

# [AI2613 Lecture 7] Doob Martingale, Azuma-Hoeffding, McDiarmid

April 6, 2023

## 1 Hoeffding's Inequality

We introduced the following Hoeffding's inequality to bound the concentration for the sum of a sequence independent random variables.

**Theorem 1 (Hoeffding's Inequality)** *Let  $X_1, \dots, X_n$  be independent random variables where each  $X_i \in [a_i, b_i]$  for certain  $a_i \leq b_i$  with probability 1. Let  $X = \sum_{i=1}^n X_i$  and  $\mu \triangleq \mathbf{E}[X] = \sum_{i=1}^n \mathbf{E}[X_i]$ , then*

$$\Pr[|X - \mu| \geq t] \leq 2 \exp\left(-\frac{2t^2}{\sum_{i=1}^n (b_i - a_i)^2}\right)$$

for all  $t \geq 0$ .

Before proving Theorem 1 in Section 3, we see a practical application of Hoeffding's inequality.

**Example 1 (Meal Delivery)** *During the quarantine of our campus, the professors deliver meals for students using their private cars or trikes. Then a practical problem is how to estimate the amount of meals on a trike conveniently (See the [news](#)).*

*Assume there are  $n$  boxes of meal on the trike ( $n \geq 200$  and is unknown for us). Let  $X_i$  be the weight of the  $i$ -th box of meal. Assume that  $X_1, X_2, \dots, X_n$  are i.i.d. random variables, each  $X_i \in [250, 350]$  (unit: gram) and  $\mu = \mathbf{E}[X_i] = 300$ . Let  $S$  be the total weight of the meal boxes on the trike, that is,  $S = \sum_{i=1}^n X_i$ . We can weigh the meal boxes and use  $\hat{n} = \frac{S}{\mu}$  as an estimator for  $n$ . Now we compute how accurate this estimator is.*

*Let  $\delta \geq 0$  be a constant. By Hoeffding's inequality,*

$$\Pr[|\hat{n} - n| > \delta n] = \Pr[|S - \mu n| > \delta \mu n] \leq 2 \exp\left\{-\frac{2\delta^2 \mu^2 n^2}{\sum_{i=1}^n (350 - 250)^2}\right\}. \quad (1)$$

*Plugging  $\mu = 300$ ,  $\delta = 0.05$  and  $n \geq 200$  into Equation (1), by direct calculation, we have*

$$\Pr[\hat{n} \in [0.95n, 1.05n]] \geq 1 - 2.4682 \times 10^{-4}.$$

## 2 Concentration on Martingale

We consider the balls-in-a-bag problem. There are  $g$  green balls and  $r$  red balls in a bag. These balls are the all same except for the color. We want to estimate the ratio  $\frac{r}{r+g}$  by drawing balls. There are two scenarios.

- Draw balls with replacement. Let  $X_i = 1$  [the  $i$ -th ball is red]. Let  $X = \sum_{i=1}^n X_i$ . Then clearly each  $X_i \sim \text{Ber}\left(\frac{r}{r+g}\right)$  and  $\mathbf{E}[X] = n \cdot \frac{r}{r+g}$ . Since all  $X_i$ 's are independent, we can directly apply Hoeffding's inequality and obtain

$$\Pr[|X - \mathbf{E}[X]| \geq t] \leq 2 \exp\left(-\frac{2t^2}{n}\right).$$

- Draw balls without replacement. Again we let  $Y_i = 1$  [the  $i$ -th ball is red], then unlike the case of drawing with replacement, variables in  $\{Y_i\}$  are dependent. Let  $Y = \sum_{i=1}^n Y_i$ . We first calculate  $\mathbf{E}[Y]$ .

