Chapter 1

Concentration Inequalities

- Chernoff Bound
- Sub-Gaussian RV (Azuma-Hoeffding)
- Sub-Exponential RV (Bernstein)
- Johnson-Lindenstrauss
- Lipschitz functions of random variables
  - McDiarmid
  - Gaussian Concentration
- Robust Mean Estimation

$\boxed{\text{Lecture 3:}}$ Chernoff Bound & SubGaussian RV

Let $X_1 \ldots X_n$ be r.v.'s with $\mathbb{E}X_i = 0$.

<u>Q:</u> How big is $\left|\sum_{i=1}^{n} X_i\right|$ typically?

In general, can be $O(n)$.

But if $X_1, \ldots, X_n$ are pairwise independent, then Chebychev gives:

$$P\left(\left|\sum_{i=1}^{n} X_i\right| \geq t\right) \leq \frac{\sum_{i=1}^{n} Var(X_i)}{t^2}$$

So $P\left(\left|\sum_{i=1}^{n} X_i\right| \geq \lambda \sqrt{\sum_{i=1}^{n} Var(X_i)}\right) \leq \frac{1}{\lambda^2}$

<u>Conclusion:</u> If $Var(X_i) \leq \sigma^2$, then

$$P\left(\left|\sum_{i=1}^{n} X_i\right| \leq 2\sigma\sqrt{n}\right) \geq \frac{3}{4}$$

<u>Q:</u> When can we expect to replace $\frac{1}{\lambda^2}$ by $e^{-\lambda}$ or $e^{-\lambda^2}$?

<u>Motivating example:</u>

$$P[\sup_{i\in I} X_i \geq t] \leq \sum_{i\in I} P[X_i \geq t]$$

If $|I|$ is <u>huge</u>, need $P[X_i \geq t]$ to be small.

<u>Chernoff Method:</u>

Let X be r.v. with $\mu = \mathbb{E}X < \infty$. Then for all $\lambda \geq 0$, we have

$$P[X-\mu \geq t] = \mathbb{P}[e^{\lambda(X-\mu)} \geq e^{\lambda t}] \overset{\text{Markov}}{\leq} \frac{\mathbb{E}e^{\lambda(X-\mu)}}{e^{\lambda t}}$$

$$\Rightarrow \log \mathbb{P}[X-\mu \geq t] \leq \inf_{\lambda\geq 0}\{\log \mathbb{E}e^{\lambda(X-\mu)} - \lambda t\}$$

$$= -\sup_{\lambda \geq 0}\{\lambda t - \log \mathbb{E}e^{\lambda(X-\mu)}\}$$

Define for any function $\psi : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$, the <u>Fenchel conjugate</u>

$$\psi^*(t) = \sup_\lambda \{ t\lambda - \psi(\lambda) \}$$

Let's look at the main example

$$\psi_X(\lambda) = \log \mathbb{E} e^{\lambda(X-\mu)}$$

For all $\lambda \in \mathbb{R}$, observe

$$\psi_X(\lambda) = \log \mathbb{E} e^{\lambda(X-\mu)} \geq \mathbb{E} \log e^{\lambda(X-\mu)} = 0$$

So when $\lambda < 0$ and $t > 0$, we have

$$\lambda t - \psi(\lambda) \leq 0 = 0 \cdot t - \psi(0).$$

Therefore for $t \geq 0$, equality holds:

$$\psi_X^*(t) = \sup_{\lambda \geq 0} \{ t\lambda - \psi(\lambda) \}$$

We therefore arrive at the bound:

Chernoff Bound:

$$\mathbb{P}[X-\mu \geq t] \leq \exp(-\psi_X^*(t))$$

where $\psi_X(\lambda) := \log(\mathbb{E} e^{\lambda(X-\mu)})$.

Ex: Let $X \sim N(\mu, \sigma^2)$. Then

$$\mathbb{E} e^{\lambda(X-\mu)} = e^{\frac{\sigma^2\lambda^2}{2}} \quad \forall_{\lambda \in \mathbb{R}}.$$

Then $\psi_X^*(t) = \sup_\lambda \lambda t - \frac{\sigma^2\lambda^2}{2} = \frac{t^2}{2\sigma^2}$

So $\mathbb{P}[X \geq \mu + t] \leq e^{-\frac{t^2}{2\sigma^2}} \quad \forall t > 0$

Defn: X with mean $\mu = \mathbb{E}X$ is sub-Gaussian with parameter $\sigma > 0$ if

$$\mathbb{E} e^{\lambda(X-\mu)} \leq e^{\frac{\sigma^2\lambda^2}{2}} \quad \forall_{\lambda \in \mathbb{R}}$$

If X is sub-Gaussian, so is $-X$.

$$\Rightarrow \mathbb{P}[|X-\mu| \geq t\sigma] \leq 2e^{-\frac{t^2}{2}}$$

Lemma: (Bounded RV)

Suppose X is supported on $[a,b]$. Then X is $\frac{b-a}{2}$ sub-Gaussian.

pf: Set $y = X - \mu$ and define

$$f(\lambda) = \log(\mathbb{E}\exp(\lambda y))$$

Then
$$f'(\lambda) = \frac{\mathbb{E}\, y\exp(\lambda y)}{\mathbb{E}\exp(\lambda y)}$$

$$f''(\lambda) = \frac{\mathbb{E}\, y^2\exp(\lambda y)}{\mathbb{E}\exp(\lambda y)} - \left[\frac{\mathbb{E}\, y\exp(\lambda y)}{\mathbb{E}\exp(\lambda y)}\right]^2$$

Define measure $dm = \frac{\exp(\lambda y)\,dy}{\mathbb{E}\exp(\lambda y)}$

Then
$$f''(\lambda) = \mathrm{Var}_m(y)$$
$$= \inf_{w \in \mathbb{R}} \mathbb{E}_m\left[(y-w)^2\right]$$
$$\leq \mathbb{E}_m\left(y - \frac{a+b}{2} + \mu\right)^2 \leq \frac{(b-a)^2}{4}$$

Taylor theorem $\Rightarrow f(\lambda) \leq f(0) + f'(0)\lambda + \frac{(b-a)^2}{8}\lambda^2$ ∎