Entropy method, log-Sobolev inequalities, Gaussian concentration

Recall: For $Y \geq 0$, $\operatorname{Ent} Y = \mathbb{E} Y \log Y - \mathbb{E} Y \log \mathbb{E} Y$

Entropy method: Exponential tail bounds for $Z = f(X_1, \ldots, X_n)$ by bounding entropy of $Y = e^{\lambda Z}$.

Lemma (Herbst argument): Suppose, for some $\sigma^2 > 0$,

$$\operatorname{Ent} e^{\lambda Z} \leq \frac{\lambda^2 \sigma^2}{2} \mathbb{E} e^{\lambda Z} \quad \forall \lambda > 0$$

Then $\mathbb{E} e^{\lambda(Z - \mathbb{E}Z)} \leq e^{\frac{\lambda^2 \sigma^2}{2}}$ $\forall \lambda > 0$, and

$$\mathbb{P}[Z - \mathbb{E}Z \geq t] \leq e^{-\frac{t^2}{2\sigma^2}} \quad \forall t \geq 0$$

Proof: Let $F(\lambda) = \mathbb{E} e^{\lambda Z}$, $F'(\lambda) = \mathbb{E} Z e^{\lambda Z}$. Then

$$\operatorname{Ent} e^{\lambda Z} = \mathbb{E} \lambda Z e^{\lambda Z} - \mathbb{E} e^{\lambda Z} \log \mathbb{E} e^{\lambda Z}$$

$$= \lambda F'(\lambda) - F(\lambda) \log F(\lambda)$$

Thus $$\frac{\sigma^2}{2} \geq \frac{\operatorname{Ent} e^{\lambda Z}}{\lambda^2 \mathbb{E} e^{\lambda Z}} = \frac{1}{\lambda} \cdot \frac{F'(\lambda)}{F(\lambda)} - \frac{1}{\lambda^2} \log F(\lambda)$$

$$= \frac{d}{d\lambda} \left[\frac{1}{\lambda} \log F(\lambda) \right] \quad \forall \lambda > 0.$$

By l'Hopital, $\lim_{\lambda \to 0} \frac{\log F(\lambda)}{\lambda} = \lim_{\lambda \to 0} \frac{d}{d\lambda} \log F(\lambda) = \mathbb{E} Z$.

So for any $\lambda > 0$,

$$\frac{1}{\lambda}\log F(\lambda) - \mathbb{E}Z = \int_0^\lambda \frac{d}{du}\left[\frac{1}{u}\log F(u)\right] \le \int_0^\lambda \frac{\sigma^2}{2} = \frac{\lambda\sigma^2}{2}$$

$$\Rightarrow F(\lambda) \le e^{\lambda \mathbb{E}Z + \frac{\lambda^2\sigma^2}{2}} \Rightarrow \mathbb{E}e^{\lambda(Z-\mathbb{E}Z)} \le e^{\frac{\lambda^2\sigma^2}{2}}.$$

$\blacksquare$

Log-Sobolev inequalities

By tensorization of entropy, for $Z = f(X_1, \ldots, X_n)$ w/ independent X_i's,

$$\mathrm{Ent}\, e^{\lambda Z} \le \sum_{i=1}^{n} \mathbb{E}\, \mathrm{Ent}^{(i)} e^{\lambda Z}$$

LSI: For certain distributions X_i and functions f,

$$\mathrm{Ent}\left[f(X_i)^2\right] \le C \cdot \mathbb{E}\left[\text{"squared gradient" of } f(X_i)\right].$$

Thm: Let $X_1, \ldots, X_n \overset{iid}{\sim}$ Rademacher, $f: \{\pm 1\}^n \to \mathbb{R}$,

$$\mathrm{grad}\, f(x) = \{D_i f(x)\}_{i=1}^n = \left\{\frac{f(x_1, \ldots, x_i, \ldots, x_n) - f(x_1, \ldots, -x_i, \ldots, x_n)}{2}\right\}_{i=1}^n.$$

Then $\mathrm{Ent}\left[f(X_1, \ldots, X_n)^2\right] \le 2\,\mathbb{E}\|\mathrm{grad}\, f(X_1, \ldots, X_n)\|_2^2$.

