A large-deviation inequality for vector-valued martingales

Thomas P. Hayes

revised July 26, 2005

Abstract

Let $\mathbf{X} = (X_0, \ldots, X_n)$ be a discrete-time martingale taking values in any real Euclidean space such that $X_0 = 0$ and for all n, $\|X_n - X_{n-1}\| \leq 1$. We prove the large deviation bound

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

This upper bound is within a constant factor, e^2, of the Azuma-Hoeffding Inequality for real-valued martingales. This improves an earlier result of O. Kallenberg and R. Sztencel (1992).

Our inequality holds even for "very-weak martingales," namely, discrete stochastic processes $\mathbf{X}$ which satisfy, for every n,

$$\mathrm{E}\left(X_n \mid X_{n-1}\right) = X_{n-1} \tag{1}$$

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

More generally, we prove that, for every very-weak martingale $\mathbf{X}$ in $\mathbb{R}^d$, there exists a martingale $\mathbf{Y}$ in $\mathbb{R}^d$ 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-Hoeffding Inequality) *Let* $\mathbf{X} = (X_0, X_1, \ldots, X_n)$ *be a real-valued martingale such that* $X_0 = 0$ *and for every* i, $|X_i - X_{i-1}| \leq 1$. *Then, for every* $a > 0$,

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

Our first objective is to replace "real-valued martingale" with "vector-valued martingale." Throughout this paper, $\mathbb{E}$ will denote any real Euclidean space (of finite or infinite dimension), and $\|\cdot\|$ will denote the Euclidean norm.

We will use the following definition in this paper.

Definition 1.2 Let $\mathbf{X} = (X_i : \Omega \to \mathbb{E})$ be a sequence of random vectors taking values in $\mathbb{E}$, such that $X_0 = 0$, and for every $i \geq 1$, $\mathrm{E}\left(\|X_i\|\right) < \infty$ and $\mathrm{E}\left(X_i \mid X_0, X_1, \ldots, X_{i-1}\right) = X_{i-1}$. 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 $\mathrm{E}\left(X_i \mid X_0, X_1, \ldots, X_{i-1}\right)$ to $\mathrm{E}\left(X_i \mid X_{i-1}\right)$.

Definition 1.3 Let $\mathbf{X} = (X_i : \Omega \to \mathbb{E})$ be a sequence of random vectors taking values in $\mathbb{E}$, such that $X_0 = 0$, and for every $i \geq 1$, $\mathrm{E}\left(\|X_i\|\right) < \infty$ and $\mathrm{E}\left(X_i \mid X_{i-1}\right) = X_{i-1}$. 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, introduced by P. Nelson [15] (see also J. Berman [6]).

Definition 1.4 Let $\mathbf{X} = (X_i : \Omega \to \mathbb{E})$ be a sequence of random vectors taking values in $\mathbb{E}$, such that $X_0 = 0$, and for every $j < i$, $\mathrm{E}\left(\|X_i\|\right) < \infty$ and $\mathrm{E}\left(X_i \mid X_j\right) = X_j$. 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. 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 the Azuma-Hoeffding inequality holds with real-valued very-weak martingales in place of martingales. A nice presentation may be found in Alon, Spencer, and Erdős [2, Appendix A]; the reader should note that what that text calls "martingales" are actually "very-weak martingales" in our terminology.

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-Hoeffding 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 real Euclidean space.

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

1.1 Overview of Results

Our first tool is the discrete-time case of a theorem of Kallenberg and Sztencel [11, Theorem 3.1], stating that for many purposes, "local martingales" (cf. [12, Definition 1.5.15, p. 36]) in $\mathbb{E}$ are equivalent to local martingales in $\mathbb{R}^2$.

Proposition 1.5 (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 $\|X_i - X_{i-1}\|$.*

We give a simple direct proof of this result for discrete martingales and show that the result applies even to discrete very-weak martingales, in the following strong sense:

Proposition 1.6 *Let $\mathbf{X}$ be a very-weak martingale in $\mathbb{E}$. Then there exists a (strong) martingale $\mathbf{Y}$ in $\mathbb{R}^2$ such that for every $i \geq 1$, the triples $(\|Y_{i-1}\|, \|Y_i\|, \|Y_i - Y_{i-1}\|)$ and $(\|X_{i-1}\|, \|X_i\|, \|X_i - X_{i-1}\|)$ share the same distribution.*

The proof is in Section 2.

