Efron-Stein, Poincaré inequalities, tensorization of entropy

$X_1, \dots, X_n$ independent r.v.'s. How to bound $\mathrm{Var}\, f(X_1, \dots, X_n)$?

Thm (Efron-Stein): If $X_1, \dots, X_n$ independent, $Z = f(X_1, \dots, X_n)$ with $\mathrm{Var}\, Z < \infty$, then

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

where $\mathrm{Var}^{(i)} Z = \mathrm{Var}[Z \mid \underbrace{X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n}_{\equiv X^{(i)}}]$

This is a "tensorization" property of variance, and reduces bounds for $\mathrm{Var}\, Z$ to variance bounds for <u>univariate</u> functions.

Proof: Recall the martingale

$$M_i = \mathbb{E}[Z \mid X_1, \dots, X_i], \qquad \Delta_i = M_i - M_{i-1}$$

Since $\mathbb{E}\Delta_i \Delta_j = 0 \ \forall i \ne j$,

$$\mathrm{Var}\, Z = \mathbb{E}(Z - \mathbb{E}Z)^2 = \mathbb{E}\Big(\sum_{i=1}^{n} \Delta_i\Big)^2 = \sum_{i=1}^{n} \mathbb{E}\Delta_i^2.$$

Apply $\Delta_i^2 = \left(\mathbb{E}[Z - \mathbb{E}^{(i)} Z \mid X_1, \dots, X_i]\right)^2$

$$\le \mathbb{E}\left[(Z - \mathbb{E}^{(i)} Z)^2 \mid X_1, \dots, X_i\right]$$

$$\Rightarrow \operatorname{Var} Z \le \sum_{i=1}^n \mathbb{E}(Z - \mathbb{E}^{(i)} Z)^2 = \sum_{i=1}^n \mathbb{E} \operatorname{Var}^{(i)} Z. \qquad \blacksquare$$

Equivalent forms: Set $v = \sum_{i=1}^n \mathbb{E} \operatorname{Var}^{(i)} Z$.

① Recall $\operatorname{Var} Y = \frac{1}{2}\mathbb{E}(Y-Y')^2$ for Y' an independent copy of Y.

$$\Rightarrow v = \frac{1}{2}\sum_{i=1}^n \mathbb{E}(Z - Z_i')^2, \text{ where}$$

$$Z_i' = f(X_1, \dots, X_{i-1}, \underbrace{X_i'}_{\text{independent copy of } X_i}, X_{i+1}, \dots, X_n)$$

② By symmetry $Z - Z_i' \overset{d}{=} -(Z - Z_i')$, so

$$v = \sum_{i=1}^n \mathbb{E}(Z - Z_i')_+^2 = \sum_{i=1}^n \mathbb{E}(Z - Z_i')_-^2.$$

Example: Suppose f has bounded differences property, $\|D_i f\|_\infty < \infty$.

Then $\operatorname{Var} Z \le \sum_{i=1}^n \|D_i f\|_\infty^2$. (Weaker than tail bound of Lecture 1.)

Example: Recall Rademacher complexity: $\varepsilon_1, \dots, \varepsilon_n \overset{iid}{\sim}$ Rademacher,

$$Z = f(\varepsilon_1, \dots, \varepsilon_n) = \sup_{t \in T} \varepsilon^T t, \qquad T \subseteq \mathbb{R}^n.$$

Let $t^* = t^*(\varepsilon)$ be point where sup is attained, for ε.

$$Z - Z_i' \le \varepsilon^T t^* - (\varepsilon_1, \dots, \varepsilon_i', \dots, \varepsilon_n)^T t^* = (\varepsilon_i - \varepsilon_i') t_i^*$$

$$\Rightarrow \mathbb{E}(Z - Z_i')_+^2 \le \mathbb{E}(\varepsilon_i - \varepsilon_i')_+^2 t_i^{*2} \le \mathbb{E}\, 2 t_i^{*2}$$

$$\Rightarrow \operatorname{Var} Z \le \sum_{i=1}^n \mathbb{E}(Z-Z_i')_+^2 \le 2\mathbb{E}\|t^*\|_2^2 \le 2\sup_{t\in T}\|t\|_2^2.$$

(Better than bounded differences bound.)

