CONCENTRATION INEQ.

LECTURE 1

<u>Q1</u>: You toss a coin 10,000 times. How many H's do you see?

<u>Q2</u>: Coupon collector problem: N coupons, we need to collect them all. How many coupons do we need to sample to collect all N distinct coupons?

<u>Q3</u>: Largest common subsequence problem:
$(X_1, X_2, X_3, \dots, X_n)$ independent
$(Y_1, Y_2, Y_3, \dots, Y_n)$
What is the largest k s.t. $\exists$ indices $i_1, i_2, \dots, i_k$ and $j_1, j_2, \dots, j_k$ s.t. $X_{i_1} = Y_{j_1}, \dots, X_{i_k} = Y_{j_k}$
→ increasing
↳ increasing.

<u>Q1</u>: possible answer: 5,000

$$X_i = \begin{cases} 1 & \text{if the coin lands H} \\ 0 & \text{otherwise.} \end{cases}$$

$$S = \sum_{i=1}^{10,000} X_i \rightarrow \mathbb{E}[S] = \sum_{i=1}^{10,000} \mathbb{E}[X_i] = 5000.$$

$$\mathbb{P}(S = 5{,}000) = \binom{10{,}000}{5{,}000} \frac{1}{2^{10{,}000}} \simeq 0.008.$$

Possible answer: <u>Weak law of large numbers</u>
Let X_i be iid r.v.'s with finite expectation and finite second moment. Then for every $\varepsilon > 0$,

$$\mathbb{P}\left(\left| \frac{\sum X_i}{n} - \mu \right| \geq \varepsilon \right) \xrightarrow{n \to \infty} 0, \quad \text{where } \mu = \mathbb{E}[X_1].$$

→ For large enough n, # heads lies in $[n(\tfrac{1}{2} - \varepsilon), n(\tfrac{1}{2} + \varepsilon)]$

Asymptotic result (holds when $n \to \infty$).

Possible ans: <u>Central Limit Theorem</u>:
Let X_i be iid r.v.'s with finite mean and second moment. Then

$$\frac{\sum_{i=1}^{n} (X_i - \mu)}{\sqrt{n}\, \sigma} \xrightarrow{d} \mathcal{N}(0,1).$$

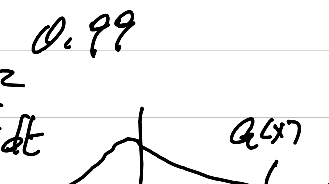
Here $\sigma^2 = Var(X_1)$.

$\sum_{i=1}^{n} (X_i - \mu)$ has deviations of the order $\sqrt{n} \cdot \sigma$.

Suppose we pretend 10,000 is big:

$$\sum_{i=1}^{10,000} X_i \in \left[5{,}000 - Q^{-1}(0.005) \cdot \frac{100}{2},\ 5000 + Q^{-1}(0.005) \cdot \frac{100}{2} \right]$$
$$\simeq [5{,}000 \pm 128] \text{ w.p. } 0.99$$

$$Q(x) = \mathbb{P}(Z \geq x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{t^2}{2}} dt$$
$Z \sim N(0,1)$



Non-asymptotic answers to Q1.
<u>Chebyshev's inequality</u>: Let X be any r.v. with mean μ, variance σ^2, then $\mathbb{P}(|X - \mu| \geq t) \leq \frac{\sigma^2}{t^2}$.

$$\mathbb{P}\left(\left| \sum_{i=1}^{10,000} X_i - 5{,}000 \right| \geq t \right) \leq \frac{10{,}000 \times 1/4}{t^2} = \frac{2500}{t^2}$$

If $t = 500$, the RHS is 0.01.
$\sum X_i \in [4500, 5500]$ with prob. 0.99.

Chernoff inequality (later on).

<u>Q2</u>: The number of samples $S = \sum_{i=1}^{N} X_i$, where $X_i \sim Geom(i/N)$.

$$\mathbb{E}[S] = \sum (N/i) = N\left(\sum 1/i\right) \simeq N \log N.$$

To solve Q1, Q2 we'll develop Chernoff Cramer method.

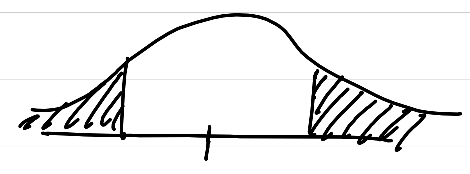
<u>Q3</u>: $f(X_1, \dots, X_n, Y_1, \dots, Y_n)$
"Talagrand's Principle" Any "smooth" function of independent r.v.'s "concentrates" around its mean.

<u>Modules:</u>

I: Chernoff-Cramer method (Sums).
II: Stein method (Bound Var $(f(X_1, \dots, X_n))$)
III: Entropy method (Bounds on the MGFs of $f(X_1, \dots, X_n)$).
IV: Transport method (Bounds on MGF but different technique).

CONCENTRATION INEQ.
LECTURE 2

Chernoff-Cramer method



$\mathbb{E}Z$

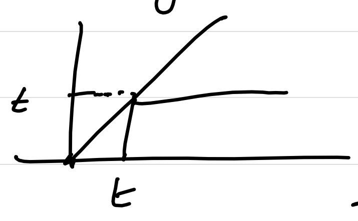
right tail bound

$\mathbb{P}(Z - \mathbb{E}Z \geq t)$

$\mathbb{P}(Z - \mathbb{E}Z \leq -t)$.

Theorem (Markov's Inequality) Let Y be a non-negative random variable with finite $\mathbb{E}Y$. Then for any $t>0$: $\mathbb{P}(Y \geq t) \leq \frac{\mathbb{E}[Y]}{t}$.

Proof:



$$y \geq t \cdot \mathbb{1}\{y \geq t\}$$

$$\Rightarrow \mathbb{E}[Y] \geq t\,\mathbb{P}(Y \geq t)$$

$$\Rightarrow \mathbb{P}(Y \geq t) \leq \frac{\mathbb{E}[Y]}{t}$$ $\square$

Suppose $\phi: \mathbb{R} \to \mathbb{R}_+$ is non-decreasing.

$\mathbb{P}(Y \geq t) \leq \mathbb{P}(\phi(Y) \geq \phi(t))$

(Let Y be a real-valued random variable)

$$\leq \frac{\mathbb{E}[\phi(Y)]}{\phi(t)}.$$

$Y = |Z - \mathbb{E}Z|$ for a r.v. Z.

Choose $\phi(t) = t^2$, and conclude

$$\mathbb{P}(|Z - \mathbb{E}Z| \geq t) \leq \frac{\mathbb{E}[|Z - \mathbb{E}Z|^2]}{t^2} = \frac{Var(Z)}{t^2} \quad \text{(Chebyshev inequality)}$$

we could pick $\phi(t) = t^q$ for any $q>0$ to conclude $\mathbb{P}(|Z - \mathbb{E}Z| \geq t) \leq \frac{\mathbb{E}[|Z - \mathbb{E}Z|^q]}{t^q}$

The bound with $q=2$ is more popular because $Var\left(\sum_{i=1}^n X_i\right) = \sum Var(X_i)$,

for $X_1, X_2, \dots, X_n$ independent.

To prove WLLN, note that

$$\mathbb{P}\left(\left|\frac{\sum (X_i - \mu)}{n}\right| \geq t\right) \leq \frac{\sigma^2 n}{n^2 t^2} = \frac{\sigma^2}{n t^2}$$

$\mathbb{E}X_i = \mu$

"Tensorisation"

Chernoff-Cramer method

Consider $\phi(t) = e^{\lambda t}$ for $\lambda > 0$.

$$\mathbb{P}(Z \geq t) \leq \frac{\mathbb{E}[e^{\lambda Z}]}{e^{\lambda t}}$$

Define $F(\lambda) = \mathbb{E}[e^{\lambda Z}]$ which is called the moment generating function (MGF) of Z.

$\psi_Z(\lambda) := \log \mathbb{E}[e^{\lambda Z}]$

$$F(\lambda) = \mathbb{E}\left[1 + \lambda Z + \frac{\lambda^2 Z^2}{2!} + \dots\right] = \sum_{i=0}^{\infty} \frac{\lambda^i \mathbb{E}[Z^i]}{i!}$$

If $X_1, X_2, \dots, X_n$ are independent, and $Z = \sum X_i$, then

$$\psi_Z(\lambda) = \log \mathbb{E}[e^{\lambda Z}] = \log \mathbb{E}[e^{\lambda \sum X_i}]$$

$$= \log\left(\prod_{i=1}^n \mathbb{E}[e^{\lambda X_i}]\right)$$

$$= \sum_{i=1}^n \log \mathbb{E}[e^{\lambda X_i}] = \sum_{i=1}^n \psi_{X_i}(\lambda)$$

Coming back to the earlier bound

$\mathbb{P}(Z \geq t) \leq \frac{\mathbb{E}[e^{\lambda Z}]}{e^{\lambda t}}$ for any $\lambda > 0$

We can infimise the RHS to get

$$\mathbb{P}(Z \geq t) \leq \inf_{\lambda \geq 0} e^{(\psi_Z(\lambda) - \lambda t)}$$

Define $\psi_Z^*(t) := \sup_{\lambda \geq 0} \lambda t - \psi_Z(\lambda)$, and

write $\mathbb{P}(Z \geq t) \leq e^{-\psi_Z^*(t)}$

This is the Chernoff bound, ψ_Z^* is the Chernoff-Cramer transform of ψ_Z.

Properties of ψ_Z and ψ_Z^*

① ψ_Z is convex and infinitely differentiable on $(0, b)$, where $b = \sup\{\lambda : \psi_Z(\lambda) < \infty\}$.

(Smoothness follows from infinite differentiability of the mgf where it is defined)

Convexity:

$f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta) f(y)$

$F(\theta x + (1-\theta)y) = \mathbb{E}[e^{\theta x Z} \cdot e^{(1-\theta)yZ}]$

Hölder's ineq $\mathbb{E}[|XY|] \leq \mathbb{E}[|X|^p]^{1/p} \cdot \mathbb{E}[|Y|^q]^{1/q}$

where $1/p + 1/q = 1$.

Choose $1/p = \theta$, $1/q = 1-\theta$ to conclude.

② $\psi_Z^* \geq 0$, and it is convex

(Follows from definition)

③ Suppose $t > \mathbb{E}[Z]$, then $\psi_Z^*(t) = \sup_{\lambda} \lambda t - \psi_Z(\lambda)$

$\mathbb{P}(Z - \mathbb{E}[Z] \geq t')$

we show that if $\lambda < 0$, then $\lambda t - \psi_Z(\lambda) \leq 0$

$\mathbb{E}[e^{\lambda Z}] \geq e^{\lambda \mathbb{E}[Z]}$ (by Jensen).

so $\psi_Z(\lambda) \geq \lambda \mathbb{E}[Z]$

$$\Rightarrow \lambda t - \psi_Z(\lambda) \leq \lambda(t - \mathbb{E}[Z]) \leq 0$$

CONCENTRATION INEQ.

LECTURE 3

<u>Example:</u> $Z \sim N(0, v)$. We want to upper bound $\mathbb{P}(Z \geq t)$ for $t > 0$.

$$\mathbb{E}[e^{\lambda Z}] = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi v}} e^{-\frac{t^2}{2v}} e^{\lambda t} \, dt$$

$$= \cdots = e^{v\lambda^2/2}$$

$$\psi_Z^*(t) = \sup_{\lambda \geq 0} \lambda t - \frac{\lambda^2 v}{2} \quad (t > 0 = \mathbb{E}[Z]$$

$\Rightarrow$ can ignore constraint $)$. $\Rightarrow \psi_Z^*(\lambda) = \sup_{\lambda} \lambda t - \frac{\lambda^2 v}{2}$

$\Rightarrow t - \lambda v = 0 \Rightarrow \lambda = t/v$ is the optimiser.

Plug in, $\psi_Z^*(t) = \frac{t^2}{v} - \frac{t^2}{v2} = \frac{t^2}{2v}$

$$\mathbb{P}(Z \geq t) \leq \exp(-t^2/2v)$$

Sub. Gaussian r.v.'s

<u>Definition</u> a r.v. Y with $\mathbb{E}[Y] = 0$ is sub-Gaussian with variance parameter v if $\psi_Y(\lambda) \leq \frac{\lambda^2 v}{2}$ for all $\lambda \in \mathbb{R}$.

The set of sub. Gaussian r.v. with variance parameter v is $\mathcal{G}(v)$.

