LaTeX Template for STAT 547C Final Project

Benjamin Bloem-Reddy

December 10, 2020

1 Background

Inequalities in statistics provide a means of bounding measures and quantities. They are usually used to specify bounds on quantities when these bounds are particularly difficult or intractable to compute. Inequalities play an important role in the algorithm of statistical learning and machine learning. They are involved in underpinning methods or approaches used in actual cases.

Here are several famous concentration inequalities involved in statistical learning.

- Markov's Inequality
- Chebychev's Inequality
- Bounded Differences Inequality

In this project, I will focus on the Bounded Differences Inequality.

1.1 Chebychev's Inequality

For some $a \in X$ where $X \subseteq R$, let f be a non-negative function such that $\{f(x) \geq b, forallx \geq a\}$, where $b \in Y$ where $Y \subseteq R$. Then the following ineuqality holds,

$$P(x \geq a) \leq \frac{\mathbb{E}f(x)}{b} \tag{1}$$

1.2 Chernoff's bound

Suppose the function f is monotonically increasing. Thus, for every $x \geq a$, $f(x) \geq f(a)$. Substitute $b = f(a)$ in Chebychev's Inequality, and we have:

$$P(x \geq a) \leq \frac{\mathbb{E}f(x)}{f(a)} \tag{2}$$

Markov's Inequality says that

$$P(x \geq a) \leq \frac{\mathbb{E}x}{a} \tag{3}$$

It holds for $a > 0$ and nonnegetive x,

$$P(|x - \mathbb{E}(x)| \geq a) \leq \frac{\mathbb{E}\left\{|x - \mathbb{E}(x)|^2\right\}}{a^2} = \frac{\mathrm{Var}\{x\}}{a^2} \tag{4}$$

The Chernoff's bound is then,

$$P(x \geq a) \leq \frac{\mathbb{E}e^{sx}}{e^{sa}} \tag{5}$$

1.3 Chernoff technique

[**Bartlett:2020**] For $t > 0$:

$$P(X - \mathbb{E}X \geq t) \leq \inf_{\lambda > 0} e^{-\lambda t} M_{X-\mu}(\lambda) \tag{6}$$

where $M_{X-\mu}(\lambda) = \mathbb{E}\exp(\lambda(X-\mu))($ for $\mu = \mathbb{E}X)$ is the moment-generating function of $X - \mu$.
By using Hoeffding's Inequality, we will have
[**Bartlett:2020**] For a random variable $X \in [a, b]$ with $\mathbb{E}X = \mu$ and $\lambda \in \mathbb{R}$,

$$\ln M_{X-\mu}(\lambda) \leq \frac{\lambda^2(b-a)^2}{8} \tag{7}$$

1.4 Sub-Gaussian Random Variables

[**Bartlett:2020**] X is sub-Gaussian with parameter σ^2 if, for all $\lambda \in \mathbb{R}$,

$$\ln M_{X-\mu}(\lambda) \leq \frac{\lambda^2(\sigma)^2}{2} \tag{8}$$

Examples:

- $X \sim N(\mu, \sigma^2)$ is sub-Gaussian with parameter σ^2;
- $X \in [a, b]$ is sub-Gaussian with parameter $(b-a)^2/4$;
- X_i is independent, sub-Gaussian with parameters σ_i^2 implies $\sum_i X_i$ is sub-Gaussian with parameter $\sum_i \sigma_i^2$.

1.5 Pre-Gaussian Random Variables

[**Bartlett:2020**] X is pre-Gaussian with parameters (σ^2, b) if, for all $|\lambda| < 1/b$,

$$\log M_{X-\mu}(\lambda) \leq \frac{\lambda^2\sigma^2}{2} \tag{9}$$

Example:

- Sub-Gaussian X with parameter σ^2 is pre-Gaussian with parameters (σ^2, b) for all $b > 0$.