For every  $i \geq 1$ ,  $\mathbf{E}[Y_i]$  is the probability that the  $i$ -th draw is a red ball. Note that drawing without replacement is equivalent to first drawing a uniform permutation of  $r + g$  balls and drawing each ball one by one in that order. Therefore, the probability of  $Y_i = 1$  is  $\frac{r \cdot (r+g-1)!}{(r+g)!} = \frac{r}{r+g}$ . So we have  $\mathbf{E}[Y] = n \cdot \frac{r}{r+g}$ .

However, since  $\{Y_i\}$  are dependent, we cannot apply Hoeffding's inequality directly. This motivate us to generalize it by removing the requirement of independence.

### 2.1 Azuma-Hoeffding's Inequality

**Theorem 2 (Azuma-Hoeffding's Inequality)** *Let  $\{Z_n\}_{n \geq 0}$  is a martingale with respect to a filtration  $\{\mathcal{F}_n\}$ . If for every  $i \geq 1$ ,  $|Z_i - Z_{i-1}| \leq c_i$  with probability 1, then*

$$\Pr[|Z_n - Z_0| \geq t] \leq 2 \exp\left(-\frac{2t^2}{\sum_{i=1}^n c_i^2}\right).$$

Azuma-Hoeffding's inequality generalizes Hoeffding's inequality by letting  $Z_n = \sum_{i=1}^n (X_i - \mathbf{E}[X_i])$  and  $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$ .

The proof of this theorem is in Section 3. The requirement of martingale in Theorem 2 seems to be even harder to satisfy than the requirement of independence. However, in many cases, we can construct a doob martingale to apply the Azuma-Hoeffding's inequality.

**Definition 3 (Doob Martingale, Doob Sequence)** *Let  $X_1, \dots, X_n$  be a sequence of (unnecessarily independent) random variables and  $f(\bar{X}_{1,n}) = f(X_1, \dots, X_n) \in \mathbb{R}$  be a function. For  $i \geq 0$ , Let  $Z_i \triangleq \mathbf{E}\left[f(\bar{X}_{1,n}) \mid \bar{X}_{1,i}\right]$ . Then we call  $\{Z_n\}_{n \geq 0}$  a Doob martingale or a Doob sequence.*

It is easy to verify that  $\{Z_n\}_{n \geq 0}$  in Definition 3 is indeed a martingale w.r.t.  $\{X_n\}$  by seeing

$$\mathbf{E}\left[Z_i \mid \bar{X}_{1,i-1}\right] = \mathbf{E}\left[\mathbf{E}\left[f(\bar{X}_{1,n}) \mid \bar{X}_{1,i}\right] \mid \bar{X}_{1,i-1}\right] = \mathbf{E}\left[f(\bar{X}_{1,n}) \mid \bar{X}_{1,i-1}\right] = Z_{i-1}.$$

Let  $\mathcal{F} = \sigma(\bar{X}_{1,i})$ . We can see that  $Z_i$  is  $\mathcal{F}_i$  measurable by definition. Moreover, we know that  $Z_0 = \mathbf{E} \left[ f(\bar{X}_{1,n}) \right]$  and  $Z_n = f(\bar{X}_{1,n})$ .

Recall the balls-in-a-bag problem we discussed above. Define  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  by letting  $f(y_1, y_2, \dots, y_n) = \sum_{i=1}^n y_i$ . Then in the drawing without replacement scenario,  $Y = \sum_{i=1}^n Y_i = f(Y_1, Y_2, \dots, Y_n)$ . Now we construct the Doob martingale for  $f$ .

Let  $Z_i = \mathbf{E} \left[ f(\bar{Y}_{1,n}) \mid \bar{Y}_{1,i} \right]$ . We know that  $Z_0 = \mathbf{E} \left[ f(\bar{Y}_{1,n}) \right] = \mathbf{E} [Y] = n \cdot \frac{r}{r+g}$  and  $Z_n = f(\bar{Y}_{1,n})$ . In order to apply Azuma-Hoeffding, we need to bound the *width* of the martingale  $|Z_i - Z_{i-1}|$ .

