Week 6 – Lecture 28

$$\Pr\left\{\hat{\mu} \geq \mu + \varepsilon\right\} \lesssim \sqrt{\frac{\sigma^2}{2\pi n \varepsilon^2}} \exp\left\{\frac{-n\varepsilon^2}{2\sigma^2}\right\}$$

For sufficiently large n
[using CLT]

$$\Pr\left\{\hat{\mu} \geq \mu + \varepsilon\right\} \leq \frac{\sigma^2}{n\varepsilon^2}$$ [using Chebyshev]

SubGaussian Random Variables

A r.v. X is σ-subgaussian if $\forall \lambda \in \mathbb{R}$ it holds that

$$\mathbb{E}\left[e^{\lambda X}\right] \leq \exp\left(\lambda^2\sigma^2/2\right) \quad \text{——— (1)}$$

Let $X \sim N(\mu, \sigma^2)$

$$\mathbb{E}\left[e^{\lambda X}\right] = \exp\left\{\mu\lambda + \lambda^2\sigma^2/2\right\}$$

If $\mu = 0$, then:

$$\mathbb{E}\left[e^{\lambda X}\right] = \exp\left\{\lambda^2\sigma^2/2\right\}$$

Thus $N(0, \sigma^2)$ is σ-subgaussian

Also,
$$\log \mathbb{E}\left[e^{\lambda X}\right] \leq \lambda \sigma^2/2$$

If $X \sim \exp(\mu)$
$$\mathbb{E}\left[e^{\lambda X}\right] = \mu \int_0^\infty e^{\lambda x} e^{-\mu x}\, dx$$
$$= \left(\frac{\mu}{\lambda - \mu}\right) e^{(\lambda-\mu)x} \Big|_0^\infty$$

If $\lambda > \mu$ then this quantity $\sim \infty$

If $\lambda < \mu$ then this quantity $= 0$

$\therefore$ $X \sim \exp(\mu)$ is <u>not</u> σ- subgaussian r.v.

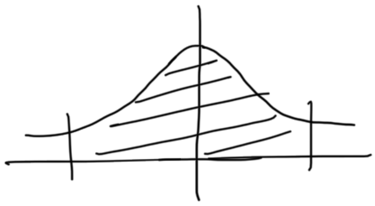
R.V. X is called heavy tailed if
$$\log \mathbb{E}\left[e^{\lambda X}\right] = \infty \quad \forall \lambda > 0$$
otherwise it is called <u>light - tailed</u>.

Focus will be on <u>light-tailed r.v.</u>

$X \sim \mathcal{N}(0, \sigma^2)$ is σ- subgaussian r.v.

support of this is $-\infty$ to $+\infty$ (unbounded support)

Should be <u>light -tailed</u> which means :



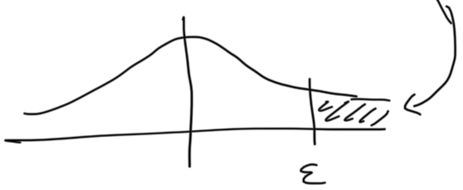
As we go towards the tails the probability reduces, probability content is less in the tails

Thus subgaussian r.v. can have unbounded support but should be light-tailed.

Theorem: If X is σ-subgaussian

then $\forall\ \varepsilon \geq 0$

$$\Pr\{X \geq \varepsilon\} \leq \exp\{-\varepsilon^2/2\sigma^2\} \qquad (2)$$



This probability is upper bounded

Proof:

$$\Pr\{X \geq \varepsilon\} = \Pr\{e^X \geq e^\varepsilon\}$$

X can be +ve or −ve

e^X is +ve r.v.

$$= \Pr\{e^{\lambda X} \geq e^{\lambda \varepsilon}\}$$

$$\leq \frac{\mathbb{E}[e^{\lambda X}]}{e^{\lambda \varepsilon}} \qquad \text{[Markov inequality]}$$

Now if X is σ-subgaussian, then

$$\leq \exp\left\{ \lambda^2 \sigma^2/2 - \lambda \varepsilon \right\}$$ is true for any $\lambda > 0$

[from eq 1
defn of subgaussian]

Optimizing for λ and plugging in the value

$$\leq \exp\left\{ -\varepsilon^2/2\sigma^2 \right\}$$ [optimal value of $\lambda = \varepsilon/\sigma^2$]

Hence proved

Theorem:

$$\Pr\left\{ X \leq -\varepsilon \right\} \leq \exp\left\{ -\varepsilon^2/2\sigma^2 \right\} \quad \text{——} \quad 2a$$

$\therefore$

$$\Pr\left\{ |X| \geq \varepsilon \right\} = \Pr\left\{ X \geq \varepsilon \right\} + \Pr\left\{ X \leq -\varepsilon \right\}$$

$$\Pr\{|X| \geq \varepsilon\} = 2\exp\left\{ -\varepsilon^2/2\sigma^2 \right\} \quad \text{——} \quad 3$$

Let $\varepsilon = \sqrt{2\sigma^2 \log 1/\delta}$, then:

