Lecture 07: Independent Bounded Differences Inequality

Setting

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

Bounded Differences

Definition (Bounded Differences)

Let function $f : \Omega \to \mathbb{R}$. The function f has *bounded differences*, if for all $x, x' \in \Omega$, $i \in [n]$, and x and x' differ only at i-th coordinate, the output of the function $\left|f(x) - f(x')\right| \leqslant c_i$.

Bounded Difference Inequality

We will state the following bound without proof

Theorem (Bounded Difference Inequality)

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

We can apply the same inequality to $-f$ and deduce that

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

Intuition: $f(\mathbb{X})$ is concentrated around $\mathbb{E}\left[f(\mathbb{X})\right]$ within a radius of $t \approx \sqrt{n}$

Examples

- Prove Chernoff-Hoeffding's bound as a corollary of this result
- Let $\mathcal{G}_{n,p}$ be a 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, one indicator variable for each edge being included. Note that the chromatic number of the graph is a function with bounded difference.
- Several graph properties like number of connected components
- Longest increasing subsequence
- Max load in balls-and-bins experiments
- Max load in the power-of-two-choices is *not* bounded difference function

Applicability and Meaningfulness of the Bounds

- Although the theorem is applicable, the bound it produces might not be applicable
- The bound says that the probability mass is concentrated within $\approx \sqrt{n}$ of the expected value $\mathbb{E}\left[f(\mathbb{X})\right]$
- If $\mathbb{E}\left[f(\mathbb{X})\right] := \mu$ is $\omega(\sqrt{n})$ then the theorem gives a good bound. The distribution is concentrated within $o(\mu)$ from the average μ. This, we will consider a *good concentration bound*
- If $\mathbb{E}\left[f(\mathbb{X})\right]$ is $O(\sqrt{n})$ then the theorem does not give a good bound. For example, longest increasing subsequence, max-load in balls-and-bins

Hamming Distance

Definition (Hamming Distance)

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

$$d_H(x, x') := \left| \left\{ i : x_i \neq x'_i \right\} \right|$$

- The Hamming distance counts the number of indices where x and x' differ
- Let $A \subseteq \Omega$ and $d_H(x, A) := \min_{y \in A} d_H(x, y)$. Intuitively, $d_H(x, A) \geqslant t$ implies that x is t-far from every point in A

Definition

The set A_k is defined as

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

Distance from Dense Sets

Lemma

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

Intuition

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

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

  That is, nearly all points lie within $t \approx \sqrt{n}$ distance from the set A
- Note that this result hold for *all* sets A

Proof based on the Bounded Differences Inequality I

- Note that $d_H(\cdot, A)$ is a bounded difference function with $c_i = 1$
- For $\mu = \mathbb{E}\left[d_H(\mathbb{X}, A)\right]$, consider the inequality

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

- We use $t = \mu$ and we get:

$$\mathbb{P}\left[d_H(\mathbb{X}, A) \leqslant 0\right] \leqslant \exp\left(-2\mu^2/n\right)$$

- Note that
  $\mathbb{P}\left[d_H(\mathbb{X}, A) \leqslant 0\right] = \mathbb{P}\left[d_H(\mathbb{X}, A) = 0\right] = \mathbb{P}\left[\mathbb{X} \in A\right] =: \nu$
- Now, we can relate μ and ν:

$$\mu \leqslant \sqrt{\frac{n}{2}\log(1/\nu)}$$

Proof based on the Bounded Differences Inequality II

- Now, we apply the other inequality

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

- By change of variables, we have

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

- Case 1: $t \geqslant 2\mu$. For this case, we can conclude that $t/2 \leqslant (t-\mu)$. So, we have:

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

- Case 2: $0 \leqslant t \leqslant 2\mu$. For this case, we can conclude that $\mathbb{P}[\mathbb{X} \in A] \leqslant \exp(-2\mu^2/n) \leqslant \exp(-t^2/2n)$
- Therefore, the two cases imply

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

- This inequality, implies, for all t, that

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

An Application

(Slightly weaker-version of) Chernoff-bound for $B(n, 1/2)$.

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