Lecture 4

1)

CS 59000 Reinforcement Learning (RL)

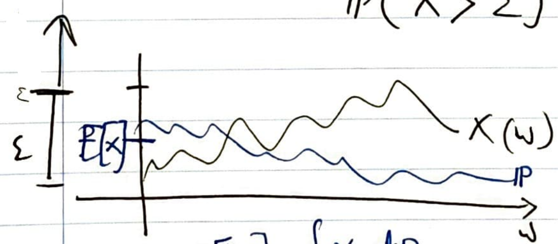
Agenda:

- Markov Inequality
- Hoeffding Inequality
- Bernstein Inequality
- Martingale sequences
- Maximal inequality
- Azuma & Freedman Inequality.

This lecture is about concentration inequalities.

Theorem: [Markov' Inequality]. Consider a random variable $X: \Omega \to \mathbb{R}^+$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, then for $\varepsilon > 0$,

$$\mathbb{P}(X > \varepsilon) \leq \frac{E[X]}{\varepsilon}$$



$$E[X] = \int X \, d\mathbb{P}$$

Proof:

$$E[X] = \int_{\Omega} X \, d\mathbb{P} = \underbrace{\int_{\mathbb{R}^+} X \, dP_X}_{\infty} \to \text{Radon Nikodym derivative}$$

$$= \underbrace{\int_{\mathbb{R}} x \, h(x) \, \underbrace{dx}_{\text{Lebesgue}}}_{\text{change of measure}} = \int_0^{\infty} x \, h(x) \, dx$$

2)

$$= \underbrace{\int_0^\infty x\, h(x)\, I(x<\varepsilon)\, dx}_{\text{Non-negative}} + \int_0^\infty x\, h(x)\, I(x\geq\varepsilon)\, dx$$

$$\geq \int_0^\infty x\, h(x)\, I(x>\varepsilon)\, dx \geq \int_0^\infty \varepsilon\, h(x)\, I(x>\varepsilon)\, dx$$

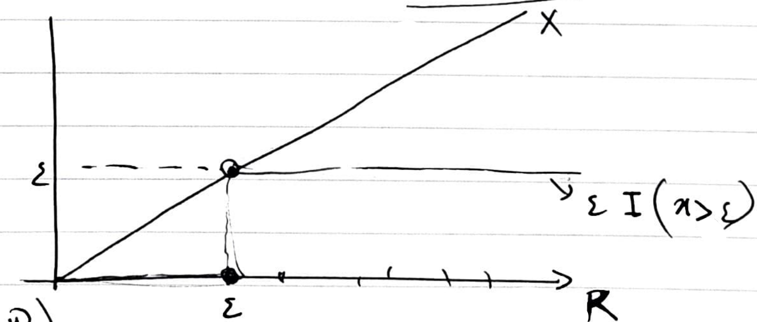
$$= \varepsilon \int_0^\infty h(x)\, I(x>\varepsilon)\, dx$$

$$= \varepsilon\, \mathbb{P}(X>\varepsilon)$$

$$\rightarrow \quad E[X] \geq \varepsilon\, \mathbb{P}(X>\varepsilon)$$

$$\hookrightarrow \quad \mathbb{P}(X>\varepsilon) \leq \frac{E[X]}{\varepsilon}$$

Proof by picture: $E[X] = \int X \, d\mathbb{P}_X$



$(\Omega, \mathcal{F}, \mathbb{P})$

$(R, \mathcal{B}(R), P_X)$

In the class, we mentioned this holds a.s., but it holds always under the definition provided.

$$E[X] = \int X \, d\mathbb{P}_X \geq \int_R \varepsilon\, I(x>\varepsilon)\, dP_X$$

$$= \varepsilon\, \mathbb{P}(X>\varepsilon)$$

3)

Theorem [Chebyshev's Inequality]: consider a random variable $X : \Omega \to \mathbb{R}$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. For any K, and $\varepsilon > 0$, we have:

$$\mathbb{P}\left(|X - E[X]| > \varepsilon\right) \leq \frac{1}{\varepsilon^K} E\left(|X - E[X]|^K\right)$$

for any K, such that $|X - E[X]|^K$ is integrable.