Proof: In 1-dimension ($n=1$), let $a = f(+1)$, $b = f(-1)$.

$$\mathrm{Ent}\, f(X)^2 = \mathbb{E}\, f(X)^2 \log f(X)^2 - \mathbb{E} f(X)^2 \log \mathbb{E} f(X)^2$$

$$= \frac{1}{2}\left(a^2 \log a^2 + b^2 \log b^2\right) - \frac{a^2+b^2}{2} \log \frac{a^2+b^2}{2}.$$

$$\mathbb{E}\left(\text{grad } f(X)\right)^2 = \frac{1}{4}(a-b)^2.$$

Suffices to show $\text{Ent } f(X)^2 \le 2 \cdot \mathbb{E}\left(\text{grad } f(X)\right)^2$ for $a \ge b \ge 0$.

Fixing $b \ge 0$, can check that

$$h(a) = \frac{1}{2}\left(a^2 \log a^2 + b^2 \log b^2\right) - \frac{a^2+b^2}{2} \log \frac{a^2+b^2}{2} - \frac{1}{2}(a-b)^2$$

satisfies $h(b) = 0$, $h'(b) = 0$, $h''(a) \le 0$ $\forall a \ge b \ge 0$.

$\Rightarrow h(a) \le 0$ $\forall a \ge b$, as desired.

For general n, by tensorization,

$$\text{Ent } f(X_1, \dots, X_n)^2 \le \sum_{i=1}^{n} \mathbb{E} \underbrace{\text{Ent}^{(i)} f(X_1, \dots, X_n)^2}_{\le 2 \mathbb{E}^{(i)} D_i f(X_1, \dots, X_n)^2} \le 2 \mathbb{E} \|\text{grad } f\|_2^2.$$

$\blacksquare$

Thm: Let $X_1, \dots, X_n \overset{iid}{\sim}$ Rademacher, $Z = f(X_1, \dots, X_n)$.

Suppose $\sum_{i=1}^{n} \left(D_i f(x_1, \dots, x_n)\right)_+^2 \le \sigma^2$ for all $x_1, \dots, x_n \in \{\pm 1\}$.

Then $\mathbb{P}[Z - \mathbb{E}Z \ge t] \le e^{-t^2/4\sigma^2}$.

Proof: Let $g(X_1,, X_n) = \mathbb{E} e^{\frac{\lambda}{2} f(X_1,, X_n)}$, $\lambda > 0$.

$$\text{Ent } e^{\lambda Z} = \text{Ent } g(X_1,, X_n)^2$$
$$\leq 2 \cdot \mathbb{E} \|\text{grad } g(X_1,, X_n)\|_2^2$$
$$= \frac{1}{2} \sum_{i=1}^{n} \mathbb{E} \left(e^{\frac{\lambda}{2} f(X_1,, X_n)} - e^{\frac{\lambda}{2} f(X_1,, -X_i,, X_n)} \right)^2$$
$$= \sum_{i=1}^{n} \mathbb{E} \left(e^{\frac{\lambda}{2} f(X_1,, X_n)} - e^{\frac{\lambda}{2} f(X_1,, -X_i,, X_n)} \right)_+^2.$$

By convexity, $e^{\frac{z}{2}} - e^{\frac{y}{2}} \leq \frac{z-y}{2} e^{\frac{z}{2}}$ for $z \geq y$.

$$\Rightarrow \text{Ent } e^{\lambda Z} \leq \sum_{i=1}^{n} \mathbb{E}\left[\lambda^2 (D_i f(X_1,, X_n))_+^2 e^{\lambda Z} \right] \leq \lambda^2 \sigma^2 \mathbb{E} e^{\lambda Z}.$$

By Herbst argument, $\mathbb{P}[Z - \mathbb{E}Z \geq t] \leq e^{-\frac{t^2}{4\sigma^2}}$. $\blacksquare$

Cor: Let $X_1, \ldots, X_n \overset{iid}{\sim}$ Rademacher, $Z = f(X_1,, X_n)$.

