18.656 Notes

Lecturer:

Andrew Liu

Spring 2024

Last updated on Saturday 10^{th} February, 2024.

Contents

1 February 8, 2024

1.1 Recap

Definition 1.1

A random variable X with mean $\mu = \mathbb{E}[X]$ is **σ-sub-Gaussian** if $\mathbb{E}[e^{\lambda(X-\mu)}] \leq e^{\lambda^2\sigma^2/2}$ for all $\lambda \in \mathbb{R}$.
Equality if and only if $X \sim \mathcal{N}(\mu, \sigma^2)$.

Example 1.2

Given r.v. $X \in [a,b]$, $\mathbb{E}[X] = \mu$. Then, X is sub-Gaussian with $\sigma = (b-a)$.

It turns out that we can also show $\sigma = (b-a)/2$, but we won't show this during lecture, since the technique used to show the weaker result is more interesting.

Proof. Let $\tilde{X}$ be i.i.d. to X. Then,

$$\begin{aligned}\mathbb{E}_X[e^{\lambda(X-\mu)}] &= \mathbb{E}_X[e^{\lambda(X-\mathbb{E}_{\tilde{X}}[\tilde{X}])}] \\ &\leq \mathbb{E}_{X,\tilde{X}}[e^{\lambda(X-\tilde{X})}],\end{aligned}$$

by Jensen's inequality. Since $X, \tilde{X}$ are i.i.d., $(X-\tilde{X})$ has a distribution symmetric around 0. Now, we also have that

$$(X-\tilde{X}) \overset{\text{dist}}{=} \varepsilon(X-\tilde{X}),$$

where $\varepsilon \in \{\pm 1\}$ with equal probability (also called a **Rademacher** random variable). Therefore,

$$\begin{aligned}\mathbb{E}_{\tilde{X}}[e^{\lambda(X-\mu)}] &\leq \mathbb{E}_{\tilde{X},X}[\mathbb{E}_\varepsilon[e^{\lambda\varepsilon(X-\tilde{X})}]] \\ &\leq \mathbb{E}_{X,\tilde{X}}[e^{\lambda^2(X-\tilde{X})^2/2}],\end{aligned}$$

since ε is 1-sub-gaussian. Finally, since X is bounded,

$$\mathbb{E}_{X,\tilde{X}}[e^{\lambda^2(X-\tilde{X})^2/2}] \leq e^{\lambda^2(b-a)^2/2}.$$

□

1.2 More on sub-Gaussians

Definition 1.3 (Addition property of Gaussians)
Given $X_1 \sim \mathcal{N}(0, \sigma_1^2)$ and $X_2 \sim \mathcal{N}(0, \sigma_2^2)$,

$$X_1 + X_2 \sim \mathcal{N}(0, \sigma_1^2 + \sigma_2^2).$$

Claim 1.4 (Addition property of sub-Gaussians)
Given $X_i \sim \sigma_i$-sub-Gaussian, $i \in \{1, 2\}$, then $X_1 + X_2$ is $\sqrt{\sigma_1^2 + \sigma_2^2}$-sub-Gaussian.

Proof.

$$\begin{aligned}
\mathbb{E}_{X_1,X_2}[e^{\lambda(X_1+X_2)}] &= \mathbb{E}_{X_1,X_2}[e^{\lambda X_1} e^{\lambda X_2}] \\
&= \mathbb{E}_{X_1}[e^{\lambda X_1}]\mathbb{E}_{X_2}[e^{\lambda X_2}] \\
&\leq e^{\lambda^2\sigma_1^2/2} e^{\lambda^2\sigma_2^2/2} = e^{\lambda^2(\sigma_1^2+\sigma_2^2)/2}.
\end{aligned}$$

□

A consequence of this fact is that given $X_i \sim \sigma$-sub-Gaussian, i.i.d. with zero-mean, then

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i \sim \sigma\text{-sub-Gaussian},$$

or equivalently

$$\frac{1}{n} \sum_{i=1}^{n} X_i \sim \frac{\sigma}{\sqrt{n}}\text{-sub-Gaussian}.$$

Example 1.5 (Survey sampling)
Two candidates for an election A and B.

Sample people $i = 1, \ldots, n$, giving responses $X_i = 1$ if A or 0 if B. Let μ^* be the actual fraction of people who will vote A. Let $\hat{\mu}$ be our estimator for μ^*:

$$\hat{\mu} = \sum_{i=1}^{n} X_i.$$

We can construct a confidence interval $\hat{\mathcal{I}}$ for our estimator, and we would like to know at what point we can say

$$\mathbb{P}[\hat{\mathcal{I}} \ni \mu^*] \geq 1 - \delta.$$

For example, with $\delta = 0.02$, interval width of 0.03, we require $n \approx 10000$ to make this guarantee.

We can model $X_i \sim \textsc{Bern}(\mu^*)$. Since $X_i \in [0,1]$, our earlier result shows that X_i is $1/2$-sub-Gaussian. Using additivity and i.i.d., our sample mean $\hat{\mu}$ is $1/(2\sqrt{n})$-sub-Gaussian. Thus,

$$\mathbb{E}[e^{\lambda(\hat{\mu}-\mu^*)}] \leq e^{\lambda^2/2 \cdot 1/(4n)} = e^{\lambda^2/(8n)}.$$

Using Chernoff,

$$\mathbb{P}[|\hat{\mu} - \mu^*| \geq s] \leq 2e^{2ns^2},$$

for some $s > 0$. So for some fixed δ, we can make a guarantee about interval width $s = \sqrt{\log(2/\delta)/(2n)}$.

Theorem 1.6 (Hoeffding bound)
Let $X_i \sim \sigma_i$-sub-Gaussian, with $\mathbb{E}[X_i] = \mu_i$ and all independent. Then

$$\mathbb{P}\left[\sum_{i=1}^{n}(X_i - \mu_i) \geq t\right] \leq e^{-t^2/(2\sum_{i=1}^{n}\sigma_i^2)}.$$

Example 1.7
$X \sim \mathcal{N}(0,1)$ is 1-sub-Gaussian. Let $Y = X^2$ is not sub-Gaussian.

$$\mathbb{E}[e^{\lambda X^2}] = \frac{1}{\sqrt{1-\lambda}},$$

$\lambda \in (0,1)$.

Despite not being sub-Gaussian, it is close. Consider an n-dimensional Gaussian, $X = (X_1, \ldots, X_n)$, where each $X_i \sim \mathcal{N}(0,1)$. Then $\mathbb{E}[\|X\|_2^2/n] = \mathbb{E}[\sum X_i^2/n] = 1$, and we can further show that

$$\mathbb{P}\left[\left|\frac{\|X\|^2}{n^2} - 1\right| \geq \delta\right] \leq 2e^{-cn\delta^2},$$

for all $\delta \in (0,1)$. This looks very similar to the sub-gaussian tail bound from earlier, but will only hold for small delta. For larger delta, the tail bound becomes linear in δ (because X is not truly sub-Gaussian).