

CS265/CME309: Randomized Algorithms and Probabilistic Analysis

Lecture #16: Martingales, the Doob Martingale, and Azuma-Hoeffding Tail Bounds

Gregory Valiant*, Updated by Mary Wootters

November 28, 2023

1 Introduction

In this lecture we introduce the concept of a Martingale, which can be used to analyze many random processes. The technical core of this lecture will be the Azuma-Hoeffding tail bound, which gives a Chernoff style bound on tail probabilities of martingales. These bounds were independently proved by Azuma and Hoeffding in the 1960's [1, 2]. Crucially, these bounds will let us often effortlessly prove strong concentration results for sums of certain *dependent* random variables and other stochastic processes. The proof technique for the Azuma-Hoeffding bounds follow the same general approach as Chernoff bounds: applying Markov's inequality to the moment generating function for the random variable in question. Our analysis of this moment generating function, however, will crucially leverage the properties of martingales, as opposed to the properties of sums of independent random variables.

Definition 1. *A sequence of real-valued random variables, $\{Z_t\}$ is a martingale with respect to the sequence (of not necessarily real valued random variables) $\{X_t\}$ if the following conditions hold for all t :*

1. Z_t is a function of X_0, \dots, X_t ,
2. $\mathbf{E}[|Z_t|] < \infty$,
3. $\mathbf{E}[Z_t | X_0, \dots, X_{t-1}] = Z_{t-1}$.

Given this setting, letting $Y_t = Z_t - Z_{t-1}$, often refers to the sequence $\{Y_t\}$ as the sequence of martingale differences. In many cases, $\{X_t\}$ is a martingale with respect to itself, in the sense that $\mathbf{E}[X_t | X_0, \dots, X_{t-1}] = X_{t-1}$. In such cases, one generally just says that $\{X_t\}$ “is a martingale”.

*©2019, Gregory Valiant. Not to be sold, published, or distributed without the authors' consent.

The above definition can be slightly generalized via the notion of *filters* and σ -fields, though we will not get bogged down in that. For our purposes, the above definition will be sufficient. The first condition, that Z_t is a function of X_0, \dots, X_t is not too restrictive: if we would like Z_t to be a randomized function, we can regard Z_t as being a deterministic function of X_0, \dots, X_t in addition to some random bits b_0^t, \dots, b_m^t and then define Z_t as a martingale with respect to the sequence $\{X'_t\}$ where $X'_t = (X_t, b_0^t, \dots, b_m^t)$. Hence we can assume that Z_t is a deterministic function of the X_0, \dots, X_t . The above definition can be applied both when we have a finite set of Z_t 's, as well as the infinite case.

1.1 The Doob Martingale

One reason why martingales are such a useful concept is that we can obtain a martingale from *any* real-valued random variable.

Definition 2. *Given a random variable, A , and a sequence of random variables, $\{X_t\}$ defined over a common probability space (i.e. quantities such as $\Pr[A = 1|X_4 = 7]$ are well defined notions), the Doob martingale of A with respect to $\{X_t\}$ is the sequence $\{Z_t\}$ defined by*

$$Z_t = \mathbf{E}[A|X_0, \dots, X_t].$$

To see why the above definition yields that $\{Z_t\}$ is a martingale with respect to $\{X_t\}$, provided $\mathbf{E}[|Z_t|] < \infty$, we simply apply the definition of Z_t and simplify:

$$\mathbf{E}[Z_t|X_0, \dots, X_{t-1}] = \mathbf{E}[\mathbf{E}[A|X_0, \dots, X_t]|X_0, \dots, X_{t-1}] = \mathbf{E}[A|X_0, \dots, X_{t-1}] = Z_{t-1}.$$

In the above, we leveraged the fact that for any random variables X, Y, Z , by the definition of conditional probability,

$$\mathbf{E}[\mathbf{E}[X|Y, Z]|Y] = \mathbf{E}[X|Y],$$

since the conditioning on Z is integrated/marginalized out by the outer expectation.

One of the most common ways that the Doob martingale is employed is to analyze properties of some random variable that depends on a random object—for example, the number of empty bins if we randomly toss m balls into n bins, or the chromatic number of an Eros-Renyi random graph, etc. In those settings, the X_t 's sequentially reveal information about the random object and the quantity in question, until there is no more randomness. The following examples illustrate this use-case of the Doob martingale.

