{p+1}.$$

Hence, by inequality (2.5) from Lemma 1, we obtain (2.6):

$$\mathbb{E}[\mathcal{O}_{n_0}^{p+1}] \leqslant \mathbb{E}[\mathbb{m}_{n_0}^{p+1}] \leqslant (p+1)\frac{cq}{q-1}\zeta(q-p-1; n_0) \leqslant cq\zeta(q-p-1; n_0).$$

We apply the Markov inequality and obtain (2.7), which we further optimize with respect to p

$$\mathbb{P}(\mathcal{O}_{n_0} \geqslant k) \leqslant \mathbb{P}(\mathbb{m}_{n_0} \geqslant k) \leqslant \inf_{p \in [0, q-2)} k^{-(p+1)} cq\zeta(q-p-1; n_0), \qquad k \geqslant 1.$$

After an optimization in p which is given in Appendix B we obtain that for any $0 \leqslant p < q-2$ and $c_1 = cqe^{(q-2)\psi(n_0)} n_0^{q-2}$, where $\psi(n_0) := \Gamma'(n_0)/\Gamma(n_0)$, it follows

$$\mathbb{P}(\mathcal{O}_{n_0} \geqslant k) \leqslant \mathbb{P}(\mathbb{m}_{n_0} \geqslant k) \leqslant c_1 \cdot k^{-(q-1)} \cdot (\ln(k) + \frac{1}{n_0} - \psi(n_0)) \qquad \text{for } k \geqslant e^{\frac{1}{q-2}+\psi(n_0)}.$$

Example 2 (**Exponential probability decay**). Assume $\mathbb{P}(A_m) \leqslant cb^m$ for all $m \geqslant n_0$, for some given constants $n_0 \in \mathbb{N}$, $b \in (0,1)$ and $c > 0$. Then we have

$$\mathbb{E}[b^{-p\mathcal{O}_{n_0}}] \leqslant \mathbb{E}[b^{-p\mathbb{m}_{n_0}}] \leqslant 1 + \frac{cb^{n_0-1}}{1-b^{1-p}} \tag{2.10}$$

and for all $k \geqslant 1$

$$\mathbb{P}(\mathcal{O}_{n_0} \geqslant k) \leqslant \mathbb{P}(\mathbb{m}_{n_0} \geqslant k) \leqslant 2e^{\frac{9}{8}} \cdot \big[k(cb^{n_0-1}+1)+1\big] \cdot b^k. \tag{2.11}$$

This is seen as follows. For any $a_n = b^{-pm}$, $p \in (0,1)$, we have

$$\sum_{m=n}^{\infty} cb^m = \frac{c}{1-b} b^n$$

and by (2.3)

$$\mathcal{S}_{a,n_0}(N) = \sum_{n=0}^{N-1} b^{-p(n_0+m)} = b^{-pn_0}\frac{(b^{-p})^N - 1}{b^{-p}-1},$$

such that

$$b^{-pN} = b^{pn_0}(b^{-p}-1)\cdot \mathcal{S}_{a,n_0}(N) + 1.$$

Consequently we have

$$\begin{aligned}\mathbb{E}[b^{-p\mathcal{O}_{n_0}}] &\leqslant \mathbb{E}[b^{-p\mathfrak{m}_{n_0}}] \leqslant b^{pn_0}(b^{-p}-1)\cdot \mathbb{E}[\mathcal{S}_{a,n_0}(\mathcal{O})] + 1 \leqslant b^{pn_0}(b^{-p}-1)\cdot \sum_{n=n_0}^{\infty} b^{-pn}\sum_{m=n}^{\infty} cb^m + 1\\ &\leqslant 1 + \frac{cb^{n_0-1}}{1-b^{1-p}}.\end{aligned}$$

This shows (2.10). Markov's inequality and Högele and Steinicke (2023, Lemma 5) yield (2.11).

Example 3 (**Weibull type probability decay**). Assume $\mathbb{P}(A_m) \leqslant cb^{m^\alpha}$ for all $m \geqslant n_0$, for some given constants $n_0 \in \mathbb{N}$, $b \in (0,1), \alpha \in (0,1)$ and $c > 0$. Then for all $p \in (0,1)$ there is a constant $K = K(b,p,\alpha,n_0) > 0$ given below such that

$$\mathbb{E}[b^{-p(\mathcal{O}_{n_0}+n_0-1)^\alpha}] \leqslant \mathbb{E}[b^{-p(\mathfrak{m}_{n_0}+n_0-1)^\alpha}] \leqslant K,$$

such that for all $k \geqslant 1$ we have

$$\mathbb{P}(\mathcal{O}_{n_0} \geqslant k) \leqslant \mathbb{P}(\mathfrak{m}_{n_0} \geqslant k) \leqslant b^{p(k-1)^\alpha}K.$$

Further optimization of the rate yields the existence of positive constants $d = d(c,\alpha,\beta,n_0), D = D(c,\alpha,\beta,n_0,p) > 0$ such that for all $k \geqslant 2$

$$\mathbb{P}(\mathcal{O}_{n_0} \geqslant k) \leqslant \mathbb{P}(\mathfrak{m}_{n_0} \geqslant k) \leqslant (d + D(k-1)^{2-\alpha})b^{(k-1)^\alpha}.$$

This is seen as follows. For any sequence $a_n = b^{-pm^\alpha}$, $p \in (0,1)$, the integral comparison test yields

$$\sum_{m=n+1}^{\infty} cb^{m^\alpha} \leqslant c\int_{m=n}^{\infty} e^{-|\ln(b)|x^\alpha}dx.$$

For $t = |\ln(b)|x^\alpha$ that is $x = \big(\frac{t}{|\ln(b)|}\big)^{\frac{1}{\alpha}}$ and $dx = \frac{1}{\alpha|\ln(b)|}\Big(\frac{t}{|\ln(b)|}\Big)^{\frac{1}{\alpha}-1}dt$ such that

$$\begin{aligned}\int_n^\infty e^{-|\ln(b)|x^\alpha}dx &= \frac{1}{\alpha|\ln(b)|}\int_{|\ln(b)|n^\alpha}^{\infty} e^{-t}\big(\frac{t}{|\ln(b)|}\big)^{\frac{1}{\alpha}-1}dt\\ &\leqslant \frac{1}{\alpha^2|\ln(b)|^{\frac{1}{\alpha}}}e^{-|\ln(b)|n^\alpha}(|\ln(b)|n^\alpha)^{\frac{1-\alpha}{\alpha}} = \frac{1}{\alpha^2|\ln(b)|}e^{-|\ln(b)|n^\alpha}n^{1-\alpha}.\end{aligned}$$

Hence for all $p \in (0,1)$ we have $a_n = b^{-pn^\alpha}$

$$\sum_{n=n_0}^{\infty} a_n \sum_{m=n}^{\infty}\mathbb{P}(A_m) \leqslant c\sum_{n=n_0}^{\infty} b^{-(1-p)n^\alpha}n^{1-\alpha} =: K(b,p,\alpha,n_0). \tag{2.12}$$

Finally, by (2.3)

$$\mathcal{S}_a(N) = \sum_{n=n_0}^{N+n_0-1} a_n = \sum_{n=0}^{N-1} b^{-p(n+n_0)^\alpha} \geqslant b^{-p(N+n_0-1)^\alpha},$$

such that

$$\mathbb{E}[b^{-p(\mathfrak{m}_{n_0}+n_0-1)^\alpha}] \leqslant K(b,p,\alpha,n_0), \qquad \text{and} \qquad \mathbb{P}(\mathfrak{m}_{n_0} \geqslant k) \leqslant \inf_{p\in(0,1)} b^{p(k-1)^\alpha}K(b,p,\alpha,n_0).$$

By Lemma 5 there are positive constants $d, D \in \mathbb{N}$ such that for $k \geqslant 2$

$$\mathbb{P}(\mathcal{O}_{n_0} \geqslant k) \leqslant \mathbb{P}(\mathrm{m}_{n_0} \geqslant k) \leqslant \inf_{p\in(0,1)} b^{p(k-1)^\alpha} \sum_{n=n_0}^{\infty} c\left(1 + \frac{1+\frac{\frac{1}{\alpha}-1}{|\ln(b)|}}{\alpha|\ln(b)|} n^{1-\alpha}\right) b^{(1-p)n^\alpha}$$
$$\leqslant (d + D(k-1)^{2-\alpha}) b^{(k-1)^\alpha}.$$

The precise values of d and D are given in the proof of Lemma 5 in Appendix A.

Lemma 2 (**Tradeoff between almost sure error tolerance and mean deviation frequency**). *Given a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, and a Polish space $\mathcal{X}$ with a complete metric d on $\mathcal{X}$ which generates the topology. We consider a sequence of random vectors $(X_n)_{n\geqslant n_0}$ for some $n_0 \in \mathbb{N}$, $X_n : \Omega \to \mathcal{X}$, $n \geqslant n_0$, and a random vector $X : \Omega \to \mathcal{X}$. Assume that X_n converges to X as $n \to \infty$ in probability, that is, for any fixed $\delta > 0$ we have*

$$p(\delta, n) := \mathbb{P}(\mathrm{d}(X_n, X) > \delta) \to 0, \qquad \text{as } n \to \infty.$$

Then we have the following tradeoff: For any positive, nonincreasing sequence $\epsilon = (\varepsilon_n)_{n\geqslant n_0}$, and any positive, nondecreasing sequence $a = (a_n)_{n\geqslant n_0}$ such that

$$K(a, \epsilon, n_0) := \sum_{n=n_0}^{\infty} a_n \sum_{m=n}^{\infty} p(\varepsilon_m, m) < \infty, \tag{2.13}$$

it follows

$$\limsup_{n\to\infty} \mathrm{d}(X_n, X) \cdot \varepsilon_n^{-1} \leqslant 1 \qquad \mathbb{P}\text{-a.s.}, \tag{2.14}$$

and

$$\mathbb{E}[\mathcal{S}_{a,n_0}(\mathcal{O}_{\epsilon,n_0})] \leqslant \mathbb{E}[\mathcal{S}_{a,n_0}(\mathrm{m}_{\epsilon,n_0})] \leqslant K(a, \epsilon, n_0), \tag{2.15}$$

where

$$\mathcal{O}_{\epsilon,n_0}(\omega) := \sum_{n=n_0}^{\infty} \mathbf{1}\{\mathrm{d}(X_n(\omega), X(\omega)) > \varepsilon_n\}, \qquad \omega \in \Omega, \tag{2.16}$$

$$\mathrm{m}_{\epsilon,n_0}(\omega) := \max\{n - n_0 \geqslant 0 \mid \mathrm{d}(X_n(\omega), X(\omega)) > \varepsilon_n\}, \qquad \omega \in \Omega, \tag{2.17}$$

and $\mathcal{S}_{a,n_0}$ is defined in (2.3). Further, by Markov's inequality, we obtain

$$\mathbb{P}(\mathcal{O}_{\epsilon,n_0} \geqslant k) \leqslant \mathbb{P}(\mathrm{m}_{\epsilon,n_0} \geqslant k) \leqslant \frac{K(a, \epsilon, n_0)}{\mathcal{S}_{a,n_0}(k)}, \qquad k \geqslant 1.$$

Definition 2. In the situation of Lemma 2, we call $\mathcal{O}_{\epsilon,n_0}$ defined by (2.16) the **overlap statistic** or **deviation frequency** and $\mathrm{m}_{\epsilon,n_0}$ the **modulus of a.s. convergence**. We call the relation (2.14) an **a.s. error tolerance of order** ϵ, while the relation (2.15) is referred to as **mean deviation frequency (MDF) bound of order $\mathcal{S}_{a,n_0}$.**

Remark 2.

(a) Note that for different sequences of error tolerances $\epsilon = (\varepsilon_n)_{n\geqslant 0}$ we obtain different rates $p_n = \mathbb{P}(A_n(\varepsilon_n))$ as a function of n. The tradeoff between (2.14) and (2.15) is quantified by the play between $\epsilon = (\varepsilon_n)$ and $a = (a_n)$ by the finiteness of the constant (2.13).

(b) In this context Lemma 2 generalizes the classical first Borel-Cantelli lemma (Kallenberg, 2002, Theorem 2.18) as follows:
For any positive, nonincreasing sequence $\epsilon = (\varepsilon_n)_{n\geqslant n_0}$ such that

$$K_0(\epsilon, n_0) := \sum_{n=n_0}^{\infty} p(\varepsilon_n, n) < \infty$$

we have the error tolerance

$$\limsup_{n\to\infty} \mathrm{d}(X_n, X)\cdot \varepsilon_n^{-1} \leqslant 1 \qquad \mathbb{P}\text{-a.s.}, \tag{2.18}$$

and the mean deviation frequency of order 1

$$\mathbb{E}[\mathcal{O}_{\epsilon,n_0}] = K_0(\epsilon, n_0).$$

In particular, the classical first Borel-Cantelli lemma does not yield information about the modulus of convergence $\mathbb{m}_{\epsilon,n_0}$.

(c) For a constant sequence $\varepsilon_n = \varepsilon > 0$, $n \geqslant n_0$, we denote the same overlap statistic in a slight abuse of notation by $\mathcal{O}_\varepsilon$, which coincides with the notation of Estrada and Högele (2022).

(d) For fixed $\varepsilon_n = \varepsilon > 0$ the rate $p(\varepsilon, n) \to 0$ is the fastest possible among all nonincreasing sequences, which translates to the largest possible finite moments of the (random) overlap count in terms of $\mathbb{E}[\mathcal{S}_{a,n_0}(\mathcal{O}_\varepsilon)] < \infty$.

(e) For any $\epsilon = (\varepsilon_n)_{n\in\mathbb{N}}$ such that $p(\varepsilon_n, n)$ is close to not being summable (such as for instance $\frac{1}{n^\theta}$, $\theta > 1$ or $\frac{1}{n\ln^\theta(n+1)}$, $\theta > 1$), the usual Borel-Cantelli lemma implies a close to optimal almost sure error tolerance, however, the MDF bound is maximal, exhibiting linear decay at best, since by Markov's inequality

$$\mathbb{P}(\mathcal{O}_{\epsilon,n_0} \geqslant k) \leqslant k^{-1}\cdot \mathbb{E}[\mathcal{O}_\epsilon].$$

The proof of Lemma 2 is based on Lemma 1.

Proof of Lemma 2: For any positive sequence $\epsilon = (\varepsilon_n)_{n\geqslant n_0}$ we consider the events

$$A_n := \{\mathrm{d}(X_n, X) > \varepsilon_n\}, \qquad n \geqslant n_0,$$

and by Lemma 1(b)(i) the respective overlap representation

$$\mathbb{m}_{\epsilon,n_0} := \sum_{n=n_0}^\infty \mathbf{1}\Big(\bigcup_{m=n}^\infty A_m\Big).$$

Since $K(\inf_n a_n, \epsilon) \leqslant K(a, \epsilon) < \infty$ by hypothesis, we may apply Lemma 1. Then $\mathcal{O}_{\varepsilon,n_0} \leqslant \mathbb{m}_{\varepsilon,n_0}$ and the usual Borel-Cantelli lemma yields

$$0 = \mathbb{P}(\mathrm{d}(X_n, X) > \varepsilon_n \text{ infinitely often }) = \mathbb{P}(\limsup_{n\to\infty} \mathrm{d}(X_n, X)\cdot\varepsilon_n^{-1} > 1),$$

and implies (2.14). Furthermore, (2.13) and Lemma 1 implies (2.15). This finishes the proof. □

In Section 3, 4 and 5 we infer a.s. MDF convergence results for various classes of martingales and sequences of martingale differences of interest with the help of Lemma 2.

3. The tradeoff in almost sure martingale convergence theorems

In the sequel we quantify the martingale convergence theorems with the help of Lemma 2.

3.1. *The tradeoff for martingales bounded in* $\mathbf{L^p}$*,* $\mathbf{p \geqslant 2}$.

We start with one of the most classical martingale convergence results in L^2 is due to Pythagoras' theorem.

