Lec 5

Page 1

CS 59000 – RL

Concentration Inequalities.

Agenda: –

- Hoeffding's inequalities
- Bernstien's inequalities
- Azuma's
- Freedman's

Def: A random variable X on $(\Omega, \mathcal{F}, \mathbb{P})$ is sub–Gaussian if there is a positive number R such that

$$E\left[e^{\lambda(X - E[X])}\right] \leq e^{\frac{R\lambda^2}{2}} \quad \text{for all } \lambda \in \mathbb{R}$$

R is refered to as sub-Gaussian constant.

$$X \to \mathbb{P}\left((|X - E[X]|^k) > \varepsilon\right) \leq \frac{E\left[|X - E(X)|^k\right]}{\varepsilon}$$

in the proof of ~~Chebyshevs~~

Now let's use exponential.

$$\mathbb{P}(X - E[X] \geq \varepsilon) = \mathbb{P}\left(e^{\lambda(X - E(X))} \geq e^{\lambda\varepsilon}\right) \quad \text{for } \lambda > 0$$

$$\leq \frac{E\left[e^{\lambda(X - E(X))}\right]}{e^{\lambda\varepsilon}}$$

if the random variable is sub-Gaussian

$$\leq \frac{e^{\frac{R\lambda^2}{2}}}{e^{\lambda\varepsilon}}$$

2)

$$\mathbb{P}(X - E[X] \geq \varepsilon) \leq e^{\frac{R\lambda^2}{2} - \lambda\varepsilon} \qquad \lambda > 0$$

$$R\lambda = \varepsilon \to \lambda = \frac{\varepsilon}{R}$$

$$\mathbb{P}(X - E[X] \geq \varepsilon) \leq e^{\frac{R\varepsilon^2}{2R^2} - \frac{\varepsilon^2}{R}} = e^{-\frac{\varepsilon^2}{2R}}$$

HW: $\mathbb{P}(X - E[X] \leq -\varepsilon) \leq e^{-\frac{\varepsilon^2}{2R}}$

$$\hookrightarrow \mathbb{P}(|X - E[X]| \geq \varepsilon) \leq 2e^{-\frac{\varepsilon^2}{2R}} \quad (*)$$

Theorem: [Hoeffding's Inequality]
Let $X_1, \ldots, X_n$ be n independent random variables on $(\Omega, \mathcal{F}, \mathbb{P})$, which are sub-Gaussian with constants $R_1 \ldots R_n$
then

$$\mathbb{P}\left(\left|\sum X_i - \sum E[X_i]\right| > \varepsilon\right) \leq 2\exp\left(\frac{-\varepsilon^2}{2\sum_{i=1}^{n} R_i}\right)$$

---

proof: $\sum X_i - \sum E[X_i]$

$$E\left[\exp\left(\lambda\left(\sum_{i=1}^{n} - \sum_{i=1}^{n} E[X_i]\right)\right)\right] =$$

$$= \prod_{i=1}^{n} E\left[\exp(\lambda(X_i - E[X_i]))\right] \leq \prod_{i=1}^{n} \exp\left(\frac{R_i\lambda^2}{2}\right)$$

$$\Rightarrow \quad = \exp\left(\frac{(\sum R_i)\lambda^2}{2}\right)$$

3)

therefore $\sum X_i - \sum E(X_i)$ is a sub-Gaussian random variable with sub-Gaussian constant $\sum_{i=1}^{n} R_i$

using the $\circledast$ → result in Hoeffding statement

Using Hoeffding inequality for $\sum X_i - E(X_i)$ we have

$$\mathbb{P}\left(\left|\sum X_i - \sum E(X_i)\right| \geq \varepsilon\right) \leq \underbrace{2\exp\left(\frac{-\varepsilon^2}{2\sum R_i}\right)}_{=\delta}$$

$$\rightarrow 2\exp\left(\frac{-\varepsilon^2}{2\sum R_i}\right) = \delta \rightarrow \frac{+\varepsilon^2}{2\sum R_i} = \log\left(\frac{2}{\delta}\right)$$

$$\rightarrow \varepsilon = \sqrt{2\left(\sum R_i\right)\log\left(\frac{2}{\delta}\right)}$$