If $\|\text{grad } f(x_1,, x_n)\|_2^2 \leq \sigma^2$ for all $x_1,, x_n \in \{\pm 1\}^n$,

then $\mathbb{P}[|Z - \mathbb{E}Z| \geq t] \leq 2e^{-t^2/4\sigma^2}$.

Proof: In this case, $\sum_{i=1}^{n} (D_i f(x_1,, x_n))_+^2$, $\sum_{i=1}^{n} (D_i f(x_1,, x_n))_-^2 \leq \sigma^2$.

Apply previous result to both f and $-f$. $\blacksquare$

Example: $\varepsilon_1, \ldots, \varepsilon_n \overset{iid}{\sim}$ Rademacher, $Z = f(\varepsilon) = \sup_{t \in T} \varepsilon^T t$.

Recall $D_i f(\varepsilon)_+ \leq |t_i^*(\varepsilon)|$, so $\sum_{i=1}^n D_i f(\varepsilon)_+^2 \leq \|t^*(\varepsilon)\|_2^2 \leq \underbrace{\sup_{t \in T} \|t\|_2^2}_{\equiv \sigma^2}$.

Then $\mathbb{P}[Z \geq \mathbb{E}Z + u] \leq e^{-u^2/4\sigma^2}$.

Thm (Gaussian LSI): Let $X_1, \ldots, X_n \overset{iid}{\sim} N(0,1)$,

$f: \mathbb{R}^n \to \mathbb{R}$ weakly differentiable. Then

$$\operatorname{Ent} f(X_1, \ldots, X_n)^2 \leq 2\mathbb{E}\|\nabla f(X_1, \ldots, X_n)\|_2^2.$$

Proof: In 1-dimension ($n=1$): Assume $f: \mathbb{R} \to \mathbb{R}$ has compact support, twice continuously differentiable.

Introduce $\varepsilon_1, \ldots, \varepsilon_m \overset{iid}{\sim}$ Rademacher, $S_m = \frac{\varepsilon_1 + \ldots + \varepsilon_m}{\sqrt{m}}$.

By previous LSI on hypercube:

$$\operatorname{Ent} f(S_m)^2 \leq 2\mathbb{E}\|\operatorname{grad} f(S_m)\|_2^2$$

Recall $\lim_{m \to \infty} \mathbb{E}\|\operatorname{grad} f(S_m)\|_2^2 = \mathbb{E} f'(X)^2$ from proof of Poincaré inequality, Lecture 4. Then

$$\mathrm{Ent}\, f(X)^2 = \lim_{m\to\infty} \mathrm{Ent}\, f(S_m)^2 \le 2\mathbb{E} f'(X)^2.$$

Approximate general weakly differentiable $f : \mathbb{R} \to \mathbb{R}$ by sequence of smooth, compactly supported functions.

Then in higher dimensions $f : \mathbb{R}^n \to \mathbb{R}$, by tensorization,

$$\mathrm{Ent}\, f(X_1,,X_n)^2 \le \sum_{i=1}^n \mathbb{E} \underbrace{\mathrm{Ent}^{(i)} f(X_1,,X_n)^2}_{\le 2\mathbb{E}^{(i)}(\partial_i f)^2} \le 2\mathbb{E}\|\nabla f(X_1,,X_n)\|_2^2$$

∎

Thm (Tsirelson-Ibragimov-Sudakov inequality): If $X_1,,X_n \overset{iid}{\sim} \mathcal{N}(0,1)$, $f : \mathbb{R}^n \to \mathbb{R}$ is L-Lipschitz, (i.e. $|f(x)-f(y)| \le L\|x-y\|_2$ for all $x,y \in \mathbb{R}^n$), and $Z = f(X_1,,X_n)$, then $Z - \mathbb{E}Z$ is L^2-subgaussian and

$$\mathbb{P}[|Z-\mathbb{E}Z| \ge t] \le 2e^{-\frac{t^2}{2L^2}} \quad \forall t \ge 0.$$

Remark: This result is dimension-free.