Kallenberg and Sztencel were also able to prove the following generalization of the Azuma-Hoeffding Inequality. Again, we only state the version for discrete martingales with uniformly bounded differences. (See the remarks preceding their continuous version, [11, Theorem 5.3].)

Proposition 1.7 (Kallenberg, Sztencel) *Let $\mathbf{X}$ be a martingale in $\mathbb{E}$, such that for every i, $\|X_i - X_{i-1}\| \leq 1$. Then*

$$\Pr\left[\|X_n\| \geq a\right] \;=\; O\left(\left(1 + \frac{a}{\sqrt{n}}\right) e^{-a^2/2n}\right). \tag{3}$$

(The constant implicit in the big-Oh notation is absolute, perhaps approximately 2.)

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

Theorem 1.8 *Let* $\mathbf{X}$ *be a very-weak martingale taking values in* $\mathbb{E}$ *such that* $X_0 = 0$ *and for every* i, $\|X_i - X_{i-1}\| \leq 1$. *Then, for every* $a > 0$,

$$\Pr\left[\|X_n\| \geq a\right] \; < 2e^{1-(a-1)^2/2n} < 2e^2 e^{-a^2/2n}. \tag{4}$$

Our next result shows that, for many purposes (such as Theorem 1.8), a given very-weak martingale may be replaced by an "equivalent" strong martingale.

Theorem 1.9 *Let* $\mathbf{X}$ *be a very-weak martingale taking values in* $\mathbb{R}^d$. *Then there exists a (strong) 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 theorem, combined with Proposition 1.6, allows us to automatically extend a certain class of results about strong martingales in $\mathbb{R}^2$ to apply to very-weak martingales in an arbitrary Euclidean space. 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. We will assume some familiarity with basic terminology of finite abelian groups and characters; we recommend Babai's lecture notes [5] for a concise introduction, but any abstract algebra textbook should do.

Definition 1.10 Let G be a finite abelian group. A homomorphism $\chi : G \to \mathbb{C}^\times$, from G into the multiplicative group of complex units is called a *character* of G. The all-ones character $\chi_0 : G \to 1$ often plays a special role, and is known as the *principal character.* Let f be a function on a finite abelian group G. For a given character χ, the *Fourier coefficient of f corresponding to* χ is the average $\frac{1}{|G|}\sum_{x \in G} \chi(x) f(x)$.

Note that when f_S is the characteristic function of a subset $S \subseteq G$, the Fourier coefficient corresponding to the principal character is the relative size of the subset, $\chi_0(f) = |S|/|G|$. We are interested in proving upper bounds on the size of non-principal Fourier coefficients, motivating the following definition.

Definition 1.11 (see [5]) Let G be a finite abelian group, and let $f : G \to \mathbb{C}$ be any function. For every character $\chi : G \to \mathbb{C}^\times$, we set

$$\begin{aligned} \Phi_\chi(f) &:= \left| \sum_{x \in G} \chi(x) f(x) \right|, \\ \text{and denote } \Phi(f) &:= \max_{\chi \neq \chi_0} \Phi_\chi(f) \end{aligned}$$

where the maximum is taken over all characters $\chi : G \to \mathbb{C}^\times$ *except* the principal character $\chi_0 \equiv 1$. In the special case where $f_S : G \to \{0, 1\}$ is the characteristic function of a subset $S \subseteq G$, we write $\Phi(S)$ in place of $\Phi(f_S)$.

László Babai suggested using Chernoff bounds to show, when S is a random subset of G, that usually $\Phi(S) = O(\sqrt{n \ln n})$, i. e., "almost all subsets are nice." We prove

Proposition 1.12 *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 posed the same question when the size of the subset is fixed. This case is more interesting, because the Fourier coefficients can no longer be viewed as sums of n independent random variables. Nevertheless, using martingales, we can prove a similar upper bound on $\Phi(S)$:

Theorem 1.13 *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 $|S| = k$,*

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

Remark 1.14 This result raises the following open question (as posed by L. Babai): construct an explicit family of sets of size $m < n^{1-c}$, for which $\Phi(S)$ is close to the probabilistic upper bound of Theorem 1.13. A challenge case is to construct S with $|S| = n^{2/3}$ and $\Phi(S) < n^{1/2-\epsilon}$. Note that the estimate of Proposition 1.12 can be reached (even improved on) by explicit construction; namely, the set of quadratic residues in $GF(n)$, for n a prime power, and more generally, cyclotomic sets S, always satisfy $\Phi(S) < \sqrt{n}$ (cf. [5, Section 6]).

