High-dimensional Probability

Zhiyuan Zhan
<thaleszhan@gmail.com>

April 26, 2025

Contents

Chapter 1

Basic Inequalities

1.1 Introduction

1. Motivation problems:

   Example 1.1.1. (1) Coupon Collector Problem: Let X_i the color of the i-th coupon and $X_i \in [k]$. Let

   $$f(x_1, \cdots, x_n) := \text{ the number of distinctly } x_1, \ldots, x_n$$

   The problem is to find a integer n such that

   $$\mathbb{P}(f(X_1, \cdots, X_n) = k) \text{ is large } \quad \Leftrightarrow \; \mathbb{P}(f(X_1, \cdots, X_n) < k) \text{ is small}$$

   (2) Given two $i.i.d.$ sequences $X_1^n = (X_1, \cdots, X_n)$ and $Y_1^n = (Y_1, \cdots, Y_n)$, what is the length of a longest common subsequence?

2. Concentration inequalities: Chernoff type bounds; Tensorization techniques (Martingale, Efron-Stein, Entropy); Isoperimetric inequalities; Transportation methods; Talagrand's convex distance inequalities.

3. Applications: Johnson-Lindenstrauss lemma; Hypercontractivity; Blowing-up lemma; Problem as above.

1.2 Cramér-Chernoff Method

1. The Markov's inequality: For $X \geq 0$ and $\varepsilon > 0$, we have

   $$\mathbb{P}(X \geq \varepsilon) \leq \frac{\mathbb{E}[X]}{\varepsilon}$$

   which can be extended to any non-negative, non-decreasing function ϕ,

   $$\mathbb{P}(X \geq \varepsilon) = \mathbb{P}(\phi(X) \geq \phi(\varepsilon)) \leq \frac{\mathbb{E}[\phi(X)]}{\phi(\varepsilon)}$$

2. Chernoff bound: Let $\phi(x) = e^{\lambda x}$ for $\lambda \geq 0$. Then for any X and t,

   $$\mathbb{P}(X \geq t) = \mathbb{P}(\phi(X) \geq \phi(t)) \leq e^{-\lambda t}\mathbb{E}\left[e^{\lambda X}\right] = \exp(-(\lambda t - \psi_X(\lambda)))$$

where $\psi_X(\lambda) := \log \mathbb{E}\left[e^{\lambda X}\right]$. Therefore,

$$\begin{aligned}\mathbb{P}(X \geq t) &\leq \inf_{\lambda \geq 0} \exp(-(\lambda t - \psi_X(\lambda))) \\ &= \exp\left(-\sup_{\lambda \geq 0}(\lambda t - \psi_X(\lambda))\right) \\ &= \exp\left(-\psi_X^*(t)\right)\end{aligned}$$

where $\psi_X^*(t) := \sup_{\lambda \geq 0}(\lambda t - \psi_X(\lambda))$. In general, the Chernoff bound is

$$\mathbb{P}(X \geq t) \leq \exp\left(-\psi_X^*(t)\right)$$

Remark. (1) Equivalently, for $\delta \geq 0$,

$$\mathbb{P}(X \geq (\psi_X^*)^{-1}(\log 1/\delta)) \leq \delta$$

(2) $\psi_X(\lambda) = \log \mathbb{E}\left[e^{\lambda X}\right]$ is called the cumulant generating function of X at λ.

(3) $\psi_X^*(t) = \sup_{\lambda \geq 0}(\lambda t - \psi_X(\lambda))$ is called the Cramér transform of X at t.

3. Properties of ψ_X and ψ_X^*: Suppose that there is a $\lambda > 0$ such that $\psi_X(\lambda) < \infty$. Let $b > 0$ be the supremum over all such λ.

   (1) ψ_X is smooth over $(0, b)$.

   (2) ψ_X is convex on $(0, b)$ and strictly convex if X is not a constant random variable, because

   $$\begin{aligned}\psi_X(\theta\lambda_1 + (1-\theta)\lambda_2) &= \log \mathbb{E}\left[e^{\theta\lambda_1 X + (1-\theta)\lambda_2 X}\right] \\ &\leq \log \mathbb{E}\left[e^{\lambda_1 X}\right]^{\theta} \mathbb{E}\left[e^{\lambda_2 X}\right]^{(1-\theta)} \\ &= \theta\psi_X(\lambda_1) + (1-\theta)\psi_X(\lambda_2)\end{aligned}$$

   by Hölder's inequality (strictly convex by the "=" condition in Hölder's inequality).

   (3) ψ_X^* is convex on its domain because ψ_X^* is the max of linear functions.

   (4) ψ_X^* is nonnegative because $\psi_X(0) = 0$.

   (5) If $\mathbb{E}[X]$ is finite, then by Jensen's inequality

   $$\psi_X(\lambda) = \log \mathbb{E}\left[e^{\lambda X}\right] \geq \lambda \mathbb{E}[X] \;\Rightarrow\; \lambda \mathbb{E}[X] - \psi_X(\lambda) \leq 0$$

   Therefore, for any $t \geq \mathbb{E}[X]$ and $\lambda \leq 0$,

   $$\lambda t - \psi_X(\lambda) \leq \lambda \mathbb{E}[X] - \psi_X(\lambda) \leq 0$$

   It follows that for any $t \geq \mathbb{E}[X]$,

   $$\psi_X^*(t) = \sup_{\lambda \in \mathbb{R}}(\lambda t - \psi_X(\lambda))$$

   the Fenchel dual of ψ_X.

Example 1.2.1. Examples of Chernoff method:

(1) $X \sim \mathcal{N}(0, \sigma^2)$:

$$\mathbb{E}\left[e^{\lambda x}\right] = e^{\frac{\lambda^2\sigma^2}{2}}$$

For any $t \geq \mathbb{E}[X] = 0$,

$$\psi_X^*(t) = \sup_{\lambda \in \mathbb{R}}\left(\lambda t - \frac{\lambda^2\sigma^2}{2}\right) = \frac{t^2}{2\sigma^2}$$

Therefore, for any $t \geq 0$

$$\mathbb{P}(X \geq t) \leq e^{-\frac{t^2}{2\sigma^2}}$$

(2) $X \sim \text{Poi}(v)$: Note that $v > 0$.

$$\mathbb{E}\left[e^{\lambda x}\right] = e^{-v} \sum_{k=0}^{\infty} e^{\lambda k} \frac{v^k}{k!} = e^{(e^\lambda - 1)v} \ \Rightarrow\ \psi_X(\lambda) = (e^\lambda - 1)v$$

Therefore,

$$\psi_X^*(t) = \sup_{\lambda \geq 0} (\lambda t - v e^\lambda + v)$$

For $t < 0$, clearly $\psi_X^*(t) = 0$. For $t \geq 0$, because $\frac{d}{d\lambda}(\lambda t - v e^\lambda + v) = 0$ has solution $\lambda^* = \log(t/v)$. Therefore, for $t \geq v$,

$$\psi_X^*(t) = t\log(t/v) - t + v = vh\left(\frac{t}{v} - 1\right), \quad h(x) := (1+x)\log(1+x) - x$$

Therefore, for any $t \geq v$,

$$\mathbb{P}(X \geq t) \leq e^{-v} \exp\left(-t\left(\log\frac{t}{v} - 1\right)\right) \approx \mathcal{O}\left(e^{-t\log t}\right)$$

Remark. The RHS < 1 if and only if $\log\frac{t}{v} > 1 - \frac{v}{t}$. Then $t > 7.3$. Otherwise, this is not useful. Besides,

$$h(x) = (1+x)\log(1+x) - x \geq \frac{x^2}{2(1+x/3)}$$

So for $t \geq v$,

$$\psi_X^*(t) \geq \frac{v}{2} \frac{(t/v - 1)^2}{1 + (t/v-1)/3}$$

(3) $X \sim \text{Ber}(p)$ with $p < \frac{1}{2}$:

$$\psi_X(\lambda) = \log pe^\lambda + 1 - p$$

For $0 < t \leq 1$, we have

$$\psi_X^*(t) = (1-t)\log\frac{1-t}{1-p} + t\log\frac{t}{p} = D\left(\text{Ber}(t) \parallel \text{Ber}(p)\right) = D\left(t \parallel p\right)$$

and $t > 1$, $\psi_X^*(t) = \infty$. Then for $0 < t \leq 1$,

$$\mathbb{P}(X \geq t) = \exp(-D(t \parallel p))$$

(4) Sum of independent random variables: $Z = X_1 + \cdots + X_n$ implies

$$\psi_Z(\lambda) = \sum_i \psi_{X_i}(\lambda)$$

If $X_1, \cdots, X_n$ is *i.i.d.* as X, then

$$\psi_Z^*(t) = \sup_{\lambda \geq 0}(t\lambda - n\psi_X(\lambda)) = n\psi_X^*\left(\frac{t}{n}\right)$$