<u>Verify</u>

(1) If $Y \in \mathcal{G}(v)$ then $\mathbb{P}(Y \geq t) \leq e^{-\frac{t^2}{2v}}$, $\mathbb{P}(Y \leq -t) \leq e^{-t^2/2v}$

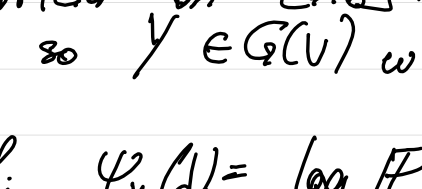
(2) If $Y_i \in \mathcal{G}(v_i)$ and independent, then $\sum_{i=1}^n Y_i \in \mathcal{G}(\sum_{i=1}^n v_i)$

(3) If $Y \in \mathcal{G}(v)$, then $Var(Y) \leq v$.

<u>Theorem:</u> The following are equivalent for suitable v, b, c, d.

1. $Y \in \mathcal{G}(v)$
2. $\max\{\mathbb{P}(Y \geq t), \mathbb{P}(Y \leq t)\} \leq e^{-\frac{t^2}{2b}}$ $\forall t > 0$.
3. $\mathbb{E}[Y^{2q}] \leq q! \, C^q$ for all $q \geq 1$.
4. $\mathbb{E}[e^{dY^2}] \leq 2$ (No proof).

Bounded random variables



$\mathbb{E}Y = 0$

<u>Lemma</u> (Hoeffding's lemma) Let Y be supported on $[a, b]$. Then $\psi_Y''(\lambda) \leq \frac{(b-a)^2}{4}$ and so $Y \in \mathcal{G}(v)$ with $v = \frac{(b-a)^2}{4}$.

<u>Proof:</u> $\psi_Y(\lambda) = \log \mathbb{E}[e^{\lambda Y}]$.

$$\psi_Y'(\lambda) = \frac{\mathbb{E}[Y \cdot e^{\lambda Y}]}{\mathbb{E}[e^{\lambda Y}]}, \quad \psi_Y''(\lambda) = \frac{\mathbb{E}[e^{\lambda Y}] \cdot \mathbb{E}[Y^2 e^{\lambda Y}]}{\mathbb{E}[e^{\lambda Y}]^2} - \frac{(\mathbb{E}[Y \cdot e^{\lambda Y}])^2}{\mathbb{E}[e^{\lambda Y}]^2}$$

suppose $Y \sim P$

$$\psi_Y''(\lambda) = \int y^2 \underbrace{\frac{e^{\lambda y}}{\mathbb{E}[e^{\lambda Y}]} dP(y)}_{dQ(y)} - \left(\int y \underbrace{\frac{e^{\lambda y}}{\mathbb{E}[e^{\lambda Y}]} dP}_{dQ(y)}\right)^2$$

$\psi_Y''(\lambda) = Var(Y)$ when $Y \sim Q$, and observe that Q is supported on $[a, b]$.

If $Y \in [a, b]$ a.s., then $Var(Y)$

$$= Var\left(Y - \frac{(a+b)}{2}\right) \leq \mathbb{E}\left[\left(Y - \left(\frac{a+b}{2}\right)\right)^2\right] \leq \frac{(b-a)^2}{4}$$

To finish the last part, observe that

$$\psi_Y(\lambda) = \psi_Y(0) + \lambda \psi_Y'(0) + \frac{\lambda^2}{2} \psi_\lambda''(\theta), \quad \theta \in (0, \lambda).$$

$$= \frac{\lambda^2}{2} \psi_\lambda''(\theta) \leq \frac{\lambda^2}{2} \frac{(b-a)^2}{4}. \qquad \square$$

<u>Theorem</u> (Hoeffding's inequality). Let Y_i be ind. r.v. supported on $[a_i, b_i]$. Then

$$\mathbb{P}\left(\sum_{i=1}^n (Y_i - \mathbb{E}Y_i) \leq t\right) \leq \exp\left(\frac{-2t^2}{\sum (b_i - a_i)^2}\right)$$

<u>Theorem</u> <u>Bennett's Inequality</u>

For $1 \leq i \leq n$, let X_i be independent r.v. satisfying $\mathbb{E}[X_i] = 0$ and $Var(X_i) = \sigma_i^2$ and let $v = \sum \sigma_i^2$. Also, assume $|X_i| \leq C$ a.s. for all i. Then

$$\mathbb{P}\left(\sum_{i=1}^n X_i \geq t\right) \leq \exp\left(-\frac{v}{c^2} h_1\left(\frac{ct}{v}\right)\right)$$

where $h_1(x) = (1+x)\log(1+x) - x$ for $x > 0$.

$$h_1(x) \geq \frac{x^2}{2(1 + x/3)}$$

$$\Rightarrow \mathbb{P}\left(\sum_{i=1}^n X_i \geq t\right) \leq \exp\left(-\frac{t^2}{2(v + ct/3)}\right)$$

<u>Example:</u> $X_i \sim Bern(p_n)$ be independent for $1 \leq i \leq n$.

(Hoeffding) $\mathbb{P}\left(\sum_i (X_i - \mathbb{E}X_i) \geq t\right) \leq e^{-\frac{2t^2}{n}}$

(Bennett) $\mathbb{P}\left(\sum (X_i - \mathbb{E}X_i) \geq t\right) \leq \exp\left(-\frac{t^2}{np_n(1-p_n) + t/3}\right)$

$v = \sum p_n(1-p_n) = np_n(1-p_n)$.

If $p_n \ll 1$, say $p_n = \frac{1}{\sqrt{n}}$, Hoeffding is the same, Bennett's will be

$$e^{-\frac{t^2}{(\sqrt{n} + t/3)}}$$

LECTURE 4

Bennett's inequality:
(Proof)

$$\mathbb{E}[e^{\lambda X_i}] = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \mathbb{E}[X_i^k]$$

$$\le 1 + \sum_{k=2}^{\infty} \frac{\lambda^k}{k!} \mathbb{E}[c^{k-2} X_i^2]$$

$$= 1 + \sum_{k=2}^{\infty} \frac{\lambda^k}{k!} c^{k-2} \sigma_i^2$$

$$= 1 + \frac{\sigma_i^2}{c^2}(e^{\lambda c} - \lambda c - 1)$$

$(1+x) \le e^x$

$$\mathbb{E}[e^{\lambda X_i}] \le \exp\left(\frac{\sigma_i^2}{c^2}(e^{\lambda c} - \lambda c - 1)\right).$$

This implies $\mathbb{E}[e^{\lambda S}] \le \exp\left(\frac{v}{c^2}(e^{\lambda c} - c\lambda - 1)\right)$

$$\psi_S(\lambda) \le \underbrace{\frac{v}{c^2}(e^{\lambda c} - \lambda c - 1)}_{\tilde{\psi}}$$

$\psi_S^*(t) \ge \tilde{\psi}^*(t)$

$\implies \mathbb{P}(S \ge t) \le \exp(-\psi_S^*(t)) \le \exp(-\tilde{\psi}^*(t))$

(Example Sheet: Calculate $\tilde{\psi}^*$)

$$= \exp\left(-\frac{v}{c^2} h_1\left(\frac{ct}{v}\right)\right) \quad \square$$

Efron - Stein Inequality

A bound on $Var(Z)$, where $Z = f(X_1, X_2, \dots, X_n)$ where X_i are independent. If $Z = \sum X_i$, then $Var(Z) = \sum Var(X_i)$. This holds even for uncorrelated X_i's.

If: $Z - \mathbb{E}Z = \sum_{i=1}^{n} \Delta_i$, where Δ_i are uncorrelated and 0-mean.

$$Var(Z) = \sum Var(\Delta_i) = \sum \mathbb{E}[\Delta_i^2]$$

Define $E_i Z = \mathbb{E}[Z \mid X_{1:i}]$

$X_{1:i} = (X_1, \dots, X_i)$.

Set $\Delta_i = E_i Z - E_{i-1} Z$

$$Z - \mathbb{E}Z = \sum_{i=1}^{n} \Delta_i, \quad \mathbb{E}\Delta_i = \mathbb{E}E_i Z - \mathbb{E}E_{i-1} Z = \mathbb{E}Z - \mathbb{E}Z = 0.$$

suppose $i < j$

$$\mathbb{E}[\Delta_i \Delta_j] = \mathbb{E}[\mathbb{E}[Z \mid X_{1:i}] \mathbb{E}[Z \mid X_{1:j}]]$$

$$= \mathbb{E}[\mathbb{E}[\Delta_i \Delta_j \mid X_{1:i}]]$$

$$= \mathbb{E}[\Delta_i \mathbb{E}[\Delta_j \mid X_{1:i}]]$$

Note that $\mathbb{E}[\Delta_j \mid X_{1:i}] =$

$$\mathbb{E}[\mathbb{E}[Z \mid X_{1:j}] \mid X_{1:i}] - \mathbb{E}[\mathbb{E}[Z \mid X_{1:j-1}] \mid X_{1:i}]$$

$$= \mathbb{E}[Z \mid X_{1:i}] - \mathbb{E}[Z \mid X_{1:i}] = 0.$$

$Var(Z) = \sum \mathbb{E}[\Delta_i^2]$, this holds regardless of independence

$\Delta_i = E_i Z - E_{i-1} Z$

Define $E^{(i)} Z = \mathbb{E}[Z \mid X_{1:i-1}, X_{i+1:n}]$.

$E_i E^{(i)} Z = \mathbb{E}[\mathbb{E}[Z \mid X^{(i)}] \mid X_{1:i}]$

$(X^{(i)} = (X_{1:i-1}, X_{i+1:n}))$.

$X_{1:i-1} = A$, $X_i = B$, $X_{i+1:n} = C$.

A, B, C are independent.

$$= \mathbb{E}[\mathbb{E}[Z \mid A, C] \mid A, B].$$

$$= \mathbb{E}[Z \mid A] = E_{i-1} Z$$

We have $Z - \mathbb{E}Z = \sum \underbrace{E_i(Z - E^{(i)}(Z))}_{\Delta_i}$

$$(\Delta_i)^2 = (E_i(Z - E^{(i)} Z))^2 \le E_i((Z - E^{(i)}(Z))^2).$$

$$Var(Y \mid X) = \mathbb{E}[(Y - \mathbb{E}[Y \mid X])^2 \mid X].$$

$$Var(Z \mid X^{(i)}) =: Var^{(i)}(Z)$$

$$= \mathbb{E}[(Z - E^{(i)} Z)^2 \mid X^{(i)}]$$

$$Var(Z) = \sum \mathbb{E}[\Delta_i^2]$$

$$\le \sum \mathbb{E}[(Z - E^{(i)} Z)^2]$$

$$= \mathbb{E}\left[\sum Var^{(i)}(Z)\right]$$

This is the Efron - Stein Inequality.

Lecture 5

Theorem (Efron–Stein Inequality) Let $X_1, X_2, \ldots, X_n$ be independent r.v.'s and let $Z = f(X_1, X_2, \ldots, X_n)$ be a square integrable function of $X = X_{1:n}$. Then

$$\mathrm{Var}(Z) \le \sum_{i=1}^n \mathbb{E}[(Z - E^{(i)}Z)^2]$$

$$E^{(i)}Z = \mathbb{E}[Z \mid X^{(i)}]$$

$$X^{(i)} = (X_{1:i-1}, X_{i+1:n})$$

$$= \mathbb{E}[\sum \mathrm{Var}^{(i)}(Z)] =: v$$

Define $X_1', X_2', \ldots, X_n'$ to be independent copies of $X_1, X_u, \ldots, X_u$. Set

$$Z_i' = f(X^{(i)}, X_i')$$

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

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

$$= \frac{1}{2}\sum_{i=1}^n \mathbb{E}[(Z - Z_i')^2]$$

here $x_+ = \max\{0, x\}$, $x_- = \max\{0, -x\}$

Also, $v = \inf_{Z_1, \ldots, Z_n} \sum_{i=1}^n \mathbb{E}[(Z - Z_i)^2]$ where

Z_i is some function of $X^{(i)}$.

Proof: We've done the first part already. For the second part, note that if X, Y are iid, then $\mathrm{Var}(X) = 1/2\, \mathbb{E}[(X-Y)^2]$

(use conditional version.)