Our proofs of Proposition 1.12 and Theorem 1.13, presented in Section 6, are a nice illustration of the convenience of having a two-dimensional version of the Azuma-Hoeffding inequality. We must in honesty point out that one can prove both results using only the (one-dimensional) Azuma-Hoeffding inequality, if we allow slightly worse constants; see Section 6 for details.

2 Reduction to Two Dimensions

In this section, we present a proof of Proposition 1.6. In light of Theorem 1.9, 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 1.9, 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 1.5. However, the proof in [11, Section 3] 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 [11, p. 223, first paragraph], but we could not find such a proof in the literature.

Proof of Proposition 1.6, modulo Theorem 1.9: Given a very-weak martingale $\mathbf{X}$ in $\mathbb{E}$, we define a very-weak martingale $\mathbf{Y}$ 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 triangle $(0, Y_{i-1}, Y_i)$ is congruent to $(0, X_{i-1}, X_i)$. Since Y_{i-1} has been defined, the problem is to embed a given ordered triangle in the plane, with the first two corners specified. There are two choices for where to place the third corner, which are mirror images of each other over the line containing the first two corners. We use a fair coin flip to assign one of these two possibilities to Y_i.

Since we defined Y_i to make the triangles $(0, Y_{i-1}, Y_i)$ and $(0, X_{i-1}, X_i)$ congruent, it follows that

$$(\|Y_{i-1}\|, \|Y_i\|, \|Y_i - Y_{i-1}\|) = (\|X_{i-1}\|, \|X_i\|, \|X_i - X_{i-1}\|) .$$

All that remains is to show 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

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

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

Since the triangles $(0, Y_{i-1}, Y_i)$ and $(0, X_{i-1}, X_i)$ are congruent, we have $\alpha_{Y,i} = \alpha_{X,i}$. Since $\alpha_{X,i}$ does not depend on the coin flips,

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

Recall that when choosing Y_i given Y_{i-1}, the two possibilities are mirror images over the line containing 0 and Y_{i-1}. This is equivalent to saying that the two possibilities for P_{Y_i} sum to the zero vector. Since we used a fair coin flip to make our selection, it follows that

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

Putting these facts together, we have

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

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

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 1.8 for the case when $\mathbf{X}$ is a strong martingale. The proof for very-weak martingales will follow by Theorem 1.9 (Section 5).

Before presenting the proof, we describe the general approach and some obstacles. The usual strategy for the proof of the Azuma-Hoeffding Inequality, Proposition 1.1, presented in [2, pp. 83–85, 233–240], is to bound $\mathrm{E}\left(e^{\lambda|X_n|}\right)$, 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 1 illustrates the situation when $\mathbf{X}$ is a perpendicular steps martingale.

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 $\|X'_0\| = \lambda$.

We will not present an explicit description of $\mathbf{X}'$, as it is not needed in 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 1. We now proceed to the proof.

Proof of Theorem 1.8, assuming X is a strong martingale: Let $\mathbf{X}$ be a given (strong) martingale in $\mathbb{E}$, such that for every i, $\|X_i - X_{i-1}\| \leq 1$.

We recursively define four 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 1.5, 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:

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

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

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

Claim 3.1 *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,

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

The claim follows by taking absolute values. ∎

Claim 3.2 *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 3.1, 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

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

Subtracting, we have

$$\begin{aligned} Y_i^2 - \|X_i\|^2 &\geq (Y_{i-1} + A_i)^2 - (\|X_{i-1}\| + A_i)^2 + (1 - A_i^2 - \|P_i\|^2) \\ &= (Y_{i-1} - \|X_{i-1}\|)(Y_{i-1} + \|X_{i-1}\| + 2A_i) + (1 - A_i^2 - \|P_i\|^2) \\ &> 0. \end{aligned}$$

The last inequality above follows because

1. $Y_{i-1} > \|X_{i-1}\|$ by induction.
2. $Y_{i-1} \geq Y_0 = 1 + \lambda^{-1} \geq 2 \geq 2|A_i|$ from the definitions.
3. $1 \geq A_i^2 + \|P_i\|^2$ as already observed.

