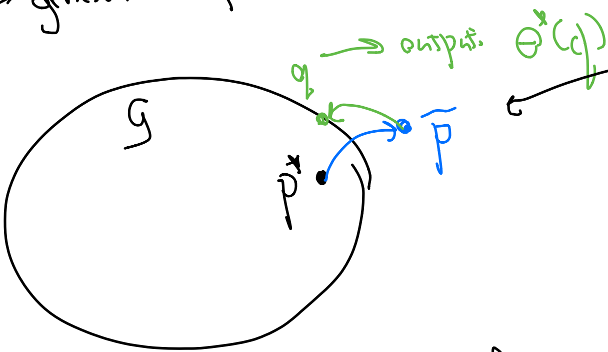
Lecture 3. Concentration inequalities

Recap last time,

- Minimum distance functionals $\quad MD(\tilde{p}) = \theta^*(q)$, where q is
  - ↳ general recipe for robust estimators



- Error: governed by modulus of continuity:

$$\mathfrak{m}(\mathcal{G}, 2\varepsilon) \triangleq \sup_{\substack{p, q \in \mathcal{G} \\ D(p,q) \leq 2\varepsilon}} L(p; \theta^*(q))$$

- → Gaussians: $\mathfrak{m}(\mathcal{G}, \varepsilon) = O(\varepsilon)$
- → bounded covariance: $\mathfrak{m}(\mathcal{G}, \varepsilon) = O(\sigma\sqrt{\varepsilon})$
- → resilient distributions

Resilience: $D = TV$, $L(p; \theta) = \|\mu(p) - \theta\|_2$

Def$^{\underline{n}}$ p is (ρ, ε)-resilient if

$$\|\mu(p) - \mu(r)\|_2 \leq \rho$$

showed this holds for $\rho = O(\sigma\sqrt{\varepsilon})$

whenever $r \leq \frac{p}{1-\varepsilon}$, i.e. r is an ε-deletion of p.

- $*$ r can be obtained by conditioning on an event of prob. $1-\varepsilon$
- $*$ $r(S) \leq \frac{p(S)}{1-\varepsilon}$ $\forall$ measurable sets S
- $*$ r is obtained by deleting ε-fraction of prob. mass and re-normalizing

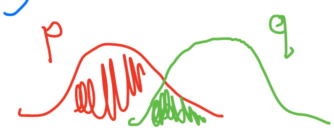
---

Lemma. If $TV(p,q) \le \epsilon$, then $\exists\, r$ such that $r \le \frac{p}{1-\epsilon}$, $r \le \frac{q}{1-\epsilon}$.

Pf. Assume p, q both have densities.

Then take $r(x) = \frac{\min(p(x), q(x))}{1 - TV(p,q)}$ $\Rightarrow$ $r(x) \le \frac{p(x)}{1-\epsilon}$, $r(x) \le \frac{q(x)}{1-\epsilon}$

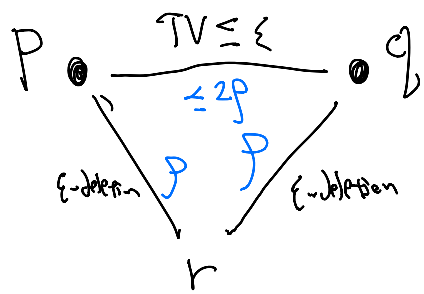
$$\Rightarrow \int \min(p(x), q(x))\, dx = 1 - TV(p,q)$$



Prop. Let $\mathcal{G}^{TV}(\rho, \epsilon)$ be the set of (ρ, ϵ)-resilient distributions.

Then $\mathfrak{m}(\mathcal{G}^{TV}(\rho, \epsilon), \epsilon) \le 2\rho$.

Pf.



