Concentration Inequi.

We want to improve estimate for $\mathbb{P}(d(\mu,\hat{\mu}_n) > \varepsilon)$. Next, we first prove deviation of general $g(X_1, \cdots X_n)$ for X_k. i.i.d. from its expected value which extends LLN if taking $g(X_1 \cdots X_n) = \bar{X}_n$.

But we also need some assumption on g and dist. of X_k:

1) Fluctuation of $g(X_1 \cdots X_n)$ can't be small if it depends strongly on single r.v. X_j.

   $\Rightarrow$ measure strength of dependence on individual r.v. using partial measure of dispersion.

2) Recall the case of LLN. We reduce Var of sum of r.v. to sum of var. of individual r.v. which requires independence of X_j.

(1) <u>Variational repre. of entropy:</u>

We use different measure of dispersion instead of var. (which's kind of L^2-concen.)

Def: $Z \geq 0$ s.t. $Z, \log Z, Z\log Z \in L^1((\Omega, \mathcal{F}, \mathbb{P}), \mathbb{R}^1)$

Entropy of Z is $\mathcal{E}(Z) = \mathbb{E}(Z\log Z) - \mathbb{E}(Z)\log\mathbb{E}(Z)$.

Set partial entropy of g is:

$g_j(x_1, \cdots x_n) = \mathcal{E}(g(x_1 \cdots x_{j-1}, X_j, x_{j+1} \cdots x_n))$ and

$\mathcal{E}_j(g)(X_1, \cdots X_n) := g_j(x_1, \cdots X_j, \cdots x_n)|_{x_1 = X_1, \cdots x_n = X_n}$

Rmk: Entropy is related to KL-divergence:

For $\mu, \nu \in \mathcal{M}_1(\mathbb{R}^n)$, $\mu \sim \nu$. Let $X \sim \mu$.

$Z = Z(X) = \frac{d\nu}{d\mu}(X) \Rightarrow \mathbb{E}^X(Z) = 1$.

$\mathcal{E}(Z) = \int \log\frac{d\nu}{d\mu}\, d\nu = d_{KL}(\nu \| \mu)$.

So: the less Z deviates from $\mathbb{E}(Z)$

the less μ & ν are separated by d_{KL}.

$\Rightarrow$ It illustrates that why we say

$\mathcal{E}$ can measure concentration of $g(X_1 \cdots X_n)$ around μ.

Lem. $Z \geq 0$. $\log Z, Z\log Z \in L^1 \Rightarrow \mathcal{E}(Z) = \sup\{\mathbb{E}(ZX)$: X is r.v. with $\mathbb{E}(e^X) = 1$, $ZX \in L^1\}$.

Pf: Set $\overline{\mathbb{P}} = e^X \mathbb{P}$. $\overline{\mathbb{P}}$ is p.m. by cond.

So $\mathcal{E}(Z) - \overline{\mathbb{E}}(ZX) =$

$\mathbb{E}^{\mathbb{P}}(e^{-X} z \log(e^{-X} z)) - \mathbb{E}^{\mathbb{P}}(e^{-X} z) \log \mathbb{E}^{\mathbb{P}}(e^{-X} z)$

≥ 0 by Jensen inequi. on $t\log t$.

And "$=$" holds when $X = \log \frac{z}{\mathbb{E}(z)}$

Thm. (Tensorization of Entropy)

$X_j \sim \mu_j \in M_1^+(\mathbb{R}^k)$. indept r.v. If $g(X_1 \dots X_n) =: z \geq 0$. z, $\log z$, $z\log z \in L^1$. Then:

$$\Sigma(g(X_1, X_2 \cdots X_n)) \leq \sum_1^n \mathbb{E}(\Sigma_j(g)(X_1 \cdots X_n))$$

Rmk: i) It means entropy $\Sigma(g)$ can be controlled by partial entropy/fluc. of individual indept. X_j. We hope change one r.v. won't change a lot on fluctuation $\Sigma(g(X_1 \dots X_n))$.

ii) Independence of X_j means dist. of $(X_1 \cdots X_n)$ is tensor product of individual r.v. X_j.

iii) We can set $g(X_1 \cdots X_n) = \sup_{V \in \mathcal{M}} |\langle \hat{L}_n \rangle(V) + h(\mu) - h(\mu \| V)|$ for i.i.d. X_k.

