Lemma 1: w.p. $1-\delta/2$:

distribution dependent

$$\boxed{\sup_{f\in\mathcal{F}} \mathbb{E}_S f(z) - \mathbb{E}_D f(z)} \le \mathbb{E}\left[\sup_{f\in\mathcal{F}} \mathbb{E}_S f(z) - \mathbb{E}_D f(z)\right] + \sqrt{\frac{\ln 4/\delta}{2n}}$$

data dependent

Lemma 2:

$$\mathbb{E}\left[\underbrace{\sup_{f\in\mathcal{F}}}_{\text{infinite collection of r.v.'s}} \mathbb{E}_S f(z) - \mathbb{E}_D f(z)\right] \le 2\,\mathrm{Rad}_n(\mathcal{F})$$

where $\mathrm{Rad}_n(\mathcal{F}) = \mathbb{E}_{S\sim D^n} \mathrm{Rad}_S(\mathcal{F})$ $\big(f(z_1)\ldots f(z_n)\big)$

$$\mathrm{Rad}_S(\mathcal{F}) = \frac{1}{n}\mathbb{E}_{\sigma\sim U(\pm 1)^n} \sup_{f\in\mathcal{F}} \sum_{i=1}^{n} f(z_i)\sigma_i$$

finite collection of r.v.'s

$X \to \{0,1\}$

Lemma 3: controlling Rademacher complexity w/ growth fn.

for any set S of size n,

$$\mathrm{Rad}_S(\mathcal{F}) \le \sqrt{\frac{2\ln(S(\mathcal{F},n))}{n}}.$$

McDiarmid's Lemma:

g is c-sensitive. $x_1 \dots x_n$ are i.i.d. D supported on V. Then, w.p. $1-\delta'$,

$$\left| g(x_1 \dots x_n) - \mathbb{E}\, g(x_1 \dots x_n) \right| \le c \cdot \sqrt{\frac{n}{2} \cdot \ln \frac{2}{\delta'}}, \quad \ll c \cdot n$$

Pf of Lemma 1

$S = \{x_1 \dots x_n\}$

$$g(x_1 \dots x_n) = \sup_{f \in F} \Big(\underbrace{\underbrace{\mathbb{E}_S f(z)}_{\frac{1}{n}\sum_{i=1}^{n} f(x_i)} - \mathbb{E}_D f(z)}_{F(f)} \Big)$$

g is c-sensitive with $c = \frac{1}{n}$.

