

A Large-Deviation Inequality for Vector-valued Martingales

Thomas P. Hayes

May 14, 2001

Abstract

Let \mathbf{X} be a discrete-time martingale taking values in \mathbb{R}^d or ℓ_2 , such that for all n , $\|X_n - X_{n-1}\| \leq 1$. We prove the large deviation bound

$$\Pr [\|X_n\| \geq a] < 2e^{1-(a-1)^2/2n}.$$

This upper bound is within a constant factor, e^2 , of Azuma's Inequality for real-valued martingales. Earlier work along these lines was done by O. Kallenberg and R. Sztencel (1992).

Our inequality also holds for a more general class of random series, namely, those which satisfy the weaker condition that for every n ,

$$\mathbb{E}(X_n | X_{n-1}) = X_{n-1} \quad (*)$$

In particular, this includes the class of *weak martingales*.

More generally, we prove that, for every random series \mathbf{X} in \mathbb{R}^d satisfying (*), there exists a martingale \mathbf{Y} such that for all n , the distribution of (Y_{n-1}, Y_n) is the same as that of (X_{n-1}, X_n) .

As an application, we answer questions posed by L. Babai about Fourier coefficients of random subsets of a finite abelian group.

1 Introduction

This paper is centered around extending the following well-known result to more general classes of random series.

Proposition 1.1 (Azuma’s Inequality) *Let \mathbf{X} be a real-valued martingale such that for every i , $|X_i - X_{i-1}| \leq 1$. Then*

$$\Pr [|X_n| \geq a] < 2e^{-a^2/2n}. \tag{1}$$

Our first objective is to replace “real-valued martingale” with “vector-valued martingale.” Throughout this paper, \mathbb{E} will denote a separable Euclidean space: either \mathbb{R}^d or ℓ_2 . $\|\cdot\|$ will denote the Euclidean norm in \mathbb{E} .

We will use the following definition in this paper.

Definition 1.1 Let $\mathbf{X} = (X_i : \Omega \rightarrow \mathbb{E})$ be a sequence of random vectors taking values in \mathbb{E} , such that $X_0 = 0$, and for every $i \geq 1$, $\mathbb{E}(\|X_i\|) < \infty$ and $\mathbb{E}(X_i | X_0, X_1, \dots, X_{i-1}) = X_{i-1}$. Then we call \mathbf{X} a *(strong) martingale in \mathbb{E}* .

Our second objective is to relax the conditioning on the expectation in this definition from $\mathbb{E}(X_i | X_0, X_1, \dots, X_{i-1})$ to $\mathbb{E}(X_i | X_{i-1})$.

Definition 1.2 Let $\mathbf{X} = (X_i : \Omega \rightarrow \mathbb{E})$ be a sequence of random vectors taking values in \mathbb{E} , such that $X_0 = 0$, and for every $i \geq 1$, $\mathbb{E}(\|X_i\|) < \infty$ and $\mathbb{E}(X_i | X_{i-1}) = X_{i-1}$. Then we call \mathbf{X} a *very-weak martingale in \mathbb{E}* .

Note that every martingale is automatically a very-weak martingale. The converse is false (see below). To explain the terminology, we recall the common definition of *weak* martingales.

Definition 1.3 Let $\mathbf{X} = (X_i : \Omega \rightarrow \mathbb{E})$ be a sequence of random variables such that $X_0 = 0$, and for every $j < i$, $\mathbb{E}(\|X_i\|) < \infty$ and $\mathbb{E}(X_i | X_j) = X_j$. Then we call \mathbf{X} a *weak martingale in \mathbb{E}* .

Every strong martingale is a weak martingale, and every weak martingale is a very-weak martingale. It is known that neither of the converses is true. In Section 4, we present a canonical example of a very-weak martingale which is not a weak martingale, and discuss some combinatorial differences between the classes.

It has been shown that Azuma’s Inequality holds for real-valued very-weak martingales (a nice presentation may be found in [ASE, Appendix A]).

We prove more generally that any very-weak martingale may be replaced by a strong martingale which has the same distributions of values and of differences. Azuma’s Inequality and other similar bounds thus *automatically* apply equally well to very-weak martingales and to martingales. This result applies to very-weak martingales taking values in any Euclidean space.

Precise statements of all our results are in the following subsection.

1.1 Overview of Results

Our first tool is essentially the discrete-time case of a theorem of Kallenberg and Sztencel [Kallenberg, Sztencel], showing that for many purposes, “local martingales” (cf. [Karatzas, Shreve]) in \mathbb{E} are equivalent to local martingales in \mathbb{R}^2 . This result also holds for (discrete-time) weak martingales and very-weak martingales.

Proposition 1.2 (Kallenberg, Sztencel) *Let \mathbf{X} be a martingale in \mathbb{E} . Then there exists a martingale \mathbf{Y} in \mathbb{R}^2 such that for every i , $\|Y_i\|$ has the same distribution as $\|X_i\|$ and $\|Y_i - Y_{i-1}\|$ has the same distribution as $\|X_i - X_{i-1}\|$.*

This result also applies to very-weak martingales (proof in Section 2).

Proposition 1.3 *Let \mathbf{X} be a very-weak martingale in \mathbb{E} . Then there exists a martingale \mathbf{Y} in \mathbb{R}^2 such that for every i , $\|Y_i\|$ has the same distribution as $\|X_i\|$ and $\|Y_i - Y_{i-1}\|$ has the same distribution as $\|X_i - X_{i-1}\|$.*

Kallenberg and Sztencel were also able to prove the following generalization of Azuma’s Inequality (again, we only state the version for discrete martingales—the same result holds for local martingales with symmetric jumps).