Theorem 1 (**Pythagoras' theorem for martingale differences**).
Given a filtered probability space $(\Omega, \mathcal{A}, \mathbb{P}, \mathbb{F})$*,* $n_0 \in \mathbb{N}_0$*, we consider a martingale* $X = (X_n)_{n\geqslant n_0}$*, with values in a Hilbert space* $(H, \langle\cdot,\cdot\rangle)$ *and which satisfies*

$$\sup_{n\geqslant n_0} \mathbb{E}[\|X_n\|^2] < \infty.$$

Then X converges a.s. and in $L^2(\Omega; H)$ to a random vector X_∞ in $L^2(\Omega; H)$. In addition, for any positive, nonincreasing sequence $\epsilon = (\varepsilon_n)_{n\geqslant n_0}$ it follows

$$\mathbb{P}(\|X_n - X_\infty\| > \varepsilon_n) \leqslant \varepsilon_n^{-2} \cdot \mathbb{E}[\|X_n - X_\infty\|^2] = \varepsilon_n^{-2} \cdot \sum_{m=n+1}^{\infty} \mathbb{E}[\|\Delta X_m\|^2] = \varepsilon_n^{-2} \cdot \pi_n, \qquad n \geqslant n_0, \quad (3.1)$$

where $\Delta X_n := X_n - X_{n-1}$ and $\pi_n = \sum_{m=n+1}^{\infty} \mathbb{E}[\|\Delta X_m\|^2]$ for $n \geqslant n_0 + 1$.
Moreover, we have the following tradeoff: For all positive, nonincreasing sequences $\epsilon = (\varepsilon_n)_{n\geqslant n_0}$ and positive, nondecreasing sequences $a = (a_n)_{n\geqslant n_0}$ such that

$$K(a, \epsilon) := \sum_{n=n_0}^{\infty} a_n \sum_{m=n}^{\infty} \varepsilon_m^{-2} \cdot \pi_m < \infty,$$

it follows

$$\limsup_{n\to\infty} \|X_n - X_\infty\| \cdot \varepsilon_n^{-1} \leqslant 1 \qquad \mathbb{P}\text{-a.s.} \quad (3.2)$$

and

$$\mathbb{E}[\mathcal{S}_{a,n_0}(\mathcal{O}_{\epsilon,n_0})] \leqslant \mathbb{E}[\mathcal{S}_{a,n_0}(\mathrm{m}_{\epsilon,n_0})] \leqslant K(a, \epsilon), \quad (3.3)$$

where $\mathcal{O}_{\epsilon,n_0}$ is given in (2.16), $\mathrm{m}_{\epsilon,n_0}$ in (2.17) and $\mathcal{S}_{a,n_0}$ in (2.3).

Proof: The proof is a straight-forward extension of the Pythagoras theorem (Williams, 1991, Subsection 14.18)

$$\mathbb{E}[|X_m - X_n|^2] = \sum_{\ell=n+1}^{m} \mathbb{E}[|X_\ell|^2]$$

to Hilbert spaces with a direct application of Lemma 2. □

Example 4 (**Centered random walk**). Consider an independent sequence of centered square integrable random variables $(\Delta X_n)_{n\geqslant 1}$ with values in some separable Hilbert space $(H, \langle\cdot,\cdot\rangle)$. Hence the process of partial sums $(X_n)_{n\in\mathbb{N}_0}$, $X_0 = 0$, $X_n := \sum_{i=1}^{n} \Delta X_i$, $n \geqslant 1$, defines a martingale with respect to the natural filtration given by $\mathcal{F}_n := \sigma(X_1, \ldots, X_n)$. If

$$\sum_{n=1}^{\infty} \mathbb{E}[\|\Delta X_n\|^2] < \infty,$$

we have that X_n converges in L^2 and a.s. For instance if $\mathrm{Var}(\Delta X_n) = n^{-q}$, $n \geqslant 1$, for some $q > 3$ we obtain that

$$\pi_n = \sum_{m=n+1}^{\infty} \mathbb{E}[\|\Delta X_m\|^2] \leqslant \int_n^{\infty} \frac{1}{x^q} dx = \frac{n^{-(q-1)}}{q-1}.$$

In particular for $\epsilon = (\varepsilon_n)_{n\in\mathbb{N}}$ with $\varepsilon_n = n^{-\alpha}$ such that $q - 1 - 2\alpha > 2$ it follows that

$$\sum_{m=n+1}^{\infty} \varepsilon_m^{-2} \pi_m = \sum_{m=n+1}^{\infty} (n+1)^{-(q-1-2\alpha)} \leqslant \frac{n^{-(q-2-2\alpha)}}{q-2-2\alpha}.$$

Hence for $a_n = (n+1)^p$, $0 < p < q - 3 - 2\alpha$ we have by (2.3) that $\mathcal{S}_{a,1}(N) = \sum_{n=1}^{N} a_n$ for $N \in \mathbb{N}$ and $\mathcal{S}_{a,1}(0) = 0$ such that

$$K := \sum_{n=1}^{\infty} a_n \sum_{m=n} \varepsilon_m^{-2} \pi_m \leqslant \frac{1}{q-2-2\alpha} \sum_{n=1}^{\infty} n^{p-(q-2-2\alpha)} < \infty$$

implies

$$\limsup_{n\to\infty} \|X_n - X\| \cdot n^{\alpha} \leqslant 1 \qquad \mathbb{P}\text{-a.s.}$$

and by Example 1, $\mathbb{E}[\mathcal{O}_\epsilon^{1+p}] \leqslant \mathbb{E}[\mathrm{m}_\epsilon^{1+p}] \leqslant q\zeta(q-p-1; n_0)$, as well as for $k \geqslant 1$ that

$$\mathbb{P}(\mathcal{O}_\epsilon \geqslant k) \leqslant \mathbb{P}(\mathrm{m}_\epsilon \geqslant k) \leqslant k^{-(p+1)} \cdot K,$$

which is further optimized in (2.8). In other words, a sufficiently fast decay of the variances translates naturally into a higher order MDF convergence.

Remark 3. By a direct application of the the Burkholder-Davis-Gundy inequality (Williams, 1991, Section (14.18)) it obvious how to generalize this result to a version for martingales which are uniformly bounded in L^r for some $r > 2$ and with rates

$$\tilde{\pi}_{n,r} := \mathbb{E}\left[\Big(\sum_{m=n+1}^{\infty} \|\Delta X_m\|^2\Big)^{\frac{r}{2}}\right].$$

Inequality (3.1) then reads

$$\mathbb{P}(\|X_n - X_\infty\| > \varepsilon_n) \leqslant \varepsilon_n^{-r} \cdot \tilde{\pi}_{n,r}, \qquad n \geqslant n_0,$$

and

$$K(a,\epsilon,r) := \sum_{n=n_0}^{\infty} a_n \sum_{m=n}^{\infty} \varepsilon_m^{-r} \cdot \tilde{\pi}_{m,r} < \infty.$$

The formulation of the tradeoff between (3.2) and (3.3) reads similar with the obvious adjustments of a and ϵ.

3.2. *The tradeoff for martingale convergence by the Azuma-Hoeffding inequality.*

The Azuma-Hoeffding inequality replaces the absolute summability of the square integrals of dX_n by the much stronger condition of a.s. summability of the squares of dX_n. As a consequence, we obtain exponential estimates. First, we consider the real-valued case, where Azuma-Hoeffding's inequality includes supermartingales.

Theorem 2 (**Azuma-Hoeffding inequality**).
Let $X = (X_n)_{n\in\mathbb{N}_0}$ be a real-valued supermartingale with super-martingale differences $(\Delta X_n)_{n\in\mathbb{N}}$. Assume that the sequence $(\Delta X_n)_{n\in\mathbb{N}}$ is bounded almost surely by positive numbers $(c_n)_{n\in\mathbb{N}}$, that is,

$$|\Delta X_n| \leqslant c_n, \qquad \mathbb{P}\text{-a.s.} \qquad \text{for all } n \in \mathbb{N}.$$

Then it follows

$$\mathbb{P}(X_n - X_0 \geqslant \varepsilon) \leqslant \exp\Big(-\tfrac{1}{2}\tfrac{\varepsilon^2}{\sum_{k=1}^n c_k^2}\Big), \qquad \text{for all } n \in \mathbb{N}.$$

The proof goes back to Azuma (1967); Hoeffding (1963). In the sequel we send $n \to \infty$ and use the tail summability in order to infer almost convergence $X_n \to X_\infty$ as $n \to \infty$, which can be quantified in terms of mean deviation frequencies.

Theorem 3 (**The tradeoff via the Azuma-Hoeffding closure**).
Assume the hypotheses of Theorem 2 and, in addition,

$$\sum_{n=1}^{\infty} c_n^2 < \infty.$$

Let $r(n) := \sum_{k=n+1}^{\infty} c_k^2$ for $n \in \mathbb{N}$. Then there exists an a.s. finite random variable X_∞ and we have $X_n \to X_\infty$ a.s. as $n \to \infty$. Then for any nonincreasing positive sequence $\epsilon = (\varepsilon_n)_{n\in\mathbb{N}}$ and any sequence of positive, nondecreasing weights $a = (a_n)_{n\in\mathbb{N}}$ such that

$$K(a,\epsilon) := 2\sum_{n=1}^{\infty} a_n \sum_{m=n}^{\infty} \exp\Big(-\frac{1}{2}\frac{\varepsilon_m^2}{r(m)}\Big) < \infty, \tag{3.4}$$

we have that

$$\limsup_{n\to\infty} |X_n - X_\infty| \cdot \varepsilon_n^{-1} \leqslant 1, \qquad \mathbb{P}\text{-}a.s.$$

and

$$\mathbb{E}\left[\mathcal{S}_{a,1}(\mathcal{O}_\epsilon)\right] \leqslant \mathbb{E}\left[\mathcal{S}_{a,1}(\mathbb{m}_\epsilon)\right] \leqslant K(a,\epsilon), \tag{3.5}$$

for $\mathcal{O}_\epsilon := \sum_{n=1}^\infty \mathbf{1}\{|X_\infty - X_n| \geqslant \varepsilon_n\}$, $\mathbb{m}_\epsilon := \max\{n-1 \geqslant 0 \mid |X_\infty - X_n| \geqslant \varepsilon_n\}$, *and* $\mathcal{S}_{a,1}$ *is defined in* (2.3).
In particular, we have:

$$\mathbb{P}(\mathcal{O}_\epsilon \geqslant k) \leqslant \mathbb{P}(\mathbb{m}_\epsilon \geqslant k) \leqslant \inf_a \ \mathcal{S}_{a,1}^{-1}(k) \cdot 2 \sum_{n=0}^\infty a_n \sum_{m=n}^\infty \exp\Big(-\frac{1}{2}\frac{\varepsilon_m^2}{r(m)}\Big), \qquad k \geqslant 1,$$

where we optimize over suitable sequences of positive, nondecreasing numbers $a = (a_n)_{n\in\mathbb{N}}$ *satisfying* (3.4).

Proof: Clearly, by the martingale convergence theorem (Williams, 1991), there is a closure X_∞ such that $X_n \to X_\infty$ a.s. For $0 \leqslant n \leqslant m$ we have by Azuma's inequality that

$$\mathbb{P}(X_m - X_n \geqslant \varepsilon_n) \leqslant \exp\left(-\frac{1}{2}\frac{\varepsilon_n^2}{\sum_{k=n}^m c_k^2}\right).$$

Sending $m \to \infty$, the left hand side converges as X converges to X_∞ in probability and we obtain by Fatou's lemma

$$\begin{aligned}\mathbb{P}(X_\infty - X_n \geqslant \varepsilon_n) &= \mathbb{E}[\liminf_{m\to\infty} \mathbf{1}\{X_m - X_n \geqslant \varepsilon_n\}] \\ &\leqslant \liminf_{m\to\infty} \mathbb{E}[\mathbf{1}\{X_m - X_n \geqslant \varepsilon_n\}] = \liminf_{m\to\infty} \exp\left(-\frac{1}{2}\frac{\varepsilon_n^2}{\sum_{k=n}^m c_k^2}\right) = \exp\left(-\frac{1}{2}\frac{\varepsilon_n^2}{r(n)}\right).\end{aligned} \tag{3.6}$$

Note that the right-hand side is strictly decreasing by hypothesis as a function of n. Whenever

$$R(n) := \sum_{\ell=n}^\infty \exp\Big(-\frac{1}{2}\frac{\varepsilon_\ell^2}{r(\ell)}\Big) < \infty$$

for some (and hence all) $n \in \mathbb{N}$ and, in addition,

$$\sum_{n=1}^\infty a_n R(n) < \infty,$$

we infer inequality (3.5) by Lemma 2. Combining the monotonicity of $\mathcal{S}_{a,n_0}$, Markov's inequality and (3.5) we have

$$\mathbb{P}(\mathcal{O}_\epsilon \geqslant k) \leqslant \mathbb{P}(\mathbb{m}_\epsilon \geqslant k) \leqslant \mathcal{S}_{a,1}(k)^{-1} \cdot \mathbb{E}[\mathcal{S}_{a,1}(\mathbb{m}_\epsilon)] \qquad k \geqslant 1.$$

A subsequent optimization over the respective sequences a yields the second statement. □

Corollary 1. *Assume the hypotheses and notation of Theorem 3. Suppose for some* $C > 0$, $q \in (0,1]$ *and* $n_0 \in \mathbb{N}_0$ *we have*

$$r(n) \leqslant \frac{C}{(n+1)^q}, \qquad \text{for all } n \geqslant n_0.$$

Then we have the following tradeoff: For all $a = (a_n)_{n\geqslant n_0}$ *positive, nondecreasing and* $\epsilon = (\varepsilon_n)_{n\geqslant n_0}$ *positive, nonincreasing such that*

$$K(a,\epsilon,n_0) := 2\sum_{n=n_0}^\infty a_n \sum_{m=n}^\infty \exp\Big(-\frac{\varepsilon_m^2 (m+1)^q}{2C}\Big) < \infty$$

we have

$$\limsup_{n\to\infty} |X_n - X| \cdot \varepsilon_n^{-1} \leqslant 1 \qquad \mathbb{P}\text{-}a.s.$$

versus

$$\mathbb{E}[\mathcal{S}_{a,n_0}(\mathcal{O}_{\epsilon,n_0})] \leqslant \mathbb{E}[\mathcal{S}_{a,n_0}(\mathrm{m}_{\epsilon,n_0})] \leqslant K(a,\epsilon).$$

Moreover, we have the following special case: For $q \in (0,1]$, $\varepsilon_n = \sqrt{\frac{2C(2+\theta)\ln(n+1)}{(n+1)^q}}$, $\theta > 0$, *and* $a_n = n^p$, $0 < p < \theta$ *we have*

$$\limsup_{n\to\infty} |X_n - X| \cdot \sqrt{\frac{(n+1)^q}{\ln(n+1)}} \leqslant \sqrt{2C(2+\theta)}, \qquad \mathbb{P}\text{-}a.s.$$

while

$$\mathbb{E}[\mathcal{O}_{\epsilon,n_0}^{p+1}] \leqslant \mathbb{E}[\mathrm{m}_{\epsilon,n_0}^{p+1}] \leqslant 2\theta\zeta(1+\theta-p; n_0). \tag{3.7}$$

For $k \geqslant e^{\psi(n_0)+\theta^{-1}}$ *it follows*

$$\mathbb{P}(\mathcal{O}_{\epsilon,n_0} \geqslant k) \leqslant \mathbb{P}(\mathrm{m}_{\epsilon,n_0} \geqslant k) \leqslant 2\theta \cdot k^{-(1+\theta)} \cdot k^{\frac{1}{\ln(k)-\psi(n_0)}} \cdot \zeta(1+\theta - \frac{1}{\ln(k)-\psi(n_0)}; n_0).$$

The proof is an application of Theorem 3 combined with Example 1 and Example 3.

For the higher dimensional, and in particular infinite-dimensional case, we state the following immediate simplification of the Azuma inequality proven in Luo (2022, Theorem 1.2).

Theorem 4 (**Azuma-Hoeffding in infinite dimensions**).
Let B *be a* p*-smooth Banach space for* $1 < p \leq 2$ *and let* $X = (X_n)_{n\in\mathbb{N}_0}$ *be a martingale with values in* B *and differences* $(\Delta X_n)_{n\in\mathbb{N}}$*. Assume that the sequence* $(\Delta X_n)_{n\in\mathbb{N}}$ *is a.s. bounded by a non-negative sequence* $(c_n)_{n\in\mathbb{N}}$*. Then, there is a constant* K*, only depending on* X *such that for all* $\varepsilon > 0$,

$$\mathbb{P}\Big(\sup_{j\in\mathbb{N}_0} \|X_j - X_0\| \geq \varepsilon\Big) \leq 2\exp\Big(-\frac{\varepsilon^p}{2K\sum_{j=1}^\infty c_j^p}\Big).$$

The according MDF martingale convergence tradeoff now takes the following form.