Therefore,

$$\mathbb{P}\left(\sum_i X_i \geq t\right) \leq \exp\left(-n\psi_X^*\left(\frac{t}{n}\right)\right)$$

e.g. $X_i \sim \text{Ber}(0)$,

$$\mathbb{P}\left(\sum_i X_i \geq t\right) \leq \exp\left(-nD\left(\frac{t}{n} \parallel p\right)\right)$$

$$\Leftrightarrow\ \mathbb{P}\left(\frac{1}{n}\sum_i X_i \geq p + \theta\right) \leq \exp\left(-nD(p+\theta \parallel p)\right)$$

(5) χ^2-distribution: $X \sim \mathcal{N}(0, \sigma^2)$ and $Y = X^2$.

$$\mathbb{E}\left[e^{\lambda Y}\right] = \frac{1}{\sqrt{1 - 2\lambda\sigma^2}}, \quad \lambda < \frac{1}{2\sigma^2}$$

otherwise, it is ∞. It implies that

$$\psi_Y(\lambda) = \frac{1}{2}\log(1 - 2\sigma^2\lambda), \quad \lambda < \frac{1}{2\sigma^2}$$

Let $Z = Y - \mathbb{E}[Y] = X^2 - \sigma^2$.

$$\psi_Z(\lambda) = -\frac{1}{2}\log(1 - 2\sigma^2\lambda) - \lambda\sigma^2, \quad \lambda < \frac{1}{2\sigma^2}$$

For $t \geq \mathbb{E}[Z] = 0$,

$$\psi_Z^*(t) = \sup_{\lambda} \frac{1}{2}\log(1 - 2\sigma^2\lambda) + \lambda(\sigma^2 + t) = \frac{1}{2}\log\left(\frac{\sigma^2}{\sigma^2 + t}\right) + \frac{t}{\sigma^2} =: \frac{1}{2}h_1\left(\frac{t}{\sigma^2}\right)$$

Likewise, for $\lambda > 0$,

$$\psi_{-Z}(\lambda) = -\frac{1}{2}\log(1 + 2\sigma^2\lambda) + \lambda\sigma^2$$

and

$$\psi_{-Z}^*(t) = \frac{1}{2}\log\left(\frac{\sigma^2}{\sigma^2 - t}\right) - \frac{t}{\sigma^2} = \frac{1}{2}h_2\left(\frac{t}{\sigma^2}\right), \quad 0 \leq t \leq \sigma^2$$

Note that

$$h_1(x) = -\log(1 + x) + x \;\Rightarrow\; \frac{1}{2}h_1(x) \geq 1 - \sqrt{1 + x} + \frac{x}{2}$$

so that

$$\psi_Z^*(t) \geq 1 + \frac{t}{\sigma^2} - \sqrt{1 + \frac{t}{\sigma^2}}$$

Similarly,

$$h_2(x) = -\log(1 - x) - x \geq \frac{x^2}{2}, \quad x \geq 0$$

so that

$$\psi_{-Z}^*(t) \geq \frac{t^2}{4\sigma^2}, \quad t \geq 0$$

1.3 Sub-Gaussian: Hoeffding

1. Sub-Gaussian random variable:

Definition 1.3.1. A random variable X with $\mathbb{E}X = 0$ is called to be sub-Gaussian if there is a $v > 0$ such that for all λ

$$\psi_X(\lambda) \leq \frac{\lambda^2 v}{2} \;\Leftrightarrow\; \mathbb{E}e^{\lambda X} \leq e^{\lambda G},\; G \sim \mathcal{N}(0, \lambda)$$

and $X \in \mathcal{G}(v)$.

Proposition 1.3.1. *(1) If $X_i \sim \mathcal{G}(v_i)$ are independent, then*

$$\sum_i X_i \sim \mathcal{G}\left(\sum_i v_i\right)$$

(2) If $X \in \mathcal{G}(v)$, then $-X \in \mathcal{G}(v)$, and for any $t \geq 0$

$$\mathbb{P}(X \geq t) \leq \exp\left(-\frac{t^2}{2v}\right), \quad \mathbb{P}(-X \geq t) \leq \exp\left(-\frac{t^2}{2v}\right)$$

(3) $X \in \mathcal{G}(v)$ implies $\operatorname{Var}(X) \leq v$ by differentiation.

Theorem 1.3.1. *Let $\mathbb{E}X = 0$. Then TFAE for chosen v, b, C, α.*

(1) $X \in \mathcal{G}(v)$.

(2) $\mathbb{P}(|X| \geq t) \leq 2e^{-bt^2}$.

(3) For any $q \in \mathbb{N}$, $\mathbb{E}\left[X^{2q}\right] \leq C^q q!$.

(4) $\mathbb{E}\left[e^{\alpha X^2}\right] \leq 2$.

Proof. (1) implies (2) is by above. For (2) $\Rightarrow$ (3), WTLG, let $b = 1$. Then

$$\begin{aligned}
\mathbb{E}\left[X^{2q}\right] &= \int_0^\infty \mathbb{P}\left\{|X|^{2q} > x\right\} dx \\
&= 2q \int_0^\infty x^{2q-1} \mathbb{P}\{|X| > x\} dx \\
&\leq 4q \int_0^\infty x^{2q-1} e^{-x^2/2} dx \\
&= 4q \int_0^\infty (2t)^{q-1} e^{-t} dt = 2^{q+1} q!
\end{aligned}$$

where the first equality is by the theory of distribution function and $x = \sqrt{2t}$ for the change of variable in the last 2 equality. For (3) $\Rightarrow$ (1), consider a copy of X denoted by X'.

$$\mathbb{E}e^{\lambda X}\mathbb{E}e^{-\lambda X} = \mathbb{E}e^{\lambda(X-X')} = \sum_{q=0}^\infty \frac{\lambda^{2q}\mathbb{E}\left[(X-X')^{2q}\right]}{(2q)!}$$

By the convexity of $x \mapsto x^{2q}$,

$$(X - X')^{2q} \leq 2^{2q-1}\left(X^{2q} + (X')^{2q}\right)$$

Therefore,

$$\mathbb{E}\left[(X-X')^{2q}\right] \leq 2^{2q-1}\left(\mathbb{E}\left[X^{2q}\right] + \mathbb{E}\left[X'^{2q}\right]\right) = 2^{2q}\mathbb{E}\left[X^{2q}\right]$$

It follows that

$$\mathbb{E}e^{\lambda X}\mathbb{E}e^{-\lambda X} = \sum_{q=0}^\infty \frac{\lambda^{2q}\mathbb{E}\left[(X-X')^{2q}\right]}{(2q)!} \leq \sum_{q=0}^\infty \frac{\lambda^{2q}2^{2q}C^q q!}{(2q)!}$$

Because $q \in \mathbb{N}$,

$$\frac{(2q)!}{q!} = \prod_{j=1}^q (q+j) \geq \prod_{j=1}^q (2j) = 2^q q!$$

Thus

$$\mathbb{E}e^{\lambda X}\mathbb{E}e^{-\lambda X} \leq \sum_{q=0}^\infty \frac{\lambda^{2q}2^q C^q}{q!} = e^{2\lambda^2 C} \;\Rightarrow\; \mathbb{E}e^{\lambda X} \leq e^{2\lambda^2 C}$$

by $\mathbb{E}e^{-\lambda X} \geq 1$ since $\mathbb{E}[X]=0$. For (3) $\Rightarrow$ (4), by setting $\alpha = 1/(2C)$

$$\mathbb{E}\exp\left(\alpha X^2\right) = \sum_{q=0}^{\infty} \frac{\alpha^q \mathbb{E}\left[X^{2q}\right]}{q!} \leq \sum_{q=0}^{\infty} 2^{-q} = 2$$

For (4) $\Rightarrow$ (3),

$$\mathbb{E}\exp\left(\alpha X^2\right) \leq 2 \;\Rightarrow\; \sum_{q=1}^{\infty} \frac{\alpha^q \mathbb{E}\left[X^{2q}\right]}{q!} \leq 1$$

It follows that

$$\mathbb{E}\left[X^{2q}\right] \leq \alpha^{-q} q!$$

□

2. Hoeffding's Inequality:

Lemma 1.3.1 (Hoeffding's lemma)**.** *Suppose a random variable X takes values in $[a,b]$ and $\mathbb{E}[X]=0$. Then $X \in \mathcal{G}\left(\frac{(b-a)^2}{4}\right)$.*

Proof. If $X \in [a,b]$, then $\left|X - \frac{a+b}{2}\right| \leq \frac{b-a}{2}$. So

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

Besides, for $\psi_X(\lambda) = \log \mathbb{E}[e^{\lambda X}]$, $\psi_X(0) = \psi'(0) = 0$ and

$$\psi''_X(\lambda) = \frac{\mathbb{E}\left[X^2 e^{\lambda X}\right]}{\mathbb{E}\left[e^{\lambda X}\right]} - \left(\frac{\mathbb{E}\left[X e^{\lambda X}\right]}{\mathbb{E}\left[e^{\lambda X}\right]}\right)^2$$

Let $\mathbb{P}$ be the probability and a new distribution $\mathbb{Q} \ll \mathbb{P}$

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = \frac{e^{\lambda X}}{\psi_X(\lambda)}$$

which also implies that $\left|X - \frac{a+b}{2}\right| \leq \frac{b-a}{2}$ in $\mathbb{Q}$. Then

$$\mathbb{E}_{\mathbb{Q}}[X^2] = \int X^2 d\mathbb{Q} = \int X^2 \frac{e^{\lambda X}}{\psi_X(\lambda)} d\mathbb{P}$$

Therefore,

$$\psi''_X(\lambda) = \mathbb{E}_{\mathbb{Q}}[X^2] - \left(\mathbb{E}_{\mathbb{Q}}[X]\right)^2 = \mathrm{Var}_{\mathbb{Q}}(X) \leq \frac{(b-a)^2}{4}$$

Then we have for any λ, by Taylor's theorem,

$$\psi_X(\lambda) \leq \frac{(b-a)^2\lambda^2}{8} \;\Rightarrow\; X \in \mathcal{G}\left(\frac{(b-a)^2}{4}\right)$$

□

Remark. This lemma is tight.