Example (1st passage percolation): G graph w/ iid edge weights $X_i \ge 0$, $\mathbb{E}X_i^2 = \sigma^2$. Fixing $u, v \in G$, let

$$Z = \inf_{\text{paths } P: u\to v} \sum_{e_i \in P} X_i.$$

Let P^* be path where inf is attained, for X.

$$(Z-Z_i')_- \le (X_i'-X_i)_+ \mathbb{1}\{e_i \in P^*\} \le X_i' \cdot \mathbb{1}\{e_i \in P^*\}$$

$$\Rightarrow \operatorname{Var} Z \le \sum_{i=1}^n \mathbb{E}(Z-Z_i')_-^2 \le \sigma^2\, \mathbb{E}[\text{length of } P^*].$$

[If e.g. $G = \mathbb{Z}^d$, $X_i \in [a,b]$ a.s., then

length of $P^* \le \frac{b}{a}\|u-v\|_1$, while $\mathbb{E}Z \ge a\|u-v\|_1$.]

Example (configuration functions): Call a property Π of $(x_1,\dots,x_n)$ <u>hereditary</u> if

$(x_1,\dots,x_n)$ satisfies Π $\Rightarrow$ all subsequences of $(x_1,\dots,x_n)$ satisfy Π.

E.g. ① $x_1,\dots,x_n$ are distinct

② $x_1, \dots, x_n$ are increasing

③ $x_1, \dots, x_n$ are <u>shattered</u> by a collection of sets $\mathcal{A}$:

$$\forall S \subseteq \{x_1, \dots, x_n\}, \ \exists A \in \mathcal{A} \text{ s.t. } A \cap \{x_1, \dots, x_n\} = S.$$

Let $Z = f(X_1, \dots, X_n)$ be longest subsequence of $(X_1, \dots, X_n)$ satisfying Π.

① # of distinct elements

② longest increasing subsequence

③ VC dimension of $\mathcal{A}$ (restricted to $\{X_1, \dots, X_n\}$)

Any such f is <u>self-bounding</u>:

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

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

Note $\mathrm{Var}^{(i)} Z \le \mathbb{E}^{(i)}(f(X) - f(X^{(i)}))^2$

$$\Rightarrow \mathrm{Var}\, Z \le \mathbb{E} \sum_{i=1}^{n} (f(X) - f(X^{(i)}))^2$$

$$\le \mathbb{E} \sum_{i=1}^{n} f(X) - f(X^{(i)}) \le \mathbb{E} f(X) = \mathbb{E} Z.$$

Poincaré inequalities

For certain distributions X_i and functions $f: \mathbb{R} \to \mathbb{R}$,

$$\mathrm{Var}\, f(X_i) \leq C \cdot \mathbb{E}\left[\text{"squared gradient" of } f(X_i)\right]$$

Thm: Let $X_1, \dots, X_n \overset{iid}{\sim}$ Rademacher, $f: \{\pm 1\}^n \to \mathbb{R}$. Define its discrete gradient $\mathrm{grad}\, f(x_1, \dots, x_n) = \left(D_i f(x_1, \dots, x_n)\right)_{i=1}^n$,

$$D_i f(x_1, \dots, x_n) = \frac{f(x_1, \dots, x_i, \dots, x_n) - f(x_1, \dots, -x_i, \dots, x_n)}{2}$$

Then $\mathrm{Var}\, f(X_1, \dots, X_n) \leq \mathbb{E}\, \|\mathrm{grad}\, f(X_1, \dots, X_n)\|_2^2$.