Proposition 1.4 (Kallenberg, Sztencel) *Let \mathbf{X} be a martingale in \mathbb{E} , such that for every i , $\|X_i - X_{i-1}\| \leq 1$. Then*

$$\Pr[\|X_n\| \geq a] = O\left((a+1)e^{-a^2/2n}\right). \quad (2)$$

By a more detailed analysis, in Section 3, we eliminate the dependence on a outside the exponent. Moreover, our inequality holds not only for martingales, but also for very-weak martingales.

Theorem 1.5 *Let \mathbf{X} be a \mathbb{E} -valued very-weak martingale such that for every i , $\|X_i - X_{i-1}\| \leq 1$. Then*

$$\Pr[\|X_n\| \geq a] < 2e^{1-(a-1)^2/2n} < 2e^2 e^{-a^2/2n} \quad (3)$$

Our next result shows that, for many purposes (such as Theorem 1.5), a given very-weak martingale may be replaced by an “equivalent” strong martingale.

Theorem 1.6 *Let \mathbf{X} be a very-weak martingale taking values in \mathbb{R}^d . Then there exists a martingale \mathbf{Y} taking values in \mathbb{R}^d such that for each $i \geq 1$, (Y_{i-1}, Y_i) has the same distribution as (X_{i-1}, X_i) .*

This result, combined with Proposition 1.3, allows us to automatically extend a certain class of theorems about strong martingales in \mathbb{R}^2 to apply to very-weak martingales in \mathbb{R}^d or ℓ_2 . The proof is in Section 5.

The question which motivated our study of martingales in higher dimensions was to find upper bounds on Fourier coefficients of random subsets of finite abelian groups.

Let f be a function on a finite abelian group G . For a given character $\chi : G \rightarrow \mathbb{C}^\times$, the Fourier coefficient of f corresponding to χ is the sum $\frac{1}{|G|} \sum_{x \in G} \chi(x) f(x)$. We are interested in upper bounds on the size of these coefficients.

Definition 1.4 Let G be a finite abelian group, and let $f : G \rightarrow \mathbb{C}$ be any function. We denote

$$\Phi(f) := \max_{\chi} \left| \sum_{x \in G} \chi(x) f(x) \right|,$$

where the maximum is taken over all characters $\chi : G \rightarrow \mathbb{C}^\times$ *except* the principal character $\chi \equiv 1$. In the special case where $f : G \rightarrow \{0, 1\}$ is the characteristic function of a subset $S \subseteq G$, we will write $\Phi(S)$ in place of $\Phi(f)$.

László Babai posed the question, “How large is $\Phi(S)$ when $S \subseteq G$ is a randomly chosen subset of G ?” We prove that $\Phi(S)$ is very likely to be small.

Theorem 1.7 *Let $\epsilon > 0$. Let G be a finite abelian group of order n . For all but a $O(n^{-\epsilon})$ fraction of subsets $S \subseteq G$,*

$$\Phi(S) < \sqrt{\frac{1 + \epsilon}{2} n \ln(n)}.$$

Babai also asked about the case when the size of the subset is specified. This question is more interesting, because the Fourier coefficients are not sums of n independent random variables. We can still prove a similar upper bound on $\Phi(S)$ however.

Theorem 1.8 *Let $\epsilon > 0$. Let G be a finite abelian group of order n . Let $k \leq n$, and let $m = \min\{k, n - k\}$. For all but an $O(n^{-\epsilon})$ fraction of subsets $S \subseteq G$ such that $|S| = k$,*

$$\Phi(S) < 2\sqrt{2(1 + \epsilon) m \ln(n)}.$$

Proofs of Theorems 1.7 and 1.8 are presented in Section 6.

2 Reduction to Two Dimensions

In this section, we present a proof of Proposition 1.3. In light of Theorem 1.6, It will suffice for us to prove that, given a very-weak martingale \mathbf{X} in \mathbb{E} , there exists a very-weak martingale \mathbf{Y} in \mathbb{R}^2 having the desired properties. Then, by Theorem 1.6, which will be proved in Section 5, \mathbf{Y} may be replaced by a strong martingale in \mathbb{R}^2 also having the desired properties.

Our proof proceeds along roughly the same lines as Kallenberg and Sztencel's proof of Proposition 1.2. However, the proof in [Kallenberg, Sztencel] is complicated by technical difficulties related to continuous time, which we do not encounter here. Kallenberg and Sztencel mention the existence of an elementary proof for discrete time, but we could not find such a proof in the literature.

Proof of Proposition 1.3, modulo Theorem 1.6: Given a very-weak martingale X in \mathbb{E} , we define a very-weak martingale Y in \mathbb{R}^2 , such that, for all i , $\|Y_i - Y_{i-1}\| = \|X_i - X_{i-1}\|$ and $\|Y_i\| = \|X_i\|$. In our construction, we will need a countable sequence of fair coin tosses, independent of \mathbf{X} and each other. Without loss of generality, we assume that \mathbf{X} is defined on a probability space containing such a sequence.

The Y_i are defined recursively, starting with $Y_0 = 0$. Suppose $i > 0$, and Y_{i-1} has been defined, satisfying the desired conditions. We define Y_i so that the ordered triangle $(-Y_i, Y_{i-1}, Y_i - Y_{i-1})$ is congruent to the ordered triangle $(-X_i, X_{i-1}, X_i - X_{i-1})$. Since Y_{i-1} has been defined, the problem is to embed a given ordered triangle in the plane, with one edge specified. There are two choices for how to embed the rest of the triangle (the cycle may be clockwise or counterclockwise). We use a coin flip to decide which orientation to use.

Since we defined Y_i to make the triangles $(-Y_i, Y_{i-1}, Y_i - Y_{i-1})$ and $(-X_i, X_{i-1}, X_i - X_{i-1})$ congruent, it is clear that $\|Y_i\|$ has the same distribution as $\|X_i\|$ and that $\|Y_i - Y_{i-1}\|$ has the same distribution as $\|X_i - X_{i-1}\|$. It remains to be checked that \mathbf{Y} is a very-weak martingale.