Corollary 1.3.1 (Hoeffding's Inequality)**.** *Consider independent random variables $X_1, \cdots, X_n$ such that $X_i \in [a_i, b_i]$ and $\mathbb{E}[X_i]=0$. Then*

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

3. Bonnett's Inequality:

Lemma 1.3.2 (Bonnett's Inequality)**.** *Consider independent random variables $X_1, \cdots, X_n$ such that $|X_i| \leq C$ and $\mathbb{E}[X_i] = 0$. Let*

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n} \mathrm{Var}(X_i)$$

Then we have

$$\mathbb{P}\left(\sum_{i=1}^{n} X_i \geq t\right) \leq \exp\left(-\frac{n\sigma^2}{C^2} h\left(\frac{tC}{\sigma^2 n}\right)\right)$$

where $h(x) = (1+x)\log(1+x) - x$.

Proof. First,

$$\begin{aligned}
\mathbb{E}\left[e^{\lambda X_i}\right] &= 1 + \sum_{k=2}^{\infty} \frac{\lambda^k \mathbb{E} X_i^k}{k!} \\
&\leq 1 + \sum_{k=2}^{\infty} \frac{\lambda^k}{k!} \mathbb{E}\left[X_i^2 \left|X_i\right|^{k-2}\right] \\
&\leq 1 + \sum_{k=2}^{\infty} \frac{\lambda^k}{k!} C^{k-2}\,\mathrm{Var}(X_i) \\
&\leq 1 + \frac{\sigma_i^2}{c^2} \sum_{k=2}^{\infty} \frac{(c\lambda)^k}{k!} = 1 + \frac{\sigma_i^2}{c^2}\left(e^{\lambda c} - 1 - \lambda c\right) \\
&\leq \exp\left(\frac{\sigma_i^2}{c^2}\left(e^{\lambda c} - 1 - \lambda c\right)\right)
\end{aligned}$$

Therefore,

$$\psi_{\sum X_i}(\lambda) \leq \frac{n\sigma^2}{c^2}\left(e^{\lambda c} - 1 - \lambda c\right)$$

Recall if $X \in \mathrm{Poi}(v)$, then $\psi_{X-v}(\lambda) = v(e^\lambda - 1 - \lambda)$ and $\psi^*_{X-v}(t) = vh(t/v)$. Therefore,

$$\psi^*_{\sum X_i}(t) \geq \frac{n\sigma^2}{c^2} h\left(\frac{tc}{n\sigma^2}\right)$$

□

Corollary 1.3.2 (Bernstein's Inequality)**.** *In the same settings,*

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

Proof. It is because $h(x) \geq \frac{x^2}{2+\frac{2}{3}x}$ □

Example 1.3.1. Let $X_i \sim \mathrm{Ber}(p)$ and $S_n = \sum_{i=1}^n X_i$. Then Hoeffding's Inequality provides

$$\mathbb{P}(S_n - np \geq t) \leq \exp\left(-\frac{2t^2}{n}\right) \Rightarrow \mathbb{P}(S_n - np \geq \sqrt{\frac{n}{2}\log\frac{1}{\delta}}) \leq \delta$$

and Bernstein's Inequality provides

$$\begin{aligned}
&\mathbb{P}(S_n - np \geq t) \leq \exp\left(\frac{-t^2}{2np(1-p) + \frac{2t}{3}}\right) \\
&\Rightarrow \mathbb{P}\left(S_n - np \geq \log(1/\delta) + 4\sqrt{np(1-p)\log(1/\delta)}\right) \leq \delta
\end{aligned}$$

1.4 Azuma and McDiarmid

1. Azuma-Hoeffding Inequality:

   Definition 1.4.1. Random variables $X_1, \cdot, X_n$ are said to constitute a multiplicative family if for every distinct $i_1 < \cdots < i_k$ with $k \leq n$,

   $$\mathbb{E}[X_{i_1} \cdots X_{i_k}] = 0$$

   Example 1.4.1. (1) Independent $X_1, \cdots, X_n$ with $\mathbb{E}[X_i] = 0$ constitute a multiplicative family.

   (2) For a martingale difference sequence $X_1, \cdots, X_n$, *i.e.*

   $$\mathbb{E}X_1 = 0,\ \mathbb{E}[X_2 \mid X_1] = 0, \cdots, \mathbb{E}[X_n \mid X_1, \cdots, X_{n-1}] = 0$$

   then by the tower property

   $$\begin{aligned}\mathbb{E}[X_{i_1} \cdots X_{i_k}] &= \mathbb{E}\left[\mathbb{E}[X_{i_1} \cdots X_{i_k} \mid X_1, \cdots, X_{i_k - 1}]\right] \\ &= \mathbb{E}\left[X_{i_1} \cdots X_{i_{k-1}} \mathbb{E}[X_{i_k} \mid X_1, \cdots, X_{i_k - 1}]\right] \\ &= 0\end{aligned}$$

   they constitute a multiplicative family.

   Lemma 1.4.1 (Azuma-Hoeffding Inequality)**.** *For a multiplicative family $X_1, \cdot, X_n$ with $|X|_i \leq c_i$ and $t > 0$,*

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

   Proof. First, let make an upper bound linear approximation to $e^{\lambda x}$.

   $$e^{\lambda x} \leq ax + b, \quad \forall\ x \in [-c, c]$$

   by setting

   $$a = \frac{e^{\lambda c} - e^{-\lambda c}}{2c}, \quad b = \frac{e^{\lambda c} + e^{-\lambda c}}{2}$$

   For $\lambda > 0$,

   $$\begin{aligned}\mathbb{E}\left[e^{\lambda \sum_{i=1}^n X_i}\right] &= \mathbb{E}\left[\Pi_{i=1}^n e^{\lambda X_i}\right] \\ &\leq \mathbb{E}\left[\Pi_{i=1}^n (a_i X_i + b_i)\right] = \mathbb{E}\left[\Pi_{i=1}^n b_i\right] \\ &= \Pi_{i=1}^n \left(\frac{e^{\lambda c_i} + e^{-\lambda c_i}}{2}\right) \\ &= \Pi_{i=1}^n \left(1 + \frac{\lambda^2 c_i^2}{2!} + \frac{\lambda^4 c_i^4}{4!} + \frac{\lambda^6 c_i^6}{6!} \cdots\right) \\ &\leq \Pi_{i=1}^n \left(1 + \frac{\lambda^2 c_i^2}{2} + \frac{\lambda^4 c_i^4/4}{2!} + \frac{\lambda^6 c_i^6/8}{3!} \cdots\right) \\ &= \Pi_{i=1}^n \exp\left(\frac{\lambda^2 c_i^2}{2}\right)\end{aligned}$$

   Therefore,

   $$\psi_{\sum X_i} \leq \frac{\lambda^2}{2} \sum_{i=1}^n c_i^2$$

   □

2. Applications: Let $Z_0, Z_1, \cdots, Z_n$ be a martingale. Then we can see

$$X_i = Z_i - Z_{i-1}$$

is a martingale difference sequence because

$$\mathbb{E}[X_i \mid X_1, \cdots, X_{i-1}] = \mathbb{E}[Z_i - Z_{i-1} \mid Z_0, \cdots, Z_{i-1}] = Z_{i-1} - Z_{i-1} = 0$$

And so they constitute a multiplicative family. Assume that $|X_i| \leq c$, then

$$\mathbb{P}(\max_i (Z_i - Z_0) \geq t) \leq \exp\left(-\frac{t^2}{2nc^2}\right)$$

by Azuma-Hoeffding Inequality. This strengthens the Doob's Inequality. Because $Z_i - Z_0$ is also a martingale, which implies that $e^{\lambda(Z_i - Z_0)}$ is a non-negative submartingale, Doob's Inequality provides

$$\mathbb{P}\left(\max_i (Z_i - Z_0) \geq t\right) = \mathbb{P}\left(\max_i e^{\lambda(Z_i - Z_0)} \geq e^{\lambda t}\right) \leq \frac{\mathbb{E}\left[e^{\lambda(Z_n - Z_0)}\right]}{e^{\lambda t}}$$

3. McDiarmid's Inequality:

Definition 1.4.2. Let S be any set. A function $f \colon S^n \to \mathbb{R}$ satisfies the Bounded Difference Property (BDP) with constants $\boldsymbol{c} = (c_1, \cdots, c_n)$ if

$$|f(x_1, \cdots, x_i, \cdots, x_n) - f(x_1, \cdots, x_i', \cdots, x_n)| \leq c_i, \quad \forall\, x_i, x_i', \ \forall\, i$$

Remark. Consider the Hamming Distance, *i.e.* for $x, y \in S^n$

$$d_{\boldsymbol{c}}(x, y) := \sum_{i=1}^{n} c_i \mathbb{1}\{x_i \neq y_i\}$$

Then f satisfies BDP if and only if

$$|f(x) - f(y)| \leq d_{\boldsymbol{c}}(x, y)$$

i.e. f is 1-Lipschitz continuous.