By definition,

$$Z_i - Z_{i-1} = \mathbf{E} \left[ f(\bar{Y}_{1,n}) \mid \bar{Y}_{1,i} \right] - \mathbf{E} \left[ f(\bar{Y}_{1,n}) \mid \bar{Y}_{1,i-1} \right].$$

If we use  $S_i$  to denote the number of red balls among the first  $i$  balls, namely  $S_i = \sum_{j=1}^i Y_j$ , then

$$\mathbf{E} \left[ f(\bar{Y}_{1,n}) \mid \bar{Y}_{1,i} \right] = \mathbf{E} \left[ f(\bar{Y}_{1,n}) \mid S_i \right] = S_i + (n-i) \cdot \frac{r - S_i}{g + r - i}.$$

Therefore  $S_i = S_{i-1} + Y_i$  and

$$\begin{aligned} Z_i - Z_{i-1} &= \left( S_i + (n-i) \cdot \frac{r - S_i}{g + r - i} \right) - \left( S_{i-1} + (n-i+1) \cdot \frac{r - S_{i-1}}{g + r - i + 1} \right) \\ &= \frac{g + r - n}{g + r - i} \left( Y_i + \frac{S_{i-1} - r}{g + r - i + 1} \right). \end{aligned}$$

Note that  $r \geq S_{i-1}$  and  $g \geq (i-1) - S_{i-1}$ , we have

$$\begin{aligned} Z_i - Z_{i-1} &\leq \frac{g + r - n}{g + r - i} \left( 1 + \frac{S_{i-1} - r}{g + r - i + 1} \right) \leq \frac{g + r - n}{g + r - i} \leq 1, \\ Z_i - Z_{i-1} &\geq \frac{g + r - n}{g + r - i} \left( \frac{S_{i-1} - r}{g + r - i + 1} \right) \geq -\frac{g + r - n}{g + r - i} \geq -1. \end{aligned}$$

Therefore  $-1 \leq X_i \leq 1$  and we can apply Azuma-Hoeffding to  $Z_n - Z_0$  to obtain

$$\Pr [|Y - \mathbf{E} [Y]| \geq t] \leq 2 \exp \left( -\frac{t^2}{2n} \right).$$

## 2.2 McDiarmids Inequality

The Doob sequence we used in the balls-in-a-bag example is a very powerful and general tool to obtain concentration bounds. For a model defined by  $n$  random variables  $X_1, \dots, X_n$  and any quantity  $f(X_1, \dots, X_n)$  that we want to estimate, we can apply the Azuma-Hoeffding inequality to the Doob sequence of  $f$ . As shown in the previous example, the quality of the bound relies on the *width* of the martingale, that is, the magnitude of  $|Z_i - Z_{i-1}|$ . To determine the width of each  $|Z_i - Z_{i-1}|$  is relatively easy if the function  $f$  and the variables  $\{X_i\}_{1 \leq i \leq n}$  enjoy certain nice properties.

**Definition 4 (c-Lipschitz Function)** A function  $f(x_1, \dots, x_n)$  satisfies  $c$ -Lipschitz condition if

$$\forall i \in [n], \forall x_1, \dots, x_n, \forall y_i : |f(x_1, \dots, x_i, \dots, x_n) - f(x_1, \dots, y_i, \dots, x_n)| \leq c.$$

The McDiarmid's inequality is the application of Azuma-Hoeffding inequality to Lipschitz  $f$  and independent  $\{X_i\}$ .