To see this, we write $Y_i = \alpha_{Y,i}Y_{i-1} + P_{Y,i}$, where $\alpha_{Y,i}Y_{i-1}$ is the component of Y_i in the direction of Y_{i-1} and $P_{Y,i}$ is perpendicular to Y_{i-1} . Similarly, break X_i up as $X_i = \alpha_{X,i}X_{i-1} + P_{X,i}$. Since \mathbf{X} is a very-weak martingale, we know that

$$\begin{aligned} X_{i-1} &= \mathbb{E}(X_i | X_{i-1}) \\ &= \mathbb{E}(\alpha_{X,i} | X_{i-1})X_{i-1} + \mathbb{E}(P_{X,i} | X_{i-1}). \end{aligned}$$

Since $P_{X,i}$ is perpendicular to X_{i-1} , it follows that $\mathbb{E}(P_{X,i} | X_{i-1}) = 0$, so $\mathbb{E}(\alpha_{X,i} | X_{i-1}) = 1$.

Since the triangles $(-Y_i, Y_{i-1}, Y_i - Y_{i-1})$ and $(-X_i, X_{i-1}, X_i - X_{i-1})$ are congruent, we have $\alpha_{Y,i} = \alpha_{X,i}$. Since $\alpha_{X,i}$ does not depend on the coin flips,

$$\mathbb{E}(\alpha_{Y,i} | Y_{i-1}) = \mathbb{E}(\alpha_{X,i} | Y_{i-1}) = \mathbb{E}(\alpha_{X,i} | X_{i-1}) = 1.$$

Since, when we are choosing the orientation of the triangle $(-Y_i, Y_{i-1}, Y_i - Y_{i-1})$ in the plane, the two possible choices result in opposite values for $P_{Y,i}$, and since we choose the orientations with equal probability, we must have

$$\mathbb{E}(P_{Y,i} | Y_{i-1}) = 0.$$

Putting these facts together, we have

$$\begin{aligned} \mathbb{E}(Y_i | Y_{i-1}) &= \mathbb{E}(\alpha_{Y,i} | Y_{i-1})Y_{i-1} + \mathbb{E}(P_{Y,i} | Y_{i-1}) \\ &= Y_{i-1}. \end{aligned}$$

Thus \mathbf{Y} is a very-weak martingale, as claimed.

We note that if \mathbf{X} is a weak martingale, then \mathbf{Y} is a weak martingale, and if \mathbf{X} is a strong martingale, then \mathbf{Y} is a strong martingale; the proofs follow the same lines as the very-weak martingale case. \blacksquare

2.1 Dimension 2 is necessary

Having seen that martingales in any dimension may be reduced to martingales in the plane, as far as large deviations are concerned, it is worth seeing that we cannot immediately reduce all the way to a martingale in \mathbb{R} .

Definition 2.1 Let \mathbf{PS} be a martingale in \mathbb{R}^2 in which each step is of unit length, and is perpendicular to the previous total. (We may take $\mathbf{PS}_1 = \pm 1$, to be determined by fair coin toss, and determine the k 'th step by tossing a coin to decide between the two possible vectors perpendicular to \mathbf{PS}_{k-1} .) We will call \mathbf{PS} the perpendicular-steps martingale.

For this martingale, the square of the distance from the origin increases by one at each step, so the deviation is always \sqrt{n} after n steps. In contrast, any unit-step martingale in \mathbb{R} has positive probability of remaining in $[-1, 1]$ for n steps.

3 A Bound on Large Deviations

In this section, we prove Theorem 1.5 for the case when X is a strong martingale. The proof for very-weak martingales will follow by Theorem 1.6 (Section 5).

Before presenting the proof, we describe the general approach and some obstacles. The strategy for the proof of Azuma’s Inequality, Proposition 1.1, presented in [ASE, pp. 83–85, 233–240], is to bound $E(e^{\lambda|X_n|})$, using the

equalities $e^{\lambda|X_n|} = \max\{e^{\pm\lambda X_n}\}$ and $e^{\lambda X_n} = \prod_{i=1}^n e^{\lambda(X_i - X_{i-1})}$. Obviously the

first equality does not generalize to higher dimensions. Nevertheless, we are able to follow a similar strategy, bounding $e^{\lambda\|X_n\|}$ using the equality $e^{\lambda\|X_n\|} =$

$$\prod_{i=1}^n e^{\lambda(\|X_i\| - \|X_{i-1}\|)}.$$

It turns out that the terms in this product have small conditional expectations when $\|X_i\|$ is sufficiently large. When $\|X_i\|$ is small, however, problems arise. We would encounter similar difficulties in the proof for martingales on the real line, if we tried to bound $\prod_{i=1}^n e^{\lambda(|X_i| - |X_{i-1}|)}$ directly.

The problematic behavior when $\|X_i\|$ is small is caused by the possibility of substantial “free progress” away from the origin. On the real line, as long as $|X_{i-1}| > 1$, the conditional expectation of $|X_i| - |X_{i-1}|$ is zero, conditioned on the history. In higher dimensions, $\|X_i\| - \|X_{i-1}\|$ may never have conditional expectation zero, as shown by the perpendicular steps martingale (Section 2.1). However, this expectation tends to zero as $\|X_i\|$ becomes large.

To handle this problem, we define a new random series \mathbf{X}' , which has a “similar shape” to \mathbf{X} , but never comes close to the origin, and which consequently never makes much free progress. On the other hand, no matter how much free progress \mathbf{X} makes, it cannot “catch up:” $\|X_i\| < \|X'_i\|$ with probability 1. Figure 3 illustrates the situation when X is a perpendicular steps martingale.