Lemma 1.4.2 (McDiarmid's Inequality)**.** *Suppose f satisfies BDP with $\boldsymbol{c} = (c_1, \cdots, c_n)$. Then if $Z = f(X_1, \cdots, X_n)$ with independent random variables $X_1, \cdots, X_n$, then*

$$\mathbb{P}(Z - \mathbb{E}Z \geq t) \leq \exp\left(-\frac{t^2}{2\sum_i c_i^2}\right)$$

similarly, because f BDP implies $-f$ BDP,

$$\mathbb{P}(-Z + \mathbb{E}Z \geq t) \leq \exp\left(-\frac{t^2}{2\sum_i c_i^2}\right)$$

Proof. Let $Y_i = \mathbb{E}[Z \mid X_1, \cdots, X_i] = g(X_1, \cdots, X_i)$ for $i \geq 1$ and $Y_0 = \mathbb{E}[Z]$. Then Y_i is a martingale, because

$$\mathbb{E}\left[\mathbb{E}[Z \mid \sigma(X_1, \cdots, X_i)] \mid \sigma(X_1, \cdots, X_{i-1})\right] = \mathbb{E}[Z \mid \sigma(X_1, \cdots, X_{i-1})]$$

can be checked by the definition of conditional expectation (Note that here we do not need the independence). Consider $Y_i - Y_{i-1}$, by the independence and the BDP of f, we have

$$\begin{aligned}
(Y_i - Y_{i-1})(x_1, \cdots, x_i) &= \mathbb{E}[Z \mid X_1 = x_1, \cdots, X_i = x_i] - \mathbb{E}[Z \mid X_1 = x_1, \cdots, X_{i-1} = x_{i-1}] \\
&= \int f(x_1, \cdots, x_i, w_{i+1}, \cdots, w_n) \frac{p(x^i, w_{i+1}^n)}{p(x^i)} dw_{i+1}^n \\
&\quad - \int f(x_1, \cdots, x_{i-1}, w_i, \cdots, w_n) \frac{p(x^{i-1}, w_i^n)}{p(x^{i-1})} dw_i^n \\
&= \int f(x_1, \cdots, x_i, w_{i+1}, \cdots, w_n) p(w_{i+1}^n) dw_{i+1}^n \\
&\quad - \int \int_{w_i} f(x_1, \cdots, x_{i-1}, w_i, \cdots, w_n) p(w_i) p(w_{i+1}^n) dw_i dw_{i+1}^n \\
&\le \int f(x_1, \cdots, x_i, w_{i+1}, \cdots, w_n) p(w_{i+1}^n) dw_{i+1}^n \\
&\quad - \int \int_{w_i} (f(x_1, \cdots, x_i, w_{i+1}, \cdots, w_n) - c_i) p(w_i) p(w_{i+1}^n) dw_i dw_{i+1}^n \\
&= c_i
\end{aligned}$$

Similarly, $Y_i - Y_{i-1} \ge -c_i$. Then by Azuma's inequality,

$$\mathbb{P}\left(Y_n - Y_0 \ge t\right) = \mathbb{P}\left(Z - \mathbb{E}Z \ge t\right) \le \exp\left(-\frac{t^2}{2\sum_i c_i^2}\right)$$

because $Y_n = \mathbb{E}[Z \mid X_1, \cdots, X_n] = Z$. □

Remark. In fact, if σ-algebra $\mathcal{H} \subset \mathcal{G}$, then we can see

$$\mathbb{E}\left[\mathbb{E}[X \mid \mathcal{G}] \mid \mathcal{H}\right] = \mathbb{E}[X \mid \mathcal{H}]$$

because for any $A \in \mathcal{H} \subset \mathcal{G}$,

$$\mathbb{E}\left[\mathbb{1}_A \mathbb{E}[X \mid \mathcal{G}]\right] = \mathbb{E}\left[\mathbb{E}[X \mathbb{1}_A \mid \mathcal{G}]\right] = \mathbb{E}[X \mathbb{1}_A]$$

1.5 Efron-Stein Inequality

1. Derivation: Let $X_1, \cdots, X_i$ be independent and $Z = f(X_1, \cdots, X_n)$. Let $Z_i = \mathbb{E}[Z \mid X_1, \cdots, X_i]$ and $Z_0 = \mathbb{E}[Z]$ and $\Delta_i = Z_i - Z_{i-1}$. Note that

$$Z - \mathbb{E}[Z] = Z_n - Z_0 = \sum_{i=1}^n \Delta_i$$

Besides, because Z_i is a martingale, Δ_i is a martingale of difference so that it is a multiplicative family. It follows that $\mathbb{E}[\Delta_i \Delta_j] = 0$ for $i \neq j$. So

$$\operatorname{Var}(Z) = \operatorname{Var}\left(\sum_{i=1}^n \Delta_i\right) = \sum_{i=1}^n \operatorname{Var}\left(\Delta_i\right) = \sum_{i=1}^n \mathbb{E}\left[\Delta_i^2\right]$$

Lemma 1.5.1. *For independent A, B, C and $Z = f(A, B, C)$, we have*

$$\mathbb{E}\left[\mathbb{E}[Z \mid AC] \mid AB\right] = \mathbb{E}[Z \mid A]$$

Proof. Because $\mathbb{E}[Z \mid AC]$ is a function of A, C, B is independent with $\sigma(\mathbb{E}[Z \mid AC], A)$,

$$\mathbb{E}\left[\mathbb{E}[Z \mid AC] \mid AB\right] = \mathbb{E}\left[\mathbb{E}[Z \mid AC] \mid A\right] = \mathbb{E}[Z \mid A] \qquad \square$$

This lemma implies that

$$\begin{aligned}\Delta_i &= \mathbb{E}[Z \mid X^i] - \mathbb{E}[Z \mid X^{i-1}] \\ &= \mathbb{E}[Z \mid X^i] - \mathbb{E}\left[\mathbb{E}[Z \mid X^{(i)}] \mid X^i\right] \\ &= \mathbb{E}\left[\left(Z - \mathbb{E}[Z \mid X^{(i)}]\right) \mid X^i\right]\end{aligned}$$

where $X^i = (X_1, \cdots, X_i)$ and $X^{(i)} = (X_1, \cdots, X_{i-1}, X_{i+i}, \cdots, X_n)$ for simplicity. It follows that

$$\begin{aligned}\mathbb{E}\left[\Delta_i^2\right] &= \mathbb{E}\left[\mathbb{E}\left[\left(Z - \mathbb{E}[Z \mid X^{(i)}]\right) \mid X^i\right]^2\right] \\ &\leq \mathbb{E}\left[\mathbb{E}\left[\left(Z - \mathbb{E}[Z \mid X^{(i)}]\right)^2 \mid X^i\right]\right] \\ &= \mathbb{E}\left[\left(Z - \mathbb{E}[Z \mid X^{(i)}]\right)^2\right]\end{aligned}$$

by Jensen's Inequality. Therefore, we have

$$\operatorname{Var}(Z) \leq \sum_{i=1}^{n} \mathbb{E}\left[\left(Z - \mathbb{E}[Z \mid X^{(i)}]\right)^2\right] = \sum_{i=1}^{n} \mathbb{E}\left[\mathbb{E}\left[\left(Z - \mathbb{E}[Z \mid X^{(i)}]\right)^2 \mid X^{(i)}\right]\right]$$

Let $\operatorname{Var}_{(i)}(Z) := \mathbb{E}\left[\left(Z - \mathbb{E}[Z \mid X^{(i)}]\right)^2 \mid X^{(i)}\right]$. Then we have

$$\operatorname{Var}(Z) \leq \sum_{i=1}^{n} \mathbb{E}\left[\operatorname{Var}_{(i)}(Z)\right]$$

which is called the Efron-Stein Inequality. There are another formulas for the Efron-Stein Inequality.