Proof: Any L-Lipschitz $f : \mathbb{R}^n \to \mathbb{R}$ is weakly differentiable, with $\|\nabla f(x)\|_2 \le L$ a.e. Then for any $\lambda \in \mathbb{R}$,

$$\mathrm{Ent}(e^{\lambda Z}) = \mathrm{Ent}(e^{\lambda f(X_1,\dots,X_n)})$$
$$\leq 2\,\mathbb{E}\|\nabla e^{\frac{\lambda}{2}f(X_1,\dots,X_n)}\|_2^2 \quad \text{by LSI}$$
$$= 2\cdot\mathbb{E}\|e^{\frac{\lambda}{2}f(X_1,\dots,X_n)}\cdot\frac{\lambda}{2}\nabla f(X_1,\dots,X_n)\|_2^2$$
$$= \frac{\lambda^2}{2}\mathbb{E}\left[e^{\lambda Z}\cdot\|\nabla f(X_1,\dots,X_n)\|_2^2\right] \leq \frac{\lambda^2 L^2}{2}\mathbb{E}e^{\lambda Z}$$

By Herbst argument (applied to both f and $-f$),

$$\mathbb{E}e^{\lambda Z} \leq e^{\frac{\lambda^2 L^2}{2}} \quad \forall \lambda\in\mathbb{R},$$

i.e. Z is L^2-subgaussian. $\blacksquare$

Example: Let $Y\sim\mathcal{N}(0,\Sigma)\in\mathbb{R}^d$. Consider

$$\|Y\|_p = \left(\sum_{i=1}^d |Y_i|^p\right)^{1/p}, \quad p\geq 1.$$

Can write $Y=\Sigma^{1/2}X$, $X\sim\mathcal{N}(0,I)$.

Set $f(X) = \|\Sigma^{1/2}X\|_p$

$$\Rightarrow |f(x)-f(x')| = \left|\|\Sigma^{1/2}x\|_p - \|\Sigma^{1/2}x'\|_p\right|$$
$$\leq \|\Sigma^{1/2}(x-x')\|_p \leq \|\Sigma^{1/2}\|_{\ell_2\to\ell_p}\cdot\|x-x'\|_2.$$

Let $L=\|\Sigma^{1/2}\|_{\ell_2\to\ell_p}$. Then $\mathbb{P}\left[\|Y\|_p \geq \mathbb{E}\|Y\|_p + t\right] \leq e^{-\frac{t^2}{2L^2}}$.

Example: Let $x_1, \dots, x_n \overset{iid}{\sim} \mathcal{N}(0, I) \in \mathbb{R}^d$, $\theta_* \in \mathbb{R}^d$ fixed,

$y_i = x_i^T \theta_*$ for each $i = 1, \dots, n$.

Fix $\theta \in \mathbb{R}^d$, loss function $\ell : \mathbb{R} \to \mathbb{R}$ with $\|\ell'\|_\infty \leq K$.

View
$$F(X) = \frac{1}{n} \sum_{i=1}^n \ell(y_i - x_i^T \theta)$$
$$= \frac{1}{n} \sum_{i=1}^n \ell(x_i^T(\theta_* - \theta))$$

as function of $X = (x_1, \dots, x_n)$.

We have
$$\|\nabla_{x_i} F(X)\|_2^2 = \left\| \frac{1}{n} \ell'(x_i^T(\theta_* - \theta)) \cdot (\theta_* - \theta) \right\|_2^2$$
$$\leq \frac{K^2}{n^2} \|\theta_* - \theta\|_2^2$$
$$\Rightarrow \|\nabla_X F(X)\|_2^2 = \sum_{i=1}^n \|\nabla_{x_i} F(X)\|_2^2 \leq \frac{K^2}{n} \|\theta_* - \theta\|_2^2$$

So $F(X)$ is $K\|\theta_* - \theta\|_2 / \sqrt{n}$ - Lipschitz.

$$\Rightarrow \mathbb{P}\left[|F(X) - \mathbb{E}F(X)| \geq t \right] \leq e^{-\frac{nt^2}{2K^2 \|\theta_* - \theta\|_2^2}}.$$

Remarks: 1. Compared w/ Poincaré inequality,

$$\operatorname{Var} f(X) \leq \mathbb{E} \|\nabla f(X_1, \dots, X_n)\|_2^2,$$

TIS inequality requires stronger uniform bound on $\|\nabla f\|_2$

rather than only in expectation, but gives stronger statement of exponential tail decay.

