13. Concentration Inequalities

$\Delta_i = E_i(Z) - E_{i-1}(Z)$ where $E_i(Z) = E(Z|X_1, \cdots X_i)$

$\sum_{i=1}^{n} \Delta_i = E(Z|X_1) - E(Z) + E(Z|X_1,X_2) - E(Z|X_1) + \cdots + E(Z|X_1 \cdots X_n) - E(Z|X_1 \cdots X_{n-1})$

$= E(Z|X_1 \cdots X_n) - E(Z) = Z - EZ$

Here we let $E(Z)=0$, so $Z = \sum_{i=1}^{n} \Delta_i$

$Var(Z) = E[(\sum_{i=1}^{n} \Delta_i)^2] = \cancel{E} \sum_{i=1}^{n} E(\Delta_i^2) + 2E(\Delta_i \Delta_j)$

$j > i \Rightarrow: E_i \Delta_j = E_i[E_j(Z) - E_{j-1}(Z)] = E_i(Z) - E_i(Z) = 0$

$Var(Z) = \sum_{i=1}^{n} E(\Delta_i^2)$

13.1 Efron-Stein inequality

$E_i(E^{(i)}Z) = \int_{\mathcal{X}^{n-i}} \int_{\mathcal{X}} f(x_1 \cdots, x_i, \cdots x_n) d\mu_i d\mu_{i+1} \cdots d\mu$

$= \int_{\mathcal{X}^{n-i+1}} f(\ ) d\mu_i \cdots d\mu_n = E_{i-1} Z$

Theorem 13.1

$\Delta_i = E_i(Z) - E_{i-1}(Z) = E_i(Z - E^{(i)}Z)$

$E\Delta_i^2 \le E(E_i[Z - E^{(i)}Z])^2)$ Jesen's inequality

$E\Delta_i^2 \le E(Z - E^{(i)}(Z))^2]$ total probability formula

$Var(Z) = E\sum(\Delta_i^2) \le E\sum(Z - E^{(i)}(Z))^2$

$E[(X-X')^2] = E(X^2 + X'^2 - 2XX') = E(X^2) + E(X'^2) - 2E(XX')$

$= E(X^2) - E^2(X) + E^2(X'^2) - E^2(X') + E^2(X) + E^2(X') - 2E(XX')$

$= \cancel{2E}\, Var(X) + Var(X') + E^2(X) + E^2(X') - 2[Cov(X,X') + E(X)E(X')]$

$= Var(X) + Var(X') + E^2(X) + E^2(X') - 2E(X)E(X')$

$= Var(X) + Var(X') + [E(X) - E(X')]^2$ $\quad X, X'$ are iid, $\therefore E(X) = E(X')$, $Var(X) = Var(X')$

$= 2Var(X)$

$\therefore Var(X) = \frac{1}{2} E[(X-X')^2]$

$\sup|f(x_1 \cdots x_n) - f(x_1, \cdots x_{i-1}, x_i', x_{i+1} \cdots x_n)| \le c_i \quad 1 \le i \le n \quad (24)$

$Var^{(i)}(Z) = \frac{1}{2} E^{(i)}[(Z - Z_i')^2] \le \frac{1}{2} c_i^2 \quad Var(Z) \le E\sum Var^{(i)}(Z) \le \frac{1}{2} c_i^2$

13.2 Concentration and logarithmic Sobolev inequalities

Entropy is a scientific concept as well as a measurable physical property that is most commonly associated with a state of disorder, randomness or uncertainty.

Def 13.5

$$D(P\|Q) = \log N - H(X) = \log N + \sum P(X)\log P$$

Since $\quad D(P\|Q) = \sum P(X)\log\frac{P(X)}{q(X)} \qquad q(X) = \frac{1}{N}$

$$D(P\|Q) = \sum P(X)\log P(X) + P(X)\log N.$$

$$D(P\|Q) = \sum P(X)\log N + \sum P(X)\log P(X) = \log N + \sum P(X)\log P(X)$$

$$= \log N - H(X)$$

Def 13.6 $\quad H(X,Y) = -\sum_{\substack{x\in X\\ y\in Y}} P(X,Y)\log P(X,Y) \quad H(Y) = -\sum_{y\in Y} P(Y)\log P(Y)$

$$H(X|Y) = -\sum_{x\in X}\sum_{y\in Y} P(X,Y)\log P(X,Y) + \sum_{y\in Y} P(Y)\log P(Y)$$

here $P(Y) = \sum_{x\in X} P(X,Y)$

So $H(X|Y) = -\sum_{x\in X}\sum_{y\in Y} P(X,Y)\log P(X,Y) + \sum_{x\in X}\sum_{y\in Y} P(X,Y)\log P(Y)$

$$= -\sum_{\substack{x\in X\\ y\in Y}} P(X,Y)\log\frac{P(X,Y)}{P(Y)} = -\sum P(X,Y)\cdot\log P(X|Y)$$

