Lecture 21: Proof of Azuma's Inequality

Recall: Azuma's Inequality

Theorem (Azuma's Inequality)

Let Ω be a sample space and p be a probability distribution over Ω. Let $\{\emptyset, \Omega\} = \mathcal{F}_0 \subset \mathcal{F}_1 \subset \cdots \subset \mathcal{F}_n$ be a filtration. Let $(\mathbb{F}_0, \mathbb{F}_1, \ldots, \mathbb{F}_n)$ be a martingale with respect to the filtration above. Suppose, for all $1 \leqslant i \leqslant n$, there exists c_i such that, for all $x \in \Omega$, we have

$$c_i \geqslant \max_{y \in \mathcal{F}_{i-1}(x)} \mathbb{F}_i(y) - \min_{y \in \mathcal{F}_{i-1}(x)} \mathbb{F}_i(y)$$

Then, the following bound holds

$$\mathbb{P}\left[\mathbb{F}_n - \mathbb{F}_0 \geqslant E\right] \leqslant \exp\left(-2E^2 \Big/ \sum_{i=1}^{n} c_i^2\right)$$

Proof I

- Let $(\Delta\mathbb{F}_1, \Delta\mathbb{F}_2, \ldots, \Delta\mathbb{F}_n)$ be the corresponding martingale difference sequence. That is, we define $\Delta\mathbb{F}_i = \mathbb{F}_i - \mathbb{F}_{i-1}$, for $1 \leqslant i \leqslant n$. Since, this is a martingale difference sequence, we have the following guarantee for all $1 \leqslant i \leqslant n$.

$$\mathbb{E}\left[\Delta\mathbb{F}_i | \mathcal{F}_{i-1}\right] = 0$$

- Note that the property of c_i can be written as follows (by subtracting $\mathbb{F}_{i-1}(x)$ from both the terms)

$$c_i \geqslant \max_{y \in \mathcal{F}_{i-1}(x)} \Delta\mathbb{F}_i(y) - \min_{y \in \mathcal{F}_{i-1}(x)} \Delta\mathbb{F}_i(y)$$

- Azuma's inequality is equivalent to proving

$$\mathbb{P}\left[\sum_{i=1}^{n} \Delta\mathbb{F}_i \geqslant E\right] \leqslant \exp\left(-2E^2 / \sum_{i=1}^{n} c_i^2\right)$$

Proof II

- Similar to the technique of proving Chernoff bound, we conclude that, for all $h > 0$, the following is true

$$\mathbb{P}\left[\sum_{i=1}^{n} \Delta\mathbb{F}_i \geqslant E\right] \leqslant \frac{\mathbb{E}\left[\exp\left(h\sum_{i=1}^{n} \Delta\mathbb{F}_i\right)\right]}{\exp(hE)}$$

- Our effort now is to upper-bound the expected value

$$\mathbb{E}\left[\exp\left(h\sum_{i=1}^{n} \Delta\mathbb{F}_i\right)\right]$$

Proof III

- Consider the following set of manipulations

$$\mathbb{E}\left[\exp\left(h\sum_{i=1}^{n}\Delta\mathbb{F}_i\right)\right]$$
$$=\mathbb{E}\left[\mathbb{E}\left[\exp\left(h\sum_{i=1}^{n}\Delta\mathbb{F}_i\right)\middle|\mathcal{F}_{n-1}\right]\right]$$
$$=\mathbb{E}\left[\exp\left(h\sum_{i=1}^{n-1}\Delta\mathbb{F}_i\right)\mathbb{E}\left[\exp\left(h\Delta\mathbb{F}_n\right)\middle|\mathcal{F}_{n-1}\right]\right]$$

The last equality is because $\exp\left(h\sum_{i=1}^{n-1}\Delta\mathbb{F}_i\right)$ is $\mathcal{F}_{n-1}$ measurable.

Proof IV

- We can apply Hoeffding's Lemma to upper bound $\mathbb{E}\left[\exp\left(h\Delta\mathbb{F}_n\right)\middle|\mathcal{F}_{n-1}\right]$ as follows

$$\mathbb{E}\left[\exp\left(h\Delta\mathbb{F}_n\right)\middle|\mathcal{F}_{n-1}\right] \leqslant \exp\left(\frac{h^2}{8}c_n^2\right)$$

- So, we obtain that

$$\mathbb{E}\left[\exp\left(h\sum_{i=1}^{n}\Delta\mathbb{F}_i\right)\right] \leqslant \exp\left(\frac{h^2}{8}c_n^2\right)\mathbb{E}\left[\exp\left(h\sum_{i=1}^{n-1}\Delta\mathbb{F}_i\right)\right]$$

- Repeatedly applying the bound to the last $\Delta\mathbb{F}_i$, we get

$$\mathbb{E}\left[\exp\left(h\sum_{i=1}^{n}\Delta\mathbb{F}_i\right)\right] \leqslant \exp\left(\frac{h^2}{8}\sum_{i=1}^{n}c_i^2\right)$$

Proof V

- So, we get that

$$\mathbb{P}\left[\sum_{i=1}^{n} \Delta\mathbb{F}_i \geqslant E\right] \leqslant \exp\left(\frac{h^2}{8}\sum_{i=1}^{n} c_i^2 - hE\right)$$

- Rest of the proof is identical to the proof of the Hoeffding's Bound. The optimal choice of h that minimizes the RHS is

$$h^* = 4E / \sum_{i=1}^{n} c_i^2$$

- Substituting this value of h, we obtain

$$\mathbb{P}\left[\sum_{i=1}^{n} \Delta\mathbb{F}_i \geqslant E\right] \leqslant \exp\left(-2E^2 / \sum_{i=1}^{n} c_i^2\right)$$

Concluding Note

- Students are highly recommended to use a representative example (as worked out in the class) to verify all the "equalities" and the "inequalities" used in the derivation of the proof