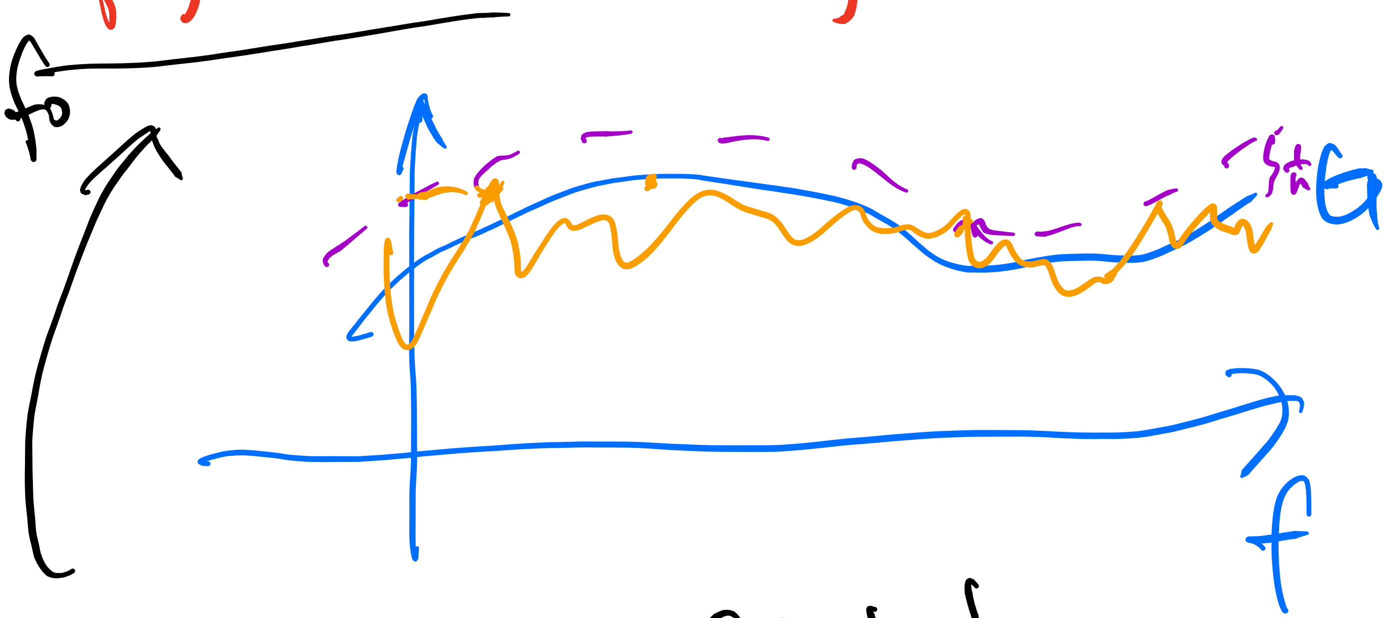
$$g(x_1 \dots x_i' \dots x_n)$$

$$G(f) = \frac{1}{n}\sum_{i=1}^{n} f(x_i) - \frac{1}{n} f(x_i) + \frac{1}{n} f(x_i') - \mathbb{E}_D f(z)$$

$\forall f \in \mathcal{F}$,

$$F(f) - G(f) \le \frac{1}{n}$$

$$\overset{-\frac{1}{n}}{\not\le} \sup_{f\in\mathcal{F}} F(f) - \sup_{f\in\mathcal{F}} G(f) \le \frac{1}{n} ?$$



$$F(f_0) \le G(f_0) + \frac{1}{n}$$

$$\le \sup_f G(f) + \frac{1}{n}$$

McDiarmid's Lemma $\Rightarrow$ $\delta' = \delta/2$

w.p. $1-\delta/2$,

$$g(Z_1 \cdots Z_n) \le \mathbb{E}\, g(Z_1 \cdots Z_n) + \sqrt{\frac{\ln 4/\delta}{2n}}.$$

$(*)$

Lemma 2 : $\mathbb{E}\left[\sup_{f\in\mathcal{F}} \underbrace{\hat{\mathbb{E}}_S f(z) - \mathbb{E}_D f(z)}_{\text{r.v.}}\right] \le 2\,\mathrm{Rad}_n(\mathcal{F})$.

infinite collection

where $\mathrm{Rad}_n(\mathcal{F}) = \mathbb{E}_{S\sim D^n} \mathrm{Rad}_S(\mathcal{F})$ $(f(z_1) \cdots f(z_n))$

$$\mathrm{Rad}_S(\mathcal{F}) = \frac{1}{n}\mathbb{E}_{\sigma\sim U(\pm 1)^n} \sup_{f\in\mathcal{F}} \sum_{i=1}^{n} f(z_i)\sigma_i$$

finite collection of r.v's.

$x \to \{0,1\}$

Pf of Lemma 2

<u>step 1</u> : symmetrization (double sampling trick)

$(f(x_1)\cdots f(x_n)\, f(x_1') \cdots f(x_n'))$

2^{2n}

$$\mathbb{E}_{S\sim D^n} \sup_{f\in\mathcal{F}}\left[\hat{\mathbb{E}}_S f(z) - \mathbb{E}_D f(z)\right] \le \mathbb{E}_{\substack{S\sim D^n\\ S'\sim D^n}} \sup_{f\in\mathcal{F}}\left[\hat{\mathbb{E}}_S f(z) - \hat{\mathbb{E}}_{S'} f(z)\right]$$

finite collection of r.v.

$S' = z_1' \cdots z_n'$

$\frac{1}{n}\sum^{n} f(z_i')$

pf of step 1:

show: $\sup_{f \in \mathcal{F}} \mathbb{E}_S f(z) - \mathbb{E}_D f(z)$

$\mathbb{E}_D f(z) = \mathbb{E}_{S' \sim D^n} \mathbb{E}_{S'} f(z) = \mathbb{E}_{S' \sim D^n} \frac{1}{n} \sum_{i=1} f(z_i)$

$= \mathbb{E}_{z \sim D} f(z)$

$$\leq \mathbb{E}_{S' \sim D^n} \sup_{f \in \mathcal{F}} \mathbb{E}_S f(z) - \mathbb{E}_{S'} f(z)$$

for all realizations of S.

observation: suppose G is a random function, then

$$\underbrace{\sup_{f \in \mathcal{F}} \mathbb{E}\, G(f)}_{f_0} \leq \mathbb{E} \sup_{f \in \mathcal{F}} G(f)$$

Pf: pick $f_0 = \arg\max_{f \in \mathcal{F}} \mathbb{E}\, G(f)$

$\mathbb{E}\, G(f_0)$

$$\leq \mathbb{E}\left[\sup_{f\in\mathcal{F}} G(f)\right]$$ ☆

taking expectation over $S\sim D^n$, we get the symmetrization lemma.

step 2: introducing random signs.