(i) Let

$$Z_i' = f(X_1, \cdots, X_{i-1}, X_i', X_{i+1}, X_n)$$

where X_i' is an independent copy of X_i. For *i.i.d.* X, Y,

$$\frac{1}{2}\mathbb{E}[(X - Y)^2] = \frac{1}{2}(\mathbb{E}[X^2] + \mathbb{E}[Y^2] - 2\mathbb{E}[X]\mathbb{E}[Y]) = \operatorname{Var}(X)$$

By this fact, we have

$$\operatorname{Var}_{(i)}(Z) = \frac{1}{2}\mathbb{E}\left[(Z_i - Z_i')^2 \mid X^{(i)}\right] = \mathcal{E}(f)$$

which induces

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

(ii) Note that

$$\operatorname{Var}(X) = \min\left\{\mathbb{E}(X - X')^2 \colon X' \text{ independent with } X,\ \mathbb{E}\left[(X')^2\right] < \infty\right\}$$

It is because that

$$\mathbb{E}(X - X')^2 = \operatorname{Var}(X) + \mathbb{E}\left[(X' - \mathbb{E}[X])^2\right]$$

Therefore,

$$\mathrm{Var}_{(i)}(Z) = \min\left\{\mathbb{E}\left[(Z - Z_i')^2 \mid X^{(i)}\right] : Z_i' \perp\!\!\!\perp Z \mid X^{i-1}, X_{i+1}^n,\ \mathbb{E}\left[(Z_i')^2\right] < \infty\right\}$$

where the infimum is achieved at $Z_i' = \mathbb{E}\left[Z \mid X^{(i)}\right]$, a function of $X^{(i)}$. So we have

$$\mathrm{Var}(Z) \leq \sum_{i=1}^n \min\left\{\mathbb{E}(Z - Z_i')^2 \colon Z_i' = g_i(X^{(i)}),\ \mathbb{E}\left[(Z_i')^2\right] < \infty\right\}$$

As a corollary, using

$$g_i(X^{(i)}) = \frac{1}{2}\left(\inf_{x_i} f(X^i, x_i, X_{i+1}^n) + \sup_{x_i} f(X^i, x_i, X_{i+1}^n)\right)$$

It implies that f satisfies the $(c_1, \cdots, c_n)$-BDP, then

$$Z - g_i(X^{(i)}) \leq c_i$$

Therefore,

$$\mathrm{Var}(Z) \leq \sum_{i=1}^n c_i^2$$

2. Concentration Bounds:

Theorem 1.5.1. *Let $X_1, \cdots, X_n$ be independent. Let $Z = f(X_1, \cdots, X_n)$ and*

$$Z_i' = f(X_1, \cdots, X_{i-1}, X_i', X_{i+1}, \cdots, X_n)$$

where X_i' is an independent copy of X_i. Suppose that

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

Then we have

$$\mathbb{P}\left(Z - \mathbb{E}[Z] > t\right) \leq 2e^{-\frac{t}{\sqrt{v}}}, \quad t \geq 0$$

Proof. I. Consider $Y = e^{\lambda(Z - \mathbb{E}[Z])/2}$.

Claim: If

$$\mathrm{Var}(Y) \leq \frac{\lambda^2}{4} v \mathbb{E}\left[e^{\lambda(Z - \mathbb{E}[Z])}\right], \quad \forall\ \lambda \geq 0$$

then

$$\psi_{Z - \mathbb{E}[Z]}(\frac{1}{\sqrt{v}}) \leq \log \frac{16}{9}$$

Proof of the Claim. First,

$$\begin{aligned}\mathrm{Var}(Y) &= \mathbb{E}\left[e^{\lambda(Z - \mathbb{E}[Z])}\right] - \left(\mathbb{E}\left[e^{\frac{\lambda(Z - \mathbb{E}[Z])}{2}}\right]\right)^2 \\ &\leq \frac{\lambda^2}{4} v \mathbb{E}\left[e^{\lambda(Z - \mathbb{E}[Z])}\right]\end{aligned}$$

Then we get

$$\left(1 - \frac{\lambda^2}{4} v\right) \mathbb{E}\left[e^{\lambda(Z - \mathbb{E}[Z])}\right] \leq \left(\mathbb{E}\left[e^{\frac{\lambda(Z - \mathbb{E}[Z])}{2}}\right]\right)^2$$

which is equivalent to

$$\psi_{Z-\mathbb{E}[Z]}(\lambda) + \log\left(1 - \frac{\lambda^2}{4}v\right) \leq 2\psi_{Z-\mathbb{E}[Z]}(\frac{\lambda}{2})$$

For simplification, let $g(\lambda) = \psi_{Z-\mathbb{E}[Z]}(\lambda)$. It follows that

$$g(\lambda) + \log\left(1 - \frac{\lambda^2}{4}v\right) \leq 2\left(-\log\left(1 - \frac{\lambda^2}{4^2}\right) + 2g(\frac{\lambda}{4})\right)$$

and so on so forth,

$$g(\lambda) \leq -\sum_{i=0}^{k} 2^i \log\left(1 - \frac{\lambda^2}{2^{2(i+1)}}v\right) + 2^{k+1}g(\frac{\lambda}{2^{k+1}})$$

Taking $k \to \infty$, consider the two terms on the RHS. First, for the second term

$$\lambda \lim_{k\to\infty} \frac{g(\frac{\lambda}{2^{k+1}})}{\lambda/2^{k+1}} = \lambda g'(0) = 0$$

Therefore,

$$\begin{aligned} g(\lambda) &\leq -\sum_{i=0}^{\infty} 2^i \log\left(1 - \frac{\lambda^2}{2^{2(i+1)}}v\right) \\ &= \sum_{i=0}^{\infty} 2^{-i}\left(-2^{2i}\log\left(1 - \frac{\lambda^2}{4\cdot 2^{2i}}v\right)\right) \end{aligned}$$

Consider the function $\log(1-cx)$ with $c > 0$,

$$-\frac{\log(1-cx)}{x} \leq -\log(1-c), \quad x \in [0,1]$$

Therefore,

$$g(\lambda) \leq -\sum_{i=0}^{\infty} 2^{-i}\log(1 - \frac{\lambda^2 v}{4}) \;\Rightarrow\; g(\sqrt{v}) \leq \log\frac{16}{9} \qquad \square$$

Under this claim, because

$$\mathbb{P}\left(Z - \mathbb{E}[Z] > t\right) \leq e^{\psi_{Z-\mathbb{E}[Z]}(\lambda)}e^{-\lambda t}$$

by taking $\lambda = \frac{1}{\sqrt{v}}$, we have

$$\mathbb{P}\left(Z - \mathbb{E}[Z] > t\right) \leq \frac{16}{9}e^{-\frac{t}{\sqrt{v}}}$$

which is desired result.

II. By the Efron-Stein Inequality,

$$\mathrm{Var}(Y) \leq \frac{1}{2}\sum_{i=1}^{n}\mathbb{E}\left[(Y - Y_i')^2\right] = \sum_{i=1}^{n}\mathbb{E}\left[(Y - Y_i')_+^2\right]$$

where

$$Y = e^{\frac{\lambda}{2}(Z-\mathbb{E}[Z])}, \quad Y_i' = e^{\frac{\lambda}{2}\left(Z_i'-\mathbb{E}[Z]\right)}$$

Note that

$$e^{\lambda x} - e^{\lambda y} \leq (x-y)\lambda e^{\lambda y}, \quad y \geq x$$

Therefore,

$$(Y - Y_i')_+^2 \leq (Z - Z_i')_+^2 \cdot \frac{\lambda^2}{4} \cdot Y^2$$

It follows that

$$\mathrm{Var}(Y) \leq \sum_{i=1}^{n}\mathbb{E}\left[(Z - Z_i')_+^2\right] \cdot \frac{\lambda^2}{4} \cdot \mathbb{E}\left[Y^2\right] \leq \frac{\lambda^2}{4}v\mathbb{E}\left[e^{\lambda(Z-\mathbb{E}[Z])}\right] \qquad \square$$

1.6 Gaussian-Poincaré Inequality

1. Poincaré Inequality: Consider a C^1 function $f\colon \mathbb{R}^n \to \mathbb{R}$, suppose

$$\int_{\mathbb{R}^n} \|\nabla f\|^2 \, d\mu < \infty$$

where μ is the Lebesgue measure. The Poincaré Inequality says

$$\int_{\mathbb{R}^n} f^2(x) d\mu \le C \int_{\mathbb{R}^n} \|\nabla f\|^2 \, d\mu$$