**Theorem 5 (McDiarmid's Inequality)** Let  $f$  be a function on  $n$  variables satisfying  $c$ -Lipschitz condition and  $X_1, \dots, X_n$  be  $n$  independent variables. Then we have

$$\Pr[|f(X_1, \dots, X_n) - \mathbf{E}[f(X_1, \dots, X_n)]| \geq t] \leq 2e^{-\frac{2t^2}{nc^2}}.$$

*Proof.* We use  $f$  and  $\{X_i\}_{i \geq 1}$  to define a Doob martingale  $\{Z_i\}_{i \geq 1}$ :

$$\forall i : Z_i = \mathbf{E}\left[f(\bar{X}_{1,n}) \mid \bar{X}_{1,i}\right].$$

Then

$$Z_i - Z_{i-1} = \mathbf{E}\left[f(\bar{X}_{1,n}) \mid \bar{X}_{1,i}\right] - \mathbf{E}\left[f(\bar{X}_{1,n}) \mid \bar{X}_{1,i-1}\right].$$

Next we try to determine the width of  $Z_i - Z_{i-1}$ . Clearly

$$Z_i - Z_{i-1} \geq \inf_x \left\{ \mathbf{E}\left[f(\bar{X}_{1,n}) \mid \bar{X}_{1,i-1}, X_i = x\right] - \mathbf{E}\left[f(\bar{X}_{1,n}) \mid \bar{X}_{1,i-1}\right] \right\},$$

and

$$Z_i - Z_{i-1} \leq \sup_y \left\{ \mathbf{E}\left[f(\bar{X}_{1,n}) \mid \bar{X}_{1,i-1}, X_i = y\right] - \mathbf{E}\left[f(\bar{X}_{1,n}) \mid \bar{X}_{1,i-1}\right] \right\}.$$

The gap between the upper bound and the lower bound is

$$\sup_{x,y} \left\{ \mathbf{E}\left[f(\bar{X}_{1,n}) \mid \bar{X}_{1,i-1}, X_i = y\right] - \mathbf{E}\left[f(\bar{X}_{1,n}) \mid \bar{X}_{1,i-1}, X_i = x\right] \right\}.$$

For every  $x, y$  and  $\sigma_1, \dots, \sigma_{i-1}$ ,

$$\begin{aligned} & \mathbf{E}\left[f(\bar{X}_{1,n}) \mid \bigwedge_{1 \leq j \leq i-1} X_j = \sigma_j, X_i = y\right] - \mathbf{E}\left[f(\bar{X}_{1,n}) \mid \bigwedge_{1 \leq j \leq i-1} X_j = \sigma_j, X_i = x\right] \\ &= \sum_{\sigma_{i+1}, \dots, \sigma_n} \left( \Pr\left[\bigwedge_{i+1 \leq j \leq n} X_j = \sigma_j \mid \bigwedge_{1 \leq j \leq i-1} X_j = \sigma_j, X_i = y\right] \cdot f(\sigma_1, \dots, \sigma_{i-1}, y, \sigma_{i+1}, \dots, \sigma_n) \right. \\ & \quad \left. - \Pr\left[\bigwedge_{i+1 \leq j \leq n} X_j = \sigma_j \mid \bigwedge_{1 \leq j \leq i-1} X_j = \sigma_j, X_i = x\right] \cdot f(\sigma_1, \dots, \sigma_{i-1}, x, \sigma_{i+1}, \dots, \sigma_n) \right) \\ & \stackrel{(\heartsuit)}{=} \sum_{\sigma_{i+1}, \dots, \sigma_n} \Pr\left[\bigwedge_{i+1 \leq j \leq n} X_j = \sigma_j\right] \cdot (f(\sigma_1, \dots, \sigma_{i-1}, y, \sigma_{i+1}, \dots, \sigma_n) - f(\sigma_1, \dots, \sigma_{i-1}, x, \sigma_{i+1}, \dots, \sigma_n)) \\ & \stackrel{(\clubsuit)}{\leq} c. \end{aligned}$$

where  $(\heartsuit)$  uses independence of  $\{X_i\}$  and  $(\clubsuit)$  uses the  $c$ -Lipschitz property of  $f$ .

Applying Azuma-Hoeffding, we have

$$\Pr [|Z_n - Z_0| \geq t] = \Pr [|f(X_1, \dots, X_n) - \mathbb{E}[f(X_1, \dots, X_n)]| \geq t] \leq 2e^{-\frac{2t^2}{nc^2}}.$$

□

Then we examine two applications of McDiarmid's inequality.