Example: $K = 2 \to \quad \leq \frac{1}{\varepsilon^2} \mathrm{Var}[X]$

Proof. Note that $|X - E[X]|^K$ is a random variable on non-negative real. Using Markov inequality on $|X - E[X]|^K$, we have

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

$$\mathbb{P}\left(|X - E[X]|^K > \varepsilon^K\right) = \mathbb{P}\left(|X - E[X]| > \varepsilon\right)$$

$$\mathbb{P}\left(|X - E[X]|^K > \varepsilon^K\right) = \mathbb{P}\left(E =: \{\omega : |X - E[X]|^K(\omega) > \varepsilon^K\}\right)$$

$$E \in \mathcal{F}$$

$$= \mathbb{P}\left(E =: \{\omega : |X - E[X]|(\omega) > \varepsilon\}\right)$$

$Y = X - E[X]$

$E = \{\omega : |Y(\omega)| > \varepsilon\}$

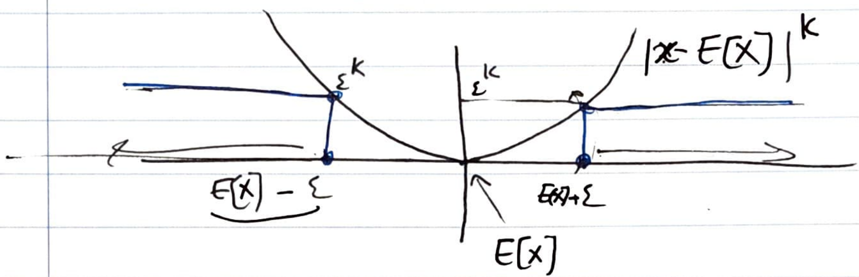
$\omega \in \Omega$

$|Y|^K > \varepsilon^K$

4)

Proof by picture

$$E\left[|X-E[X]|^k\right] = \int |X-E[X]|^k \, dP_X$$



$$\int |X-E[X]|^k \, dP_X \geq \int \varepsilon^k \, I(X - E[X] > \varepsilon) \, dP_X$$

$$+ \int \varepsilon^k \, I(X - E[X] < -\varepsilon) \, dP_X$$

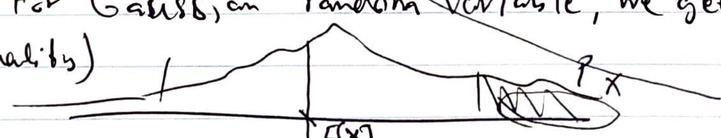
$$= \varepsilon^k \, \mathbb{P}(|X-E[X]| > \varepsilon)$$

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

$$E\left[e^{\lambda(X-E[X])}\right] \leq e^{-R\frac{\lambda^2}{2}}$$

for all $\lambda \in \mathbb{R}$

(Note: For Gaussian random variable, we get equality)



5/

An alternative proof of Markov inequality.

Theorem statement: [Markov Inequality]: Consider a random variable $X: \Omega \to R^+$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, then for $\varepsilon > 0$,

$$\mathbb{P}(X > \varepsilon) \leq \frac{E[X]}{\varepsilon}$$

Proof:
$$E[X] = \int_\Omega X \, d\mathbb{P} = \int_\Omega X \, I(X > \varepsilon) \, dP + \int_\Omega X \, I(X \leq \varepsilon) \, d\mathbb{P}$$

Note that the function $I(X > \varepsilon)$ is measurable.
and $E = \{\omega : I(X > \varepsilon) = 1\} \in \mathcal{F}$.

Therefor, we have:
$$E(X) = \int_\Omega X \, I(X > \varepsilon) \, d\mathbb{P} + \int_\Omega X \, I(X \leq \varepsilon) \, d\mathbb{P}$$

This term is non negative

$$\text{①} \geq \int_\Omega X \, I(X > \varepsilon) \, d\mathbb{P}$$

$$\text{②} \geq \varepsilon \int_\Omega I(X > \varepsilon) \, d\mathbb{P}$$

$$= \varepsilon \, \mathbb{P}(E) = \varepsilon \, \mathbb{P}(X > \varepsilon)$$

Then theorem statement follows.

6)

There was a question in the chat on how we can go from ① to ②. In other words, what are the steps used in showing

$$\int_\Omega X\, I(X>\varepsilon)\, d\mathbb{P} \geq \varepsilon \int_\Omega I(X>\varepsilon)\, d\mathbb{P}.$$

proof:

$$\int_\Omega X\, I(X>\varepsilon)\, d\mathbb{P}$$

$$= \sup_h \left\{ \int h\, d\mathbb{P} : h \text{ is simple } 0 \leq h \leq X \right\}$$

$$\geq \varepsilon \int_\Omega I(X>\varepsilon)\, dP$$

The last inequality hold since $\varepsilon I(X>\varepsilon)$ is a simple funtion where $\varepsilon I(X>\varepsilon) < X\, I(X>\varepsilon)$ there on of the Candidate in the supremum.