Proof: In 1-dimension $(n=1)$:

$$\mathrm{Var}\, f(X) = \left(\frac{f(+1) - f(-1)}{2}\right)^2 = \mathbb{E}\left(\mathrm{grad}\, f(X)\right)^2.$$

For general n, by tensorization (Efron-Stein),

$$\mathrm{Var}\, f(X_1, \dots, X_n) \leq \sum_{i=1}^n \mathbb{E}\, \underbrace{\mathrm{Var}^{(i)} f(X_1, \dots, X_n)}_{= \mathbb{E}^{(i)} D_i f(X_1, \dots, X_n)^2 \text{ by } n=1 \text{ case}} = \mathbb{E}\, \|\mathrm{grad}\, f(X_1, \dots, X_n)\|_2^2$$

$\boxtimes$

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

Previous argument shows $D_i f(\varepsilon)_+ \leq |t_i^*(\varepsilon)|$

By symmetry in law, $\mathbb{E}(D_i f(\varepsilon))_+^2 = \mathbb{E}(D_i f(\varepsilon))_-^2 = \frac{1}{2}\mathbb{E} D_i f(\varepsilon)^2$,

so $\operatorname{Var} f(\varepsilon) \le \mathbb{E}\|\operatorname{grad} f(\varepsilon)\|_2^2 \le 2\mathbb{E}\|t^*(\varepsilon)\|_2^2 \le 2\sup_{t\in T}\|t\|_2^2$

Thm (Gaussian Poincaré inequality): Let $X_1, \dots, X_n \overset{iid}{\sim} N(0,1)$, $f: \mathbb{R}^n \to \mathbb{R}$ weakly differentiable. Then

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

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

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

By previous Poincaré ineq. on hypercube,

$$\operatorname{Var}[f(S_m)] \le \sum_{i=1}^m \mathbb{E}\, D_i f(S_m)^2$$

$$= \sum_{i=1}^m \frac{1}{4}\mathbb{E}\left(f(S_m) - f\left(S_m - \frac{2\varepsilon_i}{\sqrt{m}}\right)\right)^2$$

$$\le m \cdot \frac{1}{4}\mathbb{E}\left[\left(\frac{2}{\sqrt{m}}|f'(S_m)| + \frac{2K}{m}\right)^2\right]$$

Since $S_m \xrightarrow{L} X$, by dominated convergence

$$\operatorname{Var} f(X) = \lim_{m\to\infty} \operatorname{Var} f(S_m) \le \lim_{m\to\infty} \mathbb{E} f'(S_m)^2 = \mathbb{E} f'(X)^2$$

For any weakly differentiable (i.e. absolutely cont.) $f: \mathbb{R} \to \mathbb{R}$, approximate by sequence of smooth, compactly supported functions.

For general n: By tensorization,

$$\mathrm{Var}\, f(X_1, \ldots, X_n) \le \sum_{i=1}^{n} \mathbb{E}\, \underbrace{\mathrm{Var}^{(i)} f(X_1, \ldots, X_n)}_{\le\, \mathbb{E}^{(i)} \partial_i f(X_1, \ldots, X_n)^2 \text{ by result for } n=1.} \le \mathbb{E}\, \|\nabla f(X_1, \ldots, X_n)\|_2^2.$$

$\blacksquare$

Cor: If f is L-Lipschitz (i.e. $|f(x) - f(x')| \le L \cdot \|x - x'\|_2$) and $X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(0,1)$, then $\mathrm{Var}\, f(X_1, \ldots, X_n) \le L^2$.

Note: This is dimension free, i.e. there is no explicit dependence on n.

Example (Gaussian complexity): Let $g \sim \mathcal{N}(0, I) \in \mathbb{R}^n$,

$$f(g) = \sup_{t \in T} g^T t \quad \text{for some } T \subseteq \mathbb{R}^n.$$

For each fixed t: $|g^T t - g'^T t| \le \|g - g'\|_2 \cdot \|t\|_2$, so

$$g \mapsto g^T t \text{ is } \|t\|_2\text{-Lipschitz}$$

$$\Rightarrow \left| \sup_{t \in T} g^T t - \sup_{t \in T} g'^T t \right| \le \sup_{t \in T} |g^T t - g'^T t| \le \|g - g'\|_2 \cdot \sup_{t \in T} \|t\|_2$$

So $\mathrm{Var}\, f(g) \le \sup_{t \in T} \|t\|_2^2$.