Theorem 5 (**Martingale convergence by Azuma in higher dimensions**).
Assume the hypotheses of Theorem 4. Assume that

$$\sum_{n=1}^\infty c_n^p < \infty$$

and set $r(n) := \sum_{k=n+1}^\infty c_k^p$ *for* $n \in \mathbb{N}$*. Further, assume that for a nonincreasing positive sequence* $\epsilon = (\varepsilon_n)_{n\in\mathbb{N}}$ *and a positive, nondecreasing sequence* $a = (a_n)_{n\in\mathbb{N}}$,

$$K(a,\epsilon) := 2\sum_{n=1}^\infty a_n \sum_{m=n}^\infty \exp\Big(-\frac{1}{2}\frac{\varepsilon_m^p}{r(m)}\Big) < \infty. \tag{3.8}$$

Then the following assertions hold:

(a) *There exists an a.s. finite random variable* X_∞ *and we have* $X_n \to X_\infty$ *a.s. as* $n \to \infty$.
(b) $\limsup_{n\to\infty} \|X_n - X_\infty\| \cdot \varepsilon_n^{-1} \leqslant 1, \qquad \mathbb{P}$*-a.s.*
(c) *For* $\mathcal{O}_\epsilon = \sum_{n=1}^\infty \mathbf{1}\{\|X_\infty - X_n\| \geqslant \varepsilon_n\}$, $\mathrm{m}_\epsilon = \max\{n-1 \geqslant 0 \mid \|X_\infty - X_n\| \geqslant \varepsilon_n\}$, *and* $\mathcal{S}_{a,1}$ *for* (2.3), *we get*

$$\mathbb{E}\left[\mathcal{S}_{a,1}(\mathcal{O}_\epsilon)\right] \leqslant \mathbb{E}\left[\mathcal{S}_{a,1}(\mathrm{m}_\epsilon)\right] \leqslant K(a,\epsilon).$$

(d) *In particular,*

$$\mathbb{P}(\mathcal{O}_\epsilon \geqslant k) \leqslant \mathbb{P}(\mathfrak{m}_\epsilon \geqslant k) \leqslant \inf_a \ \mathcal{S}_{a,1}^{-1}(k) \cdot 2\sum_{n=0}^{\infty} a_n \sum_{m=n}^{\infty} \exp\Big(-\frac{1}{2}\frac{\varepsilon_m^2}{r(m)}\Big), \qquad k \geqslant 1,$$

where we optimize over suitable sequences of positive, nondecreasing numbers $a = (a_n)_{n\in\mathbb{N}}$ satisfying (3.8).

Again, the proof is an application of Theorem 4 combined with Example 1 and Example 3.

Example 5 (**Exponential MDF convergence for Pólya's urn**). Consider Pólya's urn model as seen in Klenke (2008, Example 12.29) for an urn containing N balls, out of which B are black, and $N - B$ are white. Let $(\Delta Y_n)_{n\geqslant 1}$ be the sequence of independent draws from the urn, such that $\Delta Y_n = 1$ if the n-th ball is black, and $\Delta Y_n = 0$ otherwise. Also, for each draw, the ball picked returns to the urn together with an additional ball of the same color. Then, if $Y_n = \sum_{i=1}^n \Delta Y_i$, we can establish the martingale representing the proportion of black balls in the urn after n draws as $X_n := \frac{Y_n+B}{n+N}$ with $X_0 = \frac{B}{N}$. Here, the martingale differences are bounded, since

$$|X_n - X_{n-1}| = \frac{1}{n+N}\left|\Delta Y_n + \frac{1}{n-1+N}(Y_{n-1}+B)\right| \leqslant \frac{2}{n+N}. \tag{3.9}$$

Hence $X_n \to X_\infty$ a.s. and $X_\infty \sim \mathrm{Beta}(B, N-B)$. In particular, $(X_n)_{n\in\mathbb{N}_0}$ and X_∞ satisfy the conditions of Corollary 1 with

$$r(n) = \sum_{k=n}^{\infty} \frac{2}{(N+k)^2} \leqslant \frac{2}{n-1+N} \leqslant \frac{3}{n}.$$

For any $p < \frac{1}{2}$ and $\epsilon = (\varepsilon_n)_{n\in\mathbb{N}}$ with $\varepsilon_n = \sqrt{\frac{2}{3n^p}}$ we have that

$$\mathbb{P}(|X_n - X_\infty| > \varepsilon_n) \leqslant 2\exp(-n^{1-2p}).$$

By Example 3 we obtain for any $\theta \in (0,1)$ a constant $K(e^{-1}, \theta, 1-2p, 1) > 0$ given in (2.12) such that

$$\mathbb{E}[e^{\theta \mathcal{O}_{\epsilon,1}^{1-2p}}] \leqslant \mathbb{E}[e^{\theta \mathfrak{m}_{\epsilon,1}^{1-2p}}] \leqslant K(e^{-1}, \theta, 1-2p, 1) < \infty,$$

and there are constants $d, D > 0$ defined in Lemma 5 such that

$$\mathbb{P}(\mathcal{O}_{\epsilon,1} \geqslant k) \leqslant \mathbb{P}(\mathfrak{m}_{\epsilon,1} \geqslant k) \leqslant 2(d + D(k-1)^{1+2p})e^{-(k-1)^{1-2p}}, \qquad \text{for all } k \geqslant 2.$$

Example 6 (**Doubly exponential tradeoff bound for a super-critical Galton-Watson process**). Branching processes have a long history and are very well-studied objects with precisely known dynamics. For the different regimes of sub-critical and super-critical branching we quantify the a.s. MDF dynamics. Let $Z = (Z_n)_{n\in\mathbb{N}_0}$ be a Galton-Watson process with i.i.d. offspring variables $(Y_{i,n})_{i,n\geqslant 1}$ and expectation $\mathbb{E}[Y_{1,1}] = \mathfrak{m} \in [0,\infty)$, where

$$Z_{n+1} = \sum_{i=1}^{Z_n} Y_{i,n+1} \qquad Z_0 = 1. \tag{3.10}$$

We define $v := \mathrm{Var}(Y_{1,1}) \in [0,\infty]$. It is well-known, see e.g. Harris (1963, Proof of Theorem 8.1), that $X_n := \frac{Z_n}{\mathfrak{m}^n}$ defines a martingale with respect to the natural filtration. Consider a super-critical Galton-Watson process with $\mathfrak{m} > 1$, $v < \infty$ and bounded support $C := \sup(\mathrm{supp}(Y_{1,1})) < \infty$. Then

$$|\Delta Z_n| \leqslant \frac{C}{\mathfrak{m}^n} = c_n, \text{ which is clearly square summable.}$$

Hence $r(n) = \sum_{i=n}^{\infty} c_i^2 = C\frac{\mathfrak{m}^{-2n}}{\mathfrak{m}-1}$ and for $n-1 \geqslant n_0$ for some $n_0 \in \mathbb{N}_0$

$$R(n) \leqslant \frac{\mathfrak{m}-1}{C} \exp\Big(-\frac{\varepsilon_{n-1}^2(\mathfrak{m}-1)\mathfrak{m}^{n-1}}{2C}\Big)\varepsilon_{n-1}^2\mathfrak{m}^{n-1}.$$

Hence for any $\rho \in (1, \mathfrak{m})$, $\varepsilon_n(\rho) = (\frac{\rho}{\mathfrak{m}})^{\frac{n}{2}}$ and $a_n(\tilde{\rho}, \rho) = e^{\tilde{\rho}^n}$, $1 < \tilde{\rho} < \rho$, we have

$$K(\tilde{\rho}, \rho, n_0) := \sum_{n=n_0}^{\infty} a_n(\tilde{\rho}, \rho) R(n) < \infty,$$

and by (2.3), $\mathcal{S}_a(N) \geqslant e^{\tilde{\rho}^{N-1}}$ and $\mathcal{S}_a(0) = 0$. Consequently, for $\mathcal{O}_\epsilon = \sum_{n=n_0}^{\infty}\{|X_n - X_\infty| > \varepsilon_n\}$ and $\mathrm{m}_\epsilon = \max\{n - n_0 \geqslant 0 \mid |X_n - X_\infty| > \varepsilon_n\}$ we have the doubly exponential decay

$$\mathbb{P}(\mathcal{O}_\epsilon \geqslant k) \leqslant \mathbb{P}(\mathrm{m}_\epsilon \geqslant k) \leqslant e^{-\tilde{\rho}^{k-1}} K(\tilde{\rho}, \rho, n_0), \qquad k \geqslant 1,$$

which can be further optimized over suitable exponents ρ and $\tilde{\rho}$.

Example 7 (**The tradeoff for discrete stochastic integrals**). Let $\Delta = (\Delta_n)_{n\geqslant 1}$ be a sequence of i.i.d. centered random variables on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ which are bounded by a positive constant $C_1 > 0$. Define $\mathcal{F}_n := \sigma(X_k : 1 \leqslant k \leqslant n)$ for $n \geqslant 1$ and set $\mathcal{F}_0 := \{\emptyset, \Omega\}$. Let $g = (g_n)_{n\geqslant 1}$ be a sequence of random variables, uniformly bounded by a real number $C_2 > 0$ and such that g_n is $\mathcal{F}_{n-1}$-measurable. Set

$$X_0 := 0 \qquad \text{and} \qquad X_n := \sum_{k=1}^{n} g_k \frac{\Delta_k}{k} \qquad \text{for } n \geqslant 1.$$

Then, $X = (X_n)_{n\geqslant 0}$ converges to an X_∞ $\mathbb{P}$-a.s. with $c_n = \frac{C_1 \cdot C_1}{n}$ and $r(n) = (C_1 \cdot C_2)^2 \sum_{k=n}^{\infty} \frac{1}{k^2}$. It follows that there is a constant $C_3 > 0$ depending on C_1 and C_2 such that $\frac{1}{r(n)} \geqslant C_3 n$. Then the hypotheses of Corollary 1 are valid for $(X_n)_{n\in\mathbb{N}_0}$ and X_∞. In particular, Corollary 1 (a) and (b) apply.

4. The tradeoff in the strong law for martingale differences (MDs)

To obtain the deviation frequencies for the strong law of large numbers for (centered) martingales X, i.e. quantifying the $\mathbb{P}$-a.s. convergence of $\frac{X_n}{n} \to 0$, we need several estimates for the martingale's moments, and for its difference sequence. Those will result in appropriate concentration inequalities. The case of absolute, monomial moments will be covered in Theorem 6. Exponential moments will be treated afterwards using Theorem 10.

4.1. *The tradeoff in the strong law for MDs in L^p.*

The following result is an application of the Burkholder-Rosenthal inequality (see e.g. Johnson et al. (1985); Osękowski (2012); Rosenthal (1970); Talagrand (1989)). Note that the martingale differences in the subsequent result may or may not be uniformly bounded in L^p. In fact, the optimal result for uniformly bounded martingale differences in L^p is given in Subsection 4.2.

Theorem 6 (**Tradeoff for the strong law in L^p for MDs**).
Let $X = (X_n)_{n\in\mathbb{N}_0}$ be a separable Hilbert space $(H, \langle\cdot,\cdot\rangle)$-valued martingale with $X_0 = 0$ with respect to a filtration $\mathbb{F} = (\mathcal{F}_n)_{n\geq 0}$. Set $\Delta X_j := X_j - X_{j-1}$ for $j \in \mathbb{N}$. Let $p \geqslant 2$ and assume that for

$$\beta_{n,p} := \frac{1}{n}\mathbb{E}\Big[\sum_{j=1}^{n}\|\Delta X_j\|^p + \Big(\sum_{j=1}^{n}\mathbb{E}\left[\|\Delta X_j\|^2 \big| \mathcal{F}_{j-1}\right]\Big)^{\frac{p}{2}}\Big] \tag{4.1}$$

we have finiteness of the value

$$K_p := \sum_{n=1}^{\infty} \frac{\beta_{n,p}}{n^{p-1}} < \infty.$$

Then

$$\lim_{n\to\infty} \frac{X_n}{n} = 0, \quad \mathbb{P}\text{-}a.s.$$

In addition, we have the following tradeoff: For any positive, nonincreasing sequence $\epsilon = (\varepsilon_n)$ *and any positive, nondecreasing sequence* $a = (a_n)_{n\in\mathbb{N}}$ *such that*

$$K_{a,\epsilon,p} := \sum_{n=1}^{\infty} a_n \sum_{m=n}^{\infty} \frac{\beta_{m,p}}{\varepsilon_m^p m^{p-1}} < \infty$$

we have

$$\limsup_{n\to\infty} \left\|\frac{X_n}{n}\right\| \cdot \varepsilon_n^{-1} \leqslant 1 \qquad \mathbb{P}\text{-}a.s. \tag{4.2}$$

and the respective quantities

$$\mathcal{O}_\epsilon := \sum_{n=1}^{\infty} \mathbf{1}\Big\{\Big\|\frac{X_n}{n}\Big\| > \varepsilon_n\Big\} \qquad and \qquad \mathbb{m}_\epsilon := \max\Big\{n-1 \geqslant 0 \mid \Big\|\frac{X_n}{n}\Big\| > \varepsilon_n\Big\}$$

satisfy

$$\mathbb{E}[\mathcal{S}_{a,1}(\mathcal{O}_\epsilon)] \leqslant \mathbb{E}[\mathcal{S}_{a,1}(\mathbb{m}_\epsilon)] \leqslant C_p K_{a,\epsilon,r}, \tag{4.3}$$

where $\mathcal{S}_{a,1}$ *is defined in* (2.3) *for* $n_0 = 1$*.*

Proof: We use Markov's inequality

$$\mathbb{P}\Big(\frac{\|X_n\|}{n} > \varepsilon_n\Big) \leqslant \varepsilon_n^{-p} \cdot \mathbb{E}\Big[\Big(\frac{\|X_n\|}{n}\Big)^p\Big] = \frac{\mathbb{E}[\|X_n\|^p]}{(\varepsilon_n n)^p},$$

and apply the Burkholder-Rosenthal inequality from Osękowski (2012),

$$\mathbb{E}[\|X_n\|^p] \leq C_p \mathbb{E}\Bigg[\sum_{j=1}^{n} \|\Delta X_j\|^p + \Big(\sum_{j=1}^{n} \mathbb{E}\left[\|\Delta X_j\|^2 \middle| \mathcal{F}_{j-1}\right]\Big)^{\frac{p}{2}}\Bigg] = C_p n \beta_{n,p},$$

where C_p is the value stated in the assertion. □

Example 8. Taking $X_n := \sum_{i=1}^n \Delta_i$ for a centered, i.i.d. sequence $(\Delta_n)_{n\in\mathbb{N}}$, the values $\beta_{n,p}$ in (4.1) equal $\mathbb{E}\|\Delta_1\|^p + n^{\frac{p}{2}-1}(\mathbb{E}[\|\Delta_1\|^2])^{\frac{p}{2}}$. Hence our convergence condition in Theorem 6 turns to

$$K_{a,\epsilon,p} := \sum_{n=1}^{\infty} a_n \sum_{m=n}^{\infty} \frac{\mathbb{E}[\|\Delta_1\|^p] + m^{\frac{p}{2}-1}(\mathbb{E}[\|\Delta_1\|^2])^{\frac{p}{2}}}{\varepsilon_m^p m^{p-1}} < \infty,$$

which is finite whenever $\sum_{n=1}^{\infty} \frac{a_n}{n^{\frac{p}{2}-1}}$ converges (and of course the expectations above are finite). To obtain finite q-th moments of $\mathcal{O}_\epsilon$ and $\mathbb{m}_\epsilon$ for $1 < q$, choose $a_n = n^{q-1}$. Then, the condition for finiteness of $K_{a,\epsilon,p}$ is $1 < q < \frac{p}{2} - 1$, which shows that Theorem 6 includes the result of Estrada and Högele (2022, Theorem 7) (where p and q are switched and the constant C_p differs).

Note that sequences with $\mathbb{E}[|X_1|^p] < \infty$, as in this case here, are trivially bounded in L^p, such that the Baum-Katz-Nagaev type results as given in Subsection 4.2 apply.

4.2. ***The tradeoff for Baum-Katz-Nagaev type strong laws for MDs uniformly bounded in*** L^p.

We start with a version of the classical Baum-Katz-Nagaev strong law of large numbers (Baum and Katz, 1965, Theorem 3), which in general treats renormalized sums of centered i.i.d. random variables $\frac{1}{n^\alpha} X_n$, $X_n = \sum_{i=1}^n \Delta_i$ for some $\alpha \leqslant 1$ in the presence of certain finite moments $\mathbb{E}[|\Delta_i|^p], p > 2$. It is an extension of the strong law by Hsu-Robbins-Erdös (Erdös, 1949; Hsu and Robbins, 1947). Recently, these results were further improved to randomly weighted sums of random variables, see Ma and Sun (2018).

Theorem 7 (Baum-Katz-Nagaev Strong Law).
Consider an i.i.d. family of centered random variables $(\Delta_n)_{n\in\mathbb{N}}$. *Then for any* $\alpha > 1$ *and* $p > 1$ *such that* $\frac{1}{2} < \frac{\alpha}{p} \leqslant 1$ *the following statements are equivalent:*

(a) $\mathbb{E}[|\Delta_1|^p] < \infty$.