Example 3. *Consider tossing m balls uniformly at random into n bins. Let A denote the number of empty bins at the end of the process. Letting X_t denote the location that the t th ball lands, consider the Doob martingale defined by $Z_t = \mathbf{E}[A|X_1, \dots, X_t]$. Here, we are iteratively updating our expectation of A as we see where each of the n balls is thrown. $Z_0 = \mathbf{E}[A]$, and Z_n equals the actual value of A for the realization of the ball-tossing process given by the X_t 's.*

Example 4. *Consider the Erdos-Renyi random graph process $G_{n,p}$ that constructs an n node graph by independently including each edge with probability p . Let A denote the chromatic number of the resulting graph, defined as the smallest number of colors such that it is possible to color the nodes in such a way that no neighboring nodes have the same color. If we define the random variables $X_1, \dots, X_{\binom{n}{2}}$ such that X_t is the indicator for whether the t th edge is present in the graph, then we*

have the corresponding Doob martingale defined by $Z_t = \mathbf{E}[A|X_1, \dots, X_t]$. This is known as the edge exposure martingale, as we are revealing the edges one-at-a-time.

Alternately, we could define the random variables Y_1, \dots, Y_n , such that Y_t is the random variable describing the neighborhood of the t th node, namely this random variable takes one of 2^{n-1} different values, according to which of the $n-1$ potential edges are present between node t and the $n-1$ other nodes. We can now define the vertex exposure martingale Z_1, \dots, Z_n with $Z_t = \mathbf{E}[A|Y_1, \dots, Y_t]$.

[Note: this may be a slightly non-standard way of defining the vertex exposure martingale. Another way would be to define Z_t as $\mathbf{E}[A|G_t]$, where G_t is the subgraph of G induced on vertices $1, \dots, t$.]

2 The Azuma-Hoeffding Tail Bound

Here is the Azuma-Hoeffding inequality:

Theorem 1. Let $\{Z_t\}$ be a martingale with respect to $\{X_t\}$, and suppose there are constants c_1, \dots, c_n such that for all $i \leq n$, $|Z_i - Z_{i-1}| \leq c_i$. For any $\lambda > 0$,

$$\Pr[|Z_n - Z_0| \geq \lambda] \leq 2e^{-\frac{\lambda^2}{2\sum c_i^2}}.$$

Let's compare this with the standard Chernoff bounds. Suppose that X_1, \dots, X_n denote the 0/1 outcomes of n independent coin flips, and $Z_t = \mathbf{E}[\sum X_i | X_1, \dots, X_t]$ is the Doob martingale corresponding to the total number of heads out of the n tosses. We can apply the Azuma-Hoeffding inequality to $\{Z_t\}$. To do that, we need to bound $|Z_i - Z_{i-1}|$. We claim that we have $|Z_i - Z_{i-1}| \leq 1/2$. To see why, notice that

$$\begin{aligned} |Z_i - Z_{i-1}| &= \left| \mathbf{E}\left[\sum_j X_j | X_0, \dots, X_i\right] - \mathbf{E}\left[\sum_j X_j | X_0, \dots, X_{i-1}\right] \right| \\ &= |\mathbf{E}[X_i | X_i] - \mathbf{E}[X_i]| \\ &= |X_i - 1/2|, \end{aligned}$$

where above we have used that the X_j are independent and so the terms without an X_i in them cancel. Now, no matter what X_i is (either 0 or 1), we have $|X_i - 1/2| = 1/2$. This means that we may take $c_i = 1/2$ in the statement of the theorem. Applying the theorem, we get

$$\Pr[|Z_n - Z_0| \geq c\sqrt{n}] \leq 2e^{-\frac{c^2 n}{2n \cdot (1/4)}} = 2e^{-2c^2}.$$

Notice that Z_n is just equal to $\sum_i X_i$, and Z_0 is just equal to $\mathbb{E} \sum_i X_i = n/2$. So we get that

$$\Pr\left[\left|\sum_i X_i - n/2\right| \geq c\sqrt{n}\right] \leq 2e^{-2c^2},$$

which is our familiar inverse exponential tail bound, similar to what we get from a Chernoff bound. However, as we will see later, the Azuma-Hoeffding inequality can also be used in other scenarios where our original Chernoff bound doesn't.

Now, let's move on to the proof of Theorem 1. The following lemma will be central to our proof of Theorem 1.