Claim:

$$\frac{1}{n}\mathbb{E}_{\substack{S\sim D^n\\ S'\sim D^n}}\sup_{f\in\mathcal{F}}\left(\sum_{i=1}^n f(z_i)-f(z_i')\right)$$

$$=\frac{1}{n}\mathbb{E}_{\substack{S\sim D^n\\ S'\sim D^n}}\sup_{f\in\mathcal{F}}\sum_{i=1}^n \left(f(z_i)-f(z_i')\right)\sigma_i$$

For all $\sigma_1 \ldots \sigma_n \in \{\pm 1\}$.

Ex: $n=2$ $\quad \sigma_1=-1, \; \sigma_2=+1$

$$\mathbb{E}_{z_1, z_2, z_1', z_2' \sim D^4} \sup_{f \in \mathcal{F}} \left(f(z_1) - f(z_1') + f(z_2) - f(z_2') \right)$$

$$= \mathbb{E}_{z_1 z_2 z_1' z_2' \sim D^4} \sup_{f \in \mathcal{F}} \left(f(z_1') - f(z_1) + f(z_2) - f(z_2') \right)$$

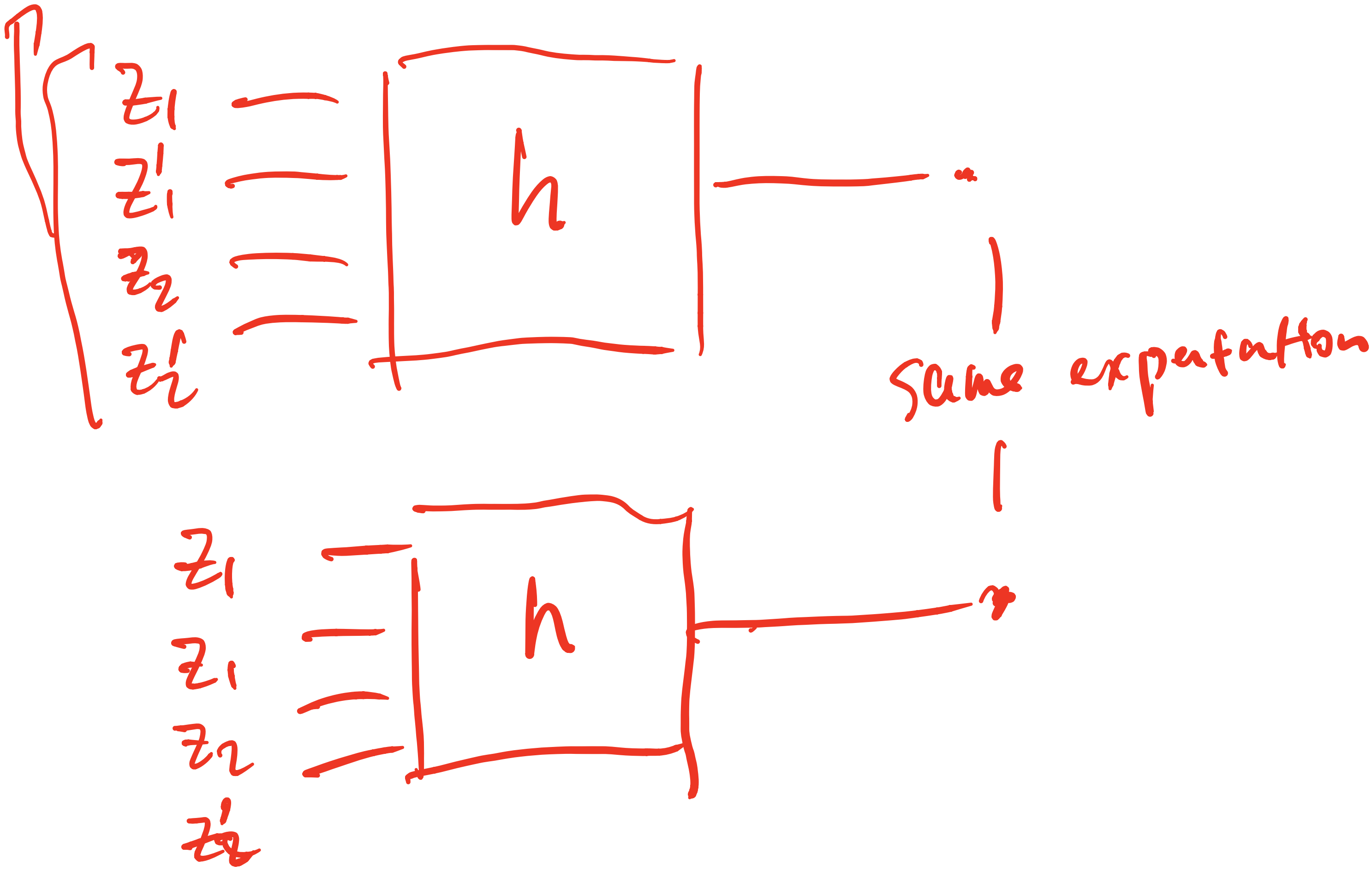
(z_1, z_1', z_2, z_2') has the same distribution