Theorem 13.7.

$$H(X_1,\cdots X_n) \le H(X_1\cdots X_{i-1}, X_{i+1},\cdots X_n) + H(X_i|X_1\cdots X_{i-1})$$

$$\sum_{i=1}^{n}\left[H(X_1,\cdots X_{i-1}, X_{i+1}\cdots X_n) + H(X_i|X_1,\cdots X_{i-1})\right]$$

$$= \sum_{i=1}^{n}\left[H(X_1,\cdots X_{i-1}, X_{i+1},\cdots X_n)\right] + \left[H(X_1) + H(X_2|X_1) + \cdots + H(X_n|X_1\cdots X_{n-1})\right]$$

$$= \sum_{i=1}^{n}\left[H(X_1,\cdots X_{i-1}, X_{i+1},\cdots X_n)\right] + \left[H(X_1,\cdots X_n)\right] \ge nH(X_1,\cdots X_n)$$

$$\sum_{i=1}^{n} H(X_1,\cdots X_{i-1}, X_{i+1}\cdots X_n) \ge (n-1)H(X_1\cdots X_n)$$

$$\therefore H(X_1,\cdots X_n) \le \frac{1}{n-1}\sum_{i=1}^{n} H(X_1\cdots X_{i-1}, X_{i+1},\cdots X_n)$$

Theorem 13.8 will be shown in Boucheron et al. 2013, Section 4.6

Theorem 13.9

to proof $E[g(X)-g(EX)] = \inf E[g(X)-g(a)-g'(a)(X-a)]$

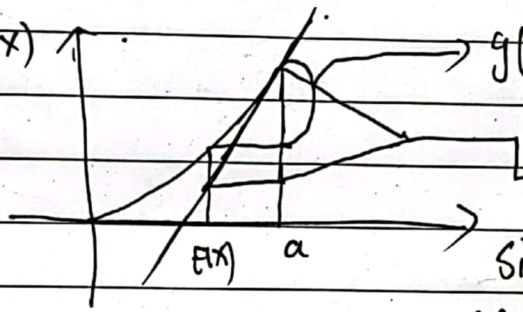
we need to proof ①: there exists a' make $E[g(X)-g(E(X))] = E[g(X)-g(a')-g'(a')(X-a')]$

② $E[g(X)-g(EX)] \le E[g(X)-g(a)-g'(a)(X-a)]$

Proof: ① when $a = E(X)$ $E[g(X)-g(EX)] = E[g(X)-g(a)-g'(a)(X-a)]$

② $E[g(X)-g(a)-g'(a)(X-a)] - E[g(X)-g(EX)]$
$= g(EX) - g(a) - g'(a)(EX-a)$

① $a > E(X)$



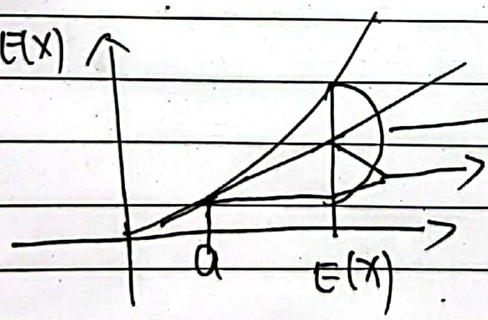
$\to g(EX) - g(a) < 0$

$\to g'(a)(E(X)-a) < 0$

Since $|g'(a)(E(X)-a)| \ge |g(E(X))-g(a)|$

$\therefore g(EX) - g(a) - g'(a)(EX-a) \ge 0$

② $a < E(X)$



$\to g(EX) - g(a) \ge 0$

$\to g'(a)(EX-a) \ge 0$

$\therefore g(EX) - g(a) - g'(a)(EX-a) \ge 0$

$\because$ g is convex and differentiable.