② LSI $\operatorname{Ent} f(X_1,.., X_n)^2 \le 2 \mathbb{E} \|\nabla f(X_1,.., X_n)\|_2^2$

implies Poincaré: Set $f = 1+\varepsilon g$, g bounded. Then

$$\operatorname{Ent} f^2 = \mathbb{E} f^2 \log f^2 - \mathbb{E} f^2 \log \mathbb{E} f^2$$

$$= 2\varepsilon^2 \operatorname{Var} g + O(\varepsilon^3) \quad \text{(check this!)}$$

$$\mathbb{E}\|\nabla f\|_2^2 = \varepsilon^2 \cdot \mathbb{E}\|\nabla g\|_2^2$$

$$LSI \Rightarrow 2\varepsilon^2 \operatorname{Var} g + O(\varepsilon^3) \le 2\varepsilon^2 \cdot \mathbb{E}\|\nabla g\|_2^2$$

$$\overset{\varepsilon \to 0}{\Longrightarrow} \operatorname{Var} g \le \mathbb{E}\|\nabla g\|_2^2 \text{ which is Poincaré.}$$

③ Poincaré + LSI for many other distributions via analysis of continuous-time Markov chains: See van Handel Chapters 2 & 3.

<u>Further applications of entropy method</u>

Lemma: Let $X_1,.., X_n$ be independent, $Z = f(X_1,.., X_n)$,

$Z_i = f^{(i)}(X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n)$ for arbitrary functions $f^{(1)}, \dots, f^{(n)}$. Then

$$\operatorname{Ent} e^{\lambda Z} \le \sum_{i=1}^{n} \mathbb{E}\left[\frac{\lambda^2}{2}(Z - Z_i)_+^2 e^{\lambda Z}\right] \quad \forall \lambda > 0.$$

Proof: By simple calculus, for $Y \ge 0$,

$$\operatorname{Ent} Y = \inf_{u > 0} \mathbb{E}\left[Y \log Y - Y \log u - Y + u\right]$$

w/ inf attained at $u^* = \mathbb{E} Y$. Apply with $Y = e^{\lambda Z}$, $u = e^{\lambda Z_i}$:

$$\operatorname{Ent} e^{\lambda Z} \le \sum_{i=1}^{n} \mathbb{E} \operatorname{Ent}^{(i)} e^{\lambda Z}$$

$$\le \sum_{i=1}^{n} \mathbb{E}\, \mathbb{E}^{(i)}\left[e^{\lambda Z} \log e^{\lambda Z} - e^{\lambda Z} \log e^{\lambda Z_i} - e^{\lambda Z} + e^{\lambda Z_i}\right]$$

$$= \sum_{i=1}^{n} \mathbb{E}\left[e^{\lambda Z}\Big(\underbrace{\lambda Z - \lambda Z_i - 1 + e^{\lambda(Z_i - Z)}}_{\text{Apply } t - 1 + e^{-t} \le \frac{t^2}{2} \ \forall t \ge 0.}\Big)\right]$$

∎

Thm: Suppose

$$\sup_{\substack{x_1, \dots, x_n \\ x_1', \dots, x_n'}} \sum_{i=1}^{n} \left(f(x_1, \dots, x_n) - f(x_1, \dots, x_i', \dots, x_n)\right)_+^2 \le \sigma^2$$

Let $X_1, \dots, X_n$ be independent, and let $Z = f(X_1, \dots, X_n)$. Then

$$\mathbb{P}\left[Z - \mathbb{E} Z \ge t\right] \le e^{-\frac{t^2}{2\sigma^2}} \quad \forall t \ge 0.$$

Proof: Take $Z_i = \inf_{x_i'} f(X_1, \ldots, X_{i-1}, x_i', X_{i+1}, \ldots, X_n)$

$$\Rightarrow \operatorname{Ent} e^{\lambda Z} \le \frac{\lambda^2 \sigma^2}{2} \mathbb{E} e^{\lambda Z} \quad \forall \lambda > 0.$$

Result follows from Herbst argument. ∎

Compare with Efron-Stein:

$$\operatorname{Var} Z \le \mathbb{E} \sum_{i=1}^n \left(f(X_1, \ldots, X_n) - f(X_1, \ldots, X_i', \ldots, X_n)\right)_+^2.$$