as (z_1', z_1, z_2, z_2'), $\overset{d}{=} D^4$

$$h(w_1, w_1', w_2, w_2') = \sup_{f \in \mathcal{F}} \left(f(w_1) - f(w_1') + f(w_2) - f(w_2') \right)$$

$$LHS = \mathbb{E}_{\cancel{z_1}, z_2, z_1', z_2'} h(z_1, z_1', z_2, z_2')$$

$$RHS = \mathbb{E}_{z_1 z_2, z_1' z_2'} h(z_1', z_1, z_2, z_2')$$



step3:

$$\frac{1}{n} \mathop{\mathbb{E}}_{\substack{S\sim D^n \\ S'\sim D^n}} \mathop{\mathbb{E}}_{\sigma\sim U\{\pm 1\}^n} \sup_{f\in\mathcal{F}} \sum_{i=1}^{n} \left(f(z_i) - f(z_i')\right) \sigma_i$$

$$\leq 2 \cdot \mathrm{Rad}_n(\mathcal{F})$$

Pf: LHS

$$= \frac{1}{n} \mathbb{E}_{S,S',\sigma} \sup_{f \in \mathcal{F}} \left[\sum_{i=1}^{n} f(z_i)\sigma_i - \sum_{i=1}^{n} f(z_i')\sigma_i \right]$$

Obs: $\sup_f \left(A(f) + B(f) \right)$

$$\leq \sup_f A(f) + \sup_f B(f)$$

Pf: f_0 is argmax

$$LHS = A(f_0) + B(f_0)$$

$$\leq \sup_f A(f) + \sup_f B(f).$$

$$\leq \frac{1}{n} \left(\mathbb{E} \sup_{f \in \mathcal{F}} \sum_{i=1}^{n} f(z_i)\sigma_i + \mathbb{E} \sup_{f \in \mathcal{F}} \sum_{i=1}^{n} f(z_i')(-\sigma_i) \right)$$

$$(-\sigma_1 \cdots -\sigma_n)$$

$$\sim U(\pm 1)^n$$

$$= 2 \cdot \mathbb{E}_{S, \sigma} \sup_f \sum_{i=1}^n f(z_i) \sigma_i$$

$$= 2 \, \mathrm{Rad}_n(\mathcal{F}) .$$

steps 1–3

$$\Rightarrow \mathbb{E} \sup_{f \in \mathcal{F}} \mathbb{E}_S f(z) - \mathbb{E}_D f(z)$$

$$\leq 2 \, \mathrm{Rad}_n(\mathcal{F})$$

<u>Lemma 3</u>: controlling Rademacher complexity w/ growth fn.

for any set S of size n,

$$\mathrm{Rad}_S(\mathcal{F}) \leq \sqrt{\frac{2 \ln(S(\mathcal{F}, n))}{n}}$$

Lemma 3 Pf:

$$\mathrm{Rad}_S(\mathcal{F}) = \frac{1}{n} \mathbb{E}_{\sigma} \sup_{f \in \mathcal{F}} \sum_{i=1}^{n} f(z_i)\,\sigma_i$$

for all

$(b_1 \cdots b_n) \in \Pi_{\mathcal{F}}(S)$

pick a f from $\mathcal{F}$ so that

It achieves this labeling;

$\mathcal{F}_S = \{$ set of representative f's $\}$

$$|\mathcal{F}_S| \overset{?}{\leq} S(\mathcal{F}, n)$$

zero mean.

