Functional inequalities and concentration

Last lecture: Gaussian Poincare inequality: For "every" $f : \mathbb{R}^n \to \mathbb{R}$

$$\mathrm{Var}[f(X)] \leq E[\|\nabla f(X)\|^2]$$

when $X \sim N(0, I^n)$.

Stronger inequality: Gaussian log-Sobolev inequality

$$\mathrm{Ent}(f^2) \leq 2E[\|\nabla f(X)\|^2] \quad \text{for "every" } f : \mathbb{R}^n \to \mathbb{R}$$

when $X \sim N(0, I_n)$

where

$$\mathrm{Ent}(f) = E\left[f(X)\log\frac{f(X)}{E[f(X)]}\right] = E[f\log f] - E[f]\log E[f]$$

Herbst argument: "Log Sobolev inequalities imply concentration".

Let $Z = f(X)$ where $f : \mathbb{R}^n \to \mathbb{R}$ is L-Lipschitz, $E[Z] = 0$,

The Herbst argument deduces a bound on the log moment generating fn. $\varphi_Z(\lambda) := \log E[e^{\lambda Z}]$ from the Log-Sobolev inequality.

Define $F(\lambda) = E[e^{\lambda Z}]$ and apply the log-Sobolev inequality with the fn. $x \mapsto \exp^{\lambda f(x)/2}$.

Then we have

$$\mathrm{Ent}(\cdot) \leq \frac{1}{2}E[\lambda^2 \exp(\lambda f(X))\|f(X)\|^2] \leq \frac{1}{2}\lambda^2 L^2 F(\lambda)$$

$$\Leftrightarrow \lambda E[Ze^{\lambda Z}] - F(\lambda)\log F(\lambda) \leq \frac{\lambda^2 L^2}{2} F(\lambda).$$

$$\lambda F'(\lambda) - F(\lambda)\log F(\lambda) \leq \frac{\lambda^2 L^2}{2} F(\lambda) \quad *$$

Dividing by $\frac{1}{\lambda^2}F(\lambda) > 0$, we get,

$$G'(\lambda) \leq 2L^2 \quad \text{where} \quad G(\lambda) = \frac{\log F(\lambda)}{\lambda}.$$

$$\left(G'(\lambda) = \frac{1}{\lambda}\frac{F'(\lambda)}{F(\lambda)} - \frac{1}{\lambda^2}\log F(\lambda)\right)$$

We now use, $\lim_{\lambda \downarrow 0} G(\lambda) = E[Z]$, so we get.

$$\log E[\exp(\lambda Z)] = F(\lambda) \leq \exp\left(\frac{\lambda^2 L^2}{2}\right). \Rightarrow Z \in \mathcal{G}(L^2)$$

$$\Rightarrow \Pr[f(x) - E[f(x)] > t] \leq \exp\left(-\frac{t^2}{2L^2}\right).$$

[Gaussian concentration]

Interestingly, such functional inequalities often have to proved only in dimension $n=1$; the extension to higher dimensions is automatic based on sub-additivity properties we will now see.

— Instead of following this approach, we will use a slightly different approach [still based on appropriate functional inequalities in $n=1$ dimension], which still uses such sub-additivity properties.

(*) → Requires some measure theoretic justifaction, but is valid, e.g. when $E[e^{\lambda Z}]$ exists in some interval $[0, b)$, and λ is in the interior of this interval.

Sub-additivity :- The Efron-Stein inequality and the sub-additivity of entropy.

Notation : We denote

$$E_X[f(X,Y)] = \int f(x,Y)\,\mu(x,Y)\,dx = \underbrace{E[X|Y]}_{\text{more robust defn.}}$$

and if $X_1 \dots X_n$ is a set of r.v.'s, $S \subseteq [n]$ then

$$E^{(S)}[f(X_1, \dots X_n)] := E[f(X_1, \dots X_n) | X_{\bar{S}}]$$

In particular,

$$E^{(i)}[f(X_1, \dots X_n)] = E[f(X_1, \dots X_n) | X_1 \dots X_{i-1}, X_{i+1}, \dots X_n]$$

For any fn Φ, and random variable Z.

define $H_\Phi(Z) := E[\Phi(Z)] - \Phi(E[Z])$.

Examples :-

$Var(Z)$: $\Phi(z) = z^2$

$$Ent(Z) = E[Z \log Z] - E[Z] \log E[Z]$$

$$\Phi(z) = Z \log Z.$$

Conditional versions :- Let $Z = f(X_1, \dots X_n)$

$$H^{(S)}_\Phi(Z) = \underbrace{E^{(S)}[\Phi(Z)] - \Phi(E^{(S)}[Z])}_{\text{"function" of } X_{\bar{S}}}.$$

$\mathrm{Var}^{(i)}(Z) = E^{(i)}[Z^2] - E^{(i)}[Z]^2$ ("fn" of $X_1 \ldots X_{i-1}, X_{i+1}, \ldots X_n$)

Efron-Stein :- Assume $X_1, X_2, \ldots X_n$ are independent. $Z = f(X_1, \ldots X_n)$

$$\mathrm{Var}(Z) \leqslant \sum_{i=1}^{n} E\left[\mathrm{Var}^{(i)}[Z]\right]$$

Sub-additivity of entropy : Same framework as above. $X_1 \ldots X_n$ independent

$$\mathrm{Ent}(Z) \leqslant \sum_{i=1}^{n} E\left[\mathrm{Ent}^{(i)}[Z]\right]$$

Convexity : H_Φ is a convex functional in both the above cases: If X, Y are independent then,

$$H_\Phi\left[E_X[f(X,Y)]\right] \leqslant E_X\left[H_\Phi^{(Y)}[f(X,Y)]\right].$$