(b) $\sum_{n=1}^{\infty} n^{\alpha-2} \cdot \mathbb{P}\Big(\frac{|X_n|}{n} > \eta n^{\frac{\alpha}{p}-1}\Big) < \infty$ *for all* $\eta > 0$.

(c) $\sum_{n=1}^{\infty} n^{\alpha-2} \cdot \mathbb{P}\Big(\max_{k\geqslant n} \frac{|X_k|}{k^{\frac{\alpha}{p}}} > \eta\Big) < \infty$ *for all* $\eta > 0$.

We use the preceding summabilities in order to obtain estimates on the mean deviation frequency.

Corollary 2. *Assume the hypotheses of Theorem 7. We define for* $\eta > 0$, *and* $\alpha, p > 1$ *and* $\epsilon(\alpha,\eta,p) = (\varepsilon_n(\alpha,\eta,p))_{n\in\mathbb{N}}$ *where* $\varepsilon_n(\alpha,\eta,p) := \eta n^{\frac{\alpha}{p}-1}$ *and some* $\varepsilon > 0$ *fixed*

$$\mathcal{O}_{\epsilon,n_0} := \sum_{n=n_0}^{\infty} \mathbf{1}\Big\{\frac{|X_n|}{n} > \varepsilon_n(\alpha,\eta,p)\Big\}, \qquad \mathfrak{m}_{\epsilon,n_0} := \max\Big\{n - n_0 \geqslant 0 \mid \frac{|X_n|}{n} > \varepsilon_n(\alpha,\eta,p)\Big\}.$$

Assume $p > 3$ *and* $\mathbb{E}[|\Delta_1|^p] < \infty$. *Then we have the following tradeoff: For any* $\alpha > 3$ *with* $\frac{1}{2} < \frac{\alpha}{p} \leqslant 1$ *and* $0 \leqslant \tilde{p} < \alpha - 3$ *and we have a constant* $C > 0$ *such that*

$$\sum_{n=n_0}^{\infty} n^{\tilde{p}} \sum_{m=n}^{\infty} \mathbb{P}\Big(\frac{|X_n|}{n} > \varepsilon_n(\alpha,\eta,p)\Big) \leqslant C(\alpha-1)\zeta(\alpha-2-\tilde{p}, n_0) < \infty,$$

we have that

$$\limsup_{n\to\infty} |\frac{X_n}{n}| \cdot \Big(\eta n^{\frac{\alpha}{p}-1}\Big)^{-1} \to 0 \qquad \mathbb{P}\text{-a.s.},$$

and

$$\mathbb{E}[\mathcal{O}_{\epsilon,n_0}^{1+\tilde{p}}] \leqslant \mathbb{E}[\mathfrak{m}_{\epsilon,n_0}^{1+\tilde{p}}] \leqslant C(\alpha-1)\zeta(\alpha-2-\tilde{p}, n_0)$$

such that

$$\mathbb{P}(\mathcal{O}_{\epsilon,n_0} \geqslant k) \leqslant \mathbb{P}(\mathfrak{m}_{\epsilon,n_0} \geqslant k) \leqslant k^{-(1+\tilde{p})} \cdot C(\alpha-1)\zeta(\alpha-2-\tilde{p}, n_0), \qquad k \geqslant 1.$$

Note that the nestedness in part (b), slightly improves our MDF result for the same value of $\alpha > 3$, while the a.s. error tolerance remains the same. An asymptotically better version for large values of k is given in Example 1.

Proof of Corollary 2: We recall Kronecker's lemma (Williams, 1991, (12.7)): For two positive sequences $(b_n)_{n\in\mathbb{N}}$ and $(c_n)_{n\in\mathbb{N}}$, where $\lim_{n\to\infty} b_n = \infty$ we have that

$$\sum_{n=1}^{\infty} \frac{c_n}{b_n} < \infty \qquad \text{implies} \quad \lim_{n\to\infty} \frac{1}{b_n} \sum_{i=1}^{n} c_i = 0.$$

Assume

$$\sum_{n=1}^{\infty} \frac{n^{\alpha-2}}{\frac{1}{p_n}} = \sum_{n=1}^{\infty} n^{\alpha-2} p_n < \infty.$$

Then Kronecker's lemma yields

$$\lim_{n\to\infty} p_n \cdot \sum_{k=1}^{n} k^{\alpha-2} \leqslant C \lim_{n\to\infty} p_n \cdot \int_1^n x^{\alpha-2}dx = \lim_{n\to\infty} \frac{p_n}{\alpha-1}(n^{\alpha-1}-1) = 0.$$

In case of $p_n = \mathbb{P}\Big(\frac{|X_n|}{n} > \eta n^{\frac{\alpha}{p}-1}\Big)$ we have for $c_n = n^{\alpha-2}$ and $b_n = p_n^{-1}$ that

$$\sum_{n=1}^{\infty} \frac{n^{\alpha-2}}{p_n^{-1}} < \infty$$

and the fact that $p_n \searrow 0$ monotonically implies

$$0 = \lim_{n\to\infty} p_n \sum_{k=1}^{n} k^{\alpha-2} \geqslant \lim_{n\to\infty} p_n \int_2^n x^{\alpha-2}dx = \lim_{n\to\infty} p_n \frac{1}{\alpha-1}(n^{\alpha-1}-2^{\alpha-1}) \geqslant 0.$$

Hence $\lim\limits_{n\to\infty} p_n n^{\alpha-1} = 0$. Therefore there exists a $C > 0$ such that

$$p_n \leqslant \frac{C}{n^{\alpha-1}} \qquad \text{for all } n \in \mathbb{N}.$$

In other words, by the summability of Theorem 7(b) there exists some $C > 0$ such that for all $n \in \mathbb{N}$

$$\mathbb{P}\Big(\frac{|X_n|}{n} > \varepsilon_n(\alpha,\eta,p)\Big)\Big) \leqslant \frac{C}{n^{\alpha-1}}. \tag{4.4}$$

Then we apply Example 1 for $\alpha > 3$. This finishes the proof. □

Remark 4. Due to the boundedness of the i.i.d. sequences $(\Delta_n)_{n\in\mathbb{N}}$ in L^q the preceding result yields for $\alpha = 1$ an improvement of the integrability of the overlap $\mathcal{O}_\varepsilon$ in Etemadi's strong law of large numbers (Estrada and Högele, 2022, Theorem 7) from moments of orders $2 \leqslant 1+p < \frac{q}{2} - 1$ to higher moments of orders $2 \leqslant 1+p < q-1$.

It is remarkable that the following result generalizes the preceding strong law to martingale differences, which are uniformly bounded in L^p. A proof is found in Stoica (2007, Theorem), see also Ma and Sun (2018).

Theorem 8 (Baum-Katz-Stoica Strong Law for MDs).
Consider a sequence $(\Delta X_n)_{n\in\mathbb{N}}$ of martingale differences bounded in L^p. Then for all $\eta > 0$ and any $\alpha > 1$ and $p > 1$ such that $\frac{1}{2} < \frac{\alpha}{p} \leqslant 1$ we have that

$$\sum_{n=1}^{\infty} n^{\alpha-2}\mathbb{P}\Big(\frac{|X_n|}{n} \geqslant \eta n^{\frac{\alpha}{p}-1}\Big) < \infty. \tag{4.5}$$

Note that by an application of Kronecker's lemma in the proof of the subsequent Corollary 2 we have the asymptotic decay $\mathbb{P}(\frac{|X_n|}{n} \geqslant \eta n^{\frac{\alpha}{p}-1}) \leqslant Cn^{-(\alpha-1)}$.

There are several extensions of this result applied to arrays of martingales in Hao and Liu (2012). In particular, there are several precise summability results for $q \geqslant 2$, however, the tradeoff relation of Lemma 2 does not apply directly.

Corollary 3. *For $\alpha > 3$, $\eta > 0$ and $p > 1$ such that $\frac{1}{2} < \frac{\alpha}{p} \leqslant 1$ we define $\epsilon = \epsilon(\alpha,\eta,p) = (\varepsilon_n(\alpha,\eta,p))_{n\in\mathbb{N}}$, $\varepsilon_n(\alpha,\eta,p) := \eta n^{\frac{\alpha}{p}-1}$ and $n_0 \in \mathbb{N}$*

$$\mathcal{O}_{\epsilon,n_0} := \sum_{n=n_0}^{\infty} \mathbf{1}\Big\{\frac{|X_n|}{n} \geqslant \varepsilon_n(\alpha,\eta,p)\Big\} \qquad \textit{and} \qquad \mathfrak{m}_{\epsilon,n_0} := \max\Big\{n - n_0 \geqslant 0 \mid \frac{|X_n|}{n} \geqslant \varepsilon_n(\alpha,\eta,p)\Big\}$$

Then for any $0 \leqslant \tilde p < \alpha - 3$ *and* $\sup_{n\in\mathbb{N}} \mathbb{E}[|\Delta X_n|^p] < \infty$ *we have a constant* $C > 0$ *such that*

$$\sum_{n=1}^{\infty} n^{\tilde p} \sum_{m=n}^{\infty} \mathbb{P}\Big(\frac{|X_n|}{n} > \varepsilon_n(\alpha,\eta,p)\Big) \leqslant C(\alpha-1)\zeta(\alpha-2-\tilde p, n_0),$$

we have

$$\frac{X_n}{n} \cdot \varepsilon_n^{-1}(\alpha,\eta,p) \to 0 \qquad \mathbb{P}\text{-a.s.},$$

and

$$\mathbb{E}[\mathcal{O}_{\epsilon,n_0}^{1+\tilde p}] \leqslant \mathbb{E}[\mathrm{m}_{\epsilon,n_0}^{1+\tilde p}] \leqslant C(\alpha-1)\zeta(\alpha-2-\tilde p, n_0).$$

In particular, we have

$$\mathbb{P}(\mathcal{O}_{\epsilon,n_0} \geqslant k) \leqslant \mathbb{P}(\mathrm{m}_{\epsilon,n_0} \geqslant k) \leqslant k^{-(\tilde p+1)} \cdot C(\alpha-1)\zeta(\alpha-2-\tilde p, n_0) \qquad \text{for } k \geqslant 1.$$

Proof of Corollary 3: The proof is similar to the proof of Corollary 2. □

Baum-Katz estimates for martingale differences in the infinite dimensional setting have been shown by Alsmeyer (1990); Dedecker and Merlevède (2008); Giraudo (2019); Hao (2013); Hao and Liu (2014), among others. We state a result from Giraudo (2019, Theorem 2.4 (3)).

Theorem 9 (Baum-Katz type estimate for Banach spaces).
Consider a martingale difference sequence $(\Delta X_n)_{n\in\mathbb{N}}$ *in a 2-smooth Banach space* B, *let* $p > 2$ *and* $\alpha \in (\frac{1}{2}, 1]$. *Assume that* $(\|\Delta X_n\|)_{n\in\mathbb{N}}$ *is identically distributed and* $\mathbb{E}[\|\Delta X_1\|^p] < \infty$. *Then there is a constant* $C(p,B)$ *such that*

$$\sum_{n=1}^{\infty} n^{p(\alpha-\frac{1}{2})-1} \mathbb{P}\left(\frac{\max_{1\leq k\leq n}\|X_k\|}{n} > \eta n^{\alpha-1}\right) < C(p,B)\frac{\mathbb{E}[\|X_1\|]^p}{\eta} \qquad \text{for all } \eta > 0.$$

The respective MDF quantification reads as follows.

Corollary 4. *With the assumptions of Theorem 9 with initial index* $n_0 \in \mathbb{N}_0$, *assume* $\alpha \in (\frac{1}{2}, 1]$, $p > \frac{2}{\alpha - \frac{1}{2}}$ *and consider* $\epsilon = (\varepsilon_n)_{n\in\mathbb{N}}$ *with* $\varepsilon_n = \varepsilon_n(\alpha,\eta,p) := \eta n^{\alpha-1}$. *Then, for any* $0 < \tilde p < p(\alpha - \frac{1}{2}) - 2$, *we have the following:*

(a) *There is a constant* $C > 0$ *such that*

$$\sum_{n=n_0}^{\infty} n^{\tilde p} \sum_{m=n}^{\infty} \mathbb{P}\Big(\frac{|X_n|}{n} > \varepsilon_n(\alpha,\eta,p)\Big) \leqslant C(p(\alpha-\tfrac{1}{2})-1)\zeta(p(\alpha-\tfrac{1}{2})-2-\tilde p, n_0).$$

(b) *We have the convergence*

$$\max_{n_0\leq k\leq n} \frac{\|X_k\|}{n} \cdot \varepsilon_n^{-1}(\alpha,\eta,p) \to 0 \qquad \mathbb{P}\text{-a.s.}.$$

(c) *The moments*

$$\mathbb{E}[\mathcal{O}_{\epsilon,n_0}^{1+\tilde p}] \leqslant \mathbb{E}[\mathrm{m}_{\epsilon,n_0}^{1+\tilde p}] \leqslant C(p(\alpha-\tfrac{1}{2})-1)\zeta(p(\alpha-\tfrac{1}{2})-2-\tilde p, n_0) \quad \text{are finite.}$$

(d) *For any* $k \geqslant 1$ *we have*

$$\mathbb{P}(\mathcal{O}_{\epsilon,n_0} \geqslant k) \leqslant \mathbb{P}(\mathrm{m}_{\epsilon,n_0} \geqslant k) \leqslant k^{-(\tilde p+1)} \cdot C(p(\alpha-\tfrac{1}{2})-1)\zeta(p(\alpha-\tfrac{1}{2})-2-\tilde p, n_0).$$

Proof: The proof is again similar to the one of Corollary 2. □

4.3. ***The tradeoff in a strong law for MDs with uniformly bounded exponential moments***.

For the exponential case we cite the following large deviations type result for martingales.

Theorem 10 (**Lesigne and Volný** (**2001**, **Theorem 3.2**)).
Let $X = (X_n)_{n\in\mathbb{N}_0}$ be a martingale with $X_0 = 0$ with respect to a filtration $\mathbb{F}$. Set $\Delta X_n := X_n - X_{n-1}$ for $n \geqslant 1$. Assume the existence of some $K > 0$ and $\lambda > 0$ such that $k \geqslant 1$, $\mathbb{E}\left[e^{\lambda|\Delta X_k|}\right] < K$. Then for any positive number $\delta \in (0,1)$ there exists a positive integer $n_0 \in \mathbb{N}$ such that for all $n \geqslant n_0$

$$\mathbb{P}\left(\left|\frac{X_n}{n}\right| > \varepsilon\right) \leqslant e^{-\frac{1-\delta}{2}\lambda^{\frac{2}{3}}\varepsilon^{\frac{2}{3}}n^{\frac{1}{3}}}. \tag{4.6}$$

In particular, $\lim_{n\to\infty} \frac{X_n}{n} = 0$ a.s.

Corollary 5. *Under the assumptions of Theorem 10, we get the following tradeoff for $\lim_{n\to\infty} \frac{X_n}{n} = 0$ $\mathbb{P}$-a.s. For any positive nonincreasing $\epsilon = (\varepsilon_n)_{n\geqslant 1}$ and $a = (a_n)_{n\in\mathbb{N}}$ positive nondecreasing such that that*

$$K(a,\epsilon,\lambda,\delta) := \sum_{n=1}^{\infty} a_n \sum_{m=n}^{\infty} e^{-\frac{1-\delta}{2}\lambda^{\frac{2}{3}}\varepsilon_m^{\frac{2}{3}}m^{\frac{1}{3}}} < \infty$$

we have that

$$\limsup_{n\to\infty} \frac{X_n}{n}\cdot\varepsilon_n^{-1} \leqslant 1 \qquad \mathbb{P}\text{-a.s.} \tag{4.7}$$

and

$$\mathbb{E}[\mathcal{S}_a(\mathcal{O}_\epsilon)] \leqslant \mathbb{E}[\mathrm{m}_a(\mathcal{O}_\epsilon)] \leqslant K(a,\epsilon,\lambda,\delta), \tag{4.8}$$

where $\mathcal{S}_a$ is defined as in Lemma 2.

Remark 5. Note that in this generality, Theorem 10 is the best one can achieve. In Lesigne and Volný (2001), the authors construct a martingale X in the context of ergodic dynamical systems such that $\mathbb{E}[e^{|\Delta X_k|}] < \infty$ for all k but still there is a constant $c > 0$ such that

$$\mathbb{P}\left(\left|\frac{X_n}{n}\right| > 1\right) > e^{-cn^{1/3}}$$

for infinitely many n.