Pf: Set $\mathcal{G}_j = \sigma(X_k,\ k \le j)$.

$\hat{\mathcal{G}}_j = \sigma(X_k,\ k \le n,\ k \ne j)$

Note $\Sigma(Z) = \mathbb{E}(Z(\log(Z) - \log \mathbb{E}(Z)))$

$= \mathbb{E}(\sum_1^n Z u_j)$

where $u_j = \log \mathbb{E}(Z|\mathcal{G}_j) - \log \mathbb{E}(Z|\mathcal{G}_{j-1})$.

With $\int_{\mathbb{R}^d} \exp(u_k(X_1 \cdots X_{k-1}, x_k))\, d\mu_k(x_k)$

$$\overset{\text{integ}}{=} \int_{\mathbb{R}^d} \frac{\int_{\mathbb{R}^{d(n-k)}} g(X_1 \cdots X_k, x_{k+1}, \cdots x_n)\, d\mu_{k+1} \cdots d\mu_n}{\int_{\mathbb{R}^{d(n-k+1)}} g(X_1 \cdots X_{k-1}, x_k \cdots x_n)\, d\mu_k \cdots d\mu_n}\, d\mu_k$$

$\overset{\text{Fubini}}{=} 1 \Rightarrow \mathbb{E}_{X_k}(\exp\{u_k(X_1 \cdots X_n)\}) = 1$.

So by the represent Lem. we have:

$\Sigma_k(Z)(X_1 \cdots X_n) \ge \mathbb{E}(Z u_k | \hat{\mathcal{G}}_k)$

$\Rightarrow \mathbb{E}(\Sigma_k(Z)(X_1 \cdots X_n)) \ge \mathbb{E}(Z u_k)$.

(2) McDiarmid's inequi. :

We want to control the outliers of r.v. X.

then: everything is close to expectation.

Lem: X is $\mathbb{R}^1$-r.v. st. $Z = e^{tZ}$ satisfies Z.

$\log z,\ z\log z \in L^1$ for $\forall \alpha > 0$. If $\mathrm{Ent}(e^{\alpha X}) \le \sigma^2 \alpha^2 \mathbb{E}(e^{\alpha X})/2$ $\forall \alpha > 0$. Then: $X - \mathbb{E}(X)$ is subgaussian with var. proxy σ^2.

Pf: $\frac{d}{d\alpha} \frac{\psi_{X-\mathbb{E}(X)}(\alpha)}{\alpha} = \frac{1}{\alpha^2} \frac{\mathrm{Ent}(e^{\alpha X})}{\mathbb{E}(e^{\alpha X})} \le \frac{\sigma^2}{2}$.

By L'Hopital: $\lim_{\alpha \to 0} \frac{\psi_{X-\mathbb{E}(X)}(\alpha)}{\alpha} = \psi'_{X-\mathbb{E}(X)}(0) = 0$

So: $\frac{\psi_{X-\mathbb{E}(X)}(\alpha)}{\alpha} = \int_0^\alpha \frac{d}{dt} \frac{\psi_{X-\mathbb{E}(X)}(t)}{t}\, dt \le \frac{\alpha\sigma^2}{2}$

Denote: $D_j^- g(x_1 \cdots x_n) := g(x_1, \cdots x_n) - \inf_{x_j \in \mathbb{R}^d} g(x_1 \cdots x_j \cdots x_n)$

$D_j^+ g(x_1 \cdots x_n) := \sup_{x_j \in \mathbb{R}^d} g(x_1 \cdots x_j \cdots x_n) - g(x_1 \cdots x_n)$

$D_j g(x_1 \cdots x_n) := D_j^- g + D_j^+ g$.