$= \mathbb{E}[(X-Y)_+^2] = \mathbb{E}[(X-Y)_-^2]$

For the third part,

$$\mathrm{Var}(X) = \inf_a \mathbb{E}[(X-a)^2]$$

$$\mathrm{Var}^{(i)}(Z) = \inf_{Z_i} \mathbb{E}[(Z - Z_i)^2 \mid X^{(i)}], \text{ where}$$

Z_i is $X^{(i)}$-measurable $\square$

Functions with bounded-differences property:

f satisfies the bounded-differences property with constants $c_1, c_2, \ldots, c_n$ if

$$\sup_{x_1, x_2, \ldots, x_n, x_i'} |f(x_1, x_2, \ldots, x_n) - f(x_1, x_2, \ldots, x_{i-1}, x_i', x_{i+1}, \ldots, x_n)| \le c_i.$$

If $Z = f(X_1, \ldots, X_n)$ where X_i are independent, we'll show that $\mathrm{Var}(Z) \le \sum \frac{c_i^2}{4}$

To see this, set

$$Z_i = \frac{\inf_{x_i} f(X^{(i)}, x_i) + \sup_{x_i} f(X^{(i)}, x_i)}{2}$$

$$v \le \sum \mathbb{E}[(Z - Z_i)^2] = \frac{\sum c_i^2}{4}$$

Example 1: $X_1, X_2, \ldots, X_n$ are independent, supp on $[0,1]$. $Z = f(X_{1:n})$ is the smallest ~~no~~ of size one bins needed to "pack" $X_1, X_2, \ldots, X_n$.

f satisfies bounded-diff. property with $c_i = 1\ \forall i$.

So $\mathrm{Var}(Z) \le n/4$.

$X_i \overset{iid}{\sim} \mathrm{Unif}([0,1])$

$\mathbb{E}[f(X_1, \ldots, X_n)] \simeq n \cdot c$

Example 2: $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \mathrm{Ber}(1/2)$

$f(X_{1:n}, Y_{1:n})$ = longest common subseq. between $X_{1:n}$ & $Y_{1:n}$.

So $\mathrm{Var}(Z) \le n/2$.

$\mathbb{E}[Z] \sim [0.75n, 0.837n]$.

Example 3: $\chi(G)$ is the smallest number of colours needed to colour vertices of a graph G s.t. no two neighbouring vertices share a colour.

Let $X_{ij} \sim$ iid $\mathrm{Ber}(p)$ for $1 \le i < j \le n$ and

$$\chi(G) = f(\{X_{ij}\}_{1 \le i < j \le n})$$

$$\mathrm{Var}(\chi(G)) \le \frac{\binom{n}{2}}{4} \sim n^2.$$

$$\mathbb{E}[\chi(G)] \simeq \frac{n}{\log n}$$

We can fix this bound by considering

$$Y_i = (X_{1,i+1}, \ldots, X_{i,i+1})$$

Observe that $Y_1, Y_2, \ldots, Y_{n-1}$ are independent and $\chi(G) = \hat{f}(Y_1, \ldots, Y_{n-1})$

Check that $\hat{f}$ also satisfies bounded differences with $c_i = 1$.

$$\Longrightarrow \mathrm{Var}(\chi(G)) \le \frac{n-1}{4}$$

Theorem (Convex Poincaré Inequality)

$X_1, X_2, \ldots, X_n$ ind over $[0,1]$.

f is a "separately convex function" over $[0,1]^n$.

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

LECTURE 6

Poincaré Inequalities:

Convex Poincaré $\qquad \mathrm{Var}(f(X)) \le \mathbb{E}[\|\nabla f(X)\|^2]$
Gaussian Poincaré

Theorem (Convex Poincaré Inequality)
Let $X_1, X_2, \dots, X_n$ be ind. supp on $[0,1]$.
Let $f : [0,1]^n \to \mathbb{R}$ be a separetly convex function whose partial derivatives exist. Then $Z = f(X_{1:n})$ satisfies $\mathrm{Var}(Z) \le \mathbb{E}[\|\nabla f(X)\|^2]$

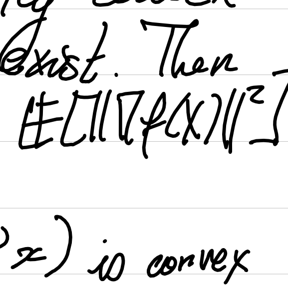
Sep. convex means $f_{x^{(i)}}(x) := f(x^{(i)}; x)$ is convex in x for each i, and every $x^{(i)}$.

Proof: $\mathrm{Var}(Z) \le \sum_{i=1}^n \mathbb{E}[(Z - Z_i)^2]$ where Z_i is $X^{(i)}$ - measurable.
$Z_i = \inf_x f(X^{(i)}, x)$

inf attained

$$0 \le Z - Z_i = f(X_1, \dots, X_n) - f(X_1, X_2, \dots, X^*, \dots, X_n)$$
$$= f(X^{(i)}, X_i) - f(X^{(i)}, X^*) \le$$

If g is a convex function, then
$g(y) \ge g(x) + g'(x) \cdot (y - x)$.



$$\le \frac{\partial f}{\partial x_i}(x) \cdot (x_i - x^*)$$

Squaring, $(Z - Z_i)^2 \le \left(\frac{\partial f}{\partial x_i}\right)^2$

$$\Rightarrow \quad \sum_i (Z - Z_i)^2 \le \|\nabla f(X)\|^2$$

$\Rightarrow$ take $\mathbb{E}$ to complete the proof.

Example: $X \in \mathbb{R}^{m \times d}$ with $\mathbb{E}[X_{ij}] = 0$, all entries ind. and supp. on $[-1, 1]$
$$\sigma_1(X) = \max_{\|v\|_2 = 1} \|Xv\| = \max_{\|u\|_2 = 1} \max_{\|v\|_2 = 1} u^T X v$$

$$\sigma_1(A + B) \le \sigma_1(A) + \sigma_1(B).$$

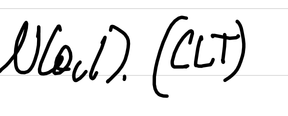
$$|\sigma_1(A) - \sigma_1(B)| \le \sigma_1(A - B) \text{ (rewriting the } \Delta\text{-ineq.)}$$

Claim: $\sigma_1(A)^2 \le \sum_{i,j} A_{ij}^2$ (Prove it using C-S).

$|\sigma_1(A) - \sigma_1(B)| \le \sigma_1(A - B) \le \|A - B\|_F$, this means that $\|\nabla \sigma_1(X)\| \le 1$. Using the convex Poincaré inequality:
$$\mathrm{Var}(\sigma_1(X)) \le 4.$$

Theorem: Let $X_1, X_2, \dots, X_n$ be iid $\sim N(0,1)$. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuously differentiable function. Then $\mathrm{Var}(f(X_{1:n})) \le \mathbb{E}[\|\nabla f(X)\|^2]$

Proof: Claim: Enough to establish the $n=1$ case.
Proof: Assume the $n=1$ case
$$\mathrm{Var}(f(X_1, X_2, \dots, X_n)) \le \sum_{i=1}^n \mathbb{E}[\mathrm{Var}^{(i)}(Z)]$$
$$\mathrm{Var}^{(i)}(Z) = \mathbb{E}[(Z - E^i Z)^2 \mid X^{(i)}]$$
$$\le \mathbb{E}\left[\frac{\partial f}{\partial x^i}(x)^2 \,\middle|\, X^{(i)}\right], \text{ (using the } n=1 \text{ case).}$$
Summing over i & taking expectations:
$$\mathrm{Var}(f(X_1, X_2, \dots, X_n)) \le \mathbb{E}[\|\nabla f(X)\|^2].$$
Let's prove the $n=1$ Poincaré inequality
Let $X_i \overset{iid}{\sim} \underbrace{\text{symmetric Ber}(1/2)}_{\text{Rademacher}}$



$S_n := \sum_{i=1}^n X_i$, then $\frac{S_n}{\sqrt{n}} \xrightarrow{d}$ to $N(0,1)$. (CLT)

$$\mathrm{Var}(f(S_n)) \le \sum_{i=1}^n \mathbb{E}[\mathrm{Var}^{(i)}(f(S_n))]$$
$$\mathrm{Var}^{(i)}(S_n) = \frac{1}{4}\left(f\left(S_n - \frac{X_i}{\sqrt{n}} + \frac{1}{\sqrt{n}}\right) - f\left(S_n - \frac{X_i}{\sqrt{n}} - \frac{1}{\sqrt{n}}\right)\right)^2$$

For the rest of the proof, assume f is twice continuously differentiable on a bounded domain.

$$f\left(S_n - \frac{X_i}{\sqrt{n}} + \frac{1}{\sqrt{n}}\right) = f(S_n) + f'(S_n)\left(\frac{1 - X_i}{\sqrt{n}}\right) + f''(\theta_1)\frac{(1 - X_i)^2}{2n}$$
$$f\left(S_n - \frac{X_i}{\sqrt{n}} - \frac{1}{\sqrt{n}}\right) = f(S_n) - f'(S_n)\left(\frac{1 + X_i}{\sqrt{n}}\right) + f''(\theta_2)\frac{(1 + X_i)^2}{2n}$$
$$\left|f\left(S_n - \frac{X_i}{\sqrt{n}} + \frac{1}{\sqrt{n}}\right) - f\left(S_n - \frac{X_i}{\sqrt{n}} - \frac{1}{\sqrt{n}}\right)\right| \le |f'(S_n)| \cdot \frac{2}{\sqrt{n}} + \frac{2K}{n}$$

where $|f''| \le K$.
Squaring, $\quad (f(\dots) - f(\dots))^2 \le f'(S_n)^2 \cdot \frac{4}{n} + \frac{4K^2}{n^2} + \frac{8K}{n^{3/2}}|f'(S_n)|$

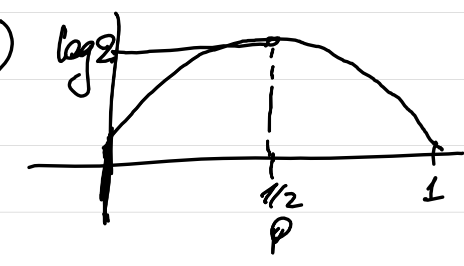
Taking $\mathbb{E}$ and summing from $i=1$ to n,
$$\mathrm{Var}(f(S_n)) \le \sum_{i=1}^n \mathbb{E}[\mathrm{Var}^{(i)} Z] \le \mathbb{E}\left[f'(S_n)^2 + \frac{4K^2}{n} + \frac{8K \cdot K}{n^{1/2}}\right]$$
Taking $n \to \infty$, CLT $\Rightarrow$
$$\mathrm{Var}(f(Z)) \le \mathbb{E}[f'(Z)^2], \; Z \sim N(0,1) \square$$

LECTURE 7

Entropy

$X \sim P_X$ on a discrete set $\mathcal{X}$, then the Shannon entropy of X is $H(X) = H(P_X) = -\sum_x P_X(x) \log P_X(x)$.
$|\mathcal{X}| < \infty$.

$H(Ber(p))$



Def (Relative entropy or Kullback-Leiber divergence). Given two prob. measures P and Q over a discrete set $\mathcal{X}$, define $D(Q \| P) = \sum_x q(x) \log \frac{q(x)}{p(x)}$, where where p and q are the pmf's of P and Q.