13.3 The Entropy method

Def 13.10. Entropy

$Ent(X) = E\Phi(X) - \Phi E(X)$ $\Phi(x) = x \cdot \log x$ for $x > 0$, $\Phi(0) = 0$

(here assuming $\Phi(X) = X^2$, then $Ent(X) = Var(X)$)

Remark 13.1

$$\mathrm{Ent}(X) = E\,g(X) - g(E(X)) = E[g(X) - g(EX)]$$

Since $E(X)$ is constant, $g(E(X))$ is constant, therefore $E[g(X)] = g(E(X))$

$$\therefore \mathrm{Ent}(X) = E[g(X) - g(E(X))] = \inf E[g(X) - g(a) - g'(a)(X-a)]$$

$$= \inf_{u>0} E[Y\log Y - u\log u - (\log u + 1)(Y-u)] = \inf E[Y(\log Y - \log u) - (Y-u)]$$

Theorem 13.11 for $E(Z)=1$, $q(x) = f(x)\cdot p(x)$ $Z = f(X)$

$$D(Q\|P) = \sum q(x)\log\frac{q(x)}{p(x)} = \sum q(x)\log f(x) = \cancel{\tfrac{1}{n}}\sum q(x)\log z$$
$$= \sum z\cdot p(x)\log z = E(Z\log Z)$$

$$D(Q\|P) \le \sum_{i=1}^{n}[D(Q\|P) - D(Q^{(i)}\|P^{(i)})] = E\sum_{i=1}^{n}\mathrm{Ent}^{(i)}(Z)$$

13.4 Gaussian Concentration inequality

Theorem 13.12

$$\mathrm{Ent}(g^2) \le 2E[\|\nabla g(X)\|^2]$$

proof is shown in Boucheron et al, 2013, Theorem 5.4.

Theorem 13.13

conditions: ① $X = (X_1, \dots X_n)$ is a vector of iid standard normal distributed random variables

② $f: R^n \to R$ is an L-Lipschitz function.

$$|f(x) - f(y)| \le L\|x - y\|$$

$$\mathrm{Ent}(e^{\lambda f}) \le 2E[\|\nabla e^{\lambda f/2}\|^2]$$

$$\nabla e^{\lambda f(x)/2} = \frac{\lambda}{2}e^{\lambda f(x)/2}\nabla f(x) \quad \therefore \|\nabla e^{\lambda f(x)/2}\|^2 = \frac{\lambda^2}{4}e^{\lambda f(x)}\cdot\|\nabla f(x)\|^2$$

Since $f(x)$ is L-Lipschitz

$$\|\nabla f(x)\|^2 \le L^2$$

$$\therefore \frac{\lambda^2}{4}e^{\lambda f(x)}\|\nabla f(x)\|^2 \le \frac{\lambda^2 L^2}{4}e^{\lambda f(x)}$$

$$\therefore \mathrm{Ent}(e^{\lambda f})\|\nabla f(x)\|^2 \le \frac{\lambda^2 L^2}{4}e^{\lambda f(x)}$$

$$\therefore \mathrm{Ent}(e^{\lambda f}) \le 2E[\|\nabla e^{\lambda f(x)/2}\|^2] \le \frac{\lambda^2 L^2}{2}E[e^{\lambda f(X)}]$$, let $g(x) = \exp(\lambda f(x)/2)$

$$\mathrm{Ent}(g^2(x)) = \mathrm{Ent}(\exp(\lambda f(x))) = E[e^{\lambda f(x)}\cdot\lambda f(x)] - E[e^{\lambda f(x)}]\cdot\log E[e^{\lambda f(x)}]$$

let $Z = f(X)$, then $\mathrm{Ent}(g(X)^2) = E[e^{\lambda Z} \cdot \lambda Z] - E[e^{\lambda Z}] \cdot \log E[e^{\lambda Z}]$

$= \lambda E[Z e^{\lambda Z}] - E[e^{\lambda Z}] \cdot \log E[e^{\lambda Z}]$

let $F(\lambda) = E e^{\lambda Z}$ $\quad = \lambda F'(\lambda) - F(\lambda) \log F(\lambda)$