Lem^2. For X $\mathbb{R}^d$-r.v. $g: \mathbb{R}^d \to \mathbb{R}^1$ measurable. If $D^- g(X)$ is $\mathbb{P}$-a.s. finite and $g(X)$, $\log g(X)$, $g(X)\log g(X) \in L^1$. Then:

$\mathrm{Ent}(e^{g(X)}) \overset{i)}{\le} \mathrm{Cov}(g(X), e^{g(X)}) \overset{ii)}{\le} \mathbb{E}(|D^- g|^2 e^g)$.

Pf: i) is from Jensen's inequi:

$\log \mathbb{E}(e^{g(X)}) \ge \mathbb{E}(g(X))$.

ii) Note $\mathbb{E}(c(e^{g(x)} - \mathbb{E}(e^{g(x)}))) = 0$.

for $\forall c$. const.

$Cov(g(x), \exp(g(x))) =$

$\mathbb{E}((g(x) - \mathbb{E}(g(x)))(e^{g(x)} - \mathbb{E}(e^{g(x)}))) =$

$\mathbb{E}((g(x) - \inf_{x\in\mathbb{R}^n} g(x))(e^{g(x)} - \mathbb{E}(e^{g(x)})))$

$\le \mathbb{E}(\square(e^{g(x)} - e^{\inf g(x)}))$

Note $e^{g(x)} - e^{\inf g(x)} = \int_{\inf g(x)}^{g(x)} e^{z}\,dz$

$\le e^{g(x)}(g(x) - \inf_{x\in\mathbb{R}^n} g(x))$

Thm. (Mc-Diarmid's inequi.)

For $X_j \sim \mu_j$. indept. r.v. $g \triangleq g(X_1, \cdots X_n)$

Set $\sigma_{\pm,n}^2 = 2\|\sum_1^n |D_j^{\pm} g|^2\|_{L^\infty(\mu)} < \infty$. Then:

$g - \mathbb{E}(g)$ is subgaussian w.r.t. var. proxy $\sigma_{-,n}^2$ and $-(g - \mathbb{E}(g))$ is subgaussian w.r.t var. proxy $\sigma_{+,n}^2$

Pf: Apply Lem.2 above:

$\mathcal{E}_j(e^{\alpha g}) \le \mathbb{E}(|D_j^{-}\alpha g|^2 e^{\alpha g} | \hat{\mathcal{F}}_j)$

By Tensorization of entropy:

$$\mathcal{E}(e^{\alpha g}) \le \sum_1^n \mathbb{E}(\mathcal{E}_j(e^{\alpha g}))$$

$$\le \sum_1^n \mathbb{E}(\mathbb{E}(|D_j^- \alpha g|^2 e^{\alpha g} \,|\, \mathcal{G}_j))$$

$$\le \alpha^2 \Big\| \sum_1^n |D_j^- g|^2 \Big\|_{L^\infty(\mu)} \mathbb{E}(e^{\alpha g}).$$

Apply Lem'. $\Rightarrow$ $g - \mathbb{E}(g)$ is subgaussian

For another part. We use $\tilde{g} = -g$.

<u>Cor.</u> By Chernov's estimate, we imply:

$$\mathbb{P}(g - \mathbb{E}(g) > \varepsilon) \le \exp(-\varepsilon^2 / 2\sigma^2_{-,n})$$

$$\mathbb{P}(g - \mathbb{E}(g) < -\varepsilon) \le \exp(-\varepsilon^2 / 2\sigma^2_{+,n})$$

<u>Rmk:</u> We consider i.i.d. model and

$g = \sup_v |(c_n \hat{L}_n(v) + h(\mu)) - d(\mu \| v)|$.

$\Rightarrow D_j^{\pm} g = D_i^{\pm} g$. $\forall i, j$. So:

$\sigma^2_{\pm,n} = 2n \|g_1^{\pm}\|^2_{L^\infty(\mu)}$. If $\sigma^2_{\pm,n} \xrightarrow{n\to\infty} 0$

Then we have "new" LLN.

