Lecture 05: Chernoff-like Concentration Bounds

This lecture note closely follows the presentation of Chapter 1 and 2 of "Concentration," by Colin McDiarmid (link)

Theorem (Chernoff Bound)

Let $X_1, \dots, X_n$ be independent binary random variables with $\Pr[X_k = 1] = p$, for every $1 \leqslant k \leqslant n$. Let $S_n = \sum X_k$. Then for any $t \geqslant 0$,

$$\Pr\left[S_n - np \geqslant nt\right] \leqslant \exp\left(-2nt^2\right)$$

Proof of Chernoff Bound

$$\begin{aligned}\Pr[S_n \geqslant np + nt] &= \Pr[\exp(hS_n) \geqslant \exp(hn(p+t))], \text{ for any } h \geqslant 0 \\ &\leqslant \frac{\mathbb{E}[\exp(hS_n)]}{\exp(hn(p+t))} \\ &\leqslant \frac{u(h)}{\exp(hn(p+t))}\end{aligned}$$

- Let $u(h)$ be an upper bound on $\mathbb{E}[\exp(hS_n)]$
- Let h^* be the value of $h > 0$ that minimizes $\frac{u(h)}{\exp(hn(p+t))}$

$$\Pr[S_n - np \geqslant nt] \leqslant \frac{u(h^*)}{\exp(h^* n(p+t))}$$

Bound on Expectation

$$\mathbb{E}[\exp(hS_n)] = \prod \exp(hS_k)$$
$$= \prod (1 - p + p\exp(h))$$
$$= (1 - p + p\exp(h))^n =: u(h)$$

$$h^* = \text{argmin}\ _h \left(\frac{1 - p + p\exp(h)}{\exp(h(p+t))} \right)^n$$

Set $\exp(h^*) = \frac{(p+t)(1-p)}{p(1-p-t)}$ to get the bound

A Useful Lemma

Lemma

Let $X_1, \ldots, X_n$ be independent such that $0 \leqslant X_k \leqslant 1$ for each k. Let $\mu = \mathbb{E}[S_n]$, $p = \mu/n$ and $\bar{p} = 1 - p$. Then for any $0 < t < \bar{p}$

$$\Pr(S_n - np \geqslant nt) \leqslant \left(\left(\frac{p}{p+t} \right)^{p+t} \left(\frac{\bar{p}}{\bar{p}-t} \right)^{\bar{p}-t} \right)^n$$

- Let $p_k = \mathbb{E}[X_k]$
- $\mathbb{E}[\exp(hX_k)] \leqslant 1 - p_k + p_k \exp(h)$, using Jensen's Inequality
- $\mathbb{E}[\exp(hS_n)] = \prod \mathbb{E}[\exp(hX_k)] \leqslant \prod (1 - p_k + p_k \exp(h)) \leqslant_{\text{AM-GM}} (1 - p + p\exp(h))^n =: u(h)$

Theorem (Hoeffding's Bound)

Let $X_1, \dots, X_n$ be independent such that $a_k \leqslant X_k \leqslant b_k$ for each k. Let $S_n = \sum X_k$ and $np = \mathbb{E}[S_n]$. Then, for any $t \geqslant 0$,

$$\Pr\left(S_n - np \geqslant nt\right) \leqslant \exp\left(-2n^2t^2 / \sum (b_k - a_k)^2\right)$$

Proof

Left as an exercise. Use the following lemma on the random variable $(X_k - \mathbb{E}[X_k])$ and apply AM-GM inequality:

Lemma

Let X be a random variable such that $\mathbb{E}[X] = 0$ and $a \leqslant X \leqslant b$. Then for any $h > 0$,

$$\mathbb{E}(\exp(hX)) \leqslant \exp(h^2(b-a)^2/8)$$

Extensions

Additional materials on the course website provide references to the following intuitions:

- Identical bounds also hold for $\Pr[\max S_k - kp \geqslant nt]$ (this uses *Doob's maximal inequality for submartingales*)
- Identical bounds also hold for random variables with "slightly lesser" independence
- Bounds for k-wise independent also exit
- Concentration bound for hypergeometric distribution (sampling with replacement) is tighter

Bounded Differences

Theorem

Let $\mathbf{X} = (X_1, \ldots, X_m)$ *be a family of independent random variables with* X_k *taking values in* Ω_k*, for each* k*. For a real valued function* f *defined on* $\prod \Omega_k$*, the following holds:*

$$\left| f(\mathsf{x}) - f(\mathsf{x}') \right| \leqslant c_k,$$

whenever x *and* x' *differ only in the* k*-th coordinate. Let* $\mu = \mathbb{E}[f(\mathbf{X})]$*. Then for any* $t \geqslant 0$*, we have:*

$$\Pr[f(\mathbf{X}) - \mu \geqslant nt] \leqslant \exp(-2n^2t^2 / \sum c_k^2)$$

Think: Concentration of longest common subsequence

- $d_H(\mathbf{x}, \mathbf{y})$ is the number of coordinates where $\mathbf{x}$ and $\mathbf{y}$ differ
- $d_H(\mathbf{x}, A)$ is the minimum distance $d_H(\mathbf{x}, \mathbf{y})$, where $\mathbf{y} \in A$

Theorem

Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a family of independent random variables with X_k taking values in Ω_k, for each k. Let A be a subset of the product space $\prod \Omega_k$. Then for any $t \geqslant 0$,

$$\Pr[\mathbf{X} \in A] \cdot \Pr[d_H(\mathbf{X}, A) \geqslant nt] \leqslant \exp(-nt^2/2)$$

General Distance

- For $\alpha = (\alpha_1, \dots, \alpha_n) \geqslant \mathbf{0}$ and $\|\alpha\|_2 = 1$, define $d_\alpha(\mathbf{x}, \mathbf{y}) = \sum_{k : x_k \neq y_k} \alpha_k$
- $d_\alpha(\mathbf{x}, A)$ is the minimum distance $d_\alpha(\mathbf{x}, \mathbf{y})$, where $\mathbf{y} \in A$

Theorem

Let $\mathbf{X} = (X_1, \dots, X_n)$ be a family of independent random variables with X_k taking values in Ω_k, for each k. Let A be a subset of the product space $\prod \Omega_k$. Then for any $t \geqslant 0$ and $\alpha \geqslant \mathbf{0}$ and $\|\alpha\|_2 = 1$,

$$\Pr[\mathbf{X} \in A] \cdot \Pr[d_\alpha(\mathbf{X}, A) \geqslant nt] \leqslant \exp(-n^2 t^2 / 2)$$

- Idea is to use a "dense set A"
- Talagrand inequality will generalize this further