Properties: (Ex Sheet 2)
1. $D(Q\|P) \geq 0$, $D(Q\|P) = 0 \Leftrightarrow Q \equiv P$.
2. $D(Q\|P)$ is convex
$(P_\lambda, Q_\lambda) = \lambda(P_1, Q_1) + (1-\lambda)(P_2, Q_2)$
$D(Q_\lambda \| P_\lambda) \leq \lambda D(Q_1\|P_1) + (1-\lambda) \cdot D(Q_2 \| P_2)$

Suppose $|\mathcal{X}|$ is finite, then
$$D(Q \| \underset{\text{uniform}}{\mathcal{U}}) = \log|\mathcal{X}| - H(Q)$$

Chain rule of Shannon entropy

$$H(Y|X) = \sum_x H(Y|X=x) P_X(x) \quad \text{Concavity of } H$$
$$= \sum_x H(P_{Y|X=x}) \cdot P_X(x) \;(\leq H(Y))$$

$$H(X,Y) = H(X) + H(Y|X) = H(Y) + H(X|Y).$$

Theorem: $H(X_1, X_2, \dots, X_n) = \sum_{i=1}^n H(X_i | X_{1:i-1})$

Proof: $H(X_1, X_2, \dots, X_n) = \mathbb{E}[-\log P_{X_{1:n}}(X_{1:n})]$
$$= \mathbb{E}[-\log \prod P_{X_i|X_{1:i-1}}(X_i|X_{1:i-1})]$$
$$= \sum_{i=1}^n \underbrace{\mathbb{E}[-\log P_{X_i|X_{1:i-1}}(X_i|X_{1:i-1})]}_{= H(X_i|X_{1:i-1})}$$

Chain rule for KL-divergence

Theorem: let P, Q be measures on $\mathcal{X}^n$
then $D(Q\|P) = \boxed{}$

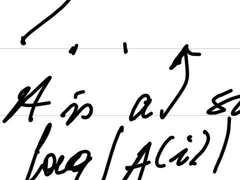
Proof: $D(Q\|P) = \sum_{x_{1:n}} q(x_{1:n}) \log\left(\frac{q(x_{1:n})}{p(x_{1:n})}\right)$
$$= \mathbb{E}_Q\left[\log\left(\frac{q(X_{1:n})}{p(X_{1:n})}\right)\right]$$
$$= \mathbb{E}_Q\left[\log\left(\prod \frac{q(X_i|X_{1:i-1})}{p(X_i|X_{1:i-1})}\right)\right]$$
$$= \sum_{i=1}^n \mathbb{E}_Q\left[\log \frac{q(X_i|X_{1:i-1})}{p(X_i|X_{1:i-1})}\right]$$
$$= \sum_{i=1}^n \mathbb{E}_{Q_{X_{1:i}}}\left[\log \frac{q(X_i|X_{1:i-1})}{p(X_i|X_{1:i-1})}\right]$$

Let's consider the i^{th} term $\sum_{x_{1:i}} q(x_{1:i}) \log \frac{q(x_i|x_{1:i-1})}{p(x_i|x_{1:i-1})}$
$$= \sum_{x_{1:i-1}} q(x_{1:i-1}) \left[\sum_{x_i} q(x_i|x_{1:i-1}) \log \frac{q(x_i|x_{1:i-1})}{p(x_i|x_{1:i-1})}\right]$$
$$= \mathbb{E}_{Q_{X_{1:i-1}}}\left[D(Q_{X_i|X_{1:i-1}} \| P_{X_i|X_{1:i-1}})\right]$$
$$:= D\left(Q_{X_i|X_{1:i-1}} \| P_{X_i|X_{1:i-1}} \,\middle|\, Q_{X_{1:i-1}}\right)$$
$$D(Q\|P) = \sum_{i=1}^n D\left(Q_{X_i|X_{1:i-1}} \| P_{X_i|X_{1:i-1}} \,\middle|\, Q_{X_{1:i-1}}\right).$$

Usually, have $P = P_1 \otimes P_2 \otimes \dots \otimes P_n$, which simplifies the formula. What if
$Q = Q_1 \otimes Q_2 \otimes \dots \otimes Q_n$.
$$D(Q\|P) = \sum_{i=1}^n D(Q_i \| P_i).$$

Theorem (Han's Inequality for Shanon Entropy)
$$H(X_{1:n}) \leq \sum \frac{H(X^{(i)})}{n-1}$$

Ex: $\mathbb{Z}^n$



$X_{1:n} \sim Unif(\text{points})$
$H(X_{1:n}) = \log|A|$
where A is a subset of $\mathbb{Z}^n$.
$$\log|A| \leq \frac{\log|A^{(i)}|}{n-1}$$
$A^{(i)}$ → projection onto plane with i^{th} coord $= 0$

$$|A| \leq \left(\prod_{i=1}^n A^{(i)}\right)^{1/n}. \quad \text{(Loomis-Whitney Ineq.)}$$

LECTURE 8

Theorem: $H(X_{1:n}) \leq \frac{1}{(n-1)} \sum_{i=1}^{n} H(X^{(i)})$

Lemma: $H(X|Y,Z) \leq H(X|Y)$

Proof:

$$LHS = \sum_{y,z} H(P_{X|Y=y,Z=z}) P_{YZ}(y,z)$$

$$= \sum_y P_Y(y) \left[\sum_z P_{Z|Y}(z|y) H(P_{X|Y=y,Z=z}) \right]$$

(Concavity of H) $\leq \sum_y P_Y(y) H\Big(\underbrace{\sum_z P_{Z|Y}(z|y) P_{X|Y=y,Z=z}}_{} \Big)$

$$= \sum_y P_Y(y) \cdot H(P_{X|Y=y})$$

$$= H(X|Y)$$

$$\sum_z \frac{\cancel{P_{Z,Y}(z,y)}}{P_Y(y)} \cdot \frac{P_{X,Y,Z=z}}{\cancel{P_{Z,Y}(z,y)}} = \frac{P_{X,Y=y}}{P_Y(y)} = P_{X|Y=y}$$

Proof of Theorem:

$$H(X_{1:n}) = H(X^{(i)}) + H(X_i | X^{(i)})$$

$$\leq H(X^{(i)}) + H(X_i | X_{1:i-1})$$

Sum over all i,

$$nH(X_{1:n}) \leq \sum_{i=1}^{n} H(X^{(i)}) + H(X_{1:n})$$

Rearrange and conclude $\square$

Theorem (Hahn's inequality for KL-divergence)
Let X be a countable set, and let P and Q be measures of X^n, and $P = P_1 \otimes \cdots \otimes P_n$.
Then $D(Q\|P) \geq \frac{1}{n-1} \sum_{i=1}^{n} D(Q_{X^{(i)}} \| P_{X^{(i)}})$

Equivalently,

$$D(Q\|P) \leq \sum_{i=1}^{n} D(Q_{X_i|X^{(i)}} \| P_{X_i} | Q_{X^{(i)}})$$

Remark: $D(Q\|P) = D(Q_{X^{(i)}} \| P_{X^{(i)}}) + D(Q_{X_i|X^{(i)}} \| \underbrace{P_{X_i|X^{(i)}}}_{= P_{X_i}} | Q_{X^{(i)}})$

Remark 2: If X is finite and $p_1, \ldots, p_n$ are uniform over X, then Hahn's inequality for H follows from KL.

Lemma: let P, Q be measures over a discrete set $X \times Y \times Z$. Then $D(Q_{Y|XZ} \| P_Y | Q_{XZ}) \geq D(Q_{Y|X} \| P_Y | Q_X)$.

Proof:

$$LHS = \sum_{x,z} Q_{XZ}(x,z) D(Q_{Y|X=x,Z=z} \| P_Y)$$

$$= \sum_x Q_X(x) \left[\sum_z Q_{Z|X}(z|x) D(Q_{Y|X=x,Z=z} \| P_Y) \right]$$

$$\geq \sum_x Q_X(x) \cdot D\Big(\sum_z \underbrace{Q_{Z|X}(z|x) Q_{Y|X=x,Z=z}}_{\text{same as previous lemma}} \| P_Y \Big)$$

$$= \sum_x Q_X(x) \cdot D(Q_{Y|X=x} \| P_Y)$$

$$= D(Q_{Y|X} \| P_Y | Q_X).$$

Proof (Han's for KL)

P is a prod. meas.

$$D(Q\|P) = D(Q_{X^{(i)}} \| P_{X^{(i)}}) + \underbrace{D(Q_{X_i|X^{(i)}} \| P_{X_i} | Q_{X^{(i)}})}$$

$X \leftrightarrow X_{1:i-1}$, $Y \leftrightarrow X_i$, $Z \leftrightarrow X_{1:n}$

lemma

$$\geq D(Q_{X^{(i)}} \| P_{X^{(i)}}) + D(Q_{X_i|X_{1:i-1}} \| P_{X_i} | Q_{X_{1:i-1}})$$

Sum over n:

$$nD(Q\|P) \geq \sum D(Q_{X^{(i)}} \| P_{X^{(i)}}) + D(Q\|P)$$

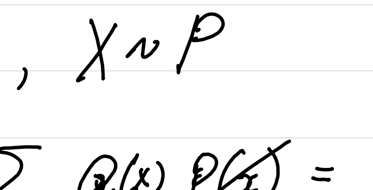
Rearrange and conclude $\square$

$$Var(Z) = \mathbb{E}[Z^2] - \mathbb{E}[Z]^2$$

$$= \mathbb{E}[\phi(Z)] - \phi(\mathbb{E}[Z]), \text{ where } \phi(\cdot) = \cdot^2$$

$Ent(Z) := \mathbb{E}[Z \log Z] - \mathbb{E}[Z] \log \mathbb{E}[Z]$ for $Z \geq 0$ a.s. $(0 \log 0 = 0)$.

$\phi(x) = x \log x$



$$Z = \frac{Q(X)}{P(X)}, \quad X \sim P$$

$$\mathbb{E}_P[Z] = \sum_x \frac{Q(x)}{\cancel{P(x)}} \cancel{P(x)} = 1$$

$$Ent(Z) = \mathbb{E}_P\left[\frac{Q(X)}{P(X)} \log \frac{Q(X)}{P(X)} \right] = D(Q\|P).$$

(Han's Ineq for Ent)

Theorem (Tensorisation of Ent)
Let $X_1, X_2, \ldots, X_n$ be ind. r.v. over $\mathcal{X}$ and let $f: \mathcal{X}^n \to [0, \infty)$.
Let $Z = f(X_1, \ldots, X_n)$. Then,

$$Ent(Z) \leq \sum_{i=1}^{n} \mathbb{E}[Ent^{(i)}(Z)],$$

where

$$Ent^{(i)}(Z) = E^{(i)}[Z \log Z] - E^{(i)} Z \log E^{(i)} Z$$

where $E^{(i)}(Z) = \mathbb{E}[Z | X^{(i)}]$

LECTURE 9

Theorem: $f : \mathcal{X}^n \to [0, \infty)$
$X_1, X_2, \dots, X_n$ independent, $Z = f(X_{1:n})$

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

where $Ent^{(i)}(Z) = E^{(i)}[Z \log Z] - E^{(i)}(Z) \log E^{(i)}(Z)$

Proof: (sketch) WLOG $Z \neq 0$.
WLOG we can assume $\mathbb{E}[Z] = 1$.
Easy to check $Ent(aZ) = a\,Ent(Z)$ for $a > 0$.

$$\sum \underbrace{f(x_1, x_2, \dots, x_n)}_{Z} P_{X_{1:n}}(x_1, \dots, x_n) = 1.$$

Define $Q(x_1, \dots, x_n) = f(x_{1:n}) P_{X_{1:n}}(x_1, \dots, x_n)$.
$Ent(Z) = D(Q \| P)$
Han's Inequality gives us:

$$\underbrace{D(Q \| P)}_{Ent(Z)} \le \sum_{i=1}^{n} \underbrace{D\left(Q_{X_i | X^{(i)}} \| P_{X_i} \mid Q_{X^{(i)}}\right)}_{\mathbb{E}[Ent^{(i)}(Z)]}$$

see ES2.

$\square$

Herbst's argument

Theorem: let Z be an integrable random variable such that for some $\nu > 0$, we have

$$Ent(e^{\lambda Z}) \le \frac{\lambda^2 \nu}{2} \mathbb{E}[e^{\lambda Z}] \text{ for all } \lambda > 0,$$

then $\psi_{Z - \mathbb{E}[Z]}(\lambda) = \log \mathbb{E}[e^{\lambda(Z - \mathbb{E}[Z])}] \le \frac{\lambda^2 \nu}{2} \ \forall \lambda > 0.$

Proof:

$$\psi_{Z - \mathbb{E}[Z]}(\lambda) = \log \mathbb{E}[e^{\lambda Z}] - \lambda \mathbb{E} Z$$

$$\psi'_{Z - \mathbb{E}[Z]}(\lambda) = \frac{\mathbb{E}[Z e^{\lambda Z}]}{\mathbb{E}[e^{\lambda Z}]} - \mathbb{E}[Z]$$

$$Ent(e^{\lambda Z}) = \mathbb{E}[e^{\lambda Z} \cdot \lambda Z] - \mathbb{E}[e^{\lambda Z}] \cdot \log \mathbb{E}[e^{\lambda Z}]$$
$$= \mathbb{E}[e^{\lambda Z}](\lambda \psi'(\lambda) - \psi(\lambda))$$

We have

$$\frac{Ent(e^{\lambda Z})}{\mathbb{E}[e^{\lambda Z}]} = \lambda \psi'(\lambda) - \psi(\lambda) \le \frac{\lambda^2 \nu}{2} \text{ for } \lambda > 0$$

This means $\underbrace{\frac{\psi'(\lambda)}{\lambda} - \frac{\psi(\lambda)}{\lambda^2}}_{\left(\frac{\psi(\lambda)}{\lambda}\right)'} \le \nu/2$

Let $\frac{\psi(\lambda)}{\lambda} = G(\lambda)$, we have $G'(\lambda) \le \nu/2$, so

$$G(\lambda) - \underbrace{G(0)}_{= \psi'(0) = 0} = \int_0^\lambda G'(t)\,dt \le \frac{\nu \lambda}{2}$$

$$\implies \frac{\psi(\lambda)}{\lambda} \le \frac{\nu \lambda}{2} \implies \psi(\lambda) \le \frac{\lambda^2 \nu}{2}$$

[Bounded differences inequality]

Theorem: Let $f : \mathcal{X}^n \to \mathbb{R}$ satisfy bounded differences property with $c_1, c_2, \dots, c_n$. Let $X_1, \dots, X_n$ be independent and $Z = f(X_{1:n})$. Then for $t \ge 0$,
$\mathbb{P}(Z - \mathbb{E}Z \ge t) \le e^{-t^2/2\nu}$ where $\nu = \frac{\sum c_i^2}{4}$
and $\mathbb{P}(Z - \mathbb{E}Z \le -t) \le e^{-t^2/2\nu}$

Proof: Step (1): Tensorisation

$$Ent(e^{\lambda Z}) \le \mathbb{E}\left[\sum Ent^{(i)}(e^{\lambda Z})\right]$$

Step (2): Lemma: Let Y be a bded r.v. on $[a, b]$. Then

$$Ent(e^{\lambda Y}) \le \mathbb{E}(e^{\lambda Y}) \cdot \frac{(b-a)^2 \cdot \lambda^2}{8}.$$

Step (3): Suppose the lemma is true,

$$Ent^{(i)}(e^{\lambda Z}) \le E^{(i)}(e^{\lambda Z}) \cdot \frac{c_i^2 \lambda^2}{8}$$

Plugging it back,

$$Ent(e^{\lambda Z}) \le \mathbb{E}(e^{\lambda Z}) \cdot \frac{\lambda^2 \nu}{2}$$

Step (4): Apply Herbst's argument to get $\psi_{Z - \mathbb{E}[Z]}(\lambda) \le \lambda^2 \nu/2$, and then use Chernoff bound.