<u>Strategy:</u> For i.i.d model. $\hat{L}_n = \sum_1^n \ell(x_j | v)$. $c_n = n^{-1}$.

i) Use Dudley's inequi. to control $\mathbb{E}(g)$.

where $g = \sup_v |Z_v|$ $\left(Z_v \triangleq d(\mu \| v) - (c_n \hat{L}_n - h(\mu)) \right)$

But note that the inequi. only works

for "$\sup Z_v$"-type (No 1·1!)

We can add v^* to $\mathcal{N}$. Let $\tilde{\mathcal{N}} = \mathcal{N} \cup \{v^*\}$.

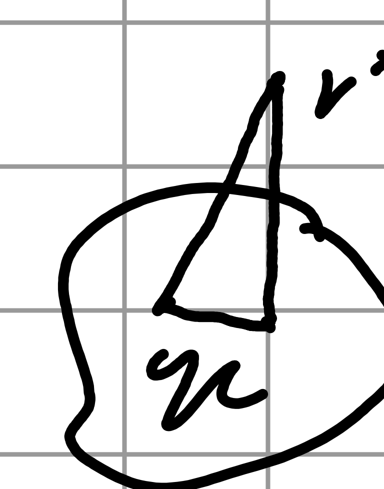
where $Z_{v^*} \equiv 0$. So $\sup_v |Z_v| = \sup_v Z_v$

<u>Problem</u>: Z_v, $v \in \tilde{\mathcal{N}}$ has to be subgaussian

w.r.t. $\bar{L}^2 \bar{\lambda}(\cdot,\cdot)^2$.

i) $\bar{\lambda}$ must be metric by extending on $\tilde{\mathcal{N}}$.

We can set $\bar{\lambda}(v, v^*)$ suff. large.



(e.g. > $\mathrm{diam}(\mathcal{N})/2$) So: $\bar{\lambda}$ will

satisfy triangle inequi.

ii) Z_v need to fulfill: $\forall v \in \mathcal{N}$. $Z_v - Z_{v^*} = Z_v$

is subgaussian w.r.t. $\bar{L}^2 \bar{\lambda}(v, v^*)^2$ in addi-

tion to $\{Z_v\}_{v \in \mathcal{N}}$ is subgaussian process.

<u>Rmk</u>: Generally, we find $\bar{\sigma}^2$ indept of v

$\in \mathcal{N}$, s.t. $\bar{\sigma}^2 \leq \bar{L}^2 \bar{\lambda}(v, v^*)^2$. And let

$\{Z_v\}_{\mathcal{N}}$ is subgaussian r.v. w.r.t. $\bar{\sigma}^2$.

$\Rightarrow$ Apply Dudley's: $\mathbb{E}(y) \leq C/\sqrt{n}$

<u>Rmk</u>: Note that though we extend $\bar{\lambda}$

to new $\bar{x}$: $\log N(\varepsilon; \bar{x}, \bar{x}) \le 1 + \log N(\varepsilon; \bar{x}, x)$ $\Rightarrow$ it won't change a lot.

ii) Apply McDiarmid's inequi. :

$\mathbb{P}(|g - \mathbb{E}(g)| > \varepsilon) \le 2\exp(-n\varepsilon^2 / 2\sigma_+^2 \vee \sigma_-^2)$.

$\mathbb{P}(\varepsilon_{sampl} > \varepsilon - c/\sqrt{n}) \le \mathbb{P}(\varepsilon_{sampl} - \mathbb{E}(\varepsilon_{sampl.}) > \varepsilon)$

$\le \exp\left(-\frac{n}{2}\left(\frac{\varepsilon - c/\sqrt{n}}{\sigma_+ \vee \sigma_-}\right)^2\right)$

e.g. let $\varepsilon = (\log(n))^p / \sqrt{n} \xrightarrow{n \to \infty} 0$.