→ There for the following hold with probability at least $(1-\delta)$

$$P\left(\left|\sum X_i - \sum E X_i\right| \leq \sqrt{2\left(\sum R_i\right)\log\left(\frac{2}{\delta}\right)}\right)$$

Note: For i.i.d. setting $R_1 = R_2 = \cdots R_n = R$

$$\left|\frac{\sum X_i}{n} - E[X]\right| \rightarrow \frac{nR}{n^2}$$

$E(X_1) = E(X_2) \cdots E(X_n) = E(X)$

$$\mathbb{P}\left(\left|\frac{\sum X_i}{n} - E(X)\right| \leq \sqrt{\frac{2R}{n}\log\left(\frac{2}{\delta}\right)}\right)$$

is at least $(1-\delta)$

4)

$$\mathbb{P}\left(\left|\frac{\sum X_i}{n} - E(X)\right| \leq \sqrt{\frac{2R}{n}\log\left(\frac{2}{\delta}\right)}\right) \geq 1-\delta$$

$\delta = \frac{1}{5} \rightarrow \log(10)$

↳ $\delta = \frac{1}{5000} \rightarrow 3\log(10)$

↳ $\delta = \frac{1}{5\times 10^{20}} \rightarrow 20\log(10)$

HW: $E\left[\max_{i=1 \ldots N} |X_i|\right] \leq ?$

Maximal Inequality

HW: For mean zero independent random variable $X_1 \ldots X_n$ such that $|X_i| \leq b$ a.s. prove similar result as standard Hoeffding ineq.

Theorem: [Bernstein's Inequality]
Let X, to be a random variable on $(\Omega, \mathcal{F}, \mathbb{P})$ such that $|X - E[X]| \leq b$ a.s.
and $Var(X) \leq R$, then

$$\mathbb{P}(|X - EX| \geq \varepsilon) \leq 2\exp\left(\frac{-\varepsilon^2}{2(\sigma^2 + \frac{b}{3}\varepsilon)}\right)$$

Bandit algorithm

↳ Martin Wainwright, High dimensional statistics

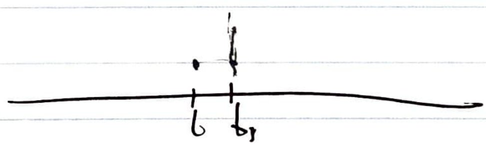
5)

Martingales:

Let $X_1, \ldots$ ..., be a sequence of random variable on $(\Omega, \mathcal{F}, \mathbb{P})$, and $\mathbb{F} = (\mathcal{F}_t)_{t=1}^{n}$ $(\mathcal{F}_i \subseteq \mathcal{F}_{i+1}) \ldots$ a filtration. A $\mathbb{F}$ adapted sequence of random variables $(X_t)_{t \in \mathbb{N}_+}$ is a $\mathbb{F}$-adapted martingale sequence if

a) $E[X_t | \mathcal{F}_{t-1}] = X_{t-1}$ a.s. $\forall t > 1$

b) X_t is integrable.



Example: Gambling in a fair Casino.
You have X_t dollars, and you play a fair game, which mean you win and lose equally likely. Then what is your expection of your pocket at time $t+1$.

$$E[X_{t+1} | \mathcal{F}_t] = X_t$$

b)

efficient.