Thm (convex Poincaré ineq.): Let $X_1, \dots, X_n$ be independent, $X_i \in [0,1]$ a.s. Let $f: \mathbb{R}^n \to \mathbb{R}$ be weakly differentiable and separately convex (i.e. convex in each coordinate $1, \dots, n$).

Then $$\mathrm{Var}\, f(X_1, \dots, X_n) \le \mathbb{E}\, \|\nabla f(X_1, \dots, X_n)\|_2^2.$$

Proof: For $n=1$: Let $x^* = \underset{x \in [0,1]}{\operatorname{argmin}}\, f(x)$.

$$\mathrm{Var}\, f(X) \le \mathbb{E}\,(f(X) - f(x^*))^2$$
$$\overset{\text{(convexity)}}{\le} \mathbb{E}\left[f'(X)^2 \cdot (X - x^*)^2\right] \le \mathbb{E}\, f'(X)^2.$$

Then by tensorization,

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

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

Let $f(X) = \sigma_{\max}(X) = \sup_{u,v : \|u\|_2 = \|v\|_2 = 1} u^T X v$.

- $|u^T X v - u^T X' v| \le \|u\|_2 \cdot \|X - X'\|_{op} \cdot \|v\|_2 \le \|X - X'\|_F$

  $\Rightarrow X \mapsto u^T X v$ is 1-Lipschitz $\Rightarrow \sigma_{\max}(X)$ is 1-Lipschitz
- $\sigma_{\max}(X)$ is supremum of linear functions, hence convex.

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

$$\Rightarrow \operatorname{Var} \sigma(X) \le (b-a)^2.$$

Tensorization of Entropy

Def: For a nonnegative r.v. Z w/ $\mathbb{E} Z \log Z < \infty$, its *entropy* is

$$\operatorname{Ent} Z = \mathbb{E} Z \log Z - (\mathbb{E} Z) \log (\mathbb{E} Z) \qquad (\ge 0 \text{ by Jensen's})$$

Interpretation: Suppose Z is defined on (Ω, P) w/ $Z \ge 0$. If $\mathbb{E} Z = 1$, then can interpret $Z(\omega) = \frac{dQ}{dP}(\omega)$ where $Q(A) = \mathbb{E}[Z(\omega) \cdot \mathbb{1}\{\omega \in A\}]$. In this case

$$\begin{aligned}
\operatorname{Ent} Z &= \mathbb{E} Z \log Z \\
&= \int \left[\frac{dQ}{dP}(\omega) \cdot \log \frac{dQ}{dP}(\omega)\right] dP(\omega) \\
&= \int dQ(\omega) \log \frac{dQ}{dP}(\omega) = D_{KL}(Q \| P).
\end{aligned}$$

For $\mathbb{E} Z = c \neq 1$, let $Z = c\tilde{Z}$ where $\mathbb{E} \tilde{Z} = 1$. Then

$$\operatorname{Ent} Z = \mathbb{E} c\tilde{Z} \log c\tilde{Z} - \mathbb{E} c\tilde{Z} \log \mathbb{E} c\tilde{Z} = c \cdot \operatorname{Ent} \tilde{Z}.$$

So Ent is a homogeneous extension of KL-divergence.

[Compare: $\mathrm{Var}\, Z = \mathbb{E} Z^2 - (\mathbb{E} Z)^2$

If $Z = \frac{dQ}{dP}(\omega)$, then $\mathbb{E} Z = 1$ and

$$\mathrm{Var}\, Z = \int \left(\frac{dQ}{dP}(\omega)\right)^2 dP(\omega) - 1 = \chi^2(Q \| P).$$

So Var is a homogeneous extension of χ^2-divergence.]