Lemma 5. Let Y denote a random variable $Y \in [-c, c]$ for some constant c , with $\mathbf{E}[Y] = 0$, then for all $t \geq 0$,

$$\mathbf{E}[e^{tY}] \leq e^{t^2 c^2 / 2}.$$

Proof. We will first prove the statement when $c = 1$, from which the claim will follow by observing that if $Y \in [-c, c]$, the random variable $Y' = Y/c$ is in the interval $[-1, 1]$, and $tY = (tc)Y'$, hence $\mathbf{E}[e^{tY}] = \mathbf{E}[e^{(tc)Y'}]$. So if we show that $\mathbf{E}[e^{tY'}] \leq e^{t^2/2}$, the claim will follow.

We now analyze $\mathbf{E}[e^{tY}]$ for a random variable Y with expectation 0 and magnitude bounded by 1. From the convexity of the exponential function, for any $x \in [-1, 1]$,

$$e^{tx} \leq \frac{(1-x)e^{t(-1)} + (1+x)e^{t(1)}}{2}.$$

Hence

$$\mathbf{E}[e^{tY}] \leq \mathbf{E}\left[\frac{(1-Y)e^{-t} + (1+Y)e^t}{2}\right] = \frac{e^{-t} + e^t}{2},$$

where the final equality is from the assumption that $\mathbf{E}[Y] = 0$. To finish the proof, we will simply re-express e^{-t} and e^t in terms of their Taylor expansions, and compare this to the Taylor expansion of $e^{t^2/2}$. To this end, we have

$$\frac{1}{2}(e^{-t} + e^t) = \frac{1}{2}\left(1 + t + \frac{t^2}{2} + \frac{t^3}{3!} + \dots + 1 - t + \frac{t^2}{2} - \frac{t^3}{3!} \dots\right) = 1 + \frac{t^2}{2} + \frac{t^4}{4!} + \frac{t^6}{6!} + \dots$$

On the other hand, the Taylor expansion of $e^{t^2/2}$ is the following:

$$e^{t^2/2} = \sum_{j \geq 0} \frac{x^{2j}}{2^j j!} = 1 + \frac{t^2}{2} + \frac{t^4}{4 \cdot 2} + \frac{t^6}{2^3 \cdot 3!} \dots,$$

hence $\frac{1}{2}(e^{-t} + e^t) \leq e^{t^2/2}$ because each term of the expansion on the left is at least the corresponding term in the expansion of the right (and both sequences converge). \square

We now prove Theorem 1.

Proof of Theorem 1. We begin with the upper bound. For any $t > 0$, by the monotonicity of the exponential, and Markov's inequality,

$$\Pr[Z_n - Z_0 \geq \lambda] = \Pr[e^{t(Z_n - Z_0)} \geq e^{t\lambda}] \leq \frac{\mathbf{E}[e^{t(Z_n - Z_0)}]}{e^{t\lambda}}.$$

Note that for any random variables A, B , $\mathbf{E}[A] = \mathbf{E}[\mathbf{E}[A|B]]$, hence

$$\mathbf{E}[e^{t(Z_n - Z_0)}] = \mathbf{E}[e^{t(Z_n - Z_{n-1} + Z_{n-1} - Z_0)}] = \mathbf{E}[\mathbf{E}[e^{t(Z_n - Z_{n-1} + Z_{n-1} - Z_0)} | X_0, \dots, X_{n-1}]].$$

Since Z_{n-1} is a function of X_0, \dots, X_{n-1} we can pull the ‘‘constant’’ $e^{t(Z_{n-1} - Z_0)}$ out from the inner expectation, yielding:

$$\mathbf{E}[\mathbf{E}[e^{t(Z_n - Z_0)} | X_0, \dots, X_{n-1}]] = \mathbf{E}[e^{t(Z_{n-1} - Z_0)} \mathbf{E}[e^{t(Z_n - Z_{n-1})} | X_0, \dots, X_{n-1}]] \leq \mathbf{E}[e^{t(Z_{n-1} - Z_0)}] e^{t^2 c_n^2 / 2},$$

where the last inequality resulted from applying Lemma 5 to the random variable $Z_n - Z_{n-1} | X_0, \dots, X_{n-1}$ which is zero in expectation and has magnitude bounded by c_n .