■

Claim 3.3 *Fix Z_{i-1}. Then the random variable $f(A_i) := e^{\lambda D_i}$ satisfies*

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

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

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

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

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

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 3.4 $\mathrm{E}\left(e^{\lambda D_n} \mid Z_{n-1}\right) \leq \cosh \lambda$.

Proof: By Fubini's Theorem,

$$\mathrm{E}\left(e^{\lambda D_n} \mid Z_{n-1}\right) = \mathrm{E}\left(\mathrm{E}\left(e^{\lambda D_n} \mid X_0, \ldots, X_{n-1}\right) \mid Z_{n-1}\right). \tag{5}$$

By Claim 3.3, we have

$$\mathrm{E}\left(e^{\lambda D_n} \mid X_0, \ldots, X_{n-1}\right) \leq \cosh \lambda + \widetilde{\operatorname{sgn}}(Z_{n-1}) \sinh(\lambda) \mathrm{E}\left(A_n \mid X_0, \ldots, X_{n-1}\right) = \cosh \lambda.$$

Because $X_0, \ldots, X_{n-1}$ determine Z_{n-1}, this upper bound applies to $\mathrm{E}\left(e^{\lambda D_n} \mid X_0, \ldots, X_{n-1}, Z_{n-1}\right)$ as well. Applying this bound to equation (5) proves the Claim. ■

By Fubini's Theorem and Claim 3.4, we deduce

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

By induction, it follows that

$$\mathrm{E}\left(e^{\lambda Z_n}\right) \leq (\cosh \lambda)^n. \tag{6}$$

Now, by Claim 3.2 and the definition of Y_n, we have:

$$e^{\lambda \|X_n\|} < e^{\lambda Y_n} = e^{\lambda(Y_0 + |Z_n|)} \leq e^{\lambda Y_0}\left(e^{\lambda Z_n} + e^{-\lambda Z_n}\right). \tag{7}$$

Combining (6) and (7), we find

$$\mathrm{E}\left(e^{\lambda \|X_n\|}\right) < e^{\lambda Y_0} 2 (\cosh \lambda)^n. \tag{8}$$

Applying Markov's Inequality and equation (8), we have

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

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

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

■

Remark 3.5 Unlike the proof of the 1-dimensional Azuma-Hoeffding inequality in [2], this proof technique apparently does not apply directly to very-weak martingales; instead we must appeal to Theorem 1.9. A careful look at the proof of Claim 3.4 shows the obstacle: we cannot replace the conditioning on $X_0, \ldots, X_{n-1}$ 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 1.8 which applies directly to the very-weak martingale case.

4 An Example of a very-weak martingale

One reason why very-weak martingales may be of interest is that they can 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 $(X_1, \ldots, X_n)$ 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 $(X_1, \ldots, X_n)$ has only $O(n^2)$ atoms. In this section, we give a very simple construction of such a very-weak martingale.

Small sample spaces have been useful in a variety of contexts, notably derandomization of algorithms using only pairwise or k-wise independent random variables [1, 3, 13, 14]. Although it is unclear how any advantage could be made of a weak martingale defined on a small sample space, we would not rule it out as a possibility.

The very-weak martingale we construct has a superficial similarity to the following familiar "coin flips" martingale, in that, if any two adjacent times are selected, and both processes observed at only those times, the processes are statistically indistinguishable.

Definition 4.1 Let $\xi_1, \xi_2, \ldots,$ be the ± 1-valued indicator variables for a sequence of fully independent coin flips. For every $n \geq 0$, let $CF_n := \sum_{i=1}^{n} \xi_i$. Then the sequence $\mathbf{CF} = (CF_0, CF_1, \ldots)$ is the "coin flips" martingale.

Construction 4.2 For each real number α in $[0, 1]$, for every $n \geq 0$, define

$$f_n(\alpha) = -n + 2 \min \left\{ k \;\middle|\; \sum_{i=0}^{k} \binom{n}{i} 2^{-n} \geq \alpha \right\}$$

The "fake coin flip" process, $\mathbf{FCF} = (\mathrm{FCF}_0, \mathrm{FCF}_1, \ldots)$ is defined by selecting α uniformly at random from $[0, 1]$, and for every $n \geq 0$, setting $\mathrm{FCF}_n = f_n(\alpha)$.

