Lecture 06: Martingale Inequalities and Talagrand Inequality

This lecture note closely follows the presentation of Chapter 3 and 4 of "Concentration," by Colin McDiarmid (link)

The lecture assumes basic familiarity with probability spaces, σ-fields and martingales. See, for example, Section 3.3 of "Concentration," by Colin McDiarmid (link)

A Useful Lemma

All results are with respect to an implicit filtration $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \cdots \subseteq \mathcal{F}_n$.

Lemma

Let $Y_1, \ldots, Y_n$ be a martingale difference sequence with $-a_k \leqslant Y_k \leqslant 1 - a_k$ for each k. Let $a = \sum a_k/n$ and $\bar{a} = 1 - a$. Then for any $0 \leqslant t \leqslant \bar{a}$,

$$\Pr(\sum Y_k \geqslant nt) \leqslant \left(\left(\frac{a}{a+t}\right)^{a+t} \left(\frac{\bar{a}}{\bar{a}-t}\right)^{\bar{a}-t} \right)^n$$

Let $S_n = S_{n-1} + Y_n$. Use the previous lecture to complete the proof from the following result:

$$\begin{aligned}
\mathbb{E}[\exp(hS_n)] &= \mathbb{E}[\exp(hS_{n-1}) \cdot \mathbb{E}[\exp(hY_n|\mathcal{F}_{n-1})]] \\
&\leqslant \mathbb{E}[\exp(hS_{n-1})\left((1-a_n)\exp(-ha_n) + a_n \exp(h(1-a_n))\right)] \\
&\leqslant \exp(-\sum ha_k) \cdot \prod (1 - a_k + a_k \exp(h))
\end{aligned}$$

Hoeffding–Azuma inequality

Theorem (Hoeffding–Azuma inequality)

Let $Y_1, \ldots, Y_n$ be a martingale difference sequence with $|Y_k| \leqslant c_k$ for each k. For any $t \geqslant 0$, we have:

$$\Pr[\sum Y_k \geqslant nt] \leqslant 2\exp(-n^2t^2/2\sum c_k^2)$$

Think: How to prove this?

Generalizations

Theorem

Let $Y_1, \ldots, Y_n$ be a martingale difference sequence with $-a_k \leqslant Y_k \leqslant 1 - a_k$ for each k. Let $a = \sum a_k/n$.

1. *For any $t \geqslant 0$*

$$\Pr[\left|\sum Y_k\right| \geqslant t] \leqslant 2\exp(-2t^2/n)$$

2. *For any $\varepsilon > 0$*

$$\Pr[\sum Y_k \geqslant \varepsilon an] \leqslant \exp\left(-\varepsilon^2 an/2(1+\varepsilon/3)\right)$$

3. *For any $\varepsilon > 0$*

$$\Pr[\sum Y_k \leqslant -\varepsilon an] \leqslant \exp(-\varepsilon^2 an/2)$$

Use the useful lemma to prove this

Talagrand's Convex Distance

- Recall definition of $d_\alpha(\cdot,\cdot)$ from the previous lecture
- $d_T(\mathbf{x}, A) = \sup\{d_\alpha(\mathbf{x}, A) \colon \alpha \geqslant 0, \|\alpha\|_2 = 1\}$

Theorem (Talagrand Inequality)

Let $\mathbf{X} = (X_1, \dots, X_n)$ be a family of independent random variables and let A be a subset of the product space. Then for any $t \geqslant 0$,

$$\Pr[\mathbf{X} \in A] \cdot \Pr[d_T(\mathbf{X}, A) \geqslant t] \leqslant \exp(-t^2/4)$$

Read the proof from the literature.

Application of Talagrand Inequality

Let X_k be a uniform random number. Let $f(\mathbf{x})$ represent the length of the longest increasing subsequence of $(x_1, \ldots, x_n)$.

- For every $\mathbf{x}$ there exists a subset $K(\mathbf{x}) \subseteq [n]$ such that $f(\mathbf{x}) = K(\mathbf{x})$
- Consider any $\mathbf{y}$
- Note that:
  $f(\mathbf{y}) \geqslant |\{i \in K(\mathbf{x}) \colon y_i = x_i\}| = f(\mathbf{x}) - |\{i \in K(\mathbf{x}) \colon y_i \neq x_i\}|$
- Let α be a $1/\sqrt{f(\mathbf{x})}$ at all indices in $K(\mathbf{x})$ and 0 elsewhere
- We have $f(\mathbf{y}) \geqslant f(\mathbf{x}) - \sqrt{f(\mathbf{x})} d_\alpha(\mathbf{x}, \mathbf{y})$

Definition (c-Configuration Function)

A c-configuration function f satisfies: For any $\mathbf{x}$ and $\mathbf{y}$, there exists α such that

$$f(\mathbf{y}) \geqslant f(\mathbf{x}) - \sqrt{c f(\mathbf{x})} d_\alpha(\mathbf{x}, \mathbf{y})$$

Concentration of c-Configuration Functions

Theorem

Let f be a c-configuration function and let m be the median for $f(\mathbf{X})$. Then for any $t \geqslant 0$,

$$\Pr[f(\mathbf{X}) \geqslant m + t] \leqslant 2\exp(-t^2/4c(m+t))$$

We will prove this using Talagrand inequality

Proof

- By definition $f(\mathbf{x}) \leqslant f(\mathbf{y}) + \sqrt{cf(\mathbf{x})}d_\alpha(\mathbf{x}, \mathbf{y})$
- Let $A_a = \{\mathbf{y} \colon f(\mathbf{y}) \leqslant a\}$
- So, $f(\mathbf{x}) \leqslant a + \sqrt{cf(\mathbf{x})}d_\alpha(\mathbf{x}, \mathbf{y})$ for any $\mathbf{y} \in A_a$
- So, $f(\mathbf{x}) \leqslant a + \sqrt{cf(\mathbf{x})}d_\alpha(\mathbf{x}, A_a) \leqslant a + \sqrt{cf(\mathbf{x})}d_T(\mathbf{x}, A_a)$
- If $f(\mathbf{x}) \geqslant a + t$ then $d_T(\mathbf{x}, A_a) \geqslant \frac{f(\mathbf{x})-a}{\sqrt{xf(\mathbf{x})}} \geqslant \frac{t}{\sqrt{c(a+t)}}$
- Therefore, $\Pr[f(\mathbf{X}) \geqslant a + t] \leqslant \Pr\left[d_T(\mathbf{X}, A_a) \geqslant \frac{t}{\sqrt{c(a+t)}}\right]$
- By Talagrand Inequality:

$$\begin{aligned}
&\Pr[f(\mathbf{X}) \leqslant a) \cdot \Pr[f(\mathbf{X}) \geqslant a + t] \\
&\leqslant \Pr[\mathbf{X} \in A_a] \cdot \Pr\left[d_T(\mathbf{X}, A_a) \geqslant \frac{t}{\sqrt{c(a+t)}}\right] \\
&\leqslant \exp(-t^2/4c(a+t))
\end{aligned}$$

- Use $a = m$ and the fact that $\Pr[\mathbf{X} \leqslant m] \geqslant 1/2$