We will not present an explicit description of \mathbf{X}' , as it is not needed in



Figure 1: At left: the given martingale, \mathbf{X} . At right: the induced random walk \mathbf{X}' , starting at $(\lambda, 0)$. \mathbf{X}' behaves like a martingale, except when this would cause $\|X'_n\| < \lambda$. Note that while \mathbf{X} has moved $\sqrt{5}$ units from the origin, $\|X'_5\|$ is still very close to $\|X'_0\| = \lambda$.

the proof. However, we will study its norm, \mathbf{Y} , which we do define. The interested reader should be able to construct a suitable definition of \mathbf{X}' from the definition of \mathbf{Y} and from Figure 3. We now proceed to the proof.

Proof of Theorem 1.5, assuming \mathbf{X} is a strong martingale: Let \mathbf{X} be a given (strong) martingale in \mathbb{E} .

We recursively define several real-valued series, \mathbf{A} , \mathbf{D} , \mathbf{Z} , and \mathbf{Y} , which will be useful for our analysis.

As in the proof of Proposition 1.2, we define

$$X_i = \alpha_{X,i} X_{i-1} + P_i,$$

where $\alpha_{X,i}$ is a real number, and P_i is the component of X_i orthogonal to X_{i-1} . Define the series $\mathbf{A} = (A_i)$ by $A_i := (\alpha_{X,i} - 1)\|X_{i-1}\|$. We observe that

$$1 \geq \|X_i - X_{i-1}\|^2 = A_i^2 + \|P_i\|^2.$$

The three random walks \mathbf{D} , \mathbf{Z} , \mathbf{Y} are defined simultaneously by induction, as follows:

$$\begin{aligned} Y_0 &= 1 + \lambda^{-1} & Z_0 &= 0 \\ D_i &= \widetilde{\text{sgn}}(Z_{i-1}) \left(\sqrt{Y_{i-1}^2 + 2Y_{i-1}A_i + 1} - Y_{i-1} \right) \\ Z_i &= Z_{i-1} + D_i \\ Y_i &= Y_0 + |Z_i|. \end{aligned}$$

With foreknowledge of the proof, we define $\lambda := (a - 1)/n$. The function $\widetilde{\text{sgn}} : \mathbb{R} \rightarrow \{-1, +1\}$ is defined by

$$\widetilde{\text{sgn}}(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ -1 & \text{if } x < 0 \end{cases}$$

Claim 3.0.1 For $1 \leq i \leq n$,

$$|Z_i| = \left| \sqrt{Y_{i-1}^2 + 2Y_{i-1}A_i + 1} - Y_0 \right|.$$

Proof: By the definitions of Z_i , D_i , and Y_i , respectively,

$$\begin{aligned} \widetilde{\text{sgn}}(Z_{i-1})Z_i &= \widetilde{\text{sgn}}(Z_{i-1})Z_{i-1} + \widetilde{\text{sgn}}(Z_{i-1})D_i \\ &= |Z_{i-1}| + \sqrt{Y_{i-1}^2 + 2Y_{i-1}A_i + 1} - Y_{i-1} \\ &= \sqrt{Y_{i-1}^2 + 2Y_{i-1}A_i + 1} - Y_0. \end{aligned}$$

The claim follows by taking absolute values. ■

Claim 3.0.2 For all i , $Y_i > \|X_i\|$.

Proof: (Induction on i .) $Y_0 = 1 + \lambda^{-1} > 0 = \|X_0\|$. Suppose the claim holds for $i - 1$. By Claim 3.0.1, we have

$$\begin{aligned} Y_i &= Y_0 + |Z_i| = Y_0 + \left| \sqrt{Y_{i-1}^2 + 2Y_{i-1}A_i + 1} - Y_0 \right| \\ &\geq \sqrt{Y_{i-1}^2 + 2Y_{i-1}A_i + 1}. \end{aligned}$$

This can be rewritten as

$$Y_i^2 \geq (Y_{i-1} + A_i)^2 + 1 - A_i^2.$$

By the definition of A_i , it is easy to see that

$$\|X_i\|^2 = (\|X_{i-1}\|^2 + A_i)^2 + 1 - A_i^2.$$

Subtracting, we have

$$Y_i^2 - \|X_i\|^2 \geq (Y_{i-1} + A_i)^2 - (\|X_{i-1}\|^2 + A_i)^2 > 0.$$

The last inequality above follows because by induction, $Y_{i-1} > \|X_{i-1}\|$, and by definition, $Y_{i-1} \geq Y_0 > 2 \geq 2|A_i|$. ■

Claim 3.0.3 Fix Z_{i-1} . Then the random variable $f(A_i) := e^{\lambda D_i}$ satisfies

$$f(A_i) \leq \cosh \lambda + A_i \widetilde{\text{sgn}}(Z_{i-1}) \sinh \lambda.$$

Proof: By definition, when $A_i = \pm 1$, $D_i = \widetilde{\text{sgn}}(Z_{i-1})A_i$. Hence $f(-1) = e^{-\lambda \widetilde{\text{sgn}}(Z_{i-1})A_i}$ and $f(1) = e^{\lambda \widetilde{\text{sgn}}(Z_{i-1})A_i}$. The secant line joining $(-1, f(-1))$ to $(1, f(1))$ has equation

$$y = \cosh \lambda + A_i \widetilde{\text{sgn}}(Z_{i-1}) \sinh \lambda.$$

Thus, to prove the claim, it suffices to show that $\frac{d^2 f}{dA_i^2} > 0$ on the interval $(-1, 1)$.