Proposition 4.3 **FCF** *is a very-weak martingale which is not a weak martingale. Moreover, for every* $i \geq 1$, $(\mathrm{FCF}_{n-1}, \mathrm{FCF}_n)$ *has the same distribution as* $(\mathrm{CF}_{n-1}, \mathrm{CF}_n)$.

Proof: Using the standard recurrence

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

we can see that, given that $\mathrm{FCF}_{n-1} = x$, there is a 1/2 chance that $\mathrm{FCF}_n = x+1$ and a 1/2 chance that $\mathrm{FCF}_n = x-1$. This proves that **FCF** is a very-weak martingale, and also, by induction, that the distribution of $(\mathrm{FCF}_{n-1}, \mathrm{FCF}_n)$ is the same as $(\mathrm{CF}_{n-1}, \mathrm{CF}_n)$.

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

$$\mathrm{E}\left(\mathrm{FCF}_n \mid \mathrm{FCF}_j = x\right) = \Theta(\sqrt{n}).$$

■

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

Proposition 4.4 *For every* n, $|\mathrm{FCF}_n - \mathrm{FCF}_{n-1}| = 1$, *and the random variable* $(\mathrm{FCF}_1, \ldots, \mathrm{FCF}_n)$ *has* $O(n^2)$ *atoms. Let* $\mathbf{X}$ *be a martingale satisfying* $|X_n - X_{n-1}| = 1$. *Then* $(X_1, \ldots, X_n)$ *has at least* 2^n *atoms.*

Proof: The atoms of $(\mathrm{FCF}_1, \ldots, \mathrm{FCF}_n)$ 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 $(X_1, \ldots, X_{n-1})$ is a disjoint union of at least two atoms of $(X_1, \ldots, X_n)$ (since the average of $X_n - X_{n-1}$ over these atoms must be zero). By induction, it follows that $(X_1, \ldots, X_n)$ has at least 2^n atoms. ■

5 Turning very-weak martingales into strong

We now present the proof of Theorem 1.9, 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 very-weak martingales whose range is a finite set and which run for only finitely many steps, second the proof for arbitrary very-weak martingales in $\mathbb{R}^d$. Most readers will probably be satisfied with the finite case, which has a simple proof.

Proof of Theorem 1.9 (case of finite range and finitely many steps): Let $\mathbf{X}$ be a very-weak martingale taking values in a real vector space V. Furthermore suppose each X_n takes values in a finite set $F \subset V$. 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 vectors in $\mathbb{R}^d$. We define

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

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

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

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 1.9 (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 [7, 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 4.2 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

$$Q_n\left[\mathbf{Y}_n \in A \times B\right] = \mathrm{E}_x\left(Q_{n-1}\left[\mathbf{Y}_{n-1} \in A \mid Y_{n-1} = x\right] P\left[X_n \in B \mid X_{n-1} = x\right]\right). \tag{9}$$

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 (9).

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

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

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 (9) are defined in terms of Radon-Nikodym derivatives (cf. [16]), 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

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

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

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

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

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

where x is selected according to μ.

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

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

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

Since the conditional probabilities on the right-hand side of (9) 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 (9) meet the regularity criteria of the Kolmogorov extension theorem [7, 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:

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

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

$$\mathrm{E}\left(\phi(\mathbf{Y}_n)\ Y_n\right) = \mathrm{E}\left(\phi(\mathbf{Y}_n)\ Y_{n-1}\right), \tag{10}$$

for every measurable function $\phi : (\mathbb{R}^d)^n \to \mathbb{R}$. 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 (10) we expand the left side using (9) (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:

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

where x is sampled according to the distribution μ introduced earlier. The right side of (10) may be expanded in the same way to obtain

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

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

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

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

6 Random Subsets have all small Fourier Coefficients

6.1 Sets of Any Size

Our first application of Theorem 1.8 is to prove Proposition 1.12, which says that almost all subsets of a finite abelian group G are "nice" (have all small Fourier coefficients). The exact result follows as an easy corollary of the following lemma.

