Lecture 11: Independent Bounded Differences Inequality

Overview

- Today, we shall see a result referred to as the "Independent Bounded Differences Inequality."
- We shall not see the proof of this result today. Something more general, called the "Azuma's Inequality," subsumes this result.
- Today, we shall see how a large class of concentration results follow due to this concentration inequality. In fact, one such consequence shall look very similar to the "Talagrand Inequality," which we shall study in the next lecture

- Let $\Omega_1, \dots, \Omega_n$ be sample spaces
- Define $\Omega := \Omega_1 \times \cdots \times \Omega_n$
- Let $f \colon \Omega \to \mathbb{R}$
- Let $\mathbb{X} = (\mathbb{X}_1, \dots, \mathbb{X}_n)$ be a random variable over Ω such that each $\mathbb{X}_i$ is independent and $\mathbb{X}_i$ is a random variable over the sample space Ω_i

Definition

A function $f \colon \Omega \to \mathbb{R}$ has *bounded differences if for all* $x, x' \in \Omega$, *there exists* $i \in \{1, \dots, n\}$ *such that* x *and* x' *differ only at the* i*-th coordinate, then the output of the function* $\left|f(x) - f(x')\right| \leqslant c_i$.

We state the following bound without proof.

Theorem (Bounded Difference Inequality)

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

Applying the same theorem to $-f$, we deduce that

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

Intuitively, if all $c_i = 1$, the random variable $f(\mathbb{X})$ is concentrated around its expected value $\mathbb{E}\left[f(\mathbb{X})\right]$ within a radius of $\sqrt{n}$

Example

- Note that the Chernoff-Hoeffding's bound is a corollary of this theorem
- Let $\mathcal{G}_{n,p}$ be a random graph over n vertices, where each edge is included in the graph independently with probability p. Note that we have m random variables and one indicator variable for each edge in the graph. Note that the chromatic number of the graph is a function with a bounded difference.
- Several graph properties like the number of connected components
- Longest increasing subsequence
- Max-load in balls-and-bins experiments
- What about the max-load in the power-of-two-choices?

Applicability and Meaningfulness of the Bounds

- Although the theorem applies to a problem, the bound that it produces might not be a meaningful bound
- The bound says that the probability mass is concentrated within $\approx \sqrt{n}$ around the expected value $\mathbb{E}\left[f(\mathbb{X})\right]$
- If the expected value $\mathbb{E}\left[f(\mathbb{X})\right]$ is $\omega(\sqrt{n})$ then the theorem gives a meaningful bound
- However, if $\mathbb{E}\left[f(\mathbb{X})\right]$ is $O(\sqrt{n})$ then the theorem does not give a meaningful bound. For example, the longest increasing subsequence, max-load in balls-and-bins experiments

Hamming Distance

Next, we shall see a powerful application of the independent bounded difference inequality. First, let us introduce the definition of Hamming Distance

Definition (Hamming Distance)

Let $x, x' \in \Omega := \Omega_1 \times \cdots \times \Omega_n$. We define

$$d_H(x, x') := \left| \{ i \colon 1 \leqslant i \leqslant n \text{ and } x_i \neq x_i' \} \right|$$

- The Hamming distance of x and x' bounds the number of indices where x and x' differ
- Let $A \subseteq \Omega$ and $d_H(x, A) := \min_{y \in S} d_H(x, y)$.

Definition

The set A_k is defined as follows

$$A_k := \{ x \in \Omega \colon d_H(x, A) \leqslant k \}$$

Distance from Dense Sets

Lemma

Let $A \subseteq \Omega$. The following bound holds.

$$\mathbb{P}\left[\mathbb{X} \in A\right] \cdot \mathbb{P}\left[d_H(\mathbb{X}, A) \geqslant E\right] \leqslant \exp(-E^2/2n)$$

Intuition

- Suppose $\mathbb{P}\left[\mathbb{X} \in A\right] = 1/2$, then we have

  $$\mathbb{P}\left[\mathbb{X} \in A_{E-1}\right] \geqslant 1 - 2\exp(-E^2/2n)$$

  That is, nearly all points lie within $E \approx \sqrt{n}$ distance from the dense set A
- Note that this result holds for <u>all dense sets</u> A

- Our objective is to prove that

$$\mathbb{P}\left[\mathbb{X} \in A\right] \cdot \mathbb{P}\left[d_H(\mathbb{X}, A) \geqslant E\right] \leqslant \exp(-E^2/2n).$$

Observe that the above inequality is a consequence of the following second inequality:

$$\min\left\{\mathbb{P}\left[\mathbb{X} \in A\right], \mathbb{P}\left[d_H(\mathbb{X}, A) \geqslant E\right]\right\} \leqslant \exp(-E^2/2n).$$

Therefore, we will prove this second inequality instead.

- Note that $d_H(\cdot, A)$ is a bounded difference function with $c_i = 1$, for $i \in \{1, \dots, n\}$
- Define $\mu = \mathbb{E}\left[d_H(\mathbb{X}, A)\right]$

Proof based on the Bounded Difference Inequality II

- Consider the inequality (using the independent bounded difference inequality for the lower tail)

$$\mathbb{P}\left[\mathbb{X} \in A\right] = \mathbb{P}\left[d_H(\mathbb{X}, A) - \mu \leqslant -\mu\right] \leqslant \exp(-2\mu^2/n).$$

  We will call this the "density bound."

- Now we are ready to prove the "second inequality."
  1. Case 1. Suppose $E \geqslant 2\mu$.

$$\begin{aligned}\min\left\{\mathbb{P}\left[\mathbb{X} \in A\right], \mathbb{P}\left[d_H(\mathbb{X}, A) \geqslant E\right]\right\} &\leqslant \mathbb{P}\left[d_H(\mathbb{X}, A) \geqslant E\right] \\ &= \mathbb{P}\left[d_H(\mathbb{X}, A) - \mu \geqslant (E - \mu)\right] \\ &\leqslant \exp(-2(E-\mu)^2/n) \\ &\quad \text{(By the upper tail bound)} \\ &\leqslant \exp(-E^2/2n). \\ &\qquad \text{(Because } E \geqslant 2\mu\text{)}\end{aligned}$$

2. Case 2. Suppose $0 \leqslant E < 2\mu$.

$$\min\left\{\mathbb{P}\left[\mathbb{X} \in A\right], \mathbb{P}\left[d_H(\mathbb{X}, A) \geqslant E\right]\right\} \leqslant \mathbb{P}\left[\mathbb{X} \in A\right]$$

$$\leqslant \exp(-2\mu^2/n)$$

(By the "density bound" inequality)

$$\leqslant \exp(-E^2/2n).$$

(Because $0 \leqslant E < 2\mu$)

- Therefore, irrespective of whether $E \geqslant 2\mu$ or $0 \leqslant E < 2\mu$, the following bound holds

$$\min\left\{\mathbb{P}\left[\mathbb{X} \in A\right], \mathbb{P}\left[d_H(\mathbb{X}, A) \geqslant E\right]\right\} \leqslant \exp(-E^2/2n).$$

This completes the proof of our result.

An Application of "Distance from Dense Sets"

(A Slightly weaker version of) Chernoff-bound

- Consider a uniform distribution over $\Omega = \{0,1\}^n$
- Let A be the set of all binary strings that have at most $n/2$ 1s. The density of this set is $\geqslant 1/2$
- A string x with $d_H(x, A) \geqslant E$ is equivalent to x having $(n/2) + E$ 1s
- So, the probability of a uniformly sampled binary string has $(n/2) + E$ 1s is at most $2\exp(-E^2/2n)$