The context of ergodic dynamical systems is another source for martingale differences, see Lesigne and Volný (2001); Volný (1989, 1993), where the following examples emerge:

Example 9. Let $(\Omega,\mathcal{F},\mathbb{P})$ be a probability space and let $T\colon \Omega \to \Omega$ be a bijective, bimeasurable, measure preserving mapping. Assume that $\mathcal{I}$ is the σ-algebra of all sets A such that $TA = A$. Assume that for all $A \in \mathcal{I}$ we have $\mathbb{P}(A) \in \{0,1\}$ (i.e. $\mathbb{P}$ is ergodic). Let $\mathcal{M}$ be a T-invariant σ-algebra, that is $\mathcal{M} \subseteq T^{-1}\mathcal{M}$. Let now $m = (m_k)_{k\geqslant 1}$ be a sequence of stationary (i.e. identically distributed) martingale differences with respect to the filtration $(T^{-n}\mathcal{M})_{n\geqslant 0}$. In Volný (1989) it is shown that then, m is of the form

$$m_k = \mathbb{E}[f|T^{k-i}\mathcal{M}] - \mathbb{E}[f|T^{k-i+1}\mathcal{M}], \quad k \geqslant 0,$$

for some $i \geqslant 0$ and $f \in L^1$.

Naturally, higher integrabilities such as L^p or exponential integrability for m are given by properties of the function f which then brings us in the situations of Theorem 6, and Theorem 10: Indeed, if $f \in L^p, p \geqslant 1$, it follows

$$\mathbb{E}[|m_k|^p]^{\frac{1}{p}} = \mathbb{E}\left[\left|\mathbb{E}[f|T^{k-i}\mathcal{M}] - \mathbb{E}[f|T^{k-i+1}\mathcal{M}]\right|^p\right]^{\frac{1}{p}} \leqslant 2\mathbb{E}[|f|^p]^{\frac{1}{p}} < \infty,$$

which is just Minkowski's inequality. For exponential moments, assume that there is $\lambda > 0$ such that $\mathbb{E}[e^{2\lambda|f|}] < \infty$. Then we have

$$\mathbb{E}[e^{\lambda|m_k|}] = \mathbb{E}\Big[e^{\lambda\left|\mathbb{E}[f|T^{k-i}\mathcal{M}]-\mathbb{E}[f|T^{k-i+1}\mathcal{M}]\right|}\Big] \leqslant \mathbb{E}\Big[e^{\lambda\left(\mathbb{E}[|f||T^{k-i}\mathcal{M}]+\mathbb{E}[|f||T^{k-i+1}\mathcal{M}]\right)}\Big]$$
$$\leqslant \frac{1}{2}\left(\mathbb{E}\Big[e^{2\lambda\left(\mathbb{E}[|f||T^{k-i}\mathcal{M}]\right.}\Big] + \mathbb{E}\Big[e^{2\lambda\left(\mathbb{E}[|f||T^{k-i+1}\mathcal{M}]\right.}\Big]\right) \leqslant \mathbb{E}[e^{2\lambda|f|}] < \infty.$$

Here, the first estimate in the second line is Young's inequality, the second one is Jensen's inequality for conditional expectations (with subsequent use of the tower property).

Theorem 10 is used in Theorem 12 in order to quantify the a.s. convergence of M-estimators in Subsection 5.3.

Remark 6. If we consider the situation of $X_n = \sum_{i=1}^n \Delta X_i$ for a centered i.i.d. sequence $(\Delta_i)_{i\in\mathbb{N}}$ with exponential moments, we obtain by Cramér's theorem a large deviations principle (LDP), with the upper bound

$$\mathbb{P}\left(\left|\frac{X_n}{n}\right| > \varepsilon\right) \leqslant \exp\Big(-n \inf_{|y|>\varepsilon} \Lambda^*_{\Delta_1}(y)\Big),$$

where the exponent is given by the good rate function

$$\Lambda^*_{\Delta_1}(y) = \inf_{t\in\mathbb{R}} ty - \Lambda_{\Delta_1}(t), \qquad \Lambda_{\Delta_1}(t) = \ln(\mathbb{E}[e^{t\Delta_1}]).$$

For further examples and comments on this setting we refer to Estrada and Högele (2022, Subsection 3.2.2). This result is used in Theorem 12.

Example 10 (**Closed martingales with exponentially integrable limit**). Let X be a centered random variable such that there is $\lambda > 0$ with $\mathbb{E}\exp(2\lambda|X|) < \infty$, and let $\mathbb{F} = (\mathcal{F}_n)_{n\geqslant 0}$ be a given filtration. Then the sequence given by $X_n := \mathbb{E}[X|\mathcal{F}_n]$ forms a martingale. We have for the differences that for $k \geqslant 1$,

$$\mathbb{E}\Big[\exp(\lambda|dX_k|)\Big] = \mathbb{E}\Big[\exp(\lambda|dX_k|)\Big] = \mathbb{E}\Big[\exp\left(\lambda\left|\mathbb{E}[X|\mathcal{F}_k] - \mathbb{E}[X|\mathcal{F}_{k-1}]\right|\right)\Big]$$
$$\leqslant \mathbb{E}\Big[\exp\left(\lambda\mathbb{E}[|X||\mathcal{F}_k] + \mathbb{E}[|X||\mathcal{F}_{k-1}]\right)\Big],$$

which, by Young's inequality, is smaller than

$$\frac{1}{2}\mathbb{E}\Big[\exp\left(2\lambda\mathbb{E}[|X||\mathcal{F}_k]\right)\Big] + \frac{1}{2}\mathbb{E}\Big[\exp\left(2\lambda\mathbb{E}[|X||\mathcal{F}_{k-1}]\right)\Big],$$

which can in turn be estimated using the conditional Jensen inequality via

$$\frac{1}{2}\mathbb{E}\Big[\exp\left(2\lambda\mathbb{E}[|X||\mathcal{F}_k]\right)\Big] + \frac{1}{2}\mathbb{E}\Big[\exp\left(2\lambda\mathbb{E}[|X||\mathcal{F}_{k-1}]\right)\Big] \leqslant \mathbb{E}\Big[\exp(2\lambda|X|)\Big] = K < \infty.$$

Hence, the martingale $(X_n)_{n\geqslant 0}$ satisfies the assumptions of Theorem 10, and we obtain the Weibull-type moments and decay rates for the overlap and the modulus. See Example 3.

5. Applications

5.1. *The tradeoff in multicolor Pólya urn models.*

In Example 5 we presented the exponential MDF convergence for the two-color Pólya's urn. However, a natural generalization involves introducing a broader range of types or colors for the balls, each with its own replacement rules. For a comprehensive survey about the applications of urn-like models, see Johnson and Kotz (1977, Chapter 5).

Definition 3. A **generalized multicolor Pólya urn process** is given by a d-dimensional Markov chain $(X_n)_{n\in\mathbb{N}_0}$, $X_n := (X_{n,1},\dots,X_{n,d})$, with transition probabilities

$$\mathbb{P}(X_{n+1} = X_n + e_i R | X_1,\dots,X_n) = \frac{X_{n,i}}{\sum_{k=1}^d X_{n,k}}, \quad i = 1,\dots,d,$$

where $e_i = (0\dots1\dots0)$ is the row unit vector with 1 in the i-th coordinate and $R = (r_{i,j})$ is a $d\times d$ deterministic matrix with integer coefficients, called replacement matrix.

In the random vector X_n, the i-th component $X_{n,i}$ represents the number of balls in the urn with color i at the n-th step of the process. Moreover, at each step a ball is drawn from the urn at random, its color is recorded and then returned along with $r_{i,j}$ additional balls of color j, $j = 1,\dots,d$. Note that any negative $r_{i,j}$ represents balls being taken away from the urn. Problems may arise when there are negative replacements, which might leave some colors to run out of balls. Therefore, Gouet (1997) introduces the notion of tenable generalized multicolor Pólya urn, which guarantees the long-term well-definedness of the process.

A **tenable** generalized multicolor Pólya urn (TGMPU) process is a generalized multicolor Pólya urn with the following additional hypotheses on R.

(a) $r_{ij} \geqslant 0$ for all $i \neq 0$.
(b) $\sum_{j=1}^d r_{ij} = s \geqslant 0$ for all $i = 1,\dots,d$.
(c) $r_{ii} < 0$ implies that r_{ii} is a divisor (modulo sign) of $r_{ki} = 1,\dots,d$.

The asymptotic behavior of such types of multicolored Pólya urn has been extensively investigated using the martingale version of the Borel-Cantelli lemma (Bagchi and Pal, 1985; Friedman, 1949; Gouet, 1989, 1993; Najock and Heyde, 1982). For the replacement matrix R, there is a natural notion of connected components and irreducibility of the submatrices of R, which underpins the fundamental theory by Seneta (2006) and allows to give a normal form of R. The long-term survival of $X_n/\sum_{i=1}^d X_{i,n}$ is known to be dominated by the irreducible components of R, whose dominant eigenvalue equals, precisely, to the row sum s. Those irreducible components are called **supercolors**. More precisely in Gouet (1997, Theorem 3.1) it is shown that the vector $X_n/\sum_{i=1}^d X_{n,i}$ converges a.s. to some lacunary random row vector X_∞ which is distributed according to a Dirichlet mixture of the dominant eigenvectors (which are nonnegative and sum up to 1) of the supercolors. All other entries, which correspond to transient states and hence irreducible components with leading eigenvalues $\tau < s$, are equal to 0 in X_∞. The parameters depend only on the initial total number of balls in each of the supercolors, the row sum s of R and the initial total number of balls in the urn.

Let $(Y_n)_{n\geqslant 1}$ be the sequence of independent draws from the urn, given by the d-dimensional vectors $Y_n = (Y_{n,1},\dots,Y_{n,d})$ representing the number of balls of each color at stage n chosen according to the replacement matrix R satisfying (a)-(c), and $T_n = \sum_{i=1}^d Y_{n,i}$. In Gouet (1997, Proposition 4.1)(i) it is shown, that up to a reordering of states, the $\mathbb{R}^d$ valued process $X_n := Y_n/T_n$ has the following shape: It is a TGMPU process with r supercolors for some $r \in \{1,\dots,d\}$. We denote the number of colors composing the i-th supercolor by d_i. Then, $X_n = (M_n, S_n), n \in \mathbb{N}$, where $(M_n)_{n\in\mathbb{N}_0}$ is a $\mathbb{R}^{\sum_{i=1}^{r+1} d_i}$-valued nonnegative martingale which converges a.s. to a random vector

$$M_\infty \sim \sum_{i=1}^{r+1} X_{\infty,i}\cdot u_i,$$

where $u_i \in \mathbb{R}^r$ consists of 0 up to the i-th entry, which is given by the dominant d_i- eigenvector of the i-th supercolor. We quantify the a.s. convergence result of $X_n = (M_n, S_n) \to (M_\infty, 0)$ for the vector-valued martingale $(M_n)_{n\in\mathbb{N}_0}$.

For convenience, we assume that the normal form of the matrix consists of a finite union of irreducible components. Under this assumption we have that $X_n = M_n$ and our setting falls under the hypotheses of Corollary 1. Note that similarly to (3.9) the increments satisfy

$$|X_n - X_{n-1}| = \frac{1}{T_n}|Y_n - Y_{n-1}| \leqslant \frac{C}{T_n} = \frac{C}{\sum_{k=1}^d Y_{0,k} + ns} \leqslant \Big(\frac{C+1}{s}\Big)\frac{1}{n} =: c_n, \qquad n \geqslant 1,$$

and the increments are almost surely square summable since for all $n \geqslant 2$ we have

$$r(n) = \sum_{k=n+1}^{\infty} c_k^2 \leqslant \Big(\frac{C+1}{s}\Big)^2 \frac{1}{n}.$$

Corollary 1 implies for each component i of $(M_n)_{n\in\mathbb{N}_0}$ the following tradeoff: For all $a_i = (a_{n,i})_{n\in\mathbb{N}_0}$ positive, nondecreasing and $\epsilon_i = (\varepsilon_{n,i})_{n\in\mathbb{N}}$ positive, nonincreasing such that for

$$K(a_i, \epsilon_i) := \sum_{n=n_0}^{\infty} a_{n,i} \sum_{m=n}^{\infty} \exp\Big(-\frac{\varepsilon_{m,i}^2(m+1)s^2}{(C+1)^2}\Big) < \infty, \tag{5.1}$$

we have

$$\limsup_{n\to\infty} |M_{n,i} - M_{\infty,i}| \cdot \varepsilon_{n,i}^{-1} \leqslant 1, \qquad \mathbb{P}\text{-a.s.},$$

and

$$\mathbb{E}[\mathcal{S}_{a,n_0,i}(\mathcal{O}_{\epsilon,n_0,i})] \leqslant \mathbb{E}[\mathcal{S}_{a,n_0,i}(\mathrm{m}_{\epsilon,n_0,i})] \leqslant K(a_i,\epsilon_i),$$

with

$$\mathbb{P}(\mathcal{O}_{\epsilon,n_0,i} \geqslant k) \leqslant \mathbb{P}(\mathrm{m}_{\epsilon,n_0,i} \geqslant k) \leqslant \mathcal{S}_{a,n_0,i}^{-1}(k) K(a_i,\epsilon_i). \tag{5.2}$$

We also refer to Franchini (2017), which derives a large deviations principle for multicolor Pólya urns, which also allows for similar (asymptotic) exponential quantifications.

5.2. *The tradeoff for the Generalized Chinese Restaurant Process (GCRP).*

In this subsection, we show how state-of-the-art results in machine learing can be further sharpened in a useful way. In Oliveira et al. (2022, Thm. 3.2) the authors show a non-asymptotic random concentration result for the GCRP. Recall that the GCRP generates a sequence of random partitions $\mathcal{P}_n$ of $[n] := \{1, \dots, n\}$ for $n \in \mathbb{N}$. Their results study the case where the growth of maximal components in $\mathcal{P}_n$ behaves like n^α, $n \in \mathbb{N}$ for a parameter $\alpha \in (0,1)$, with a particular interest in the concentration limits of the total number of components with size k in each $\mathcal{P}_n$, that is:

$$N_n(k) := |\{A \in \mathcal{P}_n : |A| = k\}|.$$

More precisely, the model is given as a Markov chain $\mathcal{P}_1, \mathcal{P}_2, \mathcal{P}_3, \dots$, where, for each $n \in \mathbb{N}$, $\mathcal{P}_n$ is a partition of $[n]$ composed by $V_n := |\mathcal{P}_n|$ disjoint parts $A_{i,n}$, $i = 1, \dots, V_n$. Then, the process will evolve following a "Chinese restaurant" metaphor. In it, $A_{i,n}$ are the tables occupied by customers 1 to n (who come in sequentially), V_n represents the total number of occupied tables and $\mathcal{P}_n$ describes the table arrangements, which follow that

(a) Customer 1 sits by herself (i.e. $\mathcal{P}_1 = \{\{1\}\}$).
(b) Given $\mathcal{P}_1, \dots, \mathcal{P}_n$, $\mathcal{P}_{n+1}$ is set up by choosing where to sit customer $n+1$. That is, all the other customers will remain in their previously assigned tables, while customer $n+1$ will sit either at an occupied table $A_{i,n}$ with probability

$$\mathbb{P}(n+1 \in A_{i,n+1} \mid \mathcal{P}_1, \dots, \mathcal{P}_n) = \frac{|A_{i,n}| - \alpha}{n+\theta}, \quad \text{for } i = 1, \dots, V_{n-1} \text{ and } \alpha, \theta \in \mathbb{R},$$

or, alternatively, sit at a new table by herself with probability

$$\mathbb{P}(n+1 \in A_{n+1,n+1} \mid \mathcal{P}_1, \dots, \mathcal{P}_n) = \frac{\alpha V_n + \theta}{n+\theta}, \quad \text{for } \alpha, \theta \in \mathbb{R}.$$

Note that for the first scenario in (b) $V_{n+1} = V_n$, while for the latter, $V_{n+1} = V_n + 1$. However, most of the results in Oliveira et al. (2022) will be set up for the normalized version V_n/φ_n, where

$$\varphi_n := \frac{\Gamma(1+\theta)}{\Gamma(1+\theta+\alpha)} \frac{\Gamma(n+\alpha+\theta)}{\Gamma(n+\theta)}.$$

In particular, this is because the limit $V_* := \lim_{n\to\infty} V_n/\varphi_n$ exists and is almost surely positive, with an explicit density. Furthermore, the authors proposed a quantification of the almost sure convergence for V_n/φ_n, which we show to be fit and quantifiable within the framework of Lemma 2.