Proof of Lemma: Recall that

$$\frac{Ent(e^{\lambda Y})}{\mathbb{E}[e^{\lambda Y}]} = \lambda \psi'(\lambda) - \psi(\lambda)$$

where $\psi(\lambda) = \log \mathbb{E}[e^{\lambda(Y - \mathbb{E}[Y])}]$

$$= \int_0^\lambda t \psi''(t)\,dt$$

By Hoeffding's Lemma, $\psi''(\lambda) \le \frac{(b-a)^2}{4}$

$$\le \int_0^\lambda t \frac{(b-a)^2}{4}\,dt = \frac{\lambda^2 (b-a)^2}{8}$$

$\square$

Log-Sobolev Inequalities

$$\frac{Ent(e^{\lambda Z})}{\mathbb{E}[e^{\lambda Z}]} \le \frac{\lambda^2 \nu}{2}$$

Poincaré $Var(f(X)) \le \mathbb{E}[\|\nabla f(X)\|^2]$ for $X \sim N(0, I)$.
$X_i \overset{iid}{\sim}$ Gaussian, or $X_i \overset{iid}{\sim} Rad(1/2)$

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

Assume $\dagger$, then choosing $Z = f(X_1, \dots, X_n)$

$$\hat{f} = e^{\lambda Z/2}$$

LECTURE 9

Log - Sobolev inequalities: (LSI)

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

$X_1, X_2, \ldots, X_n$ are independent symmetric Bernoulli.
let $f: \{-1,1\}^n \to \mathbb{R}$

$$\mathrm{Var}(f(X)) \leq \frac{1}{2} \sum_{i=1}^{n} \mathbb{E}[(Z - Z_i')_+^2], \quad Z_i' = f(X^{(i)}, X_i').$$

$$\mathrm{Var}(f(X)) \leq \frac{1}{4} \sum_{i=1}^{n} \mathbb{E}[(f(X) - f(\bar{X}^{(i)}))^2]$$

$$(Z - Z_i')^2 = \begin{cases} (f(X^{(i)}, X_i = +1) - f(X^{(i)}, X_i = -1))^2, & \text{w.p.} = 1/2 \\ 0, & \text{w.p.} = 1/2 \end{cases}$$

either both are equal or opposite

$\bar{X}^{(i)} = (X^{(i)}, -X_i)$.

$$= \frac{1}{2} \sum_{i=1}^{n} \mathbb{E}[(f(X) - f(\bar{X}^{(i)}))_+^2]$$

$\frac{f(\bar{X}^{(i)}) - f(X)}{2}$ = is like gradient $\frac{\partial f}{\partial x_i}(x_1, \ldots, x_n)$.

$$\Rightarrow \frac{1}{4} \sum \mathbb{E}[(f(X) - f(\bar{X}^{(i)}))^2] = \mathcal{E}(f)$$
$$(= \mathbb{E}[\|\nabla f(X)\|^2])$$

Theorem: (LSI for symmetric Bernoulli)

$$\mathrm{Ent}(f^2(X)) \leq 2 \cdot \mathcal{E}(f).$$

Proof: Using tensorisation of Ent.

$$\mathrm{Ent}(Z^2) \leq \mathbb{E}[\sum_{i=1}^{n} \mathrm{Ent}^{(i)}(Z^2)]$$

$$\mathrm{Ent}^{(i)}(Z^2) \leq E^{(i)}[Z^2 \log Z^2] - E^{(i)} Z^2 \log E^{(i)} Z^2.$$

If LSI is true for $n=1$, then a.s.

$$\mathrm{Ent}^{(i)}(Z^2) \leq \frac{(f(X) - f(\bar{X}^{(i)}))^2}{2}$$

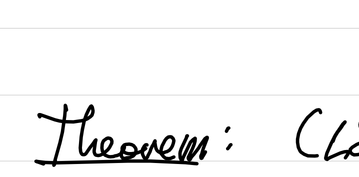
Summing over $i = 1, \ldots, n$ & taking expectations will conclude the proof.

To prove LSI for $n = 1$, need to show the following: $f(-1) = a$, $f(+1) = b$.

$$\underbrace{\frac{1}{2} a^2 \log a^2 + \frac{1}{2} b^2 \log b^2 - \frac{(a^2+b^2)}{2} \log\left(\frac{a^2+b^2}{2}\right)}_{\mathrm{Ent}(Z^2)}$$

$$\leq \frac{(b-a)^2}{2}$$

WLOG, $0 \leq b \leq a$. Consider for fixed b, the following function of a. $h: [b, \infty) \to \mathbb{R}$, $h(a) = \mathrm{LHS} - \mathrm{RHS}$.



$h(b) = 0$

$h'(b) = 0$

$h''(a) \leq 0$ for $a \in [b, \infty)$

$$h'(a) = a \log \frac{2a^2}{a^2+b^2} - (a-b).$$

$$h''(a) = 1 + \log \frac{2a^2}{a^2+b^2} - \frac{2a^2}{(a^2+b^2)} \leq 0 \quad (\log x - x \leq -1).$$

□

For asymmetric Bernoulli,

$$\mathrm{Ent}(f^2) \leq c(p) \cdot \mathcal{E}(f)$$

$$c(p) = \frac{1}{1-2p} \log\left(\frac{1-p}{p}\right)$$

Theorem (LSI for Gaussians)

Let $X_1, X_2, \ldots, X_n$ be iid $N(0,1)$ and let $f: \mathbb{R}^n \to \mathbb{R}$ be a continuously differentiable function. Then $\mathrm{Ent}(f^2) \leq 2\mathbb{E}[\|\nabla f(X)\|^2]$.

Proof (sketch): Step 1: Reduce it to the $n=1$ case by tensorisation.

Step 2: introduce $x_1, x_2, \ldots, x_n$ iid symm. Bernoulli and consider $f\left(\frac{X_1 + \cdots + X_n}{\sqrt{n}}\right)$, and use the LSI for symmetric Bernoulli and take $n \to \infty$, use CLT. □

Theorem (Gaussian concentration inequality)

Let $X_1, X_2, \ldots, X_n$ iid $N(0,1)$, let $f: \mathbb{R}^n \to \mathbb{R}$ be an L-Lipschitz function (i.e. $|f(x) - f(y)| \leq L \cdot \|x - y\|_2$).

Then $Z = f(X_1, X_2, \ldots, X_n)$ is in $G(L^2)$, i.e.

$$\mathbb{P}(Z - \mathbb{E}Z \geq t) \leq e^{-\frac{t^2}{2L^2}},$$
$$\mathbb{P}(Z - \mathbb{E}Z \leq -t) \leq e^{-\frac{t^2}{2L^2}}.$$

Proof: Apply Gaussian LSI to $e^{\lambda Z/2}$

$$\mathrm{Ent}(e^{\lambda Z}) \leq 2 \cdot \mathbb{E}[\|e^{\lambda Z/2} \cdot \lambda/2 \cdot \nabla f(X)\|^2]$$

$$(\text{Lipschitz}) = \frac{L^2 \lambda^2}{2} \mathbb{E}[e^{\lambda Z}]$$

holds $\forall \lambda \in \mathbb{R}$.

$$\Rightarrow \frac{\mathrm{Ent}(e^{\lambda Z})}{\mathbb{E}[e^{\lambda Z}]} \leq \frac{\lambda^2 L^2}{2} \Rightarrow \psi_{Z - \mathbb{E}Z}(\lambda) \leq \frac{\lambda^2 L^2}{2}$$

$$\Rightarrow Z \in G(L^2).$$

□

LECTURE 10

Recap: LSI (Bernoulli) $\mathrm{Ent}(f^2) \leq 2\mathcal{E}(f)$

$$\mathcal{E}(f) = \mathbb{E}\left[\sum_{i=1}^n \frac{(f(X) - f(\bar{X}^{(i)}))^2}{4}\right]$$

$$= \mathbb{E}\left[\sum_{i=1}^n \left(\frac{f(X) - f(\bar{X}^{(i)})}{2}\right)_+^2\right]$$

Theorem: Let $f : \{-1,1\}^n \to \mathbb{R}$ and let X_i be iid symmetric Bernoulli. Let $Z = f(X_{1:n})$ and let

$$v = \max_{x \in \{-1,1\}^n} \sum_{i=1}^n (f(x) - f(\bar{x}^{(i)}))_+^2$$

Then Z has a sub-Gaussian right tail with parameter $v/2$, i.e.

$$\mathbb{P}(Z - \mathbb{E}Z \geq t) \leq e^{-\frac{t^2}{v}}.$$

Remarks:

1. $\mathrm{Var}(Z) \leq \mathcal{E}(f) \leq \frac{v}{2}$
2. If $v = \max \sum_{i=1}^n (f(x) - f(\bar{x}^{(i)}))_-^2$, get left tail bounds ? get left tail bounds that give $\mathcal{G}(v/2)$.
3. If $v = \max \sum_{i=1}^n (f(x) - f(\bar{x}^{(i)}))^2 \Rightarrow$ right & left tail with $\mathcal{G}(v/2)$.

   More refined analysis gives $\mathcal{G}(v/4)$ (ES2).
4. If f satisfied bdd diff. property with c_i s.t. $\sum c_i^2 \leq v$.

   Bdd diff. ineq. gives $Z \in \mathcal{G}(v/4)$.
   The theorem in 3 also gives $Z \in \mathcal{G}(v/4)$ but 3 is applicable more broadly.

Proof: let $\lambda > 0$. Use LSI for $e^{\lambda Z/2}$ to get

$$\mathrm{Ent}(e^{\lambda Z}) \leq \mathbb{E}\left[\sum_{i=1}^n \left(e^{\frac{\lambda f(X)}{2}} - e^{\frac{\lambda}{2} f(\bar{X}^{(i)})}\right)_+^2\right]$$

$e^{x/2}$ is a convex function and so if $z > y$, then

$$(e^{z/2} - e^{y/2})_+ \leq (z-y)_+ \frac{e^{z/2}}{2}$$

$$\Rightarrow \leq \mathbb{E}\left[\sum_{i=1}^n (\lambda f(X) - \lambda f(\bar{X}^{(i)}))_+^2 \frac{e^{\lambda f(X)}}{4}\right]$$

$$\leq \mathbb{E}\left[e^{\lambda f(X)} \cdot \frac{\lambda^2}{4} \cdot v\right] = \frac{\mathbb{E}[e^{\lambda Z}] \cdot \lambda^2 \cdot (v/2)}{2}$$

Use Herbst's argument to get the right tail bound.

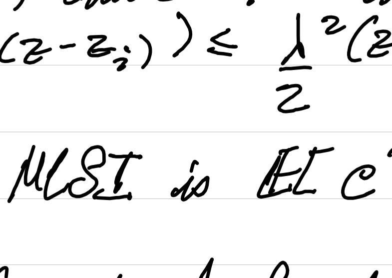
LSI "too powerful": $\mathrm{Ent}(e^{\lambda Z}) \leq \frac{\lambda^2 v}{2} \mathbb{E}[e^{\lambda Z}]$

$f = e^{\lambda Z/2}$, need Z to have a "nice" distribution.