Differentiating and simplifying, we have

$$\begin{aligned} \frac{d^2 f}{dA_i^2} &= \left(\left(\lambda \frac{dD_i}{dA_i} \right)^2 + \lambda \frac{d^2 D_i}{dA_i^2} \right) f \\ &= \left(\lambda \sqrt{Y_{i-1}^2 + 2Y_{i-1}A_i + 1} - \widetilde{\text{sgn}}(Z_{i-1}) \right) \frac{\lambda Y_{i-1}^2 f(A_i)}{(Y_{i-1}^2 + 2Y_{i-1}A_i + 1)^{3/2}} \end{aligned}$$

Since $\sqrt{Y_{i-1}^2 + 2Y_{i-1}A_i + 1} \geq Y_{i-1} - 1 \geq Y_0 - 1 \geq |\lambda|^{-1} > 1$, we see that $\frac{d^2 f}{dA_i^2} > 0$ as desired. ■

Claim 3.0.4 $\mathbb{E}(e^{\lambda D_n} \mid Z_{n-1}) \leq \cosh \lambda$.

Proof: By Fubini's Theorem,

$$\mathbb{E}(e^{\lambda D_n} \mid Z_{n-1}) = \mathbb{E}(\mathbb{E}(e^{\lambda D_n} \mid X_0, \dots, X_{n-1}) \mid Z_{n-1}). \quad (4)$$

By Claim 3.0.3, we have

$$\mathbb{E}(e^{\lambda D_n} \mid X_0, \dots, X_{n-1}) \leq \cosh \lambda + \mathbb{E}(A_n \sinh \lambda \mid X_0, \dots, X_{n-1}) = \cosh \lambda.$$

Because X_0, \dots, X_{n-1} determine Z_{n-1} , this upper bound applies to $\mathbb{E}(e^{\lambda D_n} \mid X_0, \dots, X_{n-1}, Z_{n-1})$ as well. Applying this bound to equation (4) proves the Claim. ■

By Fubini's Theorem and Claim 3.0.4, we deduce

$$\begin{aligned} \mathbb{E}(e^{\lambda Z_n}) &= \mathbb{E}(e^{\lambda Z_{n-1}} e^{\lambda D_n}) \\ &= \mathbb{E}(e^{\lambda Z_{n-1}} \mathbb{E}(e^{\lambda D_n} \mid Z_{n-1})) \\ &\leq \mathbb{E}(e^{\lambda Z_{n-1}}) \cosh \lambda \end{aligned}$$

By induction, it follows that

$$\mathbb{E} (e^{\lambda Z_n}) \leq (\cosh \lambda)^n. \quad (5)$$

Now, by Claim 3.0.2 and the definition of Y_n , we have:

$$e^{\lambda \|X_n\|} < e^{\lambda Y_n} = e^{\lambda(Y_0 + |Z_n|)} \leq e^{\lambda Y_0} (e^{\lambda Z_n} + e^{-\lambda Z_n}). \quad (6)$$

Combining (5) and (6), we find

$$\mathbb{E} (e^{\lambda \|X_n\|}) < e^{\lambda Y_0} 2 (\cosh \lambda)^n. \quad (7)$$

Applying Markov's Inequality and equation (7), we have

$$\begin{aligned} \Pr [\|X_n\| \geq a] &= \Pr [e^{\lambda \|X_n\|} \geq e^{\lambda a}] \\ &\leq \mathbb{E} (e^{\lambda \|X_n\| - \lambda a}) \\ &\leq e^{\lambda Y_0 - \lambda a} 2 (\cosh \lambda)^n. \end{aligned}$$

We will use the inequality $\cosh \lambda \leq e^{\lambda^2/2}$ (which can easily be seen by examining the power series for each side), and the definition $Y_0 = 1 + \lambda^{-1}$, to simplify the right hand side. Choosing $\lambda = (a-1)/n$ minimizes the resulting expression. Thus we have

$$\Pr [\|X_n\| \geq a] \leq 2e^{\lambda+1-\lambda a+\lambda^2 n/2} = 2e^{1-(a-1)^2/2n}.$$

■

Remark 3.1 Unlike the proof of the 1-dimensional Azuma inequality in [ASE], this proof technique apparently does not apply directly to very-weak martingales; instead we must appeal to Theorem 1.6. A careful look at the proof of Claim 3.0.4 shows the obstacle: we cannot replace the conditioning on X_0, \dots, X_{n-1} here with conditioning only on X_{n-1} , because the real number Z_{n-1} depends essentially on the entire history of \mathbf{X}_{n-1} , not just the current value X_{n-1} . We would be interested in a proof of Theorem 1.5 which applies directly to the very-weak martingale case as well.

4 An Example of a very-weak martingale

One reason why weak martingales and very-weak martingales may be of interest is that they may be realized over much smaller sample spaces than strong martingales. Indeed, if a martingale \mathbf{X} has all steps of positive length with probability one, then the random variable (X_1, \dots, X_n) induces a subalgebra with at least 2^n atoms, by an easy induction. However, there are very-weak martingales \mathbf{X} for which all steps have positive length, but (X_1, \dots, X_n) has only $O(n^2)$ atoms.

In this section, we give a very simple construction of such a very-weak martingale.

Construction 4.1 For each real number α in $[0, 1]$, define the “fake coin-flip” series by

$$\text{FCF}_i(\alpha) = -i + 2 \min \left\{ k \mid \sum_{j=0}^k \binom{i}{j} 2^{-i} \geq \alpha \right\}$$

Let α be selected uniformly at random from $[0, 1]$.

Proposition 4.1 **FCF** is a very-weak martingale which is not a weak martingale. Let **CF** be the coin-flip martingale. Then, for every n , CF_n has the same distribution as FCF_n , and $\text{CF}_n - \text{CF}_{n-1}$ has the same distribution as $\text{FCF}_n - \text{FCF}_{n-1}$.