Lemma 6.1 *Let G be a finite abelian group of order n, and let $\chi : G \to \mathbb{C}^\times$ be a non-principal character. Select a subset $S \subseteq G$ uniformly at random. Then*

$$\Pr\left[\Phi_\chi(S) \geq a\sqrt{n}\right] < 2e^2 e^{-2a^2}.$$

Proof: Let $G = \{g_1, \ldots, g_n\}$. For $0 \leq i \leq n$, let η_i be the $\{0,1\}$-valued indicator function for the event $g_i \in S$. Define $X_i := 2\mathrm{E}\left(\sum_{x\in S}\chi(x) \mid \eta_1, \ldots, \eta_i\right)$. Then $(X_0, \ldots, X_n)$ is a martingale taking values in $\mathbb{C}$. (Martingales defined in this manner are known as "Doob martingales.") Note that $X_0 = 0$ and that $|X_n| = 2\Phi_\chi(S)$. Moreover, it is easily verified that $X_i = \sum_{j=1}^{i}\chi(g_j)(2\eta_i - 1)$, which shows that $|X_i - X_{i-1}| = 1$ for every i.

Hence Theorem 1.8 implies

$$\Pr\left[\Phi_\chi(S) \geq a\sqrt{n}\right] = \Pr\left[|X_n| \geq 2a\sqrt{n}\right] < 2e^2 e^{-(2a)^2/2}.$$

■

Here is a slightly improved version of the same lemma, whose proof uses only the (weighted) real-number version of Hoeffding's inequality, Proposition 7.3). The proof is somewhat more involved.

Lemma 6.2 *Let G be a finite abelian group of order n, and let $\chi : G \to \mathbb{C}^\times$ be a non-principal character. Select a subset $S \subseteq G$ uniformly at random. Then*

$$\Pr\left[\Phi_\chi(S) \geq a\sqrt{n}\right] < 4e^{-2a^2}.$$

Proof: Let $G = \{g_1, \ldots, g_n\}$. For $0 \leq i \leq n$, let η_i be the $\{0,1\}$-valued indicator function for the event $g_i \in S$. Define $X_i := 2\mathrm{E}\left(\sum_{x\in S}\chi(x) \mid \eta_1, \ldots, \eta_i\right)$. Then $(X_0, \ldots, X_n)$ is a (Doob) martingale taking values in $\mathbb{C}$. Note that $X_0 = 0$ and that $|X_n| = 2\Phi_\chi(S)$. Moreover, it is easily verified that $X_i = \sum_{j=1}^{i}\chi(g_j)(2\eta_i - 1)$, which shows that $|X_i - X_{i-1}| = 1$ for every i.

The sequences $(\mathrm{Re}(X_0), \ldots, \mathrm{Re}(X_n))$ and $(\mathrm{Im}(X_0), \ldots, \mathrm{Im}(X_n))$ are real-valued martingales which satisfy $\mathrm{Re}(X_n)^2 + \mathrm{Im}(X_n)^2 = |X_n|^2$. Moreover, $|\mathrm{Re}(X_i) - \mathrm{Re}(X_{i-1})| = |\mathrm{Re}(\chi(g_j))| =: c_i$, and similarly $|\mathrm{Im}(X_i) - \mathrm{Im}(X_{i-1})| = |\mathrm{Im}(\chi(g_j))| =: d_i$.

Now, since $c_i^2 + d_i^2 = 1$ for every i, it follows that $\sum_{i=1}^n c_i^2 + \sum_{i=1}^n d_i^2 = n$. Hence

$$\begin{aligned}
\Pr\left[\Phi_\chi(S) \geq a\sqrt{n}\right] &= \Pr\left[|X_n| \geq 2a\sqrt{n}\right] \\
&\leq \Pr\left[|\mathrm{Re}(X_n)| \geq 2a\sqrt{\sum_{i=1}^n c_i^2}\right] + \Pr\left[|\mathrm{Im}(X_n)| \geq 2a\sqrt{\sum_{i=1}^n d_i^2}\right] \\
&< 2e^{-(2a)^2/2} + 2e^{-(2a)^2/2},
\end{aligned}$$

the last inequality following from Proposition 7.3 applied to $\mathrm{Re}(X_n)$ and to $\mathrm{Im}(X_n)$. ■

We note that a simpler version of this proof which uses only the non-weighted version of Azuma-Hoeffding, achieves a worse constant in the exponent of the upper bound.