Claim :- Convexity $\Rightarrow$ Sub-additivity

Proof :- We first derive the following identity

$$H_\Phi[f(X_S, X_{\bar{S}})] = E\left[\Phi(f(X_S, X_{\bar{S}}))\right] - \Phi\left(E[f(X_S, X_{\bar{S}})]\right)$$

$$= E\Big[\Phi(f(X_S, X_{\bar{S}})) - \Phi\left(E_{X_S}[f(X_S, X_{\bar{S}})]\right)$$

$$+ \Phi\left(E_{X_S}[f(X_S, X_{\bar{S}})]\right)$$

$$- \Phi\left(E\left[E_{X_S}[f(X_S, X_{\bar{S}})]\right]\right)\Big]$$

$$= \mathbb{E}\left[\mathbb{E}_{X_S}[\Phi(f(X_S,X_{\bar{S}}))]\right] - \mathbb{E}\left[\Phi(\mathbb{E}_{X_S}[f(X_S,X_{\bar{S}})])\right]$$
$$+ \mathbb{E}\left[\Phi(\mathbb{E}_{X_S}[f(X_S,X_{\bar{S}})])\right]$$
$$- \Phi\left(\mathbb{E}\left[\mathbb{E}_{X_S}[f(X_S,X_{\bar{S}})]\right]\right)$$
$$= \mathbb{E}_{X_{\bar{S}}}\left[H_\Phi^{(S)}(f)\right] + H_\Phi\left(\mathbb{E}_{X_S}[f(X_S,X_{\bar{S}})]\right)$$

Now, when $X_S, X_{\bar{S}}$ are independent, the convexity assumption implies,
$$H_\Phi\left(\mathbb{E}_{X_S}[f(X_S,X_{\bar{S}})]\right) \leq \mathbb{E}_{X_S}\left[H_\Phi^{(\bar{S})}[f]\right]$$

So, we have
$$H_\Phi(f) \leq \mathbb{E}\left[H_\Phi^{(S)}(f)\right] + \mathbb{E}\left[H_\Phi^{(\bar{S})}(f)\right]$$ whenever, $X_S, X_{\bar{S}}$ are independent.

In particular when $(X_i)_{i=1}^n$ have a product distribution, we get
$$H_\Phi(f) \leq \mathbb{E}\left[H_\Phi^{(1)}(f)\right] + \mathbb{E}\left[H_\Phi^{\{2,\ldots n\}}(f)\right]$$
$$\leq \mathbb{E}\left[H_\Phi^{(1)}(f)\right] + \mathbb{E}\left[H_\Phi^{(2)}(f)\right] + \mathbb{E}\left[H_\Phi^{\{3,\ldots n\}}(f)\right]$$
$$\leq \sum_{i=1}^n \mathbb{E}\left[H_\Phi^{(i)}(f)\right].$$

Both Var and Ent are convex, which we can prove by showing that they are suprema of linear functionals.

(1) <u>Variance</u>: By definition, for any r.v. [for which $\mathbb{E}[T^2]$ is well defined]
$$0 \leq \mathrm{Var}(Y-T) = \mathrm{Var}(Y) + \mathrm{Var}(T) - 2\,\mathrm{Cov}(Y,T)$$
$$\Rightarrow \mathrm{Var}(Y) \geq 2\,\mathrm{Cov}(Y,T) - \mathrm{Var}(T) \quad \forall\, T \text{ s.t. } \mathbb{E}[T^2] \text{ exists.}$$
<u>with equality</u> when <u>$T=Y$</u>.

So, $$\mathrm{Var}(Y) = \underbrace{\sup_{T: E[T^2]<\infty}}_{\text{sup over linear function.}} \underbrace{2\,\mathrm{Cov}(Y,T) - \mathrm{Var}(T)}_{\text{Linear functional of } Y}$$

$$= \sup_{T: E[T^2]<\infty,\, E[T]=0} E[T(2Y-T)].$$

<u>For entropy</u> :- Let Y be any non-negative r.v. and let T be any random variable that is zero whenever Y is. (denoted $T \ll Y$)

We use the convexity of $-\log$, which implies that

$$E\left[\frac{Y}{E[Y]} \log\left(\frac{(Y/E[Y])}{(T/E[T])}\right)\right] \geq -\log\left(E\left[\frac{T}{E[T]}\right]\right) = 0$$

Which implies, (using $E[Y]>0$)

$$\underbrace{E[Y(\log Y - \log E[Y])]}_{=\ \mathrm{Ent}(Y)} \geq E[Y(\log T - \log E[T])].$$

$\forall\ T \ll Y$ non-negative (with equality if $Y=T$)

<u>Therefore</u>,

$$\mathrm{Ent}(Y) = \sup_{\substack{T \ll Y \\ T \geq 0}} E[Y(\log T - \log E[T])]$$

$$= \sup_{\substack{U \text{ s.t.} \\ E[e^U]=1}} E[YU].$$

<u>Further</u> :- If $E[UY] \geq \mathrm{Ent}(Y)\ \forall Y$, then $E[e^U] \leq 1$, for applying to $Y = e^u$, we get

$$E[Ue^U] - E[e^u]\log E[e^u] \leq E[Ue^U]$$

$$\Rightarrow E[e^u] \leq 1$$

<u>Main result</u> :- Let P ,

$$\log E_P[\exp(\lambda(Z-\lambda Z))] = \sup_{Q \ll P} \left[E_Q Z - E_P Z - D(Q\|P)\right]$$

Using this one gets,

$$\psi_{Z-\lambda Z}(\lambda) \leqslant \frac{v\lambda^2}{2} \quad \text{under } P$$

$$\Updownarrow$$

$$E_Q Z - E_P Z \leqslant \sqrt{2 v D(Q \| P)} \quad \forall\, Q \ll P.$$