Proof: Using the standard recurrence

$$\binom{i}{j} = \binom{i-1}{j} + \binom{i-1}{j-1},$$

we can see that, given that $\text{FCF}_{i-1} = x$, there is a $1/2$ chance that $\text{FCF}_i = x + 1$ and a $1/2$ chance that $\text{FCF}_i = x - 1$. This proves that **FCF** is a very-weak martingale, and also the claims about the distributions of FCF_n and $\text{FCF}_n - \text{FCF}_{n-1}$.

To see that **FCF** is not a weak martingale, observe that if $\text{FCF}_2(\alpha) = 2$, then $\alpha \geq 3/4$, so, for instance, $\text{FCF}_6(\alpha) \geq 4$ with probability 1. More generally, for any integers $0 < x \leq j$, where $j \equiv x \pmod{2}$,

$$\text{E}(\text{FCF}_n \mid \mathbf{FCF}_j = x) = \Theta(\sqrt{n}).$$

Next, we describe a combinatorial difference between the class of very-weak martingales and that of strong martingales. ■

Proposition 4.2 *For every n , $|\text{FCF}_n - \text{FCF}_{n-1}| = 1$, and the random variable $(\text{FCF}_1, \dots, \text{FCF}_n)$ has $O(n^2)$ atoms. Let \mathbf{X} be a martingale satisfying $|X_n - X_{n-1}| = 1$. Then (X_1, \dots, X_n) has at least 2^n atoms.*

Proof: The atoms of $(\text{FCF}_1, \dots, \text{FCF}_n)$ are the intervals between consecutive values of the form

$$\sum_{j=0}^k \binom{i}{j} 2^{-i},$$

where $0 \leq i \leq n$. Since there are $\binom{n+2}{2}$ terms of this form, counting multiplicity, there are at most this many atoms.

Suppose \mathbf{X} is a martingale satisfying $|X_n - X_{n-1}| > 0$ for every n . Then every atom of (X_1, \dots, X_{n-1}) is a disjoint union of at least two atoms of (X_1, \dots, X_n) (since the average of $X_n - X_{n-1}$ over these atoms must be zero). By induction, it follows that (X_1, \dots, X_n) has at least 2^n atoms. ■

5 Turning very-weak martingales into strong

We now present the proof of Theorem 1.6, which for any very-weak martingale in \mathbb{R}^d , shows the existence of a strong martingale for which the terms X_n and the steps $X_n - X_{n-1}$ have the same distribution as the original series.

We present the proof in two parts: first the proof for discrete very-weak martingales, second the proof for arbitrary very-weak martingales in \mathbb{R}^d . Most readers will probably be satisfied with the discrete case, which has a simple proof.

Proof of Theorem 1.6 (discrete case): Let \mathbf{X} be a very-weak martingale taking values in a real vector space V . Furthermore let \mathbf{X} be discrete: each X_n takes values in a finite set. Let P be the probability measure on which \mathbf{X} is defined.

We construct \mathbf{Y} and its underlying probability measure Q recursively, starting with $Y_0 = 0$.

Assume that $\mathbf{Y}_{n-1} = (Y_0, \dots, Y_{n-1})$ has already been defined, with underlying probability measure Q_{n-1} . To define the distribution of $\mathbf{Y}_n = (Y_0, \dots, Y_n)$, we need to specify how the atoms of Q_{n-1} should be refined.

Let $(y_0, \dots, y_n) \in (\mathbb{R}^d)^n$ be any sequence of values. We define

$$Q_n[\mathbf{Y}_n = (y_0, \dots, y_n) \mid \mathbf{Y}_{n-1} = (y_0, \dots, y_{n-1})] = P[X_n = y_n \mid X_{n-1} = y_{n-1}].$$

In other words, if X_{n-1} has positive probability of equaling y_{n-1} , we set

$$Q_n[\mathbf{Y}_n = (y_0, \dots, y_n)] = \frac{P[X_{n-1} = y_{n-1}, X_n = y_n]Q_{n-1}[\mathbf{Y}_{n-1} = (y_0, \dots, y_{n-1})]}{P[X_{n-1} = y_{n-1}]},$$

and if not, all probabilities in the above equation are zero.

Intuitively, at the n 'th step, \mathbf{Y} “forgets” how it arrived at y_{n-1} , then chooses a random continuation according to \mathbf{X} . The series \mathbf{Y} thus defined obviously satisfies the desired criteria. ■

In the general case, we “construct” \mathbf{Y} by specifying its finite joint distribution functions. We employ standard techniques to demonstrate the existence of a random series with the specified distributions. The basic idea of the construction remains the same: \mathbf{Y} is obtained by modifying \mathbf{X} to “forget the history” at each step.

We now present the proof for the general case.

Proof of Theorem 1.6 (general case): We define \mathbf{Y} implicitly by specifying its distribution on finite-dimensional rectangles. The existence of a countable series \mathbf{Y} with the given distributions follows from the Kolmogorov extension theorem [Breiman, Corollary 2.19].

Let P be the probability measure on which \mathbf{X} is defined. Denote by Q the new probability measure being constructed, on which \mathbf{Y} will be defined. Note that Construction 4.1 shows that Q cannot generally be taken to be P .

Let $Y_0 = 0$. For all n , and all intervals $A \subseteq (\mathbb{R}^d)^{n-1}$, $B \subseteq \mathbb{R}^d$, define

$$Q_n[\mathbf{Y}_n \in A \times B] = E_x(Q_{n-1}[\mathbf{Y}_{n-1} \in A \mid Y_{n-1} = x]P[X_n \in B \mid X_{n-1} = x]). \tag{8}$$

Here x is an independent random variable sampled from the same distribution as X_{n-1} (and as Y_{n-1} : the distributions are the same by inductive hypothesis). Let us briefly explain the meaning of terms on the right-hand side of (8).