Theorem (Modified Log-Sobolev Inequality).
Let $X_1, \dots, X_n$ be ind, $f : \mathcal{X}^n \to \mathbb{R}$, $Z = f(X_{1:n})$.
For $1 \leq i \leq n$, let $Z_i = f_i(X^{(i)})$. Let $\phi(x) = e^x - x - 1$. Then $\forall \lambda \in \mathbb{R}$:

$$\mathrm{Ent}(e^{\lambda Z}) \leq \sum_{i=1}^n \mathbb{E}[e^{\lambda Z} \phi(-\lambda(Z - Z_i))]$$

Remark:



if $x \geq 0$, then $\phi(-x) \leq \frac{x^2}{2}$

Say $\lambda > 0$, choose Z_i so that $Z - Z_i \geq 0$.

$$\phi(-\lambda(Z - Z_i)) \leq \frac{\lambda^2}{2}(Z - Z_i)^2.$$

RHS of MLSI is $\mathbb{E}\left[e^{\lambda Z} \frac{\lambda^2}{2} \sum_{i=1}^n (Z - Z_i)^2\right]$

Lemma (Variational formula for Ent)
Let $Y \geq 0$ a.s. Then $\mathrm{Ent}(Y) = \inf_{u>0} \mathbb{E}\left[Y \log \frac{Y}{u} - (Y - u)\right]$

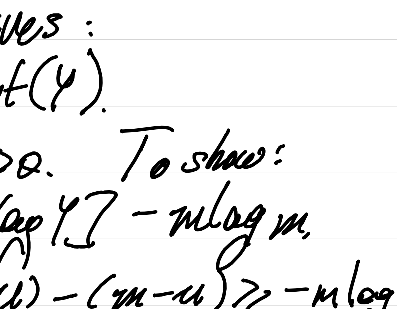
Remark: $\mathrm{Var}(Y) = \inf_u \mathbb{E}[(Y-u)^2]$

$$\mathbb{E}[\Phi(Y)] - \Phi(\mathbb{E}(Y)), \quad \Phi(x) : x^2 \to \mathrm{Var}, \quad x \log x \to \mathrm{Ent}$$

$$\|$$

$$\inf_u \mathbb{E}[\Phi(Y) - \Phi(u) - \Phi'(u)(Y - u)]$$

(Φ generally convex).



Proof of lemma: $u = \mathbb{E}Y$ gives:

$$\mathbb{E}\left[Y \log \frac{Y}{u} - (Y - u)\right] = \mathrm{Ent}(Y).$$

Suppose $\mathbb{E}[Y] = m$, fix any $u > 0$. To show:

$$\mathbb{E}\left[Y \log \frac{Y}{u} - (Y - u)\right] \geq \mathbb{E}[Y \log Y] - m \log m.$$

enough to show: $-m \log(u) - (m - u) \geq -m \log m$

$$\iff \log(m/u) \geq 1 - u/m$$

which is true since $-\log(x) \geq 1 - x$ $\square$

Proof of MLSI:
let $Y = e^{\lambda Z}$, $Y_i = e^{\lambda Z_i}$

$$\mathrm{Ent}(Y) \leq \mathbb{E}\left[\sum_{i=1}^n \mathrm{Ent}^{(i)}(Y)\right]$$

$$\leq \mathbb{E}\left[\sum_{i=1}^n \mathbb{E}^{(i)}\left[\underbrace{e^{\lambda Z} \cdot \lambda(Z - Z_i) - (e^{\lambda Z} - e^{\lambda Z_i})}_{e^{\lambda Z} \cdot \phi(-\lambda(Z - Z_i))}\right]\right]$$

$$= \sum_{i=1}^n \mathbb{E}[e^{\lambda Z} \cdot \phi(-\lambda(Z - Z_i))] \qquad \square$$

LECTURE 11

Theorem: Let $Z = f(X_{1:n})$ for independent $X_1, \dots, X_n$. Define $Z_i = \inf_{x_i} f(X^{(i)}, x_i)$. Suppose

$$\sum_{i=1}^n (Z - Z_i)^2 \le v, \text{ then for all } t > 0$$

$$\mathbb{P}(Z - \mathbb{E}Z \geq t) \le e^{-t^2/2v}$$

Proof: Let $\lambda > 0$. By MLSI,

$$\mathrm{Ent}(e^{\lambda Z}) \le \mathbb{E}\left[\sum_{i=1}^n e^{\lambda Z} \cdot \phi(-\lambda(Z - Z_i))\right]$$

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

Use Herbst's argument to conclude the proof $\square$

Theorem: Let f be a separately convex function on $[0,1]^n$. Let $X_1, \dots, X_n$ be ind. supp on $[0,1]$, let $Z = f(X_{1:n})$. Assume that f is 1-Lipschitz. Then $\mathbb{P}(Z - \mathbb{E}Z \geq t) \le e^{-t^2/2}$ for $t > 0$.

Remark: $\mathrm{Var}(Z) \le 1$ (convex Poincaré inequality).

Proof: Set $Z_i = \inf_{x_i} f(X^{(i)}, x_i)$, let x_i^* be s.t.

$$Z_i = f(X^{(i)}, x_i^*)$$

$$Z_i \geq Z + \frac{\partial f}{\partial x_i}(X) \cdot (x_i^* - x_i)$$

$$\Longrightarrow 0 \le Z - Z_i \le \frac{\partial f(X)}{\partial x_i} \cdot (x_i - x_i^*)$$

$$\Longrightarrow (Z - Z_i)^2 \le \left(\frac{\partial f}{\partial x_i}\right)^2 \cdot (x_i - x_i^*)^2 \le \left(\frac{\partial f}{\partial x_i}\right)^2$$

Summing up, $\sum_{i=1}^n (Z - Z_i)^2 \le \|\nabla f(X)\|^2 \le 1$.
Using the previous theorem, we get
$\mathbb{P}(Z - \mathbb{E}Z \geq t) \le e^{-t^2/2}$ $\square$

Transport Method

Optimal Transport

0.2	$\bullet\, x_1$	$\bullet\, y_1$	0.5
0.1	$\bullet\, x_2$	$\bullet\, y_2$	0.1
0.6	$\bullet\, x_3$	$\bullet\, y_3$	0.1
0.1	$\bullet\, x_4$	$\bullet\, y_4$	0.3

Transporting bread from x_i to y_i has a per unit cost of $c(x_i, y_i)$. A transport plan is $\pi(x_i, y_j)$ for $1 \le i \le 4$, $1 \le j \le 4$, where $\pi(x_i, y_j)$ is the amount of bread sent from x_i to y_j and $\sum_y \pi(x_i, y) = p(x_i)$, $\sum_x \pi(x, y_j) = q(y_j)$.

$$\underbrace{\min_\pi \sum_{i,j} c(x_i, y_j) \cdot \pi(x_i, y_j)}_{\text{optimal cost.}}$$

Theorem (Variational formulas for log-MGF and KL-divergence) Let Z be a real-valued r.v. on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Then

$$\log \mathbb{E}_P e^Z = \sup_{Q \ll P} \left[\mathbb{E}_Q Z - D(Q\|P)\right]$$

Conversely, if P and Q are two measures, then

$$D(P\|Q) = \sup_Z \left\{\mathbb{E}_Q Z - \log \mathbb{E}_P e^Z\right\}$$

Remark: If Z is replaced by $\lambda \cdot (Z - \mathbb{E}_P Z)$, then

$$\log \mathbb{E}_P e^{\lambda(Z - \mathbb{E}_P Z)} = \sup_{Q \ll P} \lambda(\mathbb{E}_Q Z - \mathbb{E}_P Z) - D(Q\|P).$$

Proof: Ω is discrete. Set $Q^*(\omega) = \frac{e^{Z(\omega)} P(\omega)}{\mathbb{E}_P e^Z}$

$$0 \le D(Q\|Q^*)$$

$$= \sum_{\omega \in \Omega} Q(\omega) \log \frac{Q(\omega)}{Q^*(\omega)} = \sum_\omega Q(\omega) \log\left(\frac{Q(\omega)}{P(\omega)} \frac{P(\omega)}{Q^*(\omega)}\right)$$

$$= D(Q\|P) + \sum_\omega Q(\omega) \log(\mathbb{E}_P e^Z / e^Z)$$

$$= D(Q\|P) + \log(\mathbb{E}_P e^Z) - \mathbb{E}_Q Z.$$

$$\Longrightarrow \log \mathbb{E}_P e^Z \geq \mathbb{E}_Q Z - D(Q\|P).$$

Taking supremum over Q,

$$\log \mathbb{E}_P e^Z \geq \sup_{Q \ll P} \mathbb{E}_Q Z - D(Q\|P).$$

Since Q^* achieves equality, $\log \mathbb{E}_P e^Z = \sup_{Q \ll P} \mathbb{E}Z - D(Q\|P)$

To show the second part, we have

$$D(Q\|P) \geq \mathbb{E}_P Z - \log \mathbb{E}_P e^Z$$

Taking sup over $Z \Longrightarrow D(Q\|P) \geq \sup_Z \mathbb{E}_Q Z - \log \mathbb{E}_P e^Z$

$Z(\omega) = \log \frac{Q(\omega)}{P(\omega)}$ gives equality, and so

$$D(Q\|P) = \sup_Z \mathbb{E}_Q Z - \log \mathbb{E}_P e^Z \quad \square$$

Suppose this inequality holds $\forall Q \ll P$:

$$\mathbb{E}_Q Z - \mathbb{E}_P Z \le \sqrt{2v D(Q\|P)}$$

$$\log \mathbb{E}_P e^{\lambda(Z - \mathbb{E}_P Z)} = \sup_{Q \ll P} \lambda(\mathbb{E}_Q Z - \mathbb{E}_P Z) - D(Q\|P)$$

$$\le \sup_{Q \ll P} \lambda \sqrt{2v D(Q\|P)} - D(Q\|P)$$

$$\le \text{"="} \sup_{t \geq 0} \lambda \sqrt{2vt} - t = \frac{\lambda^2 v}{2}.$$

LECTURE 12

Theorem (Marton's argument).
Suppose the following holds for all $Q \ll P$
$$\mathbb{E}_Q Z - \mathbb{E}_P Z \le \sqrt{2v D(Q\|P)} \text{ for some } v>0.$$
Then for $\lambda>0$, $\log \mathbb{E}_P e^{\lambda(Z-\mathbb{E}_P Z)} \le \frac{\lambda^2 v}{2}$, and
$\mathbb{P}(Z-\mathbb{E}_P Z \ge t) \le e^{-t^2/2v}$
Conversely, if $\log \mathbb{E}_P e^{\lambda(Z-\mathbb{E}_P Z)} \le \frac{\lambda^2 v}{2}$ for all $\lambda>0$, then $\mathbb{E}_Q Z - \mathbb{E}_P Z \le \sqrt{2v D(Q\|P)}$ for all $Q \ll P$.

Proof: $$\log \mathbb{E}_P e^{\lambda(Z-\mathbb{E}_P Z)} \le \sup_{Q \ll P} \lambda\sqrt{2vD(Q\|P)} - D(Q\|P) \quad (\lambda>0)$$
$$\le \sup_{t\ge 0} \lambda\sqrt{2vt} - t$$
$$= \frac{\lambda^2 v}{2}.$$
For the converse, wlog assume $\mathbb{E}_Q Z - \mathbb{E}_P Z > 0$.
$$D(Q\|P) \ge \lambda\cdot(\mathbb{E}_Q Z - \mathbb{E}_P Z) - \log \mathbb{E}_P e^{\lambda(Z-\mathbb{E}_P Z)}$$
$$\ge \lambda(\mathbb{E}_Q Z - \mathbb{E}_P Z) - \lambda^2 v/2 \quad \forall \lambda>0.$$
maximise RHS $\Longrightarrow D(Q\|P) \ge \frac{(\mathbb{E}_Q Z - \mathbb{E}_P Z)^2}{2v}$,
(by setting $\lambda = \frac{\mathbb{E}_Q Z - \mathbb{E}_P Z}{v}$)
$$\Longrightarrow \mathbb{E}_Q Z - \mathbb{E}_P Z \le \sqrt{2vD(Q\|P)} \qquad \square$$

$X_{1:n} \sim P = P_{X_1} \otimes \dots \otimes P_{X_n}$, $Z = f(X_{1:n})$
if $f(y) - f(x) \le \sum_{i=1}^n d(x_i, y_i) c_i$.