$[-1, +1]$.

$$\mathrm{Rad}_S(\mathcal{F}) = \frac{1}{n} \mathbb{E}_{\sigma} \sup_{f \in \mathcal{F}_S} \sum_{i=1}^{n} f(z_i) \cdot \sigma_i$$

red

Fixed

<u>Lemma</u> (Massart's Finite Lemma):

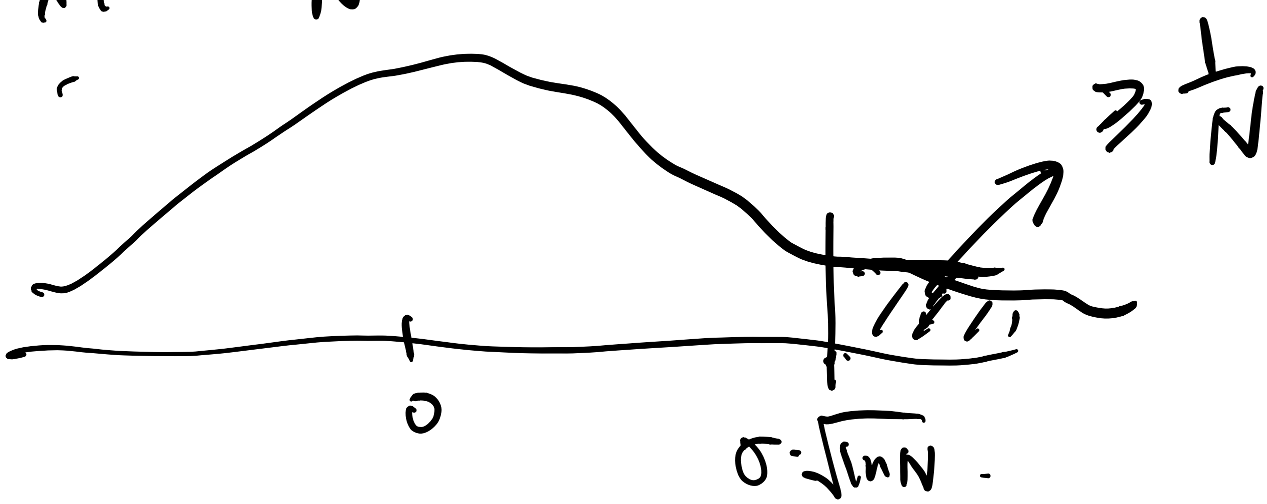
suppose $X_1 \sim X_N$ are zero mean, σ^2-SG.

then

$$\mathbb{E}\left[\max_{i=1}^{N} X_i\right] \le \sigma \cdot \sqrt{2 \ln N}.$$

This Lemma is tight up to constant,

$$X_1 \cdots X_N \overset{iid}{\sim} N(0, \sigma^2).$$



Using Massart's Lemma:

$$X_f = \sum_{i=1}^{n} \overset{1\text{-SG}}{\underbrace{\sigma_i f(z_i)}_{\in\{-1,1\}}} \quad \forall f \in \mathcal{F}_S$$

$\in [\ldots]$

X_f is σ^2-SG w/ $\sigma^2 = n$.

$$\mathbb{E} \sup_{f \in \mathcal{F}} X_f \le \sqrt{n \cdot 2 \cdot \ln S(\mathcal{F}, n)}$$

dividing by n.

$$\Rightarrow \mathrm{Rad}_S(\mathcal{F}) \le \sqrt{\frac{2 \ln S(\mathcal{F}, n)}{n}}.$$

$\forall j$

$\ge e^{tX_j}$

Pf of Massart's Lemma:

$\forall t > 0$.

$$\max_i X_i \le \frac{\ln\left(\sum_{i=1}^{N} e^{tX_i} \right)}{t}$$

$$\mathbb{E} \max_i X_i \le \frac{\mathbb{E} \ln\left(\sum_{i=1}^{N} e^{tX_i} \right)}{t}$$

$$\overset{\text{Jensen}}{\le} \frac{\ln\left(\mathbb{E} \sum_{i=1}^{N} e^{tX_i} \right)}{t}$$

SG of X_i's

$$\le \frac{\ln N}{t} + \frac{\sigma^2 t}{2}$$

choose $t = \sqrt{\frac{2\ln N}{\sigma^2}}$ to minimize RHS

$$\mathbb{E} \max_i X_i \le \sigma \cdot \sqrt{2 \ln N} .$$