6.2 Sets of Fixed Size

The proof of Proposition 1.12 did not really take full advantage of the power of Theorem 1.8, since the random variables being summed were independent, which is much more than just being a martingale. The case where the set size is fixed, requires us to analyze a much more history-dependent process. Theorem 1.13 follows easily from the following Lemma.

Lemma 6.3 *Let G be a finite abelian group of order n, and let $\chi : G \to \mathbb{C}^\times$ be a non-principal character of G. Let $k \leq n$, and let $m = \min\{k, n-k\}$. Select a subset $S \subset G$ uniformly at random subject to the constraint $|S| = k$. Then*

$$\Pr\left[\Phi_\chi(S) \geq a\sqrt{m}\right] < 2e^2 e^{-a^2/8}.$$

Proof: 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 $(g_1, g_2, \ldots, g_n)$ of the elements of G, then let $S = \{g_1, \ldots, g_k\}$. We next define a (Doob) martingale $\mathbf{Y} = (Y_0, \ldots, Y_k)$, where Y_i is defined in terms of the random variables $g_1, \ldots, g_i$, as follows.

$$Y_i \;:=\; \frac{1}{2}\,\mathrm{E}\left(\sum_{j=1}^{k} \chi(g_j) \mid g_1, \ldots, g_i\right)$$

Since for any $g, h \in G$, $|\chi(g) - \chi(h)| \leq 2$, it follows easily that $|Y_i - Y_{i-1}| \leq 1$. Hence Theorem 1.8 tells us, for every $a > 0$,

$$\begin{aligned}\Pr\left[\Phi_\chi(S) \geq a\sqrt{k}\right] &= \Pr\left[Y_k \geq a\sqrt{k}/2\right] \\ &\leq 2e^2 e^{-a^2/8}.\end{aligned}$$

Since we are assuming $m = k$, this completes the proof. ∎

It is possible to improve the bound in Theorem 1.13 to one of the form $O(e^{-a^2/2(1+3m/n)})$, by using the weighted version of Theorem 1.8 presented in the Notes as Theorem 7.4. This yields a corresponding improvement in the constant for Theorem 1.13. As in Section 6.1, it is also possible to deduce these results using only the (one-dimensional) Azuma-Hoeffding inequality, as long as one is not concerned with the constants.

7 Notes and Questions

Hoeffding's original paper [10] actually proved the following stronger version of the Azuma-Hoeffding inequality.

Theorem 7.1 *Let* $\mathbf{X} = (X_0, X_1, \ldots, X_n)$ *be a real-valued martingale such that* $X_0 = 0$ *and for every* i, $|X_i - X_{i-1}| \leq 1$. *Then, for every* $a \in [0, n]$,

$$\Pr[X_n \geq a] \leq 2^{\,n\left(H\left(\frac{a+n}{2n}\right)-1\right)} \leq \; e^{-a^2/2n}, \tag{11}$$

where $H(x) = -x \log_2(x) - (1-x)\log_2(1-x)$ *is the binary entropy function.*

The first inequality is within a factor $O(\sqrt{n})$ of equality for all values of a, and holds with equality for $a = 0$ and $a = n$. Our proof of Theorem 1.8 can easily be strengthened to the following:

Theorem 7.2 *Let* $\mathbf{X} = (X_0, X_1, \ldots, X_n)$ *be a very-weak martingale taking values in* $\mathbb{E}$ *such that* $X_0 = 0$ *and for every* i, $|X_i - X_{i-1}| \leq 1$. *Then, for every* $a \in [0, n]$,

$$\Pr[X_n \geq a] \leq e2^{\,n\left(H\left(\frac{a+n-1}{2n}\right)-1\right)} \tag{12}$$

where $H(x) = -x \log_2(x) - (1-x)\log_2(1-x)$ *is the binary entropy function.*

The proof is essentially the same as that of Theorem 1.8 up to the last paragraph, at which point, instead of approximating $\cosh(\lambda)$ by $\exp(\lambda^2/2)$, we set $\lambda := \operatorname{archtanh}((a-1)/n)$ and carefully simplify the resulting bound, along lines similar to those of Hoeffding's original proof [10]. It is not hard to show that the upper bound of Theorem 7.2 is within a factor $O(n)$ of best possible for all values of a.