Let μ be the law of X_{n-1} , i. e. the probability measure on \mathbb{R}^d defined by

$$\mu[S] = P[X_{n-1} \in S],$$

for all Borel sets $S \subseteq \mathbb{R}^d$. By inductive hypothesis, μ is also the law of Y_{n-1} .

The conditional probabilities on the right-hand side of (8) are defined in terms of Radon-Nikodym derivatives (cf. [Williams]), in a manner we will now set forth. For each interval $A \subseteq (\mathbb{R}^d)^{n-2}$, let Q_A be the probability measure on \mathbb{R}^d defined by

$$Q_A(S) = Q_{n-1}[\mathbf{Y}_{n-1} \in A \times S].$$

Then Q_A is absolutely continuous with respect to μ , and the Radon-Nikodym derivative $dQ_A/d\mu$ is a μ -measurable function f such that, for every interval $S \subseteq \mathbb{R}^d$,

$$Q_A(S) = \int_{x \in S} f(x) d\mu(x).$$

We define

$$Q_{n-1}[\mathbf{Y}_{n-1} \in A \times S \mid Y_{n-1} = x] = f(x)\psi_S(x),$$

where $\psi_S : S \rightarrow \{0, 1\}$ is the characteristic function of S . With this notation, the definition of the R-N derivative f may be rewritten as

$$Q_{n-1}[\mathbf{Y}_{n-1} \in A \times S] = \mathbb{E}_x(Q_{n-1}[\mathbf{Y}_{n-1} \in A \times S \mid Y_{n-1} = x]),$$

where x is selected according to μ .

Similarly, if P_B is the probability measure on \mathbb{R} defined by

$$P_B(S) = P[X_{n-1} \in S, X_n \in B],$$

then we denote the Radon-Nikodym derivative $dP_B/d\mu$ by $P[X_n \in B \mid X_{n-1} = x]$.

Since the conditional probabilities on the right-hand side of (8) are now seen to be random variables in the range $[0, 1]$, the right-hand side is well-defined, and is in $[0, 1]$. It is easy to check that this definition is finitely additive, and hence extends to a probability measure. Furthermore, it is easy to check that the sequence of distributions defined by (8) meet the regularity criteria of the Kolmogorov extension theorem [Breiman, Corollary 2.19], and so define a countable random series.

To verify that (Y_{n-1}, Y_n) has the same distribution as (X_{n-1}, X_n) , we just chase definitions:

$$\begin{aligned}
Q [Y_{n-1} \in A, Y_n \in B] &= Q [\mathbf{Y}_n \in (\mathbb{R}^d)^{n-2} \times A \times B] \\
&= E (Q [\mathbf{Y}_{n-1} \in (\mathbb{R}^d)^{n-2} \times A \mid Y_{n-1} = x] P [X_n \in B \mid X_{n-1} = x]) \\
&= E (\psi_A(x) P [X_n \in B \mid X_{n-1} = x]) \\
&= E (P [X_{n-1} \in A, X_n \in B \mid X_{n-1} = x]) \\
&= P [X_{n-1} \in A, X_n \in B].
\end{aligned}$$

The martingale condition $E(Y_n \mid Y_0, \dots, Y_{n-1}) = Y_{n-1}$ may be rewritten in the form

$$E(\phi(\mathbf{Y}_n) Y_n) = E(\phi(\mathbf{Y}_n) Y_{n-1}), \quad (9)$$

for every measurable function $\phi : (\mathbb{R}^d)^n \rightarrow \mathbb{R}$. By standard arguments, it suffices to prove this statement in the special case when $\phi = \psi_S$ is the characteristic function of an interval $S \subseteq (\mathbb{R}^d)^n$.

Decompose $S = A \times B$, where $A \subseteq (\mathbb{R}^d)^{n-1}$ and $B \subseteq \mathbb{R}^d$ are intervals.

To verify (9) we expand the left side using (8) (the defining recurrence for \mathbf{Y}). Writing everything explicitly in terms of Radon-Nikodym derivatives yields a rather messy intermediate expression, which may in turn be rewritten in a nice form, using Fubini's Theorem and the Dominated Convergence Theorem. The result is:

$$E(\psi_{A \times B}(\mathbf{Y}_n) Y_n) = E_x(Q [Y_{n-1} \in A \mid Y_{n-1} = x] E(\psi_B(X_n) X_n \mid X_{n-1} = x))$$

where x is sampled according to the distribution μ introduced earlier. The right side of (9) may be expanded in the same way to obtain

$$E(\psi_{A \times B}(\mathbf{Y}_n) Y_{n-1}) = E_x(Q [Y_{n-1} \in A \mid Y_{n-1} = x] E(\psi_B(X_n) X_{n-1} \mid X_{n-1} = x)).$$

Since \mathbf{X} is a very-weak martingale, we have $E(\psi_B(X_n) X_n \mid X_{n-1} = x) = E(\psi_B(X_n) X_{n-1} \mid X_{n-1} = x)$ a.e. μ . Substituting, we obtain

$$E(\psi_{A \times B}(\mathbf{Y}_n) Y_n) = E(\psi_{A \times B}(\mathbf{Y}_n) Y_{n-1}),$$

which shows that \mathbf{Y} is a martingale. ■

6 Random Subsets have all small Fourier Coefficients

In this section, we present the proofs of Theorems 1.7 and 1.8.

Proof of Theorem 1.7: Let $G = \{g_1, \dots, g_n\}$ be a finite abelian group of order n . Let $S \subseteq G$ be a subset of G selected uniformly at random, and let $\psi_S : G \rightarrow \{0, 1\}$ be the characteristic function of S . Then the values $\psi_S(g_1), \dots, \psi_S(g_n)$ are independent $\{0, 1\}$ -valued random variables.