**Example 2 (Pattern matching)** Let  $P \in \{0, 1\}^k$  be a fixed string. For a random string  $X \in \{0, 1\}^n$ , what is the expected number of occurrences of  $P$  in  $X$ ?

The expectation of occurrence times can be easily calculated using the linearity of expectations. We define  $n$  independent random variables  $X_1, \dots, X_n$ , where  $X_i$  denotes  $i$ -th character of  $X$ . Let  $Y = f(X_1, \dots, X_n)$  be the number of occurrences of  $P$  in  $X$ . Note that there are at most  $n - k + 1$  occurrences of  $P$  in  $X$ , and we can enumerate the first position of each occurrence. By the linearity of expectation, we have

$$\mathbb{E}[f] = \frac{n - k + 1}{2^k}.$$

We can then use McDiarmid's inequality to show that  $f$  is well-concentrated. To see this, we note that variables in  $\{X_i\}$  are independent and the function  $f$  is  $k$ -Lipschitz: If we change one bit of  $X$ , the number of occurrences changes at most  $k$ .

Therefore

$$\Pr [|Z_n - Z_0| \geq t] = \Pr [|f - \mathbb{E}[f]| \geq t] \leq 2e^{-\frac{2t^2}{nk^2}}.$$

Another application of McDiarmid's Inequality is to establish the concentration of chromatic number for Erdős-Rényi random graphs  $\mathcal{G}(n, p)$ .

**Example 3 (Chromatic Number of  $\mathcal{G}(n, p)$ )** Recall the notation  $\mathcal{G}(n, p)$  specifies a distribution over all undirected simple graphs with  $n$  vertices. In the model, each of the  $\binom{n}{2}$  possible edges exists with probability  $p$  independently.

For a graph  $G \sim \mathcal{G}(n, p)$ , we use  $\chi(G)$  to denote its chromatic number, the minimum number  $q$  so that  $G$  can be properly colored using  $q$  colors. There are different ways to represent  $G$  using random variables.

The most natural way is to introduce a variable  $X_e$  for every pair of vertices  $e = \{u, v\} \subseteq V$  where  $X_e = \mathbf{1}[\text{the edge } e \text{ exists in } G]$ . Then  $\{X_e\}$  are independent and the chromatic number can be written as a function  $\chi(X_{e_1}, X_{e_2}, \dots, X_{e_{\binom{n}{2}}})$ . It is easy to see that  $\chi$  is 1-Lipschitz as removing to adding one edge can only change the chromatic number by at most one. So by McDiarmid's inequality, we have

$$\Pr [|\chi - \mathbb{E}[\chi]| \geq t] \leq 2e^{-2t^2 \binom{n}{2}^{-1}}.$$

However, this bound is not satisfactory as we need to set  $t = \Theta(n)$  in order to upper bound the RHS by a constant.

We can encode the graph  $G$  in a more efficient way while reserving the Lipschitz and the independence property. Suppose the vertex set of  $G$  is  $\{v_1, \dots, v_n\}$ . We define  $n$  random variables  $Y_1, \dots, Y_n$ , where  $Y_i$  encodes the edges between  $v_i$  and  $\{v_1, \dots, v_{i-1}\}$ . Once  $Y_1, \dots, Y_n$  are given, the graph is known, so the chromatic number can be written as a function  $\chi(Y_1, \dots, Y_n)$ . Since  $Y_i$  only involves the connections between  $v_i$  and  $v_1, \dots, v_{i-1}$ , the  $n$  variables are independent.

It is also easy to see that if  $X_i$  changes, the chromatic number changes at most one. Hence  $\chi$  is 1-Lipschitz as well. Applying McDiarmid's inequality we have

$$\Pr [|\chi - \mathbb{E}[\chi]| \geq t] \leq 2e^{-\frac{2t^2}{n}}.$$

In this way, we only need  $t = \Theta(\sqrt{n})$  to bound the RHS.