Hoeffding also proved the following weighted version of the weaker inequality:

Proposition 7.3 *Let* $\mathbf{X} = (X_0, X_1, \ldots, X_n)$ *be a real-valued martingale such that for every* i, $|X_i - X_{i-1}| \leq c_i$. *Then, for every* $a \geq 0$,

$$\Pr[|X_n| \geq a] \leq \; 2\exp\left(\frac{-a^2}{2\sum_i c_i^2}\right). \tag{13}$$

Proposition 7.3 was rediscovered by Azuma [4] a few years later.

Our proof technique naturally generalizes to the following weighted version of Theorem 1.8:

Theorem 7.4 *Let* $\mathbf{X}$ *be a* $\mathbb{E}$*-valued very-weak martingale such that for every* i, $\|X_i - X_{i-1}\| \leq c_i$. *Then for every* $a > 0$,

$$\Pr[\|X_n\| \geq a] \; < 2\exp\left(\frac{-(a-Y_0)^2}{2\sum_{i=1}^n c_i^2}\right), \tag{14}$$

where $Y_0 := \max\{1 + \max c_i, 2\max c_i\}$.

This differs from the upper bound in Proposition 7.3 by a factor at most $e^{O(c_{\max}/c_{\text{avg}})}$, where $c_{\max} = \max c_i$, and $c_{\text{avg}} = \sum c_i/n$. Although this is only a constant factor when $c_{\max}/c_{\text{avg}}$ is bounded, in the general case this bound is considerably worse than that of Kallenberg and Sztencel, cited here as Proposition 1.7.

It seems likely that Theorem 1.9 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-Hoeffding Inequality holds, without the additional factor of $e^{1+(2a-1)/2n}$.

8 Acknowledgements

I am indebted to László Babai not only for inspiring our investigations by his questions about Fourier coefficients, but also for much encouragement and assistance with the preparation of this paper. I am also grateful to Todd Rowland and Varsha Dani, among others, for helpful conversations.

I would like to thank the referee for pointing out several oversights and offering helpful suggestions.

Thanks also to Geir Ove Myhr for pointing out a serious typo in the previous version.

References

[1] N. Alon, L. Babai, A. Itai. *A Fast and Simple Randomized Parallel Algorithm for the Maximal Independent Set Problem.* J. Algorithms **7** (1986) 567–583.

[2] Alon, Spencer, Erdős. *The Probabilistic Method.* John Wiley & Sons, Inc. New York.

[3] S. Arora, C. Lund, R. Motwani, M. Sudan, M. Szegedy. *Proof verification and the hardness of approximation problems.* JACM **45** (1998) 501–555.

[4] K. Azuma. *Weighted Sums of Certain Dependent Random Variables.* Tôhoku Math. Journ. **19** (1967) 357–367.

[5] 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. Technical report TR-2001-01. http://www.cs.uchicago.edu/research/publications/techreports/TR-2001-01

[6] J. Berman. *An example of a weak martingale.* Annals of Probability **4** (1976), 107–108.

[7] L. Breiman. *Probability.* Addison-Wesley Publishing Co., 1968.

[8] 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.

[9] 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.

[10] W. Hoeffding. *Probability inequalities for sums of bounded random variables.* J. Am. Stat. Assoc. **58** (1963) 13–30.

[11] O. Kallenberg, R. Sztencel. *Some dimension-free features of vector-valued martingales.* Probab. Th. Rel. Fields **88** (1991) 215–247.

[12] I. Karatzas, S.E. Shreve. *Brownian motion and stochastic calculus.* Springer-Verlag, 1988.

[13] M. Luby. *A simple parallel algorithm for the maximal independent set problem.* SIAM J. Comput., **15** (1986) 1036–1053.

[14] M. Luby, A. Wigderson. *Pairwise independence and derandomization, a survey.* ICSI TR-95-035. http://citeseer.nj.nec.com/luby95pairwise.html

[15] P. Nelson. *A class of orthogonal series related to martingales.* Annals of Mathematical Statistics **41** (1970) 1684–1694.

[16] D. Williams. *Probability with Martingales.* Cambridge University Press, 1991.