Theorem 11.
Consider a realization $(\mathcal{P}_n)_{n\in\mathbb{N}}$ of the GCRP with parameters $\alpha \in (0,1)$ and $\theta > -\alpha$. Then there exist constants $n_0 = n_0(\alpha,\theta) \in \mathbb{N}$ and $C = C(\alpha,\theta)$ such that the following holds for all $n \geqslant n_0$. For any nondecreasing positive sequence $A = (A_n)_{n\geqslant n_0}$, and nonincreasing positive sequence $\epsilon = (\varepsilon_n)_{n\geqslant n_0} > 0$ we define

$$k_{\epsilon,n} := \left\lceil \frac{\varepsilon_n n^{\frac{1}{2}\frac{\alpha}{\alpha+2}}}{\ln(n)^{\frac{1}{\alpha+2}}} \right\rceil, \quad c(\alpha,\theta) := \frac{\alpha\Gamma(1+\theta)}{\Gamma(1-\alpha)\Gamma(1+\alpha+\theta)} > 0, \quad \textit{and}$$

$$E_n(A,\epsilon) := \left\{ \forall k \in \{1,\ldots,k_{\varepsilon_n,n}\} : \left| N_n(k) - c(\alpha,\theta)\frac{\Gamma(k-\alpha)}{\Gamma(k+1)} V_* n^\alpha \right| \leqslant C \frac{\Gamma(k-\alpha)}{\Gamma(k+1)} n^\alpha \varepsilon_n^{\alpha+2} \left(1 + \frac{A_n}{\ln(n)}\right) \right\}.$$

Then we have

$$\mathbb{P}(E_n^c) \leqslant e^{-A_n}. \tag{5.3}$$

Under the additional condition that e^{-A_n} is summable we have the following tradeoff: For $\mathcal{O}_A := \sum_{n=n_0}^\infty \mathbf{1}(E_n^c)$ and $\mathrm{m}_A(\omega) := \max\{n \geqslant n_0 \mid \omega \in E_n^c\}$ and any sequence $a = (a_n)_{n\geqslant n_0}$ of nonnegative, nondecreasing weights such that

$$C_{a,A} := \sum_{n=n_0}^\infty a_n \sum_{m=n}^\infty e^{-A_m} < \infty,$$

we have that for $\mathcal{S}_{a,0}$ defined in (2.3) and calculated explicitly in Example 2 the tradeoff satisfies

$$\limsup_{n\to\infty} \sup_{k\in\{1,\ldots,k_{\varepsilon_n,n}\}} \left| N_n(k) - c(\alpha,\theta)\frac{\Gamma(k-\alpha)}{\Gamma(k+1)} V_* n^\alpha \right| \cdot \left(C \frac{\Gamma(k-\alpha)}{\Gamma(k+1)} n^\alpha \varepsilon_n^{\alpha+2} \left(1 + \frac{A_n}{\ln(n)}\right)\right)^{-1} \leqslant 1 \quad \mathbb{P}\text{-}a.s.$$

and

$$\mathbb{E}[\mathcal{S}_a(\mathcal{O}_A)] \leqslant \mathbb{E}[\mathcal{S}_a(\mathrm{m}_A)] \leqslant C_{a,A}.$$

The proof is a direct consequence of the exponential decay (5.3) and Lemma 2.

Remark 7. Important particular cases which highlight the play between the asymptotic a.s. error bound and the MDF statistics are the following. Due to the asymptotics

$$\frac{\Gamma(k_{\varepsilon_n,n}-\alpha)}{\Gamma(k_{\varepsilon_n,n}+1)} n^\alpha \qquad \text{of order} \qquad k_{\varepsilon_n,n}^{-(1+\alpha)} = \left\lceil \frac{\varepsilon_n n^{\frac{\alpha}{(2\alpha+4)}}}{(\ln(n)^{\frac{1}{\alpha+2}}} \right\rceil^{-(1+\alpha)}$$

only sequences of A with an asymptotic behavior of

$$A = O(\varepsilon_n^{\alpha+3} n^{\frac{\alpha}{(2\alpha+4)}} \ln(n)^{\frac{\alpha+1}{\alpha+2}}) \qquad \text{and} \qquad \frac{1}{A} = O((2+\delta)\ln(n)), \text{ for } \delta > 0,$$

are meaningful. For fixed $\varepsilon_n = \varepsilon > 0$, extremal cases for A are given by:

(a) $A_n = n^{\frac{\alpha}{(2\alpha+4)}} \ln(n)^{\frac{\alpha+1}{\alpha+2}}$ and $p \in (0,1)$, yield by Example 3 a constant $K(p,\alpha) > 0$ such that

$$\mathbb{E}[e^{p(\mathcal{O}_A-1)^{\frac{1}{2}\frac{\alpha}{\alpha+2}}}] \leqslant \mathbb{E}[e^{p(\mathbb{m}_A-1)^{\frac{1}{2}\frac{\alpha}{\alpha+2}}}] \leqslant K(p,\alpha),$$

and $d, D > 0$ such that for $k \geqslant 2$

$$\mathbb{P}(\mathcal{O}_A \geqslant k) \leqslant \mathbb{P}(\mathbb{m}_A \geqslant k) \leqslant (d + D(k-1)^{2-\frac{1}{2}\frac{\alpha}{\alpha+2}} e^{-p(k-1)^{\frac{1}{2}\frac{\alpha}{\alpha+2}}},$$

while there is a constant $\tilde{C} > 0$ such that

$$\limsup_{n\to\infty} \sup_{k\in\{1,\ldots,k_{\varepsilon,n}\}} \big|N_n(k) - c(\alpha,\theta)\frac{\Gamma(k-\alpha)}{\Gamma(k+1)} V_* n^\alpha\big| \leqslant \tilde{C}C\varepsilon \qquad \mathbb{P}\text{-a.s.}$$

(b) $A_n = (2+\delta)\ln(n)$. Then Example 1 yields

$$\limsup_{n\to\infty} \sup_{k\in\{1,\ldots,k_{\varepsilon_n,n}\}} \big|N_n(k) - c(\alpha,\theta)\frac{\Gamma(k-\alpha)}{\Gamma(k+1)} V_* n^\alpha\big| \cdot \left(\frac{n^{\frac{\alpha}{(2\alpha+4)}}}{(\ln(n)^{\frac{1}{\alpha+2}}}\right)^{(1+\alpha)} \leqslant C(3+\delta)\varepsilon \qquad \mathbb{P}\text{-a.s.}$$

and a constant $K(\alpha,\delta)$ such that

$$\mathbb{E}[\mathcal{O}_A^{1+\delta}] \leqslant \mathbb{E}[\mathbb{m}_A^{1+\delta}] \leqslant K(\alpha,\delta).$$

For variable error tolerance $\epsilon = (\varepsilon_n)_{n\in\mathbb{N}}$ even finer tradeoffs between the a.s. asymptotic error tolerance and the mean deviation frequency can be derived.

5.3. ***A tradeoff quantification of a.s. convergent M-estimators.***

M-estimators are one of the most elementary classes of point estimators in statistics based on the law of large numbers. So far, it was complicated to quantify the respective results for the strong law, with the results in Subsection 4.2, however, we may quantify the tradeoff between the a.s. rate of convergence and its mean deviation frequency.

Definition 4. For $\ell \in \mathbb{N}$ we call a sequence of random variables $(Y_n)_{n\in\mathbb{N}}$ weakly ℓ-stationary if $\mathbb{E}[|Y_1|^\ell] < \infty$ and

$$\mathbb{E}[Y_n^j] = \mathbb{E}[Y_1^j], \qquad \text{for all } n \in \mathbb{N}, \quad j = 1,\ldots,\ell.$$

Remark 8.

(a) The most natural example are sequences of i.i.d. random variables $(Y_i(\theta))_{i\in\mathbb{N}}$. For instance given by strongly irreducible and positive recurrent homogeneous Markov chains on a countable state space $\mathbb{S}$ starting in its dynamical equilibrium (stationary distribution) π, both of which are strictly stationary.

(b) Let us clarify the scope of the results of this section. For an i.i.d. sequence $(Y_i)_{i\in\mathbb{N}}$ with third moments we have by Kolmogorov's strong law that

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n Y_i = \mathbb{E}[Y_1].$$

More over it is clear that $(Y_i^j)_{i\in\mathbb{N}}$, $j = 1,2$ is also an i.i.d. family of random variables which has first moments. Hence, again by Kolmogorov's strong law, we obtain

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n Y_i^j = \mathbb{E}[Y_1^j].$$

However for a sequence $(\Delta_i X)_{i\in\mathbb{N}}$ of martingale differences with finite third moments we have that $X_n = \sum_{i=1}^n \Delta_i X$ is a martingale and martingale strong laws apply for the process $\frac{1}{n}X_n$ in that under the assymption of weakly 3-stationarity

$$\frac{1}{n}X_n \to \mathbb{E}[\Delta_1 X] \qquad \text{in probability, as } n \to \infty.$$

If we consider now the sequence $((\Delta_i X)^j)_{i\in\mathbb{N}}$, $j = 2, 3$ it is not any more a sequence of martingale differences and no law of large numbers can be guaranteed in general. While for $j = 2$ there still is a theory available due to the Doob-Meyer decomposition for the quadratic variation, for $j = 3$ (or even higher moments) this cannot be guaranteed in general. For this reason we present our results for independent weakly ℓ-stationary, though not necessarily strictly stationary (i.i.d.) increments.

The basic setup: Given $k \geqslant 1$ and an open bounded subset $\Theta \subseteq \mathbb{R}^\ell$ of parameters, $\theta = (\theta_1, \ldots, \theta_\ell) \in \Theta$, we consider a sequence of weakly ℓ-stationary independent random variables $(Y_i(\theta))_{i\in\mathbb{N}}$ with values in $\mathbb{R}$ and distributions $\mu_i(\theta) := \mathbb{P}_{Y_i(\theta)}$ which depend on θ. For any $1 \leqslant j \leqslant \ell$, $n \in \mathbb{N}$, we set $M^j(\theta) := \mathbb{E}[Y_n^j(\theta)]$. Note that due to the weak ℓ-stationarity these moments are well-defined, and independent of $n \in \mathbb{N}$. Now we define the complete vector of moments by

$$\theta \mapsto M(\theta) := (M_1(\theta), \ldots, M_\ell(\theta)).$$

Consider for any fixed $\theta_0 \in \Theta$ the M-estimator of θ_0 by $\hat{\theta}_n(\theta_0) := M^{-1}(\bar{X}_n(\theta_0)) \in \mathbb{R}^\ell$, $n \in \mathbb{N}$. For convenience we write $\theta_0 = (\theta_{0,1}, \ldots, \theta_{0,\ell})$ and

$$\hat{\theta}_n(\theta_0) = (\hat{\theta}_{n,1}(\theta_0), \ldots, \hat{\theta}_{n,\ell}(\theta_0)), \qquad \text{for } \theta \in \Theta.$$

For any $1 \leqslant j \leqslant \ell$ we set $X_{n,j}(\theta) := \sum_{i=1}^n Y_i^j(\theta)$ and $\bar{X}_{n,j} := \frac{X_{n,j}}{n}$ and define the complete vector of higher order sample means by

$$\bar{X}_n(\theta) := (\bar{X}_{n,1}(\theta), \ldots, \bar{X}_{n,\ell}(\theta)).$$

Assumptions:

(i) Let $\sup_{\theta\in\Theta} \mathbb{E}[|Y_i(\theta)|^q] < \infty$ for some $q > \ell$ and all $i \in \mathbb{N}$.
(ii) The mapping $\Theta \ni \theta \mapsto M(\theta) \in M(\Theta) \subseteq \mathbb{R}^\ell$ is continuous and bijective.
(iii) The inverse M^{-1} is continuously differentiable in Θ.

Reduction to the law of large numbers: We fix some $\theta_0 \in \Theta$. By (ii) and (iii) There is $\varepsilon > 0$ sufficiently small such that

$$\lambda = \lambda(\varepsilon) = \min\{|\mu|, \mu \in \operatorname{spec}(D_{\theta_0} M)\} - \varepsilon > 0, \tag{5.4}$$

where $D_{\theta_0} M$ is the Jacobi matrix of M at the foot point θ_0 and $B_\delta(x) = \{\|x - z\| < \delta\} \subseteq \mathbb{R}^\ell$. For any $\varepsilon > 0$ sufficiently small, there are $\delta_1, \delta_2 \in (0, 1)$ such that

$$\begin{aligned} A_n(\varepsilon) &:= \{\|\hat{\theta}_n - \theta_0\| \geqslant \varepsilon\} = \{M^{-1}(\bar{X}_n(\theta_0)) \in B^c_\varepsilon(\theta_0)\} = \{\bar{X}_n(\theta_0) \in M(B^c_\varepsilon(\theta_0))\} \\ &= \{\bar{X}_n(\theta_0) \in M(\Theta) \setminus M(B_\varepsilon(\theta_0))\} \subseteq \{\bar{X}_n(\theta_0) \in M(B_\varepsilon(\theta_0))^c\} \\ &\subseteq \{\bar{X}_n(\theta_0) \in \big((D_{\theta_0} M)(B_{\delta_1\varepsilon}(\theta_0))\big)^c\} \subseteq \{\bar{X}_n(\theta_0) \in (D_{\theta_0} M) B^c_{\delta_1\varepsilon}(M(\theta_0))\} \\ &\subseteq \{\bar{X}_n(\theta_0) \in B^c_{\delta_1\delta_2\cdot\lambda\cdot\varepsilon}(M(\theta_0))\} = \{\|\bar{X}_n(\theta_0) - M(\theta_0)\| \geqslant \delta_1\delta_2 \cdot \lambda \cdot \varepsilon\}. \end{aligned} \tag{5.5}$$

Consequently, for

$$\mathbb{P}(\|\hat{\theta}_n(\theta_0) - \theta_0\| > \varepsilon) \leqslant \mathbb{P}(\|\bar{X}_n(\theta_0) - M(\theta_0)\| > \delta_1\delta_2 \cdot \lambda \cdot \varepsilon). \tag{5.6}$$

We denote by $B_n(\varepsilon) := \{\|\bar{X}_n(\theta_0) - M(\theta_0)\| \geqslant \delta_1\delta_2 \cdot \lambda \cdot \varepsilon\}$ which results with the help of (5.5) in

$$\mathcal{O}_{\varepsilon,n_0} = \sum_{n=n_0}^{\infty} \mathbf{1}(A_n(\varepsilon)) \leqslant \mathrm{m}_{\varepsilon,n_0} = \sum_{n=n_0}^{\infty} \mathbf{1}\Big(\bigcup_{m\geqslant n} A_m(\varepsilon)\Big) \leqslant \sum_{n=n_0}^{\infty} \mathbf{1}\Big(\bigcup_{m\geqslant n} B_m(\varepsilon)\Big) =: \tilde{\mathrm{m}}_{\varepsilon,n_0},$$

by monotonicity. We now define for some $n_0 \in \mathbb{N}$, and some positive, nonincreasing sequence $\epsilon = (\varepsilon_n)_{n\geqslant n_0}$ the quantities

$$\mathcal{O}_{\epsilon,n_0} = \sum_{n=n_0}^{\infty} \mathbf{1}\{\|\hat{\theta}_n - \theta_0\| > \varepsilon_n\} \qquad \text{and} \qquad \mathrm{m}_{\epsilon,n_0} = \max\{n - n_0 \geqslant 0 \mid \|\hat{\theta}_n - \theta_0\| > \varepsilon_n\}.$$

In the sequel we apply the LDP as explained in Remark 6 in order to obtain a satisfying MDF quantification.