### 3 Proof

#### 3.1 Proof of Theorem 1

First, we prove the following Hoeffding's lemma which will be the main technical ingredient to prove the inequality.

**Lemma 6** Let  $X$  be a random variable with  $\mathbb{E}[X] = 0$  and  $X \in [a, b]$ . Then it holds that

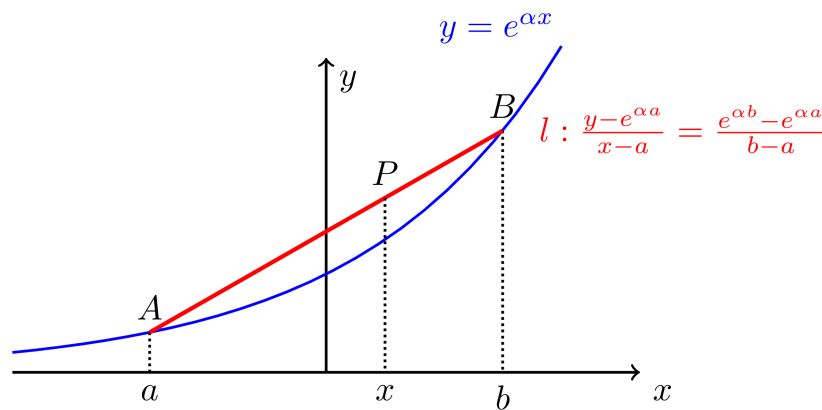
$$\mathbb{E}[e^{\alpha X}] \leq \exp\left(\frac{\alpha^2(b-a)^2}{8}\right) \text{ for all } \alpha \in \mathbb{R}.$$

*Proof.*

We first find a linear function to upper bound  $e^{\alpha x}$  so that we could apply the linearity of expectation to bound  $\mathbb{E}[e^{\alpha X}]$ . By the convexity of the exponential function and as illustrated in the figure below, we have

$$e^{\alpha x} \leq \frac{e^{\alpha b} - e^{\alpha a}}{b - a}(x - a) + e^{\alpha a}, \text{ for all } a \leq x \leq b.$$

Thus,



$$\begin{aligned}
 \mathbf{E}[e^{\alpha x}] &\leq \frac{e^{\alpha b} - e^{\alpha a}}{b - a}(-a) + e^{\alpha a} = \frac{-a}{b - a}e^{\alpha b} + \frac{b}{b - a}e^{\alpha a} \\
 &= e^{\alpha a} \left( \frac{b}{b - a} - \frac{a}{b - a}e^{\alpha(b-a)} \right) \\
 &= e^{-\theta t} (1 - \theta + \theta e^t) \qquad (\theta = -\frac{a}{b - a}, t = \alpha(b - a)) \\
 &\triangleq e^{g(t)},
 \end{aligned}$$

where

$$g(t) = -\theta t + \log(1 - \theta + \theta e^t).$$

By Taylor's theorem, for every real  $t$  there exists a  $\delta$  between 0 and  $t$  such that,

$$g(t) = g(0) + tg'(0) + \frac{1}{2}g''(\delta)t^2$$

Note that,

$$\begin{aligned}
 g(0) &= 0; \\
 g'(0) &= -\theta + \frac{\theta e^t}{1 - \theta + \theta e^t} \Big|_{t=0} \\
 &= 0; \\
 g''(\delta) &= \frac{\theta e^t (1 - \theta + \theta e^t) - \theta e^t}{(1 - \theta + \theta e^t)^2} \\
 &= \frac{(1 - \theta)\theta e^t}{(1 - \theta + \theta e^t)^2} \\
 &= \frac{(1 - \theta)\theta}{\theta^2 z + 2(1 - \theta)\theta + \frac{(1 - \theta)^2}{z}} \qquad (z = e^t) \\
 &\leq \frac{(1 - \theta)\theta}{2\theta(1 - \theta) + 2(1 - \theta)\theta} \qquad (z > 0) \\
 &= \frac{1}{4}.
 \end{aligned}$$

Thus

$$g(t) \leq 0 + t \cdot 0 + \frac{1}{2}t^2 \cdot \frac{1}{4} = \frac{1}{8}t^2 = \frac{1}{8}\alpha^2(b - a)^2.$$

Therefore,  $\mathbf{E}[e^{\alpha x}] \leq \exp\left(\frac{\alpha^2(b-a)^2}{8}\right)$  holds. □

Armed with Hoeffding's lemma, it is routine to prove Hoeffding's inequality.