Considera stock market: you hava X_t \$ and you can not predict future.

$$E\left(X_{t+1} \mid \mathcal{F}_t\right) = X_t$$

Theorem: (Maximal inequality). Let $(X_t)_{t \in \mathbb{N}}$ be a martingal sequene on $(\Omega, \mathcal{F}, \mathbb{P})$ with $X_t \geq 0$. a.s. for all t. Then

$$\mathbb{P}\left(\sup_{t \in \mathbb{N}} X_t > \varepsilon \right) \leq \frac{E[X_0]}{\varepsilon}$$

Def: (Martingal difference)

A sequence of random variable Y_t on $(\Omega, \mathcal{F}, \mathbb{P})$ with $\mathbb{F} = \{\mathcal{F}_i\}_{-\infty}^{\infty}$ a filtration of $\mathcal{F}$, is

a martingale difference if each Y_i is integrable and $E[Y_t \mid \mathcal{F}_{t-1}] = 0$

Note: $Y_t = X_t - X_{t-1}$ where X_t is $\mathcal{F}_t$-adapted martingale is a martingale sequence.

7)

Theorem. (Azuma inequality) Let $(Y_t)_{t=1}^{\infty}$ be a martingale difference on $(\Omega, \mathcal{F}, \mathbb{P})$ adapted to $(\mathcal{F}_t)_{t=1}^{\infty}$ such that the version of $E[\exp(\lambda Y_{t+1} \mid \mathcal{F}_t] \leq \exp\left(\frac{R\lambda^2}{2}\right)$ for all λ. Then

$$\mathbb{P}\left(\left|\sum_{t=1}^{n} Y_t\right| > \varepsilon\right) \leq 2\exp\left(\frac{-\varepsilon^2}{2nR}\right)$$

Theorem [Freedman inequality]. Let $(Y_t)_{t=1}^{\infty}$ be a martingale difference, such that $|Y_t| \leq b$ a.s. and let $V_n = \sum_{t=1}^{n} \mathrm{Var}(Y_t | F_{t-1})$ then

$$\mathbb{P}\left(\left|\sum_{t=1}^{n} Y_t\right| > \varepsilon \text{ and } V_n < \nu\right) \leq 2\exp\left(\frac{-\varepsilon^2}{2\nu + \frac{2\varepsilon b}{3}}\right)$$

8)

Proof of Azuma inequality

We first prove one side of the inequality and then argue about the other side.

statement $\mathbb{P}\left(\sum_{t=1}^{n} Y_t > \varepsilon\right) \leq \exp\left(\frac{-\varepsilon^2}{2nR}\right)$

Using Markov inequality, for $\lambda > 0$ we have

$$\mathbb{P}\left(\sum Y_t \geq \varepsilon\right) = \mathbb{P}\left(e^{\lambda \sum_{t=1}^{n} Y_t} \geq e^{\lambda \varepsilon}\right)$$

$$\leq \frac{E\left[e^{\lambda \sum_{t=1}^{n} Y_t}\right]}{e^{\lambda \varepsilon}}$$

$$\overset{a.s.}{=} \frac{E\left[E\left[e^{\lambda Y_n} e^{\sum_{t=1}^{n-1} Y_t} \middle| \mathcal{F}_{n-1}\right]\right]}{e^{\lambda \varepsilon}}$$

$$\overset{a.s.}{=} E\left[e^{\lambda \sum_{t=1}^{n-1} Y_t} E\left[e^{\lambda Y_n} \middle| \mathcal{F}_{n-1}\right]\right] e^{-\lambda \varepsilon}$$

$$\leq E\left[e^{\lambda \sum_{t=1}^{n-1} Y_t}\right] e^{\frac{R\lambda^2}{2}} e^{-\lambda \varepsilon}$$

$$\vdots$$

$$\leq e^{\frac{nR\lambda^2}{2} - \lambda \varepsilon}$$

Finding $\lambda > 0$ with smallest r.h.s.

$\rightarrow \inf_{\lambda > 0} \frac{nR\lambda^2}{2} - \lambda \varepsilon \xrightarrow{\text{derivative}} nR\lambda - \varepsilon = 0 \rightarrow \lambda = \frac{\varepsilon}{nR}$

$\rightarrow \mathbb{P}\left(\sum Y_t > \varepsilon\right) \leq e^{-\frac{\varepsilon^2}{2nR}}$