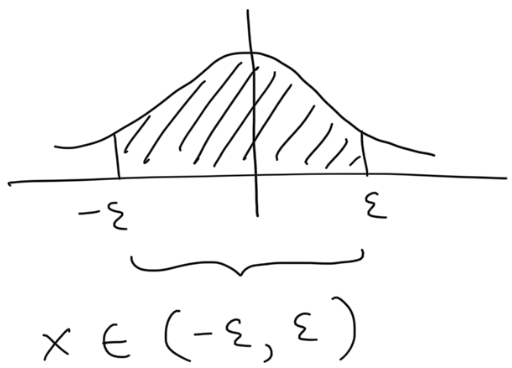
$$\Pr\left\{ X \geq \sqrt{2\sigma^2 \log 1/\delta} \right\} \leq \exp\left\{ \frac{-2\sigma^2 \log 1/\delta}{2\sigma^2} \right\}$$

$\leq \delta$

$\therefore$ For any given δ

$$x \in \left(-\sqrt{2\sigma^2 \log 1/\delta}\,,\ \sqrt{2\sigma^2 \log 1/\delta}\right)$$

will happen with probability $\underline{1 - 2\delta}$



For small δ, this probability will be large.

For small δ, value of $\varepsilon = \sqrt{2\sigma^2 \log 1/\delta}$ will be bigger and hence the interval $-\varepsilon, +\varepsilon$ will be larger, will be far apart, hence makes sense that probability will be large.

We are finally interested in knowing about the estimators.

So if r.v. are subgaussian then what will be:

$$\left[\Pr\{\hat{\mu} \geq \mu + \varepsilon\} \quad \& \text{ where } \hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i\right]$$

<u>Properties of subgaussian r.v. :</u>

Lemma: Suppose X is σ-subgaussian and X_1 & X_2 are independent and σ_1 and σ_2-subgaussian respectively, then:

1) $\mathbb{E}[X] = 0$, $Var[X] \leq \sigma^2$

2) cX is $|c|\sigma$-subgaussian $\forall\, c \in \mathbb{R}$

3) $X_1 + X_2$ is $\sqrt{\sigma_1^2 + \sigma_2^2}$-subgaussian

---

So now coming back to

$$Pr\{\hat{\mu} \geq \mu + \varepsilon\} \quad \text{where } \hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$\hat{\mu} - \mu = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)$$

Theorem: Assume $X_i - \mu$ are independent and σ-subgaussian r.v.s Then $\forall\, \varepsilon \geq 0$

$$Pr\{\hat{\mu} \geq \mu + \varepsilon\} \leq \exp\{-n\varepsilon^2/2\sigma^2\}$$

$$Pr\{\hat{\mu} \leq \mu - \varepsilon\} \leq \exp\{-n\varepsilon^2/2\sigma^2\}$$

$\therefore$ $Pr\{\hat{\mu} \geq \mu + \varepsilon\}$

$$= \Pr\left\{ \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu) \geq \varepsilon \right\}$$

$$= \Pr\left\{ \sum_{i=1}^{n} \frac{(x_i - \mu)}{n} \geq \varepsilon \right\}$$

$\therefore$ $x_i - \mu \sim \frac{\sigma}{n}$ subgaussian

We are hence taking sum of n r.v.s. with $\frac{\sigma}{n}$ - subgaussian

$$\therefore \quad \sum_{i=1}^{n} \frac{(x_i - \mu)}{n} \sim \sqrt{\sum_{i=1}^{n} (\sigma/n)^2} = \sigma/\sqrt{n}$$

$\therefore$ $\sum_{i=1}^{n} \frac{(x_i - \mu)}{n}$ is $\sigma/\sqrt{n}$ - subgaussian

Putting this into the defn of subgaussian r.v. we get

$$\Pr\{\hat{\mu} \geq \mu + \varepsilon\} \leq \exp\left\{ \frac{-\varepsilon^2}{2(\sigma/\sqrt{n})^2} \right\}$$

$$\Pr\{\hat{\mu} \geq \mu + \varepsilon\} \leq \exp\left\{ \frac{-n\varepsilon^2}{2\sigma^2} \right\}$$

Now we have a bound which is exponentially decaying in n, and this holds for <u>any n</u>.

When we had proved with CLT n had to be sufficiently large, but with subgaussian assumption sufficiently ... the bound holds for <u>any</u> n.

$$\Pr\{\hat{\mu} \geq \mu + \varepsilon\} \leq \frac{\sigma^2}{n\varepsilon^2} \quad \text{[Chebyshev]}$$

$$\leq \exp\left\{-\frac{\varepsilon^2 n}{2\sigma^2}\right\} \quad \text{[Subgaussian]}$$

this is tighter that Chebyshev

Also for $\forall\, x > 0$, $e^{-x} < \frac{1}{e' x}$

$$\therefore \Pr\{\hat{\mu} \geq \mu + \varepsilon\} \leq \frac{2}{e}\left(\frac{\sigma^2}{\varepsilon^2 n}\right)$$

$$\leq \frac{\sigma^2}{\varepsilon^2 n}$$