Let $Y_{1:n} \sim Q$, $\mathbb{E} f(Y_{1:n}) - \mathbb{E} f(X_{1:n})$
$$= \mathop{\mathbb{E}}_{\substack{\pi \sim (X_{1:n}, Y_{1:n}) \\ \pi \in \Pi(P,Q)}} [f(Y_{1:n}) - f(X_{1:n})]$$

Here π is a coupling between $X_{1:n}, Y_{1:n}$ i.e. $\pi_{X_{1:n}} = P$, $\pi_{Y_{1:n}} = Q$. Set of all couplings is $\Pi(P,Q)$.

$$\mathop{\mathbb{E}}_{\pi}[f(Y_{1:n}) - f(X_{1:n})] \le \mathop{\mathbb{E}}_{\pi}\Big[\sum_{i=1}^n d(X_i, Y_i) c_i\Big]$$
only depends on marginals $$= \sum_{i=1}^n c_i \mathop{\mathbb{E}}_{\pi} d(X_i, Y_i).$$
take inf over π $$\le \Big(\sum_{i=1}^n c_i^2\Big)^{1/2} \cdot \Big(\sum_{i=1}^n \big(\mathop{\mathbb{E}}_{\pi} d(X_i,Y_i)\big)^2\Big)^{1/2}$$
$$\mathbb{E}[f(Y_{1:n})] - \mathbb{E}[f(X_{1:n})] \le \Big(\sum_{i=1}^n c_i^2\Big)^{1/2} \cdot \Big(\inf_{\pi \in \Pi(P,Q)} \sum \mathbb{E}_\pi[d(X_i,Y_i)]^2\Big)^{1/2}$$

Suppose we can show $\uparrow \le 2C D(Q\|P)$.
then have $\mathbb{E} f(Y_{1:n}) - \mathbb{E} f(X_{1:n}) \le \sqrt{2vD(Q\|P)}$
where $v = C\sum_{i=1}^n c_i^2$

To run Marton's argument for such functions f, enough to prove:
$$\inf_{\pi} \sum_{i=1}^n \mathbb{E}_\pi[d(X_i, Y_i)]^2 \le 2C D(Q\|P)$$
for some $C>0$.

Bounded Differences inequality via the transport method.

We need to show $\inf_{\pi \in \Pi(P,Q)} \sum_{i=1}^n \mathbb{P}(X_i \ne Y_i)^2 \le \frac{1}{2} D(Q\|P)$.

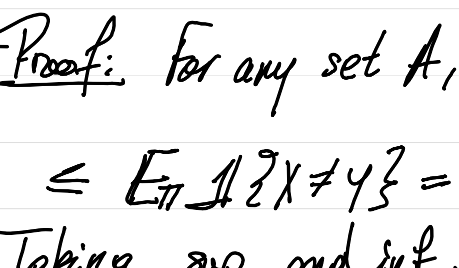
Theorem (Marton's transport cost inequality)
Let $P \sim P_{X_1} \otimes P_{X_2} \otimes \dots \otimes P_{X_n}$ and Q be an arbitrary measure s.t. $Q \ll P$, then
$$\inf_{\pi \in \Pi(P,Q)} \mathbb{P}(X_i \ne Y_i)^2 \le \frac{1}{2} D(Q\|P)$$

Remark: for $n=1$, we need $\inf_{\pi} \mathbb{P}(X_1 \ne Y_1) \le \sqrt{\frac{1}{2} D(Q\|P)}$

Lemma: P, Q are probability measures over the same space, then, $\inf_{(X,Y) \sim \pi \in \Pi(P,Q)} \mathbb{P}(X \ne Y) = d_{TV}(P,Q)$,

$$d_{TV}(P,Q) := \sup_A |P(A) - Q(A)|$$

Remark: $$d_{TV}(P,Q) = \frac{1}{2}\sum_{w\in\Omega} |P(w) - Q(w)|$$
$$= \sum_{w \in \Omega} \underbrace{(Q(w) - P(w))_+}_{(-)} = \sum_{w\in\Omega} \underbrace{(P(w) - Q(w))_+}_{(+)}$$



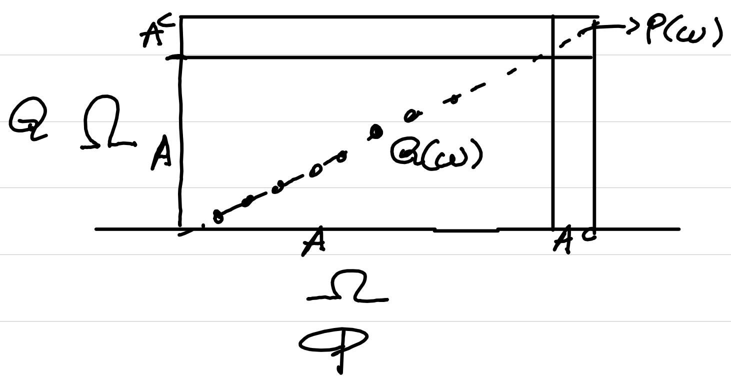
$$= 1 - \sum_w \min\{P(w), Q(w)\} \quad (\dagger)$$
$\left(|x| = x^- + x^+ \text{ and } x = x^+ - x^- \right.$
$\Rightarrow (+) = (-)$ and $(+) + (-) = 2d_{TV}$.

Proof: For any set A, $|P(A) - Q(A)| = |\mathbb{E}_\pi[\mathbb{1}(X\in A)] - \mathbb{E}_\pi[\mathbb{1}(Y \in A)]|$
$$\le \mathbb{E}_\pi \mathbb{1}\{X \ne Y\} = \mathbb{P}_\pi(X \ne Y)$$
Taking $\sup_A$ and $\inf_A$, we get $d_{TV}(P,Q) \le \inf_\pi \mathbb{P}(X \ne Y)$.

$(\dagger)$ $$\min(a,b) = \frac{b+a}{2} - \frac{|b-a|}{2}$$
$$\Rightarrow \sum_{w\in\Omega} \min\{P(w), Q(w)\} + d_{TV}(P,Q)$$
$$= \sum_{w\in\Omega} \min\{P(w), Q(w)\} + \frac{|P(w) - Q(w)|}{2}$$
$$= \sum_{w \in \Omega} \frac{P(w) + Q(w)}{2} = 1.$$

LECTURE 13

Proof: (remaining half) We want to find $\pi \in \Pi(P, Q)$ s.t. $\mathbb{P}(X \neq Y) = d_{TV}(P, Q)$.
Let $A = \{\omega : P(\omega) \geq Q(\omega)\}$.



$$\pi(\omega_1, \omega_2) = \begin{cases} Q(\omega), & (\omega,\omega) \in A \times A \\ P(\omega), & (\omega,\omega) \in A^c \times A^c \\ 0, & (\omega_1, \omega_2) \in (A^c, A) \\ \dfrac{(P(\omega_1) - Q(\omega_1)) \times (Q(\omega_2) - P(\omega_2))}{d_{TV}(P, Q)}, & (\omega_1, \omega_2) \in (A, A^c) \end{cases}$$

For this π, $\mathbb{P}(X = Y) = \sum_{\omega} \min\{P(\omega), Q(\omega)\}$

$= 1 - d_{TV}(P, Q) \Rightarrow \mathbb{P}(X \neq Y) = d_{TV}(P, Q)$ □

Lemma (Pinsker's Inequality)
$d_{TV}(P, Q)^2 \leq \frac{1}{2} D(Q \| P)$.

Proof: (Example Sheet 2) □

The above lemmas imply Marton's TCI for $n = 1$.
Assume that Marton's TCI holds for all $n \leq k$, we'll prove it for $n = k+1$. $(X_1, X_2, \dots, X_{k+1}) \sim P_{X_{1:k+1}} = P_{X_1} \otimes \dots \otimes P_{X_{k+1}}$
$(Y_1, \dots, Y_{k+1}) \sim Q_{Y_{1:k+1}}$
To show
$$\inf_{\pi \in \Pi(P_{X_{1:k+1}}, Q_{Y_{1:k+1}})} \sum_{i=1}^{k+1} \mathbb{P}(X_i \neq Y_i)^2 \leq \frac{1}{2} \cdot D(Q_{Y_{1:k+1}} \| P_{X_{1:k+1}}).$$
We know that $\exists \pi_k \in \Pi(P_{X_{1:k}}, Q_{Y_{1:k}})$ s.t.
$$\sum_{i=1}^{k} \mathbb{P}(X_i \neq Y_i)^2 \leq \frac{1}{2} D(Q_{Y_{1:k}} \| P_{X_{1:k}})$$
(by assumption).
Define $\pi \in \Pi(P_{X_{1:k+1}}, Q_{Y_{1:k+1}})$ as
$$\pi(X_{1:k+1} = x_{1:k+1}, Y_{1:k+1} = y_{1:k+1})$$
$$= \pi_k(X_{1:k} = x_{1:k}, Y_{1:k} = y_{1:k}) \times \pi_{y_{1:k}}(X_{k+1} = x_{k+1}, Y_{k+1} = y_{k+1}).$$

where $\pi_{y_{1:k}}$ is the optimal TV-coupling between $P_{X_{k+1}}$ and $Q_{Y_{k+1} | Y_{1:k} = y_{1:k}}$.
(Check coupling).

Under π, $\mathbb{P}(X_{1:k+1} = x_{1:k+1}, Y_{1:k+1} = y_{1:k+1})$
$= \mathbb{P}(X_{1:k} = x_{1:k}, Y_{1:k} = y_{1:k}) \times \mathbb{P}(X_{k+1} = x_{k+1})$
$\times \mathbb{P}(Y_{k+1} = y_{k+1} \mid Y_{1:k} = y_{1:k}, X_{k+1} = x_{k+1})$.

Under π, we have
$$\sum_{i=1}^{k} \mathbb{P}(X_i \neq Y_i)^2 + \mathbb{P}(X_{k+1} \neq Y_{k+1})^2$$
Observe $\mathbb{P}(X_{k+1} \neq Y_{k+1} \mid X_{1:k} = x_{1:k}, Y_{1:k} = y_{1:k})$
$\overset{(=)}{=} \mathbb{P}(X_{k+1} \neq Y_{k+1} \mid Y_{1:k} = y_{1:k})$
(by construction of π).
$\overset{(=)}{=} d_{TV}(P_{X_{k+1}}, Q_{Y_{k+1} | Y_{1:k} = y_{1:k}})$
$\leq \sqrt{\tfrac{1}{2} D(Q_{Y_{k+1} | Y_{1:k} = y_{1:k}} \| P_{X_{k+1}})}$.
Integrate wrt π_k, $\mathbb{P}(X_{k+1} \neq Y_{k+1})$
$$\leq \mathbb{E}_{\pi_k}\left[\sqrt{\tfrac{1}{2} D(Q_{Y_{k+1} | Y_{1:k} = y_{1:k}} \| P_{X_{k+1}})}\right]$$
By Jensen's inequality,
$$\mathbb{P}(X_{k+1} \neq Y_{k+1})^2 \leq \mathbb{E}_{\pi_k}\left[\tfrac{1}{2} D(Q_{Y_{k+1} | Y_{1:k} = y_{1:k}} \| P_{X_{k+1}})\right].$$
$$= \mathbb{E}_{Q_{Y_{1:k}}}[\cdots]$$
$$= \tfrac{1}{2} D(Q_{Y_{k+1} | Y_{1:k}} \| P_{X_{k+1}} \mid Q_{Y_{1:k}}).$$

By assumption about π_k,
$$\sum_{i=1}^{k} \mathbb{P}(X_i \neq Y_i)^2 \leq \frac{1}{2} D(Q_{Y_{1:k}} \| P_{X_{1:k}})$$
Adding and using the chain rule of KL, we conclude.

□

Def: A function $f : \mathcal{X}^n \to \mathbb{R}$ satisfies a one-sided bdd diff. property with functions $c_1, \dots, c_n$ from $\mathcal{X}^n$ to $\mathbb{R}$ if $\forall x, y \in \mathcal{X}^n$:
$$f(y) - f(x) \leq \sum_{i=1}^{n} c_i(x) \cdot \mathbb{1}\{x_i \neq y_i\}.$$

Theorem: (Talagrand's one-sided bdd differences inequality).
Let $X_1, \dots, X_n$ be independent, and f be as above. Define $v = \mathbb{E}\left[\sum_{i=1}^{n} c_i(X)^2\right]$, let $Z = f(X_{1:n})$.

Then for $\lambda > 0$, $\psi_{Z - \mathbb{E}Z}(\lambda) \leq \frac{\lambda^2 v}{2}$, and so for $t > 0$, $\mathbb{P}(Z - \mathbb{E}Z \geq t) \leq e^{-t^2/2v}$.