Iteratively applying the above process, we get that

$$\mathbf{E}[e^{t(Z_n - Z_0)}] \leq e^{\frac{t^2 \sum_{i=1}^n c_i^2}{2}}.$$

Hence putting the pieces together, and plugging in $t = \frac{\lambda}{\sum c_i^2}$, we obtain the following:

$$\Pr[Z_n - Z_0 \geq \lambda] \leq \frac{\mathbf{E}[e^{t(Z_n - Z_0)}]}{e^{-t\lambda}} \leq e^{\frac{t^2 \sum_{i=1}^n c_i^2}{2}} e^{-t\lambda} = e^{-\frac{\lambda^2}{2 \sum c_i^2}}.$$

The lower tail bound is proved analogously. □

Note: At this point we are done with the material in the video lectures. The stuff after this is meant as a reference for after class.

2.1 Applications of Azuma-Hoeffding

In many settings, the Azuma-Hoeffding tail bound will be easy to apply, because the underlying quantity we care about corresponds to a clean martingale. It is, however, always worth stopping and thinking about whether there is a *better* martingale that you could use—better in the sense that the bounds on the martingale differences, c_i , are smaller, or better in the sense that there are fewer steps in the sequence (e.g. the difference between the $\binom{n}{2}$ length edge-exposure martingale of Example 4 versus the n terms in the vertex-exposure martingale).

Proposition 6. *Letting A denote the chromatic number of an Erdos-Renyi random graph $G_{n,p}$:*

$$\Pr[|A - \mathbf{E}[A]| \geq c\sqrt{n}] \leq e^{-\frac{c^2}{2}}.$$

Proof. Recall the Doob martingale $\{Z_t\}$ corresponding to the vertex exposure martingale from Example 4. We'd like to apply Azuma's inequality to this martingale. Intuitively, we have that $|Z_t - Z_{t-1}| \leq 1$ for all t , since revealing the neighborhood of a vertex can change the expected chromatic number by at most 1 (in the best case, we save using a color, in the worst case, we need a new color just for that vertex). Formally, the way we'd see this is the following.

For any a_t (which we think of as a possible value for X_t), we write:

$$\begin{aligned} & \mathbb{E}[A | X_1, \dots, X_{t-1}, X_t = a_t] \\ &= \sum_{a_{t+1}, \dots, a_n} \mathbb{E}[A | X_{\leq t-1}, X_t = a_t, X_{\geq t+1} = a_{\geq t+1}] \Pr[X_{\geq t+1} = a_{\geq t+1} | X_{\leq t-1}, X_t = a_t] \\ &= \sum_{a_{t+1}, \dots, a_n} \mathbb{E}[A | X_{\leq t-1}, X_t = a_t, X_{\geq t+1} = a_{\geq t+1}] \Pr[X_{\geq t+1} = a_{\geq t+1}] \end{aligned}$$

using the fact that in our example, the X_t 's are independent (since they involve disjoint sets of edges). Using the same derivation for $\mathbb{E}[A | X_1, \dots, X_{t-1}, X_t]$, we can write

$$\begin{aligned} & \mathbb{E}[A | X_1, \dots, X_{t-1}, X_t = a_t] - \mathbb{E}[A | X_1, \dots, X_{t-1}, X_t] \\ &= \sum_{a_{t+1}, \dots, a_n} (\mathbb{E}[A | X_{\leq t-1}, X_t = a_t, X_{\geq t+1} = a_{\geq t+1}] - \mathbb{E}[A | X_{\leq t}, X_{\geq t+1} = a_{\geq t+1}]) \Pr[X_{\geq t+1} = a_{\geq t+1}] \end{aligned}$$

Now, we have that

$$\mathbb{E}[A|X_{\leq t-1}, X_t = a_t, X_{\geq t+1} = a_{\geq t+1}] - \mathbb{E}[A|X_{\leq t}, X_{\geq t+1} = a_{\geq t+1}] \leq 1$$

for any values of a_t, a_{t+1}, \dots, a_n . This is because the two expressions can differ only in the value of x_t , the local view of vertex t . By our observation above, if we change the local view of a single vertex, this can change the chromatic number $A = \chi(G)$ by at most 1. Thus, we conclude from the above that for any a_t ,

$$\mathbb{E}[A|X_1, \dots, X_{t-1}, X_t = a_t] - \mathbb{E}[A|X_1, \dots, X_{t-1}, X_t] \leq 1.$$

But this implies that

$$\mathbb{E}[A|X_1, \dots, X_{t-1}] - \mathbb{E}[A|X_1, \dots, X_t] \leq 1,$$

and this is the difference $Z_{t-1} - Z_t$ that we wanted to bound. (You can repeat the above derivation with the two terms switched to bound $Z_t - Z_{t-1}$ and hence $|Z_t - Z_{t-1}|$). Thus, this difference is bounded by 1. So indeed we can take $c_t = 1$ as our intuition says.