2. Gaussian-Poincaré Inequality: Consider a C^2 function $f\colon \mathbb{R}^n \to \mathbb{R}$ such that

$$\sup_x \left| \frac{\partial^2}{\partial x_i^2} f(x) \right| = K < \infty$$

Let $X \sim \mathcal{N}(0, I_n)$. Then we have

$$\operatorname{Var}(f(X)) \le \mathbb{E}\left[\|\nabla f(X)\|^2\right]$$

Proof. I. Step 1(Tensorization argument): Assume it is true for $n = 1$. For $Z = f(X) = f(X_1, \cdots, X_n)$, which are independent, we have

$$\operatorname{Var}_{(i)}(Z) = \operatorname{Var} f(x^{i-1}, X_i, x_{i+1}^n) \le \mathbb{E}\left[(\frac{\partial f}{\partial x_i})^2 \mid X^{-i} \right]$$

Then by Efron-Stein inequality, we get

$$\begin{aligned} \operatorname{Var}(Z) &\le \sum_{i=1}^n \mathbb{E}\left[\operatorname{Var}_{(i)}(Z)\right] \\ &= \sum_{i=1}^n \mathbb{E}\left[\mathbb{E}\left[(\frac{\partial f}{\partial x_i})^2 \mid X^{-i} \right] \right] = \mathbb{E}\left[\|\nabla f(X)\|^2\right] \end{aligned}$$

II. Step 2($n = 1$): Let $X \sim \mathcal{N}(0, 1)$ and $f \in C^2(\mathbb{R})$ with $\sup_x |f''(x)| = K < \infty$. Let $\varepsilon_1, \cdots, \varepsilon_n$ be *i.i.d.* such that

$$\mathbb{P}(\varepsilon_i = 1) = \mathbb{P}(\varepsilon_i = -1) = \frac{1}{2}$$

Let $S_m = \frac{1}{\sqrt{m}} \sum_{i=1}^m \varepsilon_i$. Then CLT implies that $S_m \xrightarrow{d} X$. Consider the Efron-Stein Inequality for $\operatorname{Var}(f(S_m))$,

$$\operatorname{Var}(f(S_m)) \le \sum_{j=1}^m \mathbb{E}\left[\operatorname{Var}_{(j)}(f(S_m))\right]$$

Because

$$\operatorname{Var}(g(\varepsilon_j)) = \mathbb{E}\left[g(\varepsilon_j)^2\right] - \mathbb{E}\left[g(\varepsilon_j)\right]^2 = \frac{1}{4}(g(1) - g(-1))^2$$

we have

$$\begin{aligned} \operatorname{Var}_{(j)}(f(S_m)) &= \frac{1}{4}\left(f\left(S_m - \frac{\varepsilon_j}{\sqrt{m}} + \frac{1}{\sqrt{m}} \right) - f\left(S_m - \frac{\varepsilon_j}{\sqrt{m}} - \frac{1}{\sqrt{m}} \right) \right)^2 \\ &= \frac{1}{4}\left(f'(S_m - \frac{\varepsilon_j}{\sqrt{m}}) \frac{2}{\sqrt{m}} + O\left(\frac{K}{m}\right) \right)^2 \\ &= \frac{1}{m}\left(f'(S_m - \frac{\varepsilon_j}{\sqrt{m}}) \right)^2 + o\left(\frac{K}{m}\right) \end{aligned}$$

It follows that

$$\sum_{j=1}^{m} \mathrm{Var}_{(j)}(f(S_m)) = \left(f'(S_m - \frac{\varepsilon_j}{\sqrt{m}})\right)^2 + Ko(1)$$

Therefore,

$$\mathrm{Var}(f(S_m)) \leq \left(f'(S_m - \frac{\varepsilon_j}{\sqrt{m}})\right)^2 + Ko(1)$$

Because $S_m \xrightarrow{d} X$, as $m \to \infty$

$$\mathrm{Var}(f(X)) \leq \mathbb{E}\left[f'(X)^2\right]$$

□

1.7 Entropy Method

1. Entropy: Note that

$$\mathrm{Var}(Y) = \mathbb{E}[Y^2] - (\mathbb{E}[Y])^2 = \mathbb{E}[g(Y)] - g(\mathbb{E}[Y])$$

where $g(x) = x^2$. If we let

$$h(x) = x \log x$$

which is convex, then

$$\mathrm{Ent}(Y) := \mathbb{E}[h(Y)] - h(\mathbb{E}[Y]) \geq 0$$

Lemma 1.7.1 (Herbst's Argument)**.** *For $\lambda \geq 0$, consider $Y := e^{\lambda(Z - \mathbb{E}[Z])}$ for a random variable Z. Suppose that*

$$\mathrm{Ent}(Y) \leq \frac{\lambda^2 v}{2}\mathbb{E}[Y], \quad \forall\, \lambda \geq 0$$

Then, for all $\lambda \geq 0$,

$$\psi_{Z-\mathbb{E}[Z]}(\lambda) \leq \frac{\lambda^2}{2} v$$

Proof. WTLG, let $\mathbb{E}[Z] = 0$. Let $g(\lambda) = \psi_Z(\lambda)$.

$$\begin{aligned} g'(\lambda) &= \frac{\mathbb{E}\left[Ze^{\lambda Z}\right]}{\mathbb{E}\left[e^{\lambda Z}\right]} \\ &= \frac{1}{\lambda}\frac{\mathbb{E}\left[\left(\log e^{\lambda Z}\right) e^{\lambda Z}\right]}{\mathbb{E}\left[e^{\lambda Z}\right]} \\ &= \frac{1}{\lambda}\frac{\mathbb{E}[Y \log Y]}{\mathbb{E}[Y]} \\ &= \frac{1}{\lambda}\left(\frac{\mathrm{Ent}(Y)}{\mathbb{E}[Y]} + g(\lambda)\right) \end{aligned}$$

It follows that

$$\lambda g'(\lambda) - g(\lambda) = \frac{\mathrm{Ent}(Y)}{\mathbb{E}[Y]} \leq \frac{\lambda^2}{2} v$$

Therefore,

$$\frac{d}{\lambda}\left(\frac{g(\lambda)}{\lambda}\right) \leq \frac{v}{2} \;\Rightarrow\; \frac{g(\lambda)}{\lambda} - \lim_{x \to 0}\frac{g(x)}{x} \leq \frac{\lambda}{2} v$$

But because $g(0) =$, $\lim_{x\to 0} \frac{g(x)}{x} = g'(0) = 0$,

$$g(\lambda) \leq \frac{\lambda^2}{2} v$$

□

Remark. The condition

$$\operatorname{Ent}(Y) \le \frac{\lambda^2 v}{2}\mathbb{E}[Y], \quad \forall\, \lambda \ge 0$$

is called the Herbst's condition.

2. KL divergence: Consider $\mathbb{P}$ and $\mathbb{Q}$ be two probability measures on a σ-algebra with $\mathbb{Q} \ll \mathbb{P}$. The KL divergence between $\mathbb{Q}$ and $\mathbb{P}$ is

$$D(\mathbb{Q} \parallel \mathbb{P}) := \mathbb{E}_{\mathbb{P}}\left[\frac{d\mathbb{Q}}{d\mathbb{P}} \log \frac{d\mathbb{Q}}{d\mathbb{P}}\right]$$

Otherwise, it is defined as ∞.

Lemma 1.7.2. *Let $\mathbb{P}$ be the product measure for independent random variables $X_1, \cdots, X_n$ and $f\colon \mathbb{R}^n \to \mathbb{R}$. Consider $\mathbb{Q} \ll \mathbb{P}$ with*

$$\frac{d\mathbb{Q}}{d\mathbb{P}}(x) = \frac{e^{\lambda f(x)}}{\mathbb{E}_{\mathbb{P}}\left[e^{\lambda f(X)}\right]}$$

Then

$$D(\mathbb{Q} \parallel \mathbb{P}) = \frac{\operatorname{Ent}\left(e^{\lambda f(X)}\right)}{\mathbb{E}_{\mathbb{P}}\left[e^{\lambda f(X)}\right]}$$

Proof. By definition, let $Z = f(X)$.

$$\begin{aligned} D(\mathbb{Q} \parallel \mathbb{P}) &= \mathbb{E}_{\mathbb{P}}\left[\frac{d\mathbb{Q}}{d\mathbb{P}} \log \frac{d\mathbb{Q}}{d\mathbb{P}}\right] \\ &= \mathbb{E}\left[\frac{e^{\lambda Z}}{\mathbb{E}\left[e^{\lambda Z}\right]} \log \frac{e^{\lambda Z}}{\mathbb{E}\left[e^{\lambda Z}\right]}\right] \\ &= \frac{1}{\mathbb{E}\left[e^{\lambda Z}\right]} \operatorname{Ent}\left(e^{\lambda Z}\right) \qquad \square \end{aligned}$$

Remark. Such $\mathbb{Q}$ is denoted by $\mathbb{P}^{(\lambda f)}$. Therefore, the Herbst's condition is expressed as

$$D(\mathbb{P}^{(\lambda f)} \parallel \mathbb{P}) \le \frac{\lambda^2}{2} v, \quad \forall\, \lambda \ge 0$$

In fact, in the proof of Herbst's argument, we have seen

$$D(\mathbb{P}^{(\lambda f)} \parallel \mathbb{P}) = \lambda \psi'_{Z-\mathbb{E}[Z]}(\lambda) - \psi_{Z-\mathbb{E}[Z]}(\lambda) = \lambda^2 \frac{d}{d\lambda} \frac{\psi_{Z-\mathbb{E}[Z]}(\lambda)}{\lambda}$$

It follows that

$$\psi_{Z-\mathbb{E}[Z]}(\lambda) = \lambda \int_0^\lambda \frac{D(\mathbb{P}^{(tf)} \parallel \mathbb{P})}{t^2} dt$$

Generally, for $\mathbb{Q}^{(t)} \ll \mathbb{P}$ with

$$\frac{d\mathbb{Q}^{(t)}}{d\mathbb{P}}(x) = \frac{e^{tx}}{\mathbb{E}_{\mathbb{P}}\left[e^{tX}\right]}$$

we have

$$\frac{D(\mathbb{P}^{(t)} \parallel \mathbb{P})}{t^2} = \frac{d}{dt} \frac{\psi_X(t)}{t}$$

Proposition 1.7.1. *There are basic properties for D.*

(1) $D(\mathbb{Q} \parallel \mathbb{P}) \geq 0$.