LECTURE 14/15

<u>Remark</u> (Talagrand's inequality)

$Z_i = \inf_{x_i} f(X^{(i)}, x_i)$

$\sum (Z - Z_i)^2 \le v \Rightarrow$ sub.-Gaussian rt tails with parameter v.

$\left(v_\infty := \sup_x \sum c_i(x)^2 \right)$

If instead $Z_i = \sup_{x_i} f(X^{(i)}, x_i)$, and $\sum_i (Z_i - Z)^2 \le v$, then we get left tails

For one-sided bdd diff. property: $Z_i - Z \le c_i(X)$
$\Rightarrow \sum (Z_i - Z)^2 \le v_\infty \Rightarrow$ left tails with parameter v_∞.

<u>Proof:</u> let $P = P_{X_1} \otimes \cdots \otimes P_{X_n}$. let $Y_{1:n} \sim Q$,
$f : \mathcal{X}^n \to \mathbb{R}$ $\quad (f(X_{1:n}) = Z)$

$$\mathbb{E} f(Y_{1:n}) - \mathbb{E}[f(X_{1:n})] = \mathbb{E}_\pi[f(Y_{1:n}) - f(X_{1:n})],$$

where $\pi \in \Pi(P, Q)$

$$\le \mathbb{E}_\pi\left[\sum_{i=1}^n c_i(X_{1:n}) \mathbb{1}\{X_i \ne Y_i\}\right].$$

$$= \mathbb{E}_\pi\left[\mathbb{E}_\pi\left[\sum_{i=1}^n c_i(X_{1:n}) \mathbb{1}\{X_i \ne Y_i\} \,\middle|\, X_{1:n}\right]\right]$$

$$= \mathbb{E}_\pi\left[\sum_{i=1}^n c_i(X_{1:n}) \cdot \mathbb{P}(X_i \ne Y_i \mid X_{1:n})\right]$$

where we used the notation
$\mathbb{P}(X_i \ne Y_i \mid X_{1:n}) = \mathbb{E}_\pi[\mathbb{1}\{X_i \ne Y_i\} \mid X_{1:n}]$.

Using Cauchy-Schwarz (twice),

$$\mathbb{E} f(Y_{1:n}) - \mathbb{E} f(X_{1:n}) \le \sum_{i=1}^n \left(\mathbb{E}_\pi[c_i(X)^2]\right)^{1/2} \left(\mathbb{E}_\pi[\mathbb{P}(X_i \ne Y_i \mid X)^2]\right)^{1/2}$$

$$\le \sqrt{v} \cdot \left(\sum_{i=1}^n \mathbb{E}_\pi\left[\mathbb{P}(X_i \ne Y_i \mid X_{1:n})^2\right]\right)^{1/2}$$

It's enough to show:

$$\inf_{\pi \in \Pi(P,Q)} \sum_{i=1}^n \mathbb{E}_\pi\left[\mathbb{P}(X_i \ne Y_i \mid X_{1:n})^2\right] \le 2 D(Q \| P).$$

<u>Claim:</u> (Marton's conditional transport cost inequality)

$$\inf_\pi \sum_{i=1}^n \mathbb{E}_\pi\left[\mathbb{P}(X_i \ne Y_i \mid X_{1:n})\right] \le 2 \cdot D(Q \| P)$$

<u>Lemma:</u> let P and Q be prob. measures on a common space. Then

$$\inf_{\pi \in \Pi(P,Q)} \mathbb{E}\left[\mathbb{P}(X \ne Y \mid X)^2\right] = d_2^2(Q, P),$$

where $d_2^2(Q,P)$ is called Marton's divergence given by $d_2^2(Q,P) = \sum_{\omega : P(\omega) > 0} \frac{(P(\omega) - Q(\omega))_+^2}{P(\omega)}$

<u>Proof:</u> let π be any coupling. Observe that

$$\mathbb{P}(X = Y \mid X = x) = \frac{\mathbb{P}(X = x, Y = x)}{\mathbb{P}(X = x)} \le \frac{\mathbb{P}(Y = x)}{\mathbb{P}(X = x)}$$

$$= \frac{Q(x)}{P(x)}.$$

So $\mathbb{P}(X \ne Y \mid X = x) \ge \left(1 - \frac{Q(x)}{P(x)}\right)_+$

Squaring and taking $\mathbb{E}$,

$$\mathbb{E}\left[\mathbb{P}(X \ne Y \mid X)^2\right] \ge \sum_x P(x) \frac{(P(x) - Q(x))_+^2}{P(x)^2}$$

$$= d_2^2(Q, P).$$

To show that $\exists$ a coupling that achieves the RHS bound, we guess that its the same as the optimal coupling. (Check this in Sheet 3). $\square$

<u>Lemma</u>: $d_2^2(Q, P) \le 2 D(Q \| P)$.
<u>Proof:</u> (in notes, not examinable). $\square$

The above lemmas imply Marton's C.T.C.I. for $n = 1$:

$$\inf_\pi \mathbb{E}_\pi\left[\mathbb{P}(X_1 \ne Y_1 \mid X_1^2)\right] \le d_2^2(P_{X_1}, Q_{Y_1}) \le 2D(Q_{Y_1} \| P_{X_1})$$

We'll use induction (general case). Assume M.C.T.C.I. holds for $n \le k$. We'll prove it for $n = k+1$. We need to show:

$$\inf_{\pi \in \Pi(P_{Y_{1:k+1}}, Q_{Y_{1:k+1}})} \mathbb{E}_\pi\left[\sum_{i=1}^{k+1} \mathbb{P}(X_i \ne Y_i \mid X_{1:k+1})^2\right] \le 2D(Q_{Y_{1:k+1}} \| P_{Y_{1:k+1}})$$

We know: $\exists$ a coupling $\pi_k \in \Pi(P_{X_{1:k}}, Q_{Y_{1:k}})$ s.t.

$$\mathbb{E}_\pi\left[\sum_{i=1}^k \mathbb{P}(X_i \ne Y_i \mid X_{1:k})^2\right] \le 2 \cdot D(Q_{Y_{1:k}} \| P_{1:k})$$

Define $\pi(X_{1:k+1} = x_{1:k+1}, Y_{1:k+1} = y_{1:k+1})$
$= \pi_k(X_{1:k} = x_{1:k}, Y_{1:k} = y_{1:k})$
$\quad \times \pi_{y_{1:k}}(X_{k+1} = x_{k+1}, Y_{k+1} = y_{k+1})$

where $\pi_{y_{1:k}}$ is the optimal TV coupling between $P_{X_{k+1}}$ and $Q_{Y_{k+1} \mid Y_{1:k} = y_{1:k}}$.

π has nice properties such as:
1. Marginal of π on $(X_{1:k}, Y_{1:k})$ is π_k.
2. (X_{k+1}, Y_{k+1}) depend on $(X_{1:k}, Y_{1:k})$ only through $Y_{1:k}$
3. X_{k+1} is independent of $(X_{1:k}, Y_{1:k})$.

Within the coupling π:

$$\underbrace{\sum_{i=1}^k}_{\le} + \overbrace{(k+1)\text{-th term}}^{\le} \le 2 \cdot \left[D(Q_{Y_{1:k}} \| P_{X_{1:k}}) + D(Q_{Y_{k+1} \mid Y_{1:k}} \| P_{X_{k+1}} \mid Q_{Y_{1:k}})\right]$$

LECTURE 16

We know: $\pi_k \in \Pi(P_{X_{1:k}}, Q_{Y_{1:k}})$ s.t.
$\mathbb{E}_{\pi_k}\left[\sum_{i=1}^k \mathbb{P}(X_i \neq Y_i \mid X_{1:k})^2\right] \le 2 D(Q_{Y_{1:k}} \| P_{X_{1:k}})$

Define
$$\pi(X_{1:k+1} = x_{1:k+1}, Y_{1:k} = y_{1:k+1}) = \pi_k(X_{1:k} = x_{1:k}, Y_{1:k} = y_{1:k}) \times \pi_{y_{1:k}}(X_{k+1} = x_{k+1}, Y_{k+1} = y_{k+1})$$

where $\pi_{y_{1:k}}$ is the optimal TV-coupling between $P_{X_{k+1}}$ and $Q_{Y_{k+1} \mid Y_{1:k} = y_{1:k}}$.

We need to show: $(X_{1:k+1}, Y_{1:k+1}) \sim \pi$
$$\mathbb{E}\left[\sum_{i=1}^k \mathbb{P}(X_i \neq Y_i \mid X_{1:k+1})^2\right]$$
$$+ \mathbb{E}\left[\mathbb{P}(X_{k+1} \neq Y_{k+1} \mid X_{1:k+1})^2\right]$$
$$\le 2 D(Q_{Y_{1:k}} \| P_{X_{1:k}}) + 2D(Q_{Y_{k+1} \mid Y_{1:k}} \| P_{X_{k+1}} \mid Q_{Y_{1:k}})$$

① (X_{k+1}, Y_{k+1}) depends only on $(Y_{1:k})$ given $(X_{1:k}, Y_{1:k})$.
"$X_{1:k} \to Y_{1:k} \to (X_{k+1}, Y_{k+1})$"

② X_{k+1} is independent of $(X_{1:k}, Y_{1:k})$.

③ $(X_{1:k}, Y_{1:k}) \sim \pi_k$.

$\mathbb{P}(X_i \neq Y_i \mid X_{1:k+1}) = \mathbb{P}(X_i \neq Y_i \mid X_{1:k})$ for all $1 \le i \le k$.

By the assumption for $n = k$, we conclude
$$\mathbb{E}_\pi\left[\sum_{i=1}^{k+1} \mathbb{P}(X_i \neq Y_i \mid X_{1:k+1})^2\right] \le 2D(Q_{Y_{1:k}} \| P_{X_{1:k}})$$

We know by choice of $\pi_{y_{1:k}}$ that:
$$\mathbb{E}_{\pi_{y_{1:k}}}\left[\mathbb{P}(X_{k+1} \neq Y_{k+1} \mid Y_{1:k} = y_{1:k}, X_{k+1})^2\right]$$
$$\le 2D(Q_{Y_{k+1} \mid Y_{1:k} = y_{1:k}} \| P_{X_{k+1}})$$
By the $n=1$ case of inequality

If both sides were "integrated" wrt $Q_{Y_{1:k}}$ measure,

$$LHS = \mathbb{E}_\pi\left[\mathbb{P}(X_{k+1} \neq Y_{k+1} \mid Y_{1:k}, X_{k+1})^2\right]$$

$$RHS = 2D(Q_{Y_{k+1} \mid Y_{1:k}} \| P_{X_{k+1}} \mid Q_{Y_{1:k}}).$$

LHS is not what we want. We want $\mathbb{E}_\pi\left[\mathbb{P}(X_{k+1} \neq Y_{k+1} \mid X_{1:k+1})^2\right]$

Observe that since the distribution of (X_{k+1}, Y_{k+1}) depends only on $Y_{1:k}$ given $(X_{1:k}, Y_{1:k})$
$$\mathbb{P}(X_{k+1} \neq Y_{k+1} \mid Y_{1:k}, X_{k+1}) = \mathbb{P}(X_{k+1} \neq Y_{k+1} \mid Y_{1:k}, X_{1:k+1})$$

$$\mathbb{E}\left[\mathbb{P}(X_{k+1} \neq Y_{k+1} \mid Y_{1:k}, X_{1:k+1})^2\right]$$
$$= \mathbb{E}\left[\mathbb{E}[\mathbb{1}\{X_{k+1} \neq Y_{k+1}\} \mid X_{1:k+1}, Y_{1:k}]^2\right]$$
$$= \mathbb{E}\left[\mathbb{E}\left[\mathbb{E}[\mathbb{1}\{X_{k+1} \neq Y_{k+1}\} \mid X_{1:k+1}, Y_{1:k}]^2 \mid X_{1:k+1}\right]\right]$$
$$\ge \mathbb{E}\left[\mathbb{E}\left[\mathbb{E}[\mathbb{1}\{X_{k+1} \neq Y_{k+1}\} \mid X_{1:k+1}, Y_{1:k}] \mid X_{1:k+1}\right]^2\right]$$
$$= \mathbb{E}\left[\mathbb{E}[\mathbb{1}\{X_{k+1} \neq Y_{k+1}\} \mid X_{1:k+1}]^2\right]$$
$$= \mathbb{E}\left[\mathbb{P}(X_{k+1} \neq Y_{k+1} \mid X_{1:k+1})^2\right]$$

$\square$