Hence Theorem 1 holds with $c_i = 1$ for all i , and the proposition holds. \square

Remark 7. *In the proof above, we used the fact that the X_t were independent. But Azuma's inequality doesn't require the X_t to be independent! Can we use logic like this even if the X_t are not independent? The condition that we used above was the deterministic statement that:*

$$A \text{ changes by at most } c_t \text{ when we change } X_t, \text{ no matter what the other } X_i \text{ are.} \quad (1)$$

This condition is very simple and easy to check, but it does need independence.¹ (This is often called the method of bounded differences). However, there's a similar—though less easy to check—condition that doesn't require independence of the X_i , which is that

$$\mathbb{E}[A|X_{\leq t-1}, X_t = a] - \mathbb{E}[A|X_{\leq t-1}, X_t = a'] \leq c_t \quad (2)$$

for any a, a' . For a fun exercise, check that the requirement that $|Z_t - Z_{t-1}| \leq c_t$ can be replaced by (2) in the statement of Azuma's theorem for the Doob martingale.

Note that if we had used the edge exposure martingale instead, we would have gotten the much worse bound of $\Pr[|A - \mathbb{E}[A]| \geq c\sqrt{n}] \leq e^{-\frac{c^2}{n}}$, since there we still would have had $c_i = 1$, but we would have had $\binom{n}{2} \approx n^2/2$ terms in our sequence.

The next example illustrates how one can deal with martingales that depend on additional randomness beyond the natural choice of $\{X_t\}$ by defining an alternate sequence $\{X'_t\}$ where each X'_t includes both X_t as well as some additional randomness.

Example 8. *Consider a gambling setting where, at each timestep, you can bet any amount between 0 and some “house limit” B . At each step, a fair coin is flipped, and if you correctly guess the outcome, you win your bet, otherwise you lose your bet (so the expected earnings in any round are 0). Let $\{X_t\}$ denote the outcome of the coin flip in each round. Since the expected winnings at*

¹To see that independence is required for this criterion (1) to be enough, just consider the case where either all the X_i are 0 with probability 1/2; or they are all 1 with probability 1/2, and let $A = \sum_i X_i$. Then A satisfies (1) for $X_i \in \{0, 1\}$, because if you change just one X_i , A can change by at most 1. However, clearly the conclusion of the Azuma-Hoeffding inequality does not hold, as A is not concentrated at all (it's 0 with probability 1/2 and n with probability 1/2).

each step is 0, we can define the martingale sequence $\{Z_t\}$ where Z_t is our net earnings at time t (i.e. Z_t is our net winnings or losses after t rounds). Because $|Z_t - Z_{t-1}| \leq B$, we can apply the Azuma-Hoeffding bound to conclude that

$$\Pr[|Z_t| \geq \lambda] \leq 2e^{-\frac{\lambda^2}{2tB^2}}.$$

Crucially, the above bound holds no matter what betting strategy we use. Even if our bet in the t th round depends on the outcomes in the previous $t - 1$ rounds, this still holds.

As an aside, perhaps this is why (some) casino owners can sleep soundly at night—they set the house limit B to be low enough, in comparison to the number of rounds that will be played, and the small but non-zero bias that is build into the games, so that the probability that they don't end up making money in a given week is extremely small. The inverse exponential with λ^2 in the exponent makes it okay for them to use a fairly large value B ...

One final note: if we want to be super pedantic and make sure that $\{Z_t\}$ is actually a martingale, even for a randomized betting strategy that depends on previous rounds, and on other random whims, etc., then we would say that $\{Z_t\}$ is a martingale with respect to the sequence $\{X'_t\}$, where X'_t denotes both the outcome of the t 'th round, as well as the bet we will place in the $t + 1$ 'st round. This is a perfectly valid sequence of random variables, and it is true that $\{Z_t\}$ is a martingale with respect to $\{X'_t\}$, since Z_t is a deterministic function of X'_0, \dots, X'_t and $\mathbf{E}[Z_{t+1} | X'_0, \dots, X'_t] = Z_t$.

References

- [1] Kazuoki Azuma. Weighted sums of certain dependent random variables. *Tohoku Mathematical Journal, Second Series*, 19(3):357–367, 1967.
- [2] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963.