$\|\mu(p) - \mu(r)\|_2 \le \rho$

$\|\mu(q) - \mu(r)\|_2 \le \rho$

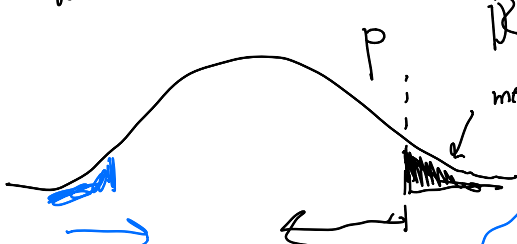
$\Rightarrow \|\mu(p) - \mu(q)\|_2 \le 2\rho$

$\Rightarrow \mathfrak{m}(\mathcal{G}, \epsilon) \le 2\rho$. ∎

---

bounded covariance $\Rightarrow$ resilient $O(\sigma\sqrt{\epsilon})$

any tail bound $\Rightarrow$ resilience

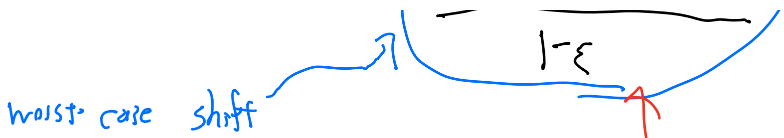


$\mathbb{R}$

mean of tail

τ_ϵ: ϵ-quantile

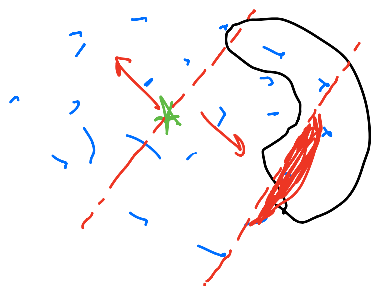
$\epsilon \cdot \mathbb{E}[X \mid X \ge \tau_\epsilon]$



$\mathbb{E}[X \mid X \geq \tau_\varepsilon]$:

- bounded variance: $\leq \sigma/\sqrt{\varepsilon} \Rightarrow \sigma\sqrt{\varepsilon}$
- sub-Gaussian: $\leq \sigma\varepsilon\sqrt{\log(1/\varepsilon)} \Rightarrow \sigma\sqrt{\log(1/\varepsilon)}$
- bounded k^{th} moments: $\leq \sigma/\varepsilon^{1/k} \Rightarrow \sigma\cdot\varepsilon^{1-1/k}$

also hold in higher dimensions



<u>Concentration inequalities</u>: Tools for getting finite-sample bounds

∞-data alg. $\Rightarrow$ show that behavior in finite samples is similar to ∞-sample behavior

"concentrates"

- Markov's inequality, Chebyshev's inequality, Chernoff bound, union bound
- Rosenthal's inequality

Two-step strategy:

1. Show something good happens for 1 random variable
2. Show good "composition" across independent random variables

---

<u>Goal</u>: Show that samples from a r.v. are close to mean

<u>Ex.</u> Suppose that there's a slot machine with expected

payout of \$0.

Q.) What's the maximum probability that you win \$100?

A.) 5%

<u>Thm.</u> (Markov) If r.v. $X \geq 0$, then $\mathbb{P}[X \geq t] \leq \frac{\mathbb{E}[X]}{t}$.

- Pretty weak
- Doesn't compose $(X_1 + X_2)$

<u>Thm</u> (Chebyshev's inequality) If $\mathbb{E}[X] = \mu$ and $\mathbb{E}[(X-\mu)^2] = \sigma^2$, then $\mathbb{P}[|X-\mu| \geq t \cdot \sigma] \leq \frac{1}{t^2}$.

<u>Pf.</u> Apply Markov to $Z = (X-\mu)^2$

- $Z \geq 0$
- $\mathbb{P}[Z \geq \underbrace{t^2 \sigma^2}_{\sigma^2)n}] \leq \frac{\mathbb{E}[Z]}{t^2 \cdot \sigma^2} = \frac{1}{t^2}$

<u>Composition.</u> If X_1 and X_2 are independent,

$$Var[X_1 + X_2] = Var[X_1] + Var[X_2] \leftarrow$$

$$Var[X_1 + \dots + X_n] = Var[X_1] + \dots + Var[X_n]$$

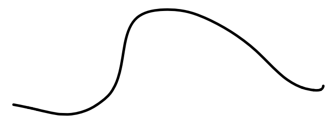
$X_1, \dots, X_n \sim p$ w/ variance σ^2

$$S = \frac{1}{n}(X_1 + \dots + X_n)$$

$$Var[S] = \frac{1}{n^2} \cdot Var[X_1 + \dots + X_n] = \frac{1}{n^2}(Var[X_1] + \dots + Var[X_n])$$

$$= \frac{1}{n^2} \cdot (n \cdot \sigma^2) = \frac{\sigma^2}{n}.$$

Markov, Chebyshev $\frac{1}{t^2}$



$\exp(-t^2)$

moment generating function

X, $m_X(\lambda) = \mathbb{E}[\exp(\lambda X)]$ $\qquad \mathbb{E}[\exp(\lambda(X-\mu))]$