(2) (Chain Rule of KL Divergence) Suppose $\mathbb{P}_{XY}$ *and* $\mathbb{Q}_{XY}$ *are probability measures on* $\mathcal{X} \times \mathcal{Y}$

$$D(\mathbb{Q}_{XY} \parallel \mathbb{P}_{XY}) = D(\mathbb{Q}_X \parallel \mathbb{P}_X) + \mathbb{E}_{\mathbb{Q}_X}\left[D(\mathbb{Q}_{Y|X} \parallel \mathbb{P}_{Y|X})\right]$$

In particular, if $\mathbb{P}_{XY} = \mathbb{P}_X\mathbb{P}_Y$ *and* $\mathbb{Q}_{XY} = \mathbb{Q}_X\mathbb{Q}_Y$*, then*

$$D(\mathbb{Q}_{XY} \parallel \mathbb{P}_{XY}) = D(\mathbb{Q}_X \parallel \mathbb{P}_X) + D(\mathbb{Q}_Y \parallel \mathbb{P}_Y)$$

Remark. We can denote

$$\mathbb{E}_{\mathbb{Q}_X}\left[D(\mathbb{Q}_{Y|X} \parallel \mathbb{P}_{Y|X})\right] = D(\mathbb{Q}_{Y|X} \parallel \mathbb{P}_{Y|X} \mid \mathbb{Q}_X)$$

Proof. (1) Let $X = \frac{d\mathbb{Q}}{d\mathbb{P}}$ and $f(t) = t\log t$ that is convex.

$$D(\mathbb{Q} \parallel \mathbb{P}) = \mathbb{E}[f(X)] \geq f(\mathbb{E}[X]) = f(1) = 0$$

(2) WTLG, assume $\mathbb{Q}_{XY} \ll \mathbb{P}_{XY}$. By the results of Disintegration Theorem,

$$\begin{aligned} D(\mathbb{Q}_{XY} \parallel \mathbb{P}_{XY}) &= \mathbb{E}_{\mathbb{Q}_{XY}}\left[\log\frac{d\mathbb{Q}_{XY}}{d\mathbb{P}_{XY}}\right] \\ &= \mathbb{E}_{\mathbb{Q}_{XY}}\left[\log\frac{d\mathbb{Q}_X}{d\mathbb{P}_X} + \log\frac{d\mathbb{Q}_{Y|X}}{d\mathbb{P}_{Y|X}}\right] \\ &= \mathbb{E}_{\mathbb{Q}_X}\left[\log\frac{d\mathbb{Q}_X}{d\mathbb{P}_X}\right] + \mathbb{E}_{\mathbb{Q}_X}\left[\mathbb{E}_{\mathbb{Q}_{Y|X}}\left[\log\frac{d\mathbb{Q}_{Y|X}}{d\mathbb{P}_{Y|X}}\right]\right] \end{aligned}$$ □

3. Tensorization: One dimensional case implies high-dimensional case.

Lemma 1.7.3 (Entropy Tensorization Lemma)**.** *Consider independent random variables* $X_1, \cdots, X_n$ *(under* $\mathbb{P} := \mathbb{P}_{X_1\cdots X_n}$*). Let* $\mathbb{Q} := \mathbb{Q}_{X_1\cdots X_n}$ *be any other distribution. Then*

$$D(\mathbb{Q} \parallel \mathbb{P}) \leq \mathbb{E}_{\mathbb{Q}_{X^{(i)}}}\left[D(\mathbb{Q}_{X_i|X^{(i)}} \parallel \mathbb{P}_{X_i|X^{(i)}})\right]$$

Note that by independence $\mathbb{P}_{X_i|X^{(i)}} = \mathbb{P}_{X_i}$.

Proof. By the chain rule, we have

$$D(\mathbb{Q} \parallel \mathbb{P}) = \sum_{i=1}^n \mathbb{E}_{\mathbb{Q}_{X^{i-1}}}\left[D\left(\mathbb{Q}_{X_i|X^{i-1}} \parallel \mathbb{P}_{X_i|X^{i-1}}\right)\right]$$

where

$$\mathbb{E}_{\mathbb{Q}_{X^0}}\left[D\left(\mathbb{Q}_{X_1|X^0} \parallel \mathbb{P}_{X_1|X^0}\right)\right] := D\left(\mathbb{Q}_{X_1} \parallel \mathbb{P}_{X_1}\right)$$

And by independence

$$D(\mathbb{Q} \parallel \mathbb{P}) = \sum_{i=1}^n \mathbb{E}_{\mathbb{Q}_{X^{i-1}}}\left[D\left(\mathbb{Q}_{X_i|X^{i-1}} \parallel \mathbb{P}_{X_i}\right)\right]$$

Consider

$$\begin{aligned} &\mathbb{E}_{\mathbb{Q}_{X^{(i)}}}\left[D(\mathbb{Q}_{X_i|X^{(i)}} \parallel \mathbb{P}_{X_i})\right] - \mathbb{E}_{\mathbb{Q}_{X^{i-1}}}\left[D\left(\mathbb{Q}_{X_i|X^{i-1}} \parallel \mathbb{P}_{X_i|X^{i-1}}\right)\right] \\ =&\mathbb{E}_{\mathbb{Q}_{X^n}}\left[\log\frac{d\mathbb{Q}_{X_i|X^{(i)}}}{d\mathbb{P}_{X_i}}\right] - \mathbb{E}_{\mathbb{Q}_{X^{i-1}}}\left[\log\frac{d\mathbb{Q}_{X_i|X^{i-1}}}{d\mathbb{P}_{X_i}}\right]\left(= \mathbb{E}_{\mathbb{Q}_{X^n}}\left[\log\frac{d\mathbb{Q}_{X_i|X^{i-1}}}{d\mathbb{P}_{X_i}}\right]\right) \\ =&\mathbb{E}_{\mathbb{Q}_{X^n}}\left[\log\frac{d\mathbb{Q}_{X_i|X^{(i)}}}{d\mathbb{Q}_{X_i|X^{i-1}}}\right] = \mathbb{E}_{\mathbb{Q}_{X^{(i)}}}\left[D(\mathbb{Q}_{X_i|X^{(i)}} \parallel \mathbb{Q}_{X_i|X^{i-1}})\right] \geq 0 \end{aligned}$$ □

Remark. Let $f(X_1, \ldots, X_n) = \frac{d\mathbb{Q}}{d\mathbb{P}}$. Then $\mathbb{E}_{\mathbb{P}}(f) = 1$. Because because

$$\frac{d\mathbb{Q}_{X_i|X^{(i)}}}{d\mathbb{P}_{X_i}}(x_i \mid x^{(i)}) = \frac{f(x_i \mid x^{(i)})}{\mathbb{E}_{\mathbb{P}_{X_i}}\left[f(x_i \mid x^{(i)})\right]}$$

we have

$$\begin{aligned}
D(\mathbb{Q} \parallel \mathbb{P}) = \operatorname{Ent}(f) &\leq \sum_{i=1}^{n} \mathbb{E}_{\mathbb{Q}_{X^{(i)}}}\left[D(\mathbb{Q}_{X_i|X^{(i)}} \parallel \mathbb{P}_{X_i|X^{(i)}})\right] \\
&= \sum_{i=1}^{n} \mathbb{E}_{\mathbb{Q}}\left[\log \frac{d\mathbb{Q}_{X_i|X^{(i)}}}{\mathbb{P}_{X_i}}\right] \\
&= \sum_{i=1}^{n} \mathbb{E}_{\mathbb{Q}}\left[\log \frac{f(x_i \mid x^{(i)})}{\mathbb{E}_{\mathbb{P}_{X_i}}\left[f(x_i \mid x^{(i)})\right]}\right] \\
&= \sum_{i=1}^{n} \mathbb{E}_{\mathbb{P}}\left[f \log \frac{f(x_i \mid x^{(i)})}{\mathbb{E}_{\mathbb{P}_{X_i}}\left[f(x_i \mid x^{(i)})\right]}\right] \\
&= \sum_{i=1}^{n} \mathbb{E}_{\mathbb{P}_{X^{(i)}}}\left[\operatorname{Ent}_{(i)}\left(f(x_i \mid x^{(i)})\right)\right]
\end{aligned}$$

where

$$\operatorname{Ent}_{(i)}\left(f(x_i \mid x^{(i)})\right) = \mathbb{E}_{\mathbb{P}_{X_i}}\left[h(f(x_i \mid x^{(i)}))\right] - h\left(\mathbb{E}_{\mathbb{P}_{X_i}}\left[f(x_i \mid x^{(i)})\right]\right)$$