Let $\chi : G \rightarrow \mathbb{C}^\times$ be a non-principal character of G . Define a sequence $\mathbf{X} = (X_0, \dots, X_n)$ by

$$X_i = \sum_{j=1}^i \chi(g_j)(2\psi_S(g_j) - 1).$$

Since the steps $X_i - X_{i-1}$ are independent random variables of unit norm, Theorem 1.5 implies

$$\Pr[|X_n| \geq a] < 2e^2 e^{-a^2/2n},$$

for every $a > 0$. Choosing $a = \sqrt{2(1 + \epsilon)n \ln(n)}$, we have

$$\Pr[|X_n| \geq a] < 2e^2 n^{-(1+\epsilon)},$$

Now, X_n is $2n$ times the Fourier coefficient of ψ_S corresponding to the character χ . To see this, recall that

$$X_n = 2 \sum_{x \in G} \chi(x) \psi_S(x) - \sum_{x \in G} \chi(x).$$

The second sum is zero, a basic property of non-principal characters (cf. [Babai]), so $X_n = 2 \sum_{x \in G} \chi(x) \psi_S(x)$. In fact, $X_i = \mathbb{E}(X_n \mid \psi_S(g_1), \dots, \psi_S(g_i))$ for each $0 \leq i < n$ (this is a standard technique for defining martingales).

Since there are $n - 1$ non-principal characters χ , and for each, the probability that the Fourier coefficient is at least a is $O(n^{-(1+\epsilon)})$, it follows that

$$\Pr[\Phi(S) \geq a] = O(n^{-\epsilon}),$$

which concludes the proof. ■

The proof of Theorem 1.7 did not really take full advantage of the power of Theorem 1.5, since the random variables being summed were independent,

a much stronger assumption than forming a martingale. The case of random sets of fixed size requires us to analyze a random walk which is much more history-dependent.

Proof of Theorem 1.8: Without loss of generality, assume $m = k \leq n/2$ (if not, replace S by $G \setminus S$; this reverses the signs of the non-principal characters, but leaves the magnitudes unchanged). Let us think of S as being chosen as follows: first, pick a random ordering (g_1, g_2, \dots, g_n) of the elements of G , then let $S = \{g_1, \dots, g_k\}$. We next define two random walks \mathbf{X} and \mathbf{Y} , where X_i and Y_i are defined in terms of the random variables g_1, \dots, g_i .

$$\begin{aligned} X_i &:= \sum_{j=1}^i \chi(g_j) \\ Y_i &:= \mathbb{E}(X_k \mid g_1, \dots, g_i) \end{aligned}$$

It is easily seen that $(Y_i/2)$ is a martingale satisfying the condition of Theorem 1.5. Therefore, for every $a > 0$,

$$\Pr[|Y_k| \geq 2a] < 2e^2 e^{-a^2/2k}.$$

Choosing $a = \sqrt{2(1 + \epsilon)k \ln(n)}$, we have

$$\Pr[|Y_k| \geq 2a] < 2e^2 n^{-(1+\epsilon)}.$$

Since $Y_k = X_k = \sum_{x \in S} \chi(x)$, $\Phi(S)$ is the maximum of the $n - 1$ random variables Y_k corresponding to the non-principal characters of G . Thus there is at most a $2e^2 n^{-\epsilon}$ probability that $\Phi(S) \geq 2a$. \blacksquare

7 Notes

It seems likely that Theorem 1.6 holds in more general contexts, such as very-weak martingales in ℓ_2 , or even in Banach spaces (using the Bochner integral to define expectation). Although we are not sure whether we can guarantee the existence of a process which has the desired distributions in these contexts, we know of no obstacle to doing so, except the abstruseness of the subject.

Regarding our upper bound for deviations of martingales in Euclidean space, it seems likely that the upper bound from the original Azuma Inequality holds, without the additional factor of $e^{1+(2a-1)/2n}$.

8 Acknowledgements

We are indebted to László Babai not only for inspiring our investigations by his questions about Fourier coefficients, but also for much encouragement and assistance in preparing this paper. We are also grateful to Todd Rowland and Varsha Dani for helpful conversations.

References

- [ASE] Alon, Spencer, Erdős. *The Probabilistic Method*. John Wiley & Sons, Inc. New York.
- [Azuma] K. Azuma. *Weighted Sums of Certain Dependent Random Variables*. Tôhoku Math. Journ. **19** (1967) 357–367.
- [Babai] L. Babai. *Fourier Transforms and Equations over Finite Abelian Groups, An introduction to the method of trigonometric sums*. University of Chicago Department of Computer Science “Lecture Notes” version 1.2, December 1989.
- [Berman] Jeremy Berman. *An example of a weak martingale*. Annals of Probability **4** (1976), 107–108.
- [Breiman] L. Breiman. *Probability*. Addison-Wesley Publishing Co., 1968.
- [Chatterji] S.D. Chatterji. *Vector-valued martingales and their applications*. Probability in Banach Spaces, Proceedings of the First International Conference on Probability in Banach Spaces. Springer Verlag, 1976.
- [Chernoff] H. Chernoff. *A measure of the asymptotic efficiency for tests of a hypothesis based on the sum of observations*. Annals of Mathematical Statistics **23** (1952) 493–509.
- [Kallenberg, Sztencel] O. Kallenberg, R. Sztencel. *Some dimension-free features of vector-valued martingales*. Probab. Th. Rel. Fields **88** (1991) 215–247.
- [Karatzas, Shreve] I. Karatzas, S.E. Shreve. *Brownian motion and stochastic calculus*. Springer-Verlag, 1988.

- [Nelson] Paul I. Nelson. *A class of orthogonal series related to martingales.*
Annals Math. Stat. **41** (1970) 1684–1694.
- [Williams] D. Williams. *Probability with Martingales.* Cambridge University
Press, 1991.