*Proof.* [Proof of Theorem 1]

First note that we can assume  $\mathbf{E}[X_i] = 0$  and therefore  $\mu = 0$  (if not so, replace  $X_i$  by  $X_i - \mathbf{E}[X_i]$ ). By symmetry, we only need to prove that  $\Pr[X \geq t] \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^n (b_i - a_i)^2}\right)$ . Since

$$\Pr[X \geq t] \stackrel{\alpha > 0}{=} \Pr[e^{\alpha X} \geq e^{\alpha t}] \leq \frac{\mathbf{E}[e^{\alpha X}]}{e^{\alpha t}}$$

and

$$\mathbf{E}[e^{\alpha X}] = \mathbf{E}\left[e^{\alpha \sum_{i=1}^n X_i}\right] = \prod_{i=1}^n \mathbf{E}[e^{\alpha X_i}],$$

applying Hoeffding's lemma for each  $\mathbf{E} [e^{\alpha X_i}]$  yields

$$\mathbf{E} [e^{\alpha X_i}] \leq \exp \left( \frac{\alpha^2 (b_i - a_i)^2}{8} \right).$$

Let  $\alpha = \frac{4t}{\sum_{i=1}^n (b_i - a_i)^2}$ , we have,

$$\begin{aligned} \Pr [X \geq t] &\leq \frac{\prod_{i=1}^n \mathbf{E} [e^{\alpha X_i}]}{e^{\alpha t}} \leq \exp \left( -\alpha t + \frac{\alpha^2}{8} \sum_{i=1}^n (b_i - a_i)^2 \right) \\ &= \exp \left( -\frac{2t^2}{\sum_{i=1}^n (b_i - a_i)^2} \right). \end{aligned}$$

□

### 3.2 Proof of Theorem 2

Now we will sketch a proof of the Azuma-Hoeffding, which is quite similar to our proof of the Hoeffding inequality.

*Proof.* [Proof of Theorem 2]

Recall when we were trying to prove the Hoeffding inequality, the most difficult part is to estimate the term

$$\mathbf{E} [e^{\alpha Z_n}] = e^{\alpha Z_0} \cdot \mathbf{E} \left[ \prod_{i=1}^n e^{\alpha(Z_i - Z_{i-1})} \right].$$

In the case of Azuma-Hoeffding, we can use the property of martingales instead of independence to obtain a bound of this term. To see this, we have

$$\begin{aligned} \mathbf{E} \left[ \prod_{i=1}^n e^{\alpha Z_i - Z_{i-1}} \right] &= \mathbf{E} \left[ \mathbf{E} \left[ \prod_{i=1}^n e^{\alpha Z_i - Z_{i-1}} \mid \mathcal{F}_{n-1} \right] \right] \\ &= \mathbf{E} \left[ \prod_{i=1}^{n-1} e^{\alpha Z_i - Z_{i-1}} \mathbf{E} [e^{\alpha Z_n - Z_{n-1}} \mid \mathcal{F}_{n-1}] \right]. \end{aligned}$$

The bounds then follows by an induction argument and a conditional expectation version of Hoeffding lemma:

$$\mathbf{E} [e^{\alpha(Z_n - Z_{n-1})} \mid \mathcal{F}_{n-1}] \leq e^{-\frac{\alpha a_i^2}{8}}.$$

The proof is almost the same as our proof of Hoeffding lemma in the last lecture. □