Theorem 12 (Method of moments: Data with uniformly bounded exponential moments). *Assume there is a constant $\gamma > 0$ such that*

$$\sup_{\theta\in\Theta} \sup_{i\geqslant n_0} \mathbb{E}\Big[e^{\gamma|Y_i(\theta)|}\Big] < \infty.$$

Then for any $0 \leqslant \alpha < \frac{1}{2}$ the choice of $\epsilon = (\varepsilon_n)_{n\in\mathbb{N}}$ $\varepsilon_n = n^{-\alpha}$, $n \in \mathbb{N}$, we have for any $\ell \in \mathbb{N}$

$$\limsup_{n\to\infty} \|\hat{\theta}(\theta_0) - \theta_0\| \cdot \varepsilon_n^{-1} \leqslant 1 \qquad \mathbb{P}\text{-a.s.}$$

and for $c_0 = (\delta_1\delta_2\lambda\gamma)^{\frac{2}{3}}$ and $p \in (0,1)$ there is a constant $K := K(\alpha, n_0) > 0$ such that the respective quantities $\mathcal{O}_{\epsilon,n_0}$ and $\mathrm{m}_{\epsilon,n_0}$ satisfy

$$\mathbb{E}\Big[\exp\big(pc_0(\mathcal{O}_{\epsilon,n_0})^{1-2\alpha}\big)\Big] \leqslant \mathbb{E}\Big[\exp\big(pc_0(\mathrm{m}_{\epsilon,n_0})^{1-2\alpha}\big)\Big] \leqslant K.$$

By Example 3 there exist constants $d, D > 0$ such that for all $k \geqslant 2$

$$\mathbb{P}(\mathcal{O}_{\epsilon,n_0} \geqslant k) \leqslant \mathbb{P}(\mathrm{m}_{\epsilon,n_0} \geqslant k) \leqslant \Big(d + D(k-1)^{1+2\alpha}\Big)\, e^{-c_0 k^{1-2\alpha}}.$$

Proof: Note that $n\bar{X}_n(\theta_0) - M(\theta_0) = X_n(\theta_0) - nM(\theta_0)$ is a random walk of centered random variables in $\mathbb{R}^\ell$. Further note that

$$\mathbb{E}[|Y_i(\theta_0)|^\ell] = \frac{\ell!}{\gamma^\ell}\mathbb{E}\Big[\frac{(\gamma|Y_i(\theta_0)|)^\ell}{\ell!}\Big] \leqslant \frac{\ell!}{\gamma^\ell} \sup_{\theta\in\Theta} \mathbb{E}[e^{\gamma|Y_i(\theta)|}].$$

By Remark 6 for any $\delta \in (0,1)$ there is $n_0 \in \mathbb{N}$ such that for $n \geqslant n_0$ we have

$$\begin{aligned}\mathbb{P}(\|\bar{X}_n(\theta_0) - M(\theta_0)\| > \varepsilon_n) &= \mathbb{P}(\sum_{j=1}^{\ell} |\bar{X}_{n,j}(\theta_0) - M_j(\theta_0)|^2 > \varepsilon_n^2) \leqslant \sum_{j=1}^{\ell} \mathbb{P}(|\bar{X}_{n,j}(\theta_0) - M_j(\theta_0)| > \frac{\varepsilon_n}{\sqrt{\ell}}) \\ &\leqslant 2\ell e^{-n \inf_{|y|>\frac{\varepsilon_n}{\sqrt{\ell}}} \Lambda^*(y)} \leqslant \ell e^{-\frac{1-\delta}{2\ell}|(\Lambda^*)''(0)|n\varepsilon_n^2},\end{aligned}$$

where Λ^* is the good rate function defined in Remark 6. Hence for $\epsilon = (\varepsilon_n)_{n\in\mathbb{N}}$ $\varepsilon_n := n^{-\alpha}$, $0 < \alpha < \frac{1}{2}$, $n \in \mathbb{N}$ (5.5) we have

$$\mathbb{P}(\|\hat{\theta}_n(\theta_0) - \theta_0\| > \varepsilon_n) \leqslant \mathbb{P}(\|\bar{X}_n(\theta_0) - M(\theta_0)\| > \delta_1\delta_2\lambda\varepsilon_n) \leqslant \ell e^{-\frac{(1-\delta)\delta_1^2\delta_2^2\lambda^2}{2\ell}|(\Lambda^*)''(0)|n^{1-2\alpha}}.$$

By Lemma 2 combined with Example 3 we have

$$\limsup_{n\to\infty} \|\bar{X}_n(\theta_0) - M(\theta_0)\| \cdot \varepsilon_n^{-1} \leqslant 1 \qquad \mathbb{P}\text{-a.s.}$$

and constants $d, D > 0$ such that for $k \geqslant 2$

$$\mathbb{P}(\mathcal{O}_{\epsilon,n_0} \geqslant k) \leqslant \mathbb{P}(\mathrm{m}_{\epsilon,n_0} \geqslant k) \leqslant (d + D(k-1)^{1+2\alpha}) e^{-\frac{(1-\delta)\delta_1^2\delta_2^2\lambda^2}{2\ell}|(\Lambda^*)''(0)|k^{1-2\alpha}}.$$

□

We stress that the condition of uniformly bounded exponential moments can be replaced by more context dependent hypotheses, such as Cesàro convergent p-th moments, uniformly bounded L^p moments, a Gärtner-Ellis LDP setting and Azuma-type setting, with respective finer MDF quantifications.

5.4. *Excursion dynamics of the Galton-Watson branching process.*

We recall and keep the notation of Example 6. In this subsection, we see that the tradeoff between error tolerance and mean deviation frequencies also works for a.s. converging processes, but also for a.s. divergence, such as for the super-critical branching processes.

5.4.1. *Sub-critical branching:* $\mathfrak{m} < 1$.

In case of sub-critical branching we know that $X_n \to 0$ almost surely. With the help of the tradeoff in Lemma 2 we may quantify the tradeoff between the number of excursions beyond a growing threshold $K(n)$.

Lemma 3. *For $\mathfrak{m} \in (0,1)$ and $K > 0$ we have for all $k \geqslant 1$, for $\mathcal{O}_K := \#\{\ell \in \mathbb{N} \mid Z_\ell \geqslant K\} = \sum_{\ell\in\mathbb{N}} \mathbf{1}\{Z_\ell \geqslant K\}$ and $\mathbb{m}_K = \max\{n \in \mathbb{N} \mid Z_n \geqslant K\}$*

$$\mathbb{P}(\mathcal{O}_K \geqslant k) \leqslant \mathbb{P}(\mathbb{m}_K \geqslant k) \leqslant \frac{2e^{\frac{9}{8}}}{\mathfrak{m}(1-\mathfrak{m})K} \cdot \Big[k\Big(\big(\frac{v}{\mathfrak{m}(1-\mathfrak{m})K} \wedge 1\big) + 1\Big) + 1\Big] \cdot \mathfrak{m}^k.$$

Proof of Lemma 3: For the case $\mathfrak{m} \in (0,1)$ it is well-known (using Markov's inequality, as well as Wald's and Blackwell-Girshwick's identities (Klenke, 2008, Theorem 5.5 and 5.10)) that for any $K > 0$ fixed

$$\mathbb{P}(Z_\ell \geqslant K) \leqslant \mathfrak{m}^\ell \cdot \Big(\frac{1}{K} \wedge \frac{v}{\mathfrak{m}(1-\mathfrak{m})K^2}\Big). \tag{5.7}$$

Hence

$$\sum_{\ell \geqslant n}^{\infty} \mathbb{P}(Z_\ell \geqslant K) \leqslant \mathfrak{m}^n \cdot \Big(\frac{1}{(1-\mathfrak{m})K} \wedge \frac{v}{\mathfrak{m}(1-\mathfrak{m})^2K^2}\Big) \tag{5.8}$$

and Example 2 yields for any $p \in (0,1)$ and $a_n = \mathfrak{m}^{-pn}$, $n \in \mathbb{N}$, that

$$\mathbb{E}[\mathfrak{m}^{-p\mathcal{O}_K}] = \mathbb{E}[e^{|\ln(\mathfrak{m})|p\mathcal{O}_K}] \leqslant \mathbb{E}[e^{|\ln(\mathfrak{m})|p\mathbb{m}_K}] \leqslant \Big(\frac{1}{1-\mathfrak{m}^{1-p}} + 1\Big) \cdot \Big(\frac{1}{\mathfrak{m}(1-\mathfrak{m})K} \wedge \frac{v}{\mathfrak{m}^2(1-\mathfrak{m})^2K^2}\Big).$$

In addition, by Högele and Steinicke (2023, Lemma 5) we have for all $k \geqslant 1$

$$\begin{aligned}\mathbb{P}(\mathcal{O}_K \geqslant k) \leqslant \mathbb{P}(\mathbb{m}_K \geqslant k) &\leqslant \inf_{0<q<|\ln(\mathfrak{m})|} e^{-qk}\mathbb{E}[e^{q\mathbb{m}_K}]\\ &\leqslant 2e^{\frac{9}{8}}\Big[k\Big(\frac{1}{\mathfrak{m}(1-\mathfrak{m})K} \wedge \frac{v}{\mathfrak{m}^2(1-\mathfrak{m})^2K^2} + 1\Big) + 1\Big] \cdot \mathfrak{m}^{k-1}.\end{aligned}$$

□

5.4.2. *Super-critical branching:* $\mathfrak{m} > 1$.

Lemma 4. *For $\mathfrak{m} > 1$ and $\mathbb{E}[Y_1^2] < \infty$ we have the following:*

(a) *For any $\varepsilon > 0$ and $\mathcal{O}_\varepsilon := \#\{\ell \in \mathbb{N} \mid |X_\ell - X_\infty| \geqslant \varepsilon\} = \sum_{\ell\in\mathbb{N}} \mathbf{1}\{|X_\ell - X_\infty| \geqslant \varepsilon\}$, and $\mathbb{m}_\varepsilon := \max\{\ell \in \mathbb{N} \mid |X_\ell - X_\infty| \geqslant \varepsilon\}$ we have*

$$\mathbb{P}(\mathcal{O}_\varepsilon \geqslant k) \leqslant \frac{2e^{\frac{9}{8}}v}{(1-\frac{1}{\mathfrak{m}})(\mathfrak{m}^2-\mathfrak{m})\varepsilon^2} \cdot (2k+1) \cdot \mathfrak{m}^{k-1}, \qquad k \geqslant 1.$$

(b) *For any* $\theta > 1$ *and* $\varepsilon_n := \sqrt{\frac{vn^\theta}{\mathfrak{m}^n(\mathfrak{m}^2-\mathfrak{m})}}$ *we have*

$$\limsup_{n\to\infty} |X_n - X_\infty| \cdot \varepsilon_n^{-1} \leqslant 1 \qquad \mathbb{P}\text{-}a.s.$$

and, for $\mathcal{O}_\epsilon := \#\{\ell \in \mathbb{N} \mid |X_\ell - X_\infty| \geqslant \varepsilon_n\} = \sum_{\ell\in\mathbb{N}} \mathbf{1}\{|X_\ell - X_\infty| \geqslant \varepsilon_n\}$ *and* $\mathbb{m}_\epsilon := \max\{n \in \mathbb{N} \mid |X_n - X_\infty| \geqslant \varepsilon_n\}$ *we get*

$$\mathbb{P}(\mathcal{O}_\epsilon \geqslant k) \leqslant \mathbb{P}(\mathbb{m}_\epsilon \geqslant k) \leqslant k^{-1} \cdot \zeta(\theta), \qquad k \geqslant 1.$$

Remark 9. Note that item (a) and (b) represent extremal cases. Case (a) counts the (random) number of infractions of a fixed error bar $\varepsilon > 0$. Case (b) instead yields close to optimal rates of convergence $\varepsilon_n \to 0$, $n \to \infty$, which are obtained only with a linear decay of the deviation frequency, that is, only after many infractions. In other words, the trespassing probabilities are barely summable. Of course, all kind of useful tradeoff regimes between (1) and (2) can be derived by the same methodology.

Proof of Lemma 4: For $\mathfrak{m} > 1$, we get that if $\mathbb{E}[Y_1^2] < \infty$, $X_n := \frac{Z_n}{\mathfrak{m}^n}$ is a martingale such that $\mathbb{E}[X_n^2] = 1 + \frac{v\mathfrak{m}^n}{\mathfrak{m}^2-m}(1-\mathfrak{m}^{-n})$ (see e.g. Harris (1963, Proof of Theorem 8.1)). Moreover

$$\mathbb{E}[(X_{n+\ell} - X_n)^2] = \frac{v\mathfrak{m}^{-n}}{\mathfrak{m}^2 - \mathfrak{m}}(1 - \mathfrak{m}^{-\ell}),$$

showing the convergence $X_n \to X_\infty$ in L^2, and, in addition,

$$\sum_{\ell=n}^{\infty} \mathbb{P}(|X_\ell - X_\infty| \geqslant \varepsilon) \leqslant \sum_{\ell=n}^{\infty} \frac{\mathbb{E}[(X_\infty - X_\ell)^2]}{\varepsilon^2} = \sum_{\ell=n}^{\infty} \frac{v\mathfrak{m}^{-\ell}}{(\mathfrak{m}^2 - \mathfrak{m})\varepsilon^2} = \frac{v\mathfrak{m}^{-n}}{(1-\frac{1}{\mathfrak{m}})(\mathfrak{m}^2 - \mathfrak{m})\varepsilon^2} < \infty. \tag{5.9}$$

Hence, Example 2 yields for any $p \in (0,1)$, $a_n = \mathfrak{m}^{pn}$, $n \in \mathbb{N}$, that

$$\mathbb{E}[\mathfrak{m}^{p\mathcal{O}_\varepsilon}] \leqslant \mathbb{E}[\mathfrak{m}^{p\mathbb{m}_\varepsilon}] \leqslant \frac{v}{(1-\frac{1}{\mathfrak{m}})(\mathfrak{m}^2 - \mathfrak{m})\varepsilon^2}\Big(\frac{1}{1-\mathfrak{m}^{p-1}} + 1\Big),$$

such that

$$\mathbb{P}(\mathcal{O}_\varepsilon \geqslant k) \leqslant \mathbb{P}(\mathbb{m}_\varepsilon \geqslant k) \leqslant \inf_{p\in(0,1)} e^{-pk}\mathbb{E}[\mathfrak{m}^{p\mathcal{O}_\varepsilon}] \leqslant \frac{2e^{\frac{9}{8}}v}{(1-\frac{1}{\mathfrak{m}})(\mathfrak{m}^2 - \mathfrak{m})\varepsilon^2} \cdot (2k+1) \cdot \mathfrak{m}^{-k}.$$

For the second statement we use the classical first Borel-Cantelli lemma in (5.9) for $n = 1$. We calculate ε_ℓ by setting

$$\frac{v\mathfrak{m}^{-\ell}}{(\mathfrak{m}^2 - \mathfrak{m})\varepsilon_\ell^2} = \ell^{-\theta}.$$

This finishes the proof. □

Appendix A. Optimal tail decay rates in case of Weibull type moments

This section gives the details of the tail optimization of $\mathcal{O}_{\epsilon,n_0}$ and $\mathbb{m}_{\epsilon,n_0}$ in Example 3.

Lemma 5. *For any* $b, \alpha \in (0,1)$, $c > 0$ *and* $n_0 \in \mathbb{N}$ *there are (explicitly known) positive constants* $d(n_0), D(n_0) \in \mathbb{N}$ *such that for* $k \geqslant 2$ *we have*

$$\inf_{p\in(0,1)} b^{p(k-1)^\alpha} \sum_{n=n_0}^{\infty} c\Bigg(1 + \frac{1 + \frac{\frac{1}{\alpha}-1}{|\ln(b)|}}{\alpha|\ln(b)|} n^{1-\alpha}\Bigg) b^{(1-p)n^\alpha} \leqslant (d + D(k-1)^{2-\alpha}) b^{(k-1)^\alpha}.$$

Proof: To estimate the desired infimum, we calculate

$$b^{p(k-1)^\alpha} K(p,\alpha) = b^{p(k-1)^\alpha} \sum_{n=n_0}^{\infty} c\left(1+\frac{1+\frac{\frac{1}{\alpha}-1}{|\ln(b)|}}{\alpha|\ln(b)|} n^{1-\alpha}\right) b^{(1-p)n^\alpha}$$

$$\leqslant b^{p(k-1)^\alpha} c\left(1+\frac{1+\frac{\frac{1}{\alpha}-1}{|\ln(b)|}}{\alpha|\ln(b)|} n_0^{1-\alpha}\right) b^{(1-p)n_0^\alpha} + b^{p(k-1)^\alpha} \sum_{n=n_0+1}^{\infty} c\left(1+\frac{1+\frac{\frac{1}{\alpha}-1}{|\ln(b)|}}{\alpha|\ln(b)|}\right) n^{1-\alpha} b^{(1-p)n^\alpha}$$

$$\leqslant b^{p(k-1)^\alpha} c\left(1+\frac{1+\frac{\frac{1}{\alpha}-1}{|\ln(b)|}}{\alpha|\ln(b)|} n_0^{1-\alpha}\right) b^{(1-p)n_0^\alpha} + b^{p(k-1)^\alpha} C \int_{n_0}^{\infty} e^{-|\ln(b)|(1-p)x^\alpha} x^{1-\alpha} dx, \tag{A.1}$$

where $C := c\left(1+\frac{1+\frac{\frac{1}{\alpha}-1}{|\ln(b)|}}{\alpha|\ln(b)|}\right)$. Substituting $t = |\ln(b)|(1-p)x^\alpha$, the integral $\int_{n_0}^{\infty} e^{-|\ln(b)|(1-p)x^\alpha} x^{1-\alpha} dx$ becomes

$$\int_{|\ln(b)|(1-p)n_0^\alpha}^{\infty} e^{-t} \frac{1}{\alpha|\ln(b)|(1-p)} \left(\frac{t}{|\ln(b)|(1-p)}\right)^{\frac{1}{\alpha}-1} \left(\frac{t}{|\ln(b)|(1-p)}\right)^{\frac{1}{\alpha}-1} dt$$

$$= \frac{1}{\alpha(|\ln(b)|(1-p))^{\frac{2}{\alpha}-1}} \int_{|\ln(b)|(1-p)n_0^\alpha}^{\infty} e^{-t} t^{\frac{2}{\alpha}-2} dt.$$