Therefore, we get

$$\operatorname{Ent}(f) \leq \sum_{i=1}^{n} \mathbb{E}_{\mathbb{P}_{X^{(i)}}}\left[\operatorname{Ent}_{(i)}\left(f(x_i \mid x^{(i)})\right)\right]$$

More generally, for any $f \geq 0$, we can construct $\mathbb{Q}$ with

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = \frac{f}{\mathbb{E}_{\mathbb{P}}[f]}$$

Because $\operatorname{Ent}(cf) = c \operatorname{Ent}(f)$, we still have

$$\operatorname{Ent}(f) \leq \sum_{i=1}^{n} \mathbb{E}_{\mathbb{P}_{X^{(i)}}}\left[\operatorname{Ent}_{(i)}\left(f(x_i \mid x^{(i)})\right)\right]$$

In particular, let $\mathbb{P}$ be the product measure for independent random variables $X_1, \cdots, X_n$, and $f\colon \mathbb{R}^n \to \mathbb{R}$, and $\mathbb{Q} \ll \mathbb{P}$ with

$$\frac{d\mathbb{Q}}{d\mathbb{P}}(x) = \frac{e^{\lambda f(x)}}{\mathbb{E}_{\mathbb{P}}\left[e^{\lambda f(X)}\right]}$$

Then

$$D(\mathbb{Q} \parallel \mathbb{P}) = \frac{\operatorname{Ent}\left(e^{\lambda f(X)}\right)}{\mathbb{E}_{\mathbb{P}}\left[e^{\lambda f(X)}\right]} \leq \sum_{i=1}^{n} \mathbb{E}\left[\frac{\operatorname{Ent}_{(i)}\left(e^{\lambda f(X)}\right)}{\mathbb{E}_{(i)}\left[e^{\lambda f(X)}\right]}\right]$$

1.8 Log-Sobolev Inequality

1. Binary case: Let

$$X = (X_1, \cdots, X_n) \sim \operatorname{Unim}\left(\{-1, 1\}^n\right)$$

Let $f\colon \{-1,1\}^n \to \mathbb{R}$ with

$$\operatorname{Ent}(f) := \mathbb{E}\left[f \log f\right] - \mathbb{E}[f] \log \mathbb{E}[f]$$

The Binary Log-Sobolev Inequality says

$$\operatorname{Ent}(f^2) \leq 2\sum_{i=1}^{n} \mathbb{E}\left[\operatorname{Var}_{(i)}(f)\right] = 2\mathcal{E}(f)$$

First, let's see how to get the concentration inequality from this. Define

$$g(x) = e^{\frac{\lambda f(x)}{2}}$$

Then by above inequality, suppose $\sum_{i=1}^{n}(Z - Z_i')_+^2 \leq v$, where $Z = f(X)$ and $Z_i' = f(X^{i-1}, X_i', X_{i+1}^n)$,

$$\begin{aligned}
\operatorname{Ent}(g^2(x)) = \operatorname{Ent}(e^{\lambda f}) &\leq 2\sum_{i=1}^{n} \mathbb{E}\left[\operatorname{Var}_{(i)}\left(e^{\frac{\lambda f}{2}}\right)\right] \\
&\leq \sum_{i=1}^{n} \frac{\lambda^2}{2}\mathbb{E}\left[e^{\lambda Z}(Z - Z_i')_+^2\right] = \frac{\lambda^2}{2}\mathbb{E}\left[e^{\lambda Z}\sum_{i=1}^{n}(Z - Z_i')_+^2\right] \\
&\leq \frac{\lambda^2}{2} v \mathbb{E}\left[e^{\lambda f}\right]
\end{aligned}$$

Therefore, we have

$$D(\mathbb{Q}^{(\lambda f)} \parallel \mathbb{P}) = \frac{\operatorname{Ent}(e^{\lambda f})}{\mathbb{E}\left[e^{\lambda f}\right]} \leq \frac{\lambda^2}{2}v$$

Then by Herbst's Argument,

$$\psi_{f(X)-\mathbb{E}[f(X)]} \leq \frac{\lambda^2}{2}v$$

It follows that

$$\mathbb{P}(f(X) - \mathbb{E}[f(X)] \geq t) \leq e^{-\frac{t^2}{2v}}$$

Proof of Binary Log-Sobolev Inequality. (I) Assume $n = 1$: For $f\colon \{-1, 1\} \to \mathbb{R}$ with $f(-1) = a$ and $f(1) = b$, we have

$$\operatorname{Ent}(f^2) = \frac{1}{2}a^2 \log a^2 + \frac{1}{2}b^2 \log b^2 - \frac{a^2 + b^2}{2} \log \frac{a^2 + b^2}{2}$$

and

$$\mathbb{E}[\operatorname{Var}_{(1)}(f)] = \operatorname{Var}(f) = \frac{1}{2}\left(a - \frac{a+b}{2}\right)^2 + \frac{1}{2}\left(b - \frac{a+b}{2}\right)^2 = \frac{(b-a)^2}{4}$$

Therefore, it is sufficient to show

$$h_b(a) = \frac{1}{2}a^2 \log a^2 + \frac{1}{2}b^2 \log b^2 - \frac{a^2 + b^2}{2} \log \frac{a^2 + b^2}{2} - \frac{(b-a)^2}{2}, \quad \forall\, a \geq b \leq 0$$

which is not difficult.

(II) By the Entropy Tensorization Lemma,

$$\operatorname{Ent}(f^2) \leq \sum_{i=1}^{n} \mathbb{E}\left[\operatorname{Ent}_{(i)}(f^2)\right]$$

Then since for $n = 1$, we already get

$$\operatorname{Ent}_{(i)}(f^2) \leq \operatorname{Var}_{(i)}(f)$$

Therefore,

$$\operatorname{Ent}(f^2) \leq \sum_{i=1}^{n} \mathbb{E}\left[\operatorname{Var}_{(i)}(f)\right]$$

□

2. Gaussian Log-Sobolev Inequality:

Theorem 1.8.1. *Let $X = (X_1, \cdots, X_n)$ be a standard normal random variable and $f\colon \mathbb{R}^n \to \mathbb{R}$ be C^1 and with bounded second derivative. Then we have*

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

Proof. I. First, by Entropy Tensorization Lemma, it is sufficient to prove the inequality for $n = 1$.

II. For $n = 1$, like the proof in the Gaussian-Poincaré Inequality, we approximate the standard Gaussian by binary distribution. Consider $\varepsilon_1, \cdots, \varepsilon_n \overset{i.i.d.}{\sim} \mathrm{Unif}(\{-1, 1\})$ and let

$$g(\varepsilon_1, \cdots, \varepsilon_n) = f\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \varepsilon_i\right)$$

For such g, consider the Binary Log-Sobolev Inequality

$$\mathrm{Ent}(g^2) \le 2 \sum_{i=1}^n \mathbb{E}\left[\mathrm{Var}_{(i)}(g)\right]$$

Similarly as the proof in the Gaussian-Poincaré Inequality,

$$\mathbb{E}\left[\mathrm{Var}_{(i)}(g)\right] \le \left(f'\left(S_n - \frac{\varepsilon_j}{\sqrt{n}}\right)\right)^2 + Ko(1)$$

Then by taking limit we have

$$\mathrm{Ent}(g^2) \le 2\mathbb{E}\left[(g')^2\right]$$

□

Corollary 1.8.1. *Let $X = (X_1, \cdots, X_n)$ be a standard normal random variable and $f\colon \mathbb{R}^n \to \mathbb{R}$ be C^1 and with bounded second derivative and $\|\nabla f\| \le 1$. Then*

$$\mathbb{P}(f(X) - \mathbb{E}[f(X)] \ge t) \le e^{-t^2/2}$$

Proof. First, by above inequality we have

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

which implies

$$\psi_{f(X)-\mathbb{E}[f(X)]} \le \frac{\lambda^2}{2}$$

by Herbst's Argument. □

Remark. $\|\nabla f\| \le 1$ can be replaced by 1-Lipschitz continuous.

Chapter 2

Isoperimetric Inequalites

Chapter 3

tmp

3.1 Log-Sobolev Inequality

1.