Thm: Let $X_1, \ldots, X_n$ be independent, $X_i \in [0,1]$, $f: \mathbb{R}^n \to \mathbb{R}$ separately convex and L-Lipschitz, $Z = f(X_1, \ldots, X_n)$. Then

$$\mathbb{P}[Z - \mathbb{E}Z \ge t] \le e^{-\frac{t^2}{2L^2}} \quad \forall t \ge 0.$$

Proof: Let $x_i^* = \operatorname{argmin}_{x_i'} f(x_1, \ldots, x_i', \ldots, x_n)$. By convexity,

$$f(x_1, \ldots, x_n) - f(x_1, \ldots, x_i^*, \ldots, x_n) \le |\partial_i f(x_1, \ldots, x_n)| \cdot \underbrace{|x_i - x_i^*|}_{\in [0,1]}$$

$$\Rightarrow \sup_{\substack{x_1, \ldots, x_n \\ x_1', \ldots, x_n'}} \sum_{i=1}^n \left(f(x_1, \ldots, x_n) - f(x_1, \ldots, x_i', \ldots, x_n)\right)_+^2$$

$$\le \sum_{i=1}^n |\partial_i f(x_1, \ldots, x_n)|^2 = \|\nabla f(x_1, \ldots, x_n)\|_2^2 \le L^2$$

so result follows from previous theorem. ∎

Compare w/ convex Poincaré: $\mathrm{Var}\, Z \le \mathbb{E}\|\nabla f(X_1, \dots, X_n)\|_2^2$.

Example: Let $X \in \mathbb{R}^{m \times n}$ have independent entries, $X_{ij} \in [a,b]$.

Recall $\sigma_{max}(X)$ is convex $(b-a)$-Lipschitz function of $\{Z_{ij}\} = \{\frac{X_{ij}-a}{b-a}\} \in [0,1]$.

$$\Rightarrow \mathbb{P}[\sigma_{max}(X) \ge \mathbb{E}\sigma_{max}(X) + t] \le e^{-\frac{t^2}{2(b-a)^2}}.$$

Note: This provides only an upper tail bound. If f is jointly convex, we'll see different method for also getting lower tail bound next lecture.

Example: $Z = f(X_1, \dots, X_n)$ is longest subsequence satisfying a hereditary property. Recall from Lecture 4

$$0 \le f(x_1, \dots, x_n) - f(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n) \le 1$$

$$\sum_{i=1}^n f(x_1, \dots, x_n) - f(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n) \le f(x_1, \dots, x_n).$$

Set $Z_i = f(X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n)$. Then

$$\mathrm{Ent}\, e^{\lambda Z} \le \sum_{i=1}^n \mathbb{E}\left[\frac{\lambda^2}{2}(Z - Z_i)_+^2 e^{\lambda Z}\right] \le \frac{\lambda^2}{2}\mathbb{E}[Z e^{\lambda Z}].$$

[Homework 5: If

$$\operatorname{Ent} e^{\lambda Z} \le \lambda^2 \left[b\,\mathbb{E}\, Z e^{\lambda Z} + (\sigma^2 - b\,\mathbb{E} Z) \cdot \mathbb{E}\, e^{\lambda Z} \right] \quad \forall \lambda \in [0, \tfrac{1}{b}),$$

then $$\log \mathbb{E}\, e^{\lambda (Z - \mathbb{E} Z)} \le \frac{\sigma^2 \lambda^2}{1 - b\lambda} \quad \forall \lambda \in [0, \tfrac{1}{b}).$$]

Setting $b = \frac{1}{2}$, $\sigma^2 = \frac{1}{2}\mathbb{E} Z$, this gives a subexponential tail

$$\mathbb{P}[Z - \mathbb{E} Z \ge t] \le e^{-c \cdot \min\left(\frac{t^2}{\mathbb{E} Z},\, t\right)}$$

If $\mathbb{E} Z \to \infty$ as $n \to \infty$, this implies $Z \le \mathbb{E} Z + O_p(\sqrt{\mathbb{E} Z})$.