By the integral substitution, (A.1) turns to

$$b^{p(k-1)^\alpha} c\left(1+\frac{1+\frac{\frac{1}{\alpha}-1}{|\ln(b)|}}{\alpha|\ln(b)|} n_0^{1-\alpha}\right) b^{(1-p)n_0^\alpha} + \frac{C\Gamma\left(\frac{2}{\alpha}-1, |\ln(b)|(1-p)n_0^\alpha\right)}{\alpha(|\ln(b)|(1-p))^{\frac{2}{\alpha}-1}} b^{p(k-1)^\alpha}, \tag{A.2}$$

where $\Gamma(x,a)$ denotes the upper incomplete Gamma function with lower limit a. Differentiating this term w.r.t. p and calculating the appearing zero p^* is quite tedious. The resulting value for p^* however is close to $1-\frac{1}{(k-1)^\alpha}$ (for high values of n_0). Subsequently, we will set p to this value. this leaves us with

$$\left(c\left(1+\frac{1+\frac{\frac{1}{\alpha}-1}{|\ln(b)|}}{\alpha|\ln(b)|} n_0^{1-\alpha}\right) b^{\left(\frac{n_0}{k-1}\right)^\alpha - 1} + \frac{C\Gamma\left(\frac{2}{\alpha}-1, |\ln(b)|\left(\frac{n_0}{k-1}\right)^\alpha\right)}{b\alpha|\ln(b)|^{\frac{2}{\alpha}-1}} (k-1)^{2-\alpha}\right) b^{(k-1)^\alpha}.$$

□

Appendix B. Optimization procedure in Example 1

Differentiating w.r.t. p yields

$$0 = \frac{d}{dp}(k^{-(p+1)} cq\zeta(q-p-1;n_0)) = -\ln(k)k^{-(p+1)} cq\zeta(q-p-1;n_0) - k^{-(p+1)} cq\zeta'(q-p-1;n_0),$$

such that $\ln(k) = -\frac{\zeta'(q-p-1;n_0)}{\zeta(q-p-1;n_0)}$. The right-hand side for large values in the ζ-function (and ζ') yields that

$$\ln(k) \approx \frac{1}{q-p-2} + \psi(n_0), \qquad \text{where} \qquad \psi(n_0) = \frac{\Gamma'(n_0)}{\Gamma(n_0)},$$

which yields the optimizer $p^* = q-2-\frac{1}{\ln(k)-\psi(n_0)}$ and hence for $k \geqslant e^{\frac{1}{q-2}+\psi(n_0)}$ we have

$$\mathbb{P}(\mathcal{O}_{n_0} \geqslant k) \leqslant \mathbb{P}(\mathrm{m}_{n_0} \geqslant k) \leqslant cq \cdot k^{-(q-1)} \cdot k^{\frac{1}{\ln(k)-\psi(n_0)}} \cdot \zeta(1+\frac{1}{\ln(k)-\psi(n_0)}; n_0). \tag{B.1}$$

Note that the right-hand side for large k behaves asymptotically as

$$cqe \cdot k^{-(q-1)} \cdot \ln(k).$$

In order to obtain an upper bound we use the integral comparison principle

$$\zeta(1+\frac{1}{\ln(k)-\psi(n_0)};n_0)=\sum_{n=n_0}^{\infty} n^{-1+\frac{1}{\ln(k)-\psi(n_0)}} \leqslant n_0^{-1+\frac{1}{\ln(k)-\psi(n_0)}}+\int_{n_0}^{\infty} x^{-1+\frac{1}{\ln(k)-\psi(n_0)}}dx$$
$$= n_0^{-1+\frac{1}{\ln(k)-\psi(n_0)}}+(\ln(k)-\psi(n_0))n_0^{\frac{1}{\ln(k)-\psi(n_0)}}.$$

For $k \geqslant e^{\frac{1}{q-2}+\psi(n_0)}$ this yields

$$\zeta(1+\frac{1}{\ln(k)-\psi(n_0)};n_0)\leqslant n_0^{q-3}+(\ln(k)-\psi(n_0))n_0^{q-2}. \tag{B.2}$$

Combining (B.1) and (B.2) we obtain (2.8): for any $0 \leqslant p < q-2$ and $c_1 = cqe^{(q-2)\psi(n_0)}n_0^{q-2}$ it follows

$$\mathbb{P}(\mathcal{O}_{n_0} \geqslant k) \leqslant \mathbb{P}(\mathrm{m}_{n_0} \geqslant k) \leqslant c_1 \cdot k^{-(q-1)} \cdot (\ln(k)+\frac{1}{n_0}-\psi(n_0)) \qquad \text{for } k \geqslant e^{\frac{1}{q-2}+\psi(n_0)}.$$

Acknowledgements

Parts of this article were written when A.S. was enrolled in the JYU Visiting Fellow program of the University of Jyväskylä, Finland. The research of M.A.H. has been supported by the project "Mean deviation frequencies and the cutoff phenomenon" (INV-2023-162-2850) of the School of Sciences (Facultad de Ciencias) at Universidad de los Andes, Bogotá, Colombia. M.A.H. thanks E. Hausenblas for the hospitality during a JESH exchange project of the Austrian Academy of Sciences exchange at Montanuniversität Leoben, Austria, where this project was started. A.S. and M.A.H. are grateful to E. Hausenblas and the work group of Applied Mathematics at MU Leoben for the scientific environment which made this collaboration possible.

The authors would like to thank the anonymous referees for many helpful suggestions, which have improved the manuscript significantly.

References

Alsmeyer, G. Convergence rates in the law of large numbers for martingales. *Stochastic Process. Appl.*, **36** (2), 181–194 (1990). DOI: 10.1016/0304-4149(90)90090-F.

Azuma, K. Weighted sums of certain dependent random variables. *Tohoku Math. J.*, **19** (3), 357 – 367 (1967). DOI: 10.2748/tmj/1178243286.

Bagchi, A. and Pal, A. K. Asymptotic normality in the generalized Pólya-Eggenberger urn model, with an application to computer data structures. *SIAM J. Algebraic Discrete Methods*, **6** (3), 394–405 (1985). DOI: 10.1137/0606041.

Baum, L. E. and Katz, M. Convergence rates in the law of large numbers. *Trans. Amer. Math. Soc.*, **120**, 108–123 (1965). DOI: 10.2307/1994170.

Billingsley, P. *Convergence of probability measures.* Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons, Inc., New York, second edition (1999). ISBN 0-471-19745-9. DOI: 10.1002/9780470316962.

Borel, E. Les probabilités dénombrables et leurs applications arithmétiques. *Rend. Circ. Mat. Palermo*, **27**, 247–271 (1909). DOI: 10.1007/BF03019651.

Cantelli, F. P. Sulla probabilità come limite della frequenza. *Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Natur. (5)*, **26** (1), 39–45 (1917).

Chandra, T. K. *The Borel-Cantelli lemma.* SpringerBriefs in Statistics. Springer, Heidelberg (2012). ISBN 978-81-322-0676-7. Chapter 2. DOI: 10.1007/978-81-322-0677-4.

Chung, K. L. and Erdös, P. On the application of the Borel-Cantelli lemma. *Trans. Amer. Math. Soc.*, **72**, 179–186 (1952). DOI: 10.2307/1990661.

Dedecker, J. and Merlevède, F. Convergence rates in the law of large numbers for Banach-valued dependent variables. *Theory Probab. Appl.*, **52** (3), 416–438 (2008). DOI: 10.1137/S0040585X97983171.

Doob, J. L. *Stochastic processes.* John Wiley & Sons, Inc., New York; Chapman & Hall, Ltd., London (1953). MR58896.

Erdös, P. On a theorem of Hsu and Robbins. *Ann. Math. Statistics*, **20**, 286–291 (1949). DOI: 10.1214/aoms/1177730037.

Estrada, L. and Högele, M. A. Moment estimates in the first Borel-Cantelli Lemma with applications to mean deviation frequencies. *Statist. Probab. Lett.*, **190**, 109636 (2022). DOI: 10.1016/j.spl.2022.109636.

Etemadi, N. An elementary proof of the strong law of large numbers. *Z. Wahrsch. Verw. Gebiete*, **55** (1), 119–122 (1981). DOI: 10.1007/BF01013465.

Fan, X., Grama, I., and Liu, Q. Exponential inequalities for martingales with applications. *Electron. J. Probab.*, **20**, no. 1, 22 (2015). DOI: 10.1214/EJP.v20-3496.

Franchini, S. Large deviations for generalized Pólya urns with arbitrary urn function. *Stochastic Process. Appl.*, **127** (10), 3372–3411 (2017). DOI: 10.1016/j.spa.2017.02.010.

Friedman, B. A simple urn model. *Comm. Pure Appl. Math.*, **2**, 59–70 (1949). DOI: 10.1002/cpa.3160020103.

Föllmer, H. and Schied, A. *Stochastic Finance: An Introduction in Discrete Time.* De Gruyter, Berlin, 5 edition (2025). DOI: 10.1515/9783111045283.

Gerber, H. U. A proof of the Schuette-Nesbitt formula for dependent events. *Actuar. Res. Clearing House*, **1**, 9–10 (1979).

Giraudo, D. Deviation inequalities for Banach space valued martingales differences sequences and random fields. *ESAIM Probab. Stat.*, **23**, 922–946 (2019). DOI: 10.1051/ps/2019016.

Gouet, R. A martingale approach to strong convergence in a generalized Pólya-Eggenberger urn model. *Statist. Probab. Lett.*, **8** (3), 225–228 (1989). `https://ideas.repec.org/a/eee/stapro/v8y1989i3p225-228.html`.

Gouet, R. Martingale functional central limit theorems for a generalized Pólya urn. *Ann. Probab.*, **21** (3), 1624–1639 (1993). DOI: 10.1214/aop/1176989134.

Gouet, R. Strong convergence of proportions in a multicolor Pólya urn. *J. Appl. Probab.*, **34** (2), 426–435 (1997). DOI: 10.2307/3215382.

Hao, S. Convergence rates in the law of large numbers for arrays of Banach valued martingale differences. *Abstr. Appl. Anal.*, pp. 26, Art. ID 715054 (2013). DOI: 10.1155/2013/715054.

Hao, S. and Liu, Q. Baum-Katz type theorems for martingale arrays. *C. R. Math. Acad. Sci. Paris*, **350** (1-2), 91–96 (2012). DOI: 10.1016/j.crma.2011.12.006.

Hao, S. and Liu, Q. Convergence rates in the law of large numbers for arrays of martingale differences. *J. Math. Anal. Appl.*, **417** (2), 733–773 (2014). DOI: 10.1016/j.jmaa.2014.03.049.

Harris, T. E. *The theory of branching processes.* Die Grundlehren der mathematischen Wissenschaften, Band 119. Springer-Verlag, Berlin; Prentice Hall, Inc., Englewood Cliffs, NJ (1963). MR163361.

Hill, T. P. A stronger form of the Borel-Cantelli lemma. *Illinois J. Math.*, **27** (2), 240–243 (1983). DOI: 10.1215/ijm/1256046493.

Hoeffding, W. Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.*, **58**, 13–30 (1963). DOI: 10.2307/2282952.

Hsu, P. L. and Robbins, H. Complete convergence and the law of large numbers. *Proc. Nat. Acad. Sci. U.S.A.*, **33**, 25–31 (1947). DOI: 10.1073/pnas.33.2.25.

Högele, M. A. and Steinicke, A. Deviation frequencies of Brownian path property approximations. *ArXiv Mathematics e-prints* (2023). arXiv: 2302.04115.

Johnson, N. L. and Kotz, S. *Urn models and their application. An approach to modern discrete probability theory.* Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons

(1977). MR488211.

Johnson, W. B., Schechtman, G., and Zinn, J. Best constants in moment inequalities for linear combinations of independent and exchangeable random variables. *Ann. Probab.*, **13** (1), 234–253 (1985). DOI: 10.1214/aop/1176993078.

Kallenberg, O. *Foundations of modern probability*. Probability and its Applications (New York). Springer-Verlag, New York, second edition (2002). ISBN 0-387-95313-2. DOI: 10.1007/978-1-4757-4015-8.

Klenke, A. *Probability theory. A comprehensive course*. Universitext. Springer-Verlag London, Ltd., London (2008). ISBN 978-1-84800-047-6. DOI: 10.1007/978-1-84800-048-3.

Lesigne, E. and Volný, D. Large deviations for martingales. *Stochastic Process. Appl.*, **96** (1), 143–159 (2001). DOI: https://doi.org/10.1016/S0304-4149(01)00112-0.

Luo, S. On Azuma-type inequalities for Banach space-valued martingales. *J. Theoret. Probab.*, **35** (2), 772–800 (2022). DOI: 10.1007/s10959-021-01086-5.

Lévy, P. *Théorie de l'addition des variables aléatoires*. Gauthier-Villars, Paris (1937).

Ma, H. and Sun, Y. Complete convergence and complete moment convergence for randomly weighted sums of martingale difference sequence. *J. Inequal. Appl.*, pp. 16, Paper No. 173 (2018). DOI: 10.1186/s13660-018-1770-3.

Mazliak, L. How Paul Lévy saw Jean Ville and martingales. *Electron. J. Hist. Probab. Stat.*, **5** (1), 1–24 (2009). HAL ID: hal-00434317.

Najock, D. and Heyde, C. C. On the number of terminal vertices in certain random trees with an application to stemma construction in philology. *J. Appl. Probab.*, **19** (3), 675–680 (1982). DOI: 10.2307/3213526.

Naor, A. On the Banach-space-valued Azuma inequality and small-set isoperimetry of Alon-Roichman graphs. *Combin. Probab. Comput.*, **21** (4), 623–634 (2012). DOI: 10.1017/S0963548311000757.

Oliveira, R. I., Pereira, A., and Ribeiro, R. Concentration in the generalized Chinese restaurant process. *Sankhya A*, **84** (2), 628–670 (2022). DOI: 10.1007/s13171-020-00210-7.

Osękowski, A. A note on the Burkholder-Rosenthal inequality. *Bull. Pol. Acad. Sci. Math.*, **60** (2), 177–185 (2012). DOI: 10.4064/ba60-2-7.

Pinelis, I. Optimum bounds for the distributions of martingales in Banach spaces. *Ann. Probab.*, **22** (4), 1679–1706 (1994). DOI: 10.1214/aop/1176988477.

Pisier, G. Martingales with values in uniformly convex spaces. *Israel J. Math.*, **20** (3-4), 326–350 (1975). DOI: 10.1007/BF02760337.

Protter, P. *Stochastic integration and differential equations. A new approach*, volume 21 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin (1990). ISBN 3-540-50996-8. DOI: 10.1007/978-3-662-02619-9.

Rosenthal, H. P. On the subspaces of L^p $(p > 2)$ spanned by sequences of independent random variables. *Israel J. Math.*, **8**, 273–303 (1970). DOI: 10.1007/BF02771562.

Seneta, E. *Non-negative matrices and Markov chains*. Springer Series in Statistics. Springer, New York (2006). ISBN 978-0387-29765-1; 0-387-29765-0. MR2209438.

Shiryaev, A. N. *Probability*, volume 95 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition (1996). ISBN 0-387-94549-0. DOI: 10.1007/978-1-4757-2539-1.

Stoica, G. Baum-Katz-Nagaev type results for martingales. *J. Math. Anal. Appl.*, **336** (2), 1489–1492 (2007). DOI: 10.1016/j.jmaa.2007.03.012.

Talagrand, M. Isoperimetry and integrability of the sum of independent Banach-space valued random variables. *Ann. Probab.*, **17** (4), 1546–1570 (1989). DOI: 10.1214/aop/1176991174.

Ville, J. *Étude critique de la notion de collectif*, volume 3 of *Monographies des Probabilités*. Gauthier-Villars, Paris (1939).

Volný, D. On non-ergodic versions of limit theorems. *Apl. Mat.*, **34** (5), 351–363 (1989). `http://eudml.org/doc/15589`.

Volný, D. Approximating martingales and the central limit theorem for strictly stationary processes. *Stochastic Process. Appl.*, **44** (1), 41–71 (1993). DOI: 10.1016/0304-4149(93)90037-5.

Williams, D. *Probability with martingales*. Cambridge Mathematical Textbooks. Cambridge University Press, Cambridge (1991). ISBN 0-521-40455-X; 0-521-40605-6. DOI: 10.1017/CBO9780511813658.

Yukich, J. E. Asymptotics for the length of a minimal triangulation on a random sample. *Ann. Appl. Probab.*, **9** (1), 27–45 (1999). DOI: 10.1214/aoap/1029962596.