Thm: If $X_1, \dots, X_n$ independent, $Z = f(X_1, \dots, X_n)$ nonnegative w/ $\mathbb{E} Z \log Z < \infty$, then

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

where
$$\mathrm{Ent}^{(i)}(Z) = \mathrm{Ent}[Z \mid X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n]$$
$$= \mathbb{E}^{(i)} Z \log Z - \mathbb{E}^{(i)} Z \log \mathbb{E}^{(i)} Z$$

Lemma (duality): Let $Z \ge 0$ be a r.v. (Ω, P), $\mathbb{E} Z \log Z < \infty$. Then

$$\mathrm{Ent}\, Z = \sup_{\substack{U : \Omega \to [-\infty, \infty) \\ \text{s.t. } \mathbb{E} e^U = 1}} \mathbb{E}\, U Z$$

Proof: By homogeneity, assume $\mathbb{E} Z = 1$.

Consider any $U:\Omega\mapsto[-\infty,\infty)$ s.t. $\mathbb{E}e^U=1$.

Let $Q(A)=\mathbb{E}[e^{U(\omega)}\mathbb{I}\{\omega\in A\}]$, so $\frac{dQ}{dP}=e^U$.

Then $1=\mathbb{E}_P Z=\mathbb{E}_Q e^{-U}Z$, and

$$\operatorname{Ent} Z=\mathbb{E}_P Z\log Z=\mathbb{E}_Q e^{-U}Z\log Z$$

$$=\mathbb{E}_Q e^{-U}Z\log e^{-U}Z+\mathbb{E}_Q e^{-U}ZU$$

$$=\operatorname{Ent}_Q[e^{-U}Z]+\mathbb{E}_P UZ$$

$$\Rightarrow \operatorname{Ent} Z-\mathbb{E}UZ=\operatorname{Ent}_Q[e^{-U}Z]\geq 0.$$

$$\Rightarrow \operatorname{Ent} Z\geq \sup_{U:\mathbb{E}e^U=1}\mathbb{E}UZ.$$ Equality holds for $e^U=Z$. ■

Remark: The supremum can be restricted to $U=f(Z)$ since equality holds for such U.

Cor (Donsker-Varadhan): For any P,Q on Ω w/ $Q\ll P$,

$$D_{KL}(Q\|P)=\sup_{\substack{U:\Omega\mapsto[-\infty,\infty)\\ \text{s.t. } \mathbb{E}_P e^U<\infty}}\mathbb{E}_Q U-\log\mathbb{E}_P e^U$$

Proof: Let $Z=\frac{dQ}{dP}(\omega)$. Then

$$D_{KL}(Q\|P)=\operatorname{Ent}_P Z=\sup_{U:\mathbb{E}_P e^U=1}\mathbb{E}_P UZ$$

$$= \sup_{U' : \mathbb{E}_P e^{U'} < \infty} \mathbb{E}_P Z\left(U' - \log \mathbb{E}_P e^{U'}\right) \qquad \left(e^U = e^{U'} / \mathbb{E}_P e^{U'}\right)$$

$$= \sup_{U' : \mathbb{E}_P e^{U'} < \infty} \mathbb{E}_Q U' - \log \mathbb{E}_P e^{U'}.$$

∎

Proof of Thm: Let $Z = f(X_1, \ldots, X_n)$, $M_i = \mathbb{E}[Z \mid X_1, \ldots, X_i]$.

$$\mathrm{Ent}\, Z = \mathbb{E} Z(\log Z - \log \mathbb{E} Z)$$

$$= \mathbb{E} \sum_{i=1}^{n} Z(\log M_i - \log M_{i-1})$$

By duality, $\mathrm{Ent}^{(i)} Z \geq \mathbb{E}^{(i)} Z\Big(\underbrace{\log M_i - \log \mathbb{E}^{(i)} M_i}_{=U,\ \mathbb{E}^{(i)} e^U = \mathbb{E}^{(i)} \frac{M_i}{\mathbb{E}^{(i)} M_i} = 1.}\Big)$

$$= \mathbb{E}^{(i)} Z(\log M_i - \log M_{i-1}).$$

$$\Rightarrow \mathrm{Ent}\, Z \leq \mathbb{E} \sum_{i=1}^{n} \mathrm{Ent}^{(i)} Z.$$

∎