1.6 Martingales

[**Bartlett:2020**] A sequence Y_n of random variables adapted to a filtration $\mathcal{F}_n$ is a martingale if, for all n,

$$\begin{aligned} \mathbb{E}\,|Y_n| &< \infty \\ \mathbb{E}\,[Y_{n+1} \mid \mathcal{F}_n] &= Y_n \end{aligned} \tag{10}$$

Note:

- $\mathcal{F}_n$ is a filtration means these σ-fields are nested: $\mathcal{F}_n \subset \mathcal{F}_{n+1}$.
- Y_n is adapted to $\mathcal{F}_n$ means that each Y_n is measurable with respect to $\mathcal{F}_n$.

1.7 Martingale Difference Sequences

[**Bartlett:2020**] A sequence D_n of random variables adapted to a filtration $\mathcal{F}_n$ is a martingale Difference sequence if, for all n,

$$\begin{aligned} \mathbb{E}\,|D_n| &< \infty \\ \mathbb{E}\,[D_{n+1} \mid \mathcal{F}_n] &= 0 \end{aligned} \tag{11}$$

2 Open questions and research directions

2.1 Open questions

In 2000, Talagrand [**talagrand:1995**] proposed more inequalities in dertermining bounds. Generally he extends the Bounded Differences Inequality to further applicable fields. For example, he extends it towards the form of integration. Besides, he also discusses the optimization (or sharpening) on the bounds based on different objects. Can we develop further specific inequalities for more narrowed objects? In other words, for some specific application and problems, can we simplify the Bounded Differences Inequality for more precise using? Also, can we use the original Bounded Differences Inequality to proof other concept?

For example the "Hamming Distance": [**hmaji:2017**]

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

$$d_H\left(x, x'\right) := \left|\left\{i : x_i \neq x_i'\right\}\right| \tag{12}$$

and define the set $A_k := \{x : x \in \Omega, d_H(x, A) \leq k\}$

Can we prove that

$$P[X \in A] \cdot P\left[d_H(X, A) \geqslant t\right] \leqslant \exp\left(-t^2/2n\right) \tag{13}$$

2.2 Doob construction

A Doob martingale is a construction of a stochastic process to approximates a given random variable. It has the properties of the martingale with respect to the given filtration.

The Doob construction can be used to help prove the Bounded Differences Inequality by following brief steps. [**Bartlett:2020**]

$$\begin{aligned} Y_i &= \mathbb{E}\left[f(X) \mid X_1^i\right] \\ D_i &= Y_i - Y_{i-1} \\ f(X) - \mathbb{E} f(X) &= \sum_{i=1}^n D_i \end{aligned} \tag{14}$$

Then

$$\begin{aligned} D_i = Y_i - Y_{i-1} &= \mathbb{E}\left[f(X) \mid X_1^i\right] - \mathbb{E}\left[f(X) \mid X_1^{i-1}\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[f(X) \mid X_1^i\right] - f(X) \mid X_1^{i-1}\right] \end{aligned} \tag{15}$$

Thus D_i is a random variable that falls in an interval of length no more than B_i.

Similarly, can we use the Doob construction in proof of extended bounded differences inequalities? For example, the inequalities proposed by Talagrand in 2020.

A Exercises

A.1 Exercise 1: Bernstein's Inequality

Prove the Bernstein's Inequality

If $\mathbb{P}(|X_i| \le c) = 1$ and $\mathbb{E}(X_i) = \mu$, then for any $\epsilon > 0$,

$$\mathbb{P}\left(\left|\bar{X}_n - \mu\right| > \epsilon\right) \le 2\exp\left\{-\frac{n\epsilon^2}{2\sigma^2 + 2c\epsilon/3}\right\} \tag{16}$$

Solution:

[**John:2008**] Suppose thath $|X| \le c$ and $\mathbb{E}(X) = 0$, for any $t > 0$,

$$\mathbb{E}(e^{tX}) \le \exp\{t^2\sigma^2(\frac{e^{tc} - 1 - tc}{(tc)^2})\} \tag{17}$$

where $\sigma^2 = Var(X)$

Proof. Let $F = \sum_{r=2}^{\infty} \frac{t^{r-2}\mathbb{E}(X^r)}{r!\sigma^2}$. Then,

$$\mathbb{E}\left(e^{tX}\right) = \mathbb{E}\left(1 + tx + \sum_{r=2}^{\infty} \frac{t^r X^r}{r!}\right) = 1 + t^2\sigma^2 F \le e^{t^2\sigma^2 F} \tag{18}$$

For $r \ge 2$, $\mathbb{E}\left(X^r\right) = \mathbb{E}\left(X^{r-2}X^2\right) \le c^{r-2}\sigma^2$.

Therefore, we have

$$F \le \sum_{r=2}^{\infty} \frac{t^{r-2}c^{r-2}\sigma^2}{r!\sigma^2} = \frac{1}{(tc)^2}\sum_{i=2}^{\infty}\frac{(tc)^r}{r!} = \frac{e^{tc} - 1 - tc}{(tc)^2} \tag{19}$$

Hence,

$$\mathbb{E}\left(e^{tX}\right) \le \exp\left\{t^2\sigma^2 \frac{e^{tc} - 1 - tc}{(tc)^2}\right\} \tag{20}$$

□

From Lemma A.1, we can assume that $\mu = 0$ for simplicity. Then

$$\mathbb{E}\left(e^{tX_i}\right) \le \exp\left\{t^2\sigma_i^2 \frac{e^{tc} - 1 - tc}{(tc)^2}\right\} \tag{21}$$

where $\sigma^2 = \mathbb{E}(X_i^2)$.

Then we have

$$\begin{aligned}\mathbb{P}\left(\bar{X}_n > \epsilon\right) &= \mathbb{P}\left(\sum_{i=1}^{n} X_i > n\epsilon\right) = \mathbb{P}\left(e^{t\sum_{i=1}^{n} X_i} > e^{tn\epsilon}\right) \\ &\le e^{-tn\epsilon}\mathbb{E}\left(e^{t\sum_{i=1}^{n} X_i}\right) = e^{-tn\epsilon}\prod_{i=1}^{n}\mathbb{E}\left(e^{tX_i}\right) \\ &\le e^{-tn\epsilon}\exp\left\{nt^2\sigma^2\frac{e^{tc} - 1 - tc}{(tc)^2}\right\}\end{aligned} \tag{22}$$

Take $t = (1/c)\log(1 + \epsilon c/\sigma^2)$ we will have"

$$\mathbb{P}\left(\bar{X}_n > \epsilon\right) \le \exp\left\{-\frac{n\sigma^2}{c^2} h\left(\frac{c\epsilon}{\sigma^2}\right)\right\} \tag{23}$$

where $h(u) = (1 + u)\log(1 + u) - u$.

By noting that $h(u) \ge u^2/(2 + 2u/3)$ for $u \ge 0$, we just show the Bernstein's Inequality.

A.2 Exercise 2

Prove that for any $z > 0$, if $m \geq 3(z_3)\ln(z+3)$, then $\frac{m}{\ln m} > z$. [**Kutin:2002**]

Solution:

Firstly, we note that

$$\frac{d}{dm}\frac{m}{\ln m} = \frac{\ln m - 1}{\ln^2 m} \tag{24}$$

Thus, the term $\frac{m}{\ln m}$ is increasing when $m > e$.

Then, given that $\ln(z+3) \geq \ln\ln(z+3)$, and $z > 0 \Rightarrow \ln(z+3) > \ln 3$. Hence,

$$\begin{aligned} \frac{m}{\ln m} &\geq \frac{3(z+3)\ln(z+3)}{\ln 3 + \ln(z+3) + \ln\ln(z+3)} \\ &> \frac{3(z+3)\ln(z+3)}{3\ln(z+3)} = z+3 > z \end{aligned} \tag{25}$$