$\Rightarrow \lambda F'(\lambda) - F(\lambda) \log F(\lambda) \le \frac{\lambda^2 L^2}{2} F(\lambda)$

$$\frac{F'(\lambda)}{\lambda F(\lambda)} - \frac{\log F(\lambda)}{\lambda^2} \le \frac{L^2}{2}$$

let $G(\lambda) = \log F(\lambda)$

$G'(\lambda) = \frac{F'(\lambda)}{F(\lambda)}$

so $\frac{G'(\lambda) \cdot F(\lambda)}{\lambda F(\lambda)} - \frac{G(\lambda)}{\lambda^2} \le \frac{L^2}{2}$

$$\frac{G'(\lambda)}{\lambda} - \frac{G(\lambda)}{\lambda^2} \le \frac{L^2}{2} \Rightarrow \frac{d}{d\lambda}\left(\frac{G(\lambda)}{\lambda}\right) \le \frac{L^2}{2}$$

since $\frac{d}{d\lambda}\left(\frac{G(\lambda)}{\lambda}\right) = \frac{\lambda G'(\lambda) - G(\lambda)}{\lambda^2} = \frac{G'(\lambda)}{\lambda} - \frac{G(\lambda)}{\lambda^2}$

$$\int_0^{\lambda_0} \frac{d}{d\lambda}\left(\frac{G\lambda}{\lambda}\right) d\lambda \le \int_0^{\lambda_0} \frac{L^2}{2} d\lambda$$

$$\left[\frac{G(\lambda)}{\lambda}\right]_0^{\lambda_0} \le \frac{L^2}{2} \lambda_0$$

$$\frac{G(\lambda_0)}{\lambda_0} \le \lim_{\lambda \to 0} \frac{G(\lambda)}{\lambda} + \frac{L^2}{2} \lambda_0$$

$\because \lim_{\lambda \to 0} \frac{G(\lambda)}{\lambda} = \frac{F'(0)}{F(0)} = E(Z)$

(L'Hospital's ~~rue~~ rule)

$\frac{G(\lambda_0)}{\lambda_0} \le \cancel{G(Z)} E(Z) + \frac{L^2}{2} \lambda_0$ ~~let $\lambda = \lambda_0$~~ replace λ_0 by λ

$G(\lambda) \le \lambda E(Z) + \frac{L^2 \lambda^2}{2} \Rightarrow F(\lambda) \le \exp\left(\lambda E Z + \frac{L^2 \lambda^2}{2}\right)$

by marcov's inequality

of interest $\Leftarrow P(Z > EZ + t) = P(\lambda Z > \lambda E Z + \lambda t) = P(e^{\lambda t} > e^{\lambda t + \lambda E Z})$

$\le \frac{E(e^{\lambda Z})}{e^{\lambda E Z + \lambda t}} = E(e^{\lambda Z}) e^{-\lambda E Z - \lambda t} = F(\lambda) e^{-\lambda E Z - \lambda t} \le e^{\lambda^2 L^2/2 - \lambda t}$

when $\lambda = \frac{t}{L^2}$, we can maximize $e^{\lambda^2 L^2/2 - \lambda t}$ to be $e^{-t^2/2L^2}$

Theorem 13.15

First according to (180), we can derive

$\text{Ent}^{(i)}(Y) \le E^{(i)}[Y(\log Y - \log u) - (Y-u)]$

$\Rightarrow \cancel{\text{Ent}^{(i)}}\ E^{(i)}[\Phi(Z)] - \Phi(E^{(i)}(Z)) \le E^{(i)}[Y(\log Y - \log u) - (Y-u)]$

$\text{Ent}^{(i)}(Y) = E^{(i)}[\Phi(Y)] - \Phi[E^{(i)}(Y)]$ $\quad \Phi(Y) = Y\log Y$

$\Rightarrow E^{(i)}(Y\log Y) - (E^{(i)}Y)\log(E^{(i)}Y) \le E^{(i)}(Y(\log Y - \log Y_i) - (Y - Y_i))$

$\Rightarrow$ let $Y = e^{\lambda Z}$, $Y_i = e^{\lambda Z_i}$ to obtain the result.

Theorem 13.16

ristiction: $\sum_{i=1}^{n}(Z - Z_i)^2 \le \nu$ (Similar to squared norm of gradient)

just based on the result of Theorem 13.15

and $\phi(-x) \le x^2/2$, where $-\cancel{x} = \lambda(Z - Z_i)$

the proof process is the same as theorem 13.13

Remark 13.5

$P(|Z - EZ| > t) = P(Z - EZ > t) + P((-Z) - E(-Z) > t)$

since $Z_i = \sup_{x_i'} f(x_1, \ldots x_{i-1}, x_i', x_{i+1}, \ldots x_n)$

then $-Z$ is the infimum of f.

so $P((-Z) - E(-Z) > t) = P(Z - EZ > t) \le e^{-t^2/2\nu}$