Compose: $m_{X_1+X_2}(\lambda) = m_{X_1}(\lambda)\, m_{X_2}(\lambda)$

Chernoff bound. $\mathbb{P}[X-\mu \geq t] \leq \inf_\lambda m_X(\lambda) e^{-\lambda t}$ $\quad (\star)$

$$Z = \exp(\lambda(X-\mu))$$

$$\mathbb{P}[X-\mu \geq t] = \mathbb{P}[Z \geq e^{\lambda t}] \leq \frac{\mathbb{E}[Z]}{e^{\lambda t}} = m_X(\lambda) e^{-\lambda t}.$$

Ex. sub-Gaussian

$$\mathbb{E}[\exp(\lambda(X-\mu))] \leq \exp\left(\frac{\sigma^2\lambda^2}{2}\right).$$

(sub-Gaussian w/ parameter σ)

e.g. Gaussian

bounded: $X \in [-M, M] \Rightarrow$ sub-Gaussian w/ parameter $\sigma \leq C \cdot M$

$$\lambda = \frac{t}{\sigma^2}$$

$$\inf_\lambda m_X(\lambda) e^{-\lambda t} \leq \inf_\lambda \exp\left(\frac{\sigma^2\lambda^2}{2} - \lambda t\right) = \exp\left(-\frac{t^2}{2\sigma^2}\right)$$

Gaussian decay

Composition. If $X_1, \ldots, X_n$ are σ-sub-Gaussian

$\Rightarrow \frac{1}{n}(X_1 + \cdots + X_n) \to \left(\frac{\sigma}{\sqrt{n}}\right)$-sub-Gaussian

s.g. def^n: $m_X(\lambda) \leq \exp\left(\frac{\sigma^2\lambda^2}{2}\right)$

$$\Rightarrow m_{X_1+\cdots+X_n}(\lambda) \leq \exp\left(\frac{\sigma^2\lambda^2}{2}\right)\exp\left(\frac{\sigma^2\lambda^2}{2}\right)\cdots\exp\left(\frac{\sigma^2\lambda^2}{2}\right)$$

$$= \exp\left(\frac{n\sigma^2\lambda^2}{2}\right)$$

$$m_{\frac{X_1+\dots+X_n}{n}}(\lambda) \le \exp\left(\frac{1}{n^2}\cdot\left(\frac{n\sigma^2\lambda^2}{2}\right)\right) = \exp\left(\frac{\sigma^2\lambda^2}{2n}\right)$$

Recap. Markov, Chebyshev, Chernoff (sub-Gaussian)

→ Show r.v. near its mean

→ Chebyshev, Chernoff ⟹ avg. within $1/\sqrt{n}$ of its mean

$1/t^2$ vs. $\exp(-t^2)$

↳ sub-exponential $\exp(-t)$

$$\mathbb{E}[\exp(\lambda(X-\mu))] \le \exp\left(\frac{\sigma_A^2\lambda^2}{2}\right).$$

$\sigma_A \le 10\cdot\sigma_B$

$\sigma_B \le 10\cdot\sigma_A$

$$\mathbb{E}\left[\exp\left(\frac{(X-\mu)^2}{\sigma_B^2}\right)\right] \le 2$$

Gaussian tails of smaller

Why not do Chebyshev with $\mathbb{E}[(X-\mu)^4]$ $\quad \frac{1}{t^2}$

$$\mathbb{E}[(X-\mu)^4] = K^4$$

$$\Rightarrow \mathbb{P}[|X-\mu| \ge t\cdot K] \le \frac{1}{t^4}$$

But doesn't compose: $\mathbb{E}[(X_1+X_2-\mu_1-\mu_2)^4] \ne \mathbb{E}[(X_1-\mu_1)^4] + \mathbb{E}[(X_2-\mu_2)^4]$

Suppose X_1, X_2 have mean 0

$$\mathbb{E}[(X_1+X_2)^4] = \mathbb{E}[X_1^4] + \mathbb{E}[X_2^4] + 4\,\cancel{\mathbb{E}[X_1]}\,\mathbb{E}[X_2^3] + \underbrace{6\,\mathbb{E}[X_1^2]\,\mathbb{E}[X_2^2]}_{\ne 0}$$