Concentration of Measure

Assaf Naor
Scribe: Lingjiong Zhu *

Fall 2008

In mathematics, concentration of measure (e.g. about a median) is a principle that is applied in measure theory, probability and combinatorics, and has consequences for other fields such as Banach space theory. Informally, it states that Lipschitz functions that depend on many parameters are almost constant.

The concentration of measure phenomenon was put forth in the early 1970s by Vitali Milman in his works on the local theory of Banach space, extending an idea going back to the work of Paul Lévy. The idea of concentration of measure (which was discovered by V.Milman) is arguably one of the great ideas of analysis in our times. While its impact on Probability is only a small part of the whole picture, this impact should not be ignored. It was further developed in the works of Milman and Mikhail Gromov, Maurey, Pisier, Schechtman, Michel Talagrand, Ledoux, and others.

In our notes, we would start with isoperimetric problems, introducing Bobkov's Inequality, Maurey-Pisier Theorem etc. Then, we would state and prove Brunn-Minkowski Inequality, Borell's Inequality, Prékopa-Leindler Inequality, and Gromov-Milman Theorem. Then, we would discuss Martingale method, Talagrand's Induction method. We will also mention Khintchine's Inequality and Kahane's Inequality. We will then move onto Spectral methods, introducing Poincaré's Inequality, log-Sobolev Inequality, Herbst's The-

*These notes are based on the same-name course given by Prof Assaf Naor at Courant Institute in the fall semester in 2008. I attended all the lectures but didn't spend enough time studying the materials at that time. However, after learning some spin glasses and other stuff in probability theory, I started to appreciate the power and beauty of concentration of measure. All the credits are due to Prof Assaf Naor whilst I'm responsible for the typos and mistakes in these notes.

orem and tensorization. Results by Gross, Schechtman-Zinn, and Bobkov-Ledoux-Maurey-Talagrand would be mentioned. Finally, we will discuss briefly the Stein's method.

Basic setting: (X, d) is a metric space, $\mathcal{F}$ the σ-algebra of Borel sets on X and μ a Borel (probability) measure on X.

Theorem 1. *(Lévy's Inequality) Consider $(S^{n-1}, \|\cdot\|_2)$, where S^{n-1} is the Euclidean sphere in $\mathbb{R}^n$ and $\|\cdot\|_2$ is the Euclidean metric. Let μ be the normalized surface area measure. If $A \subseteq S^{n-1}$ is Borel measurable and $\mu(A) \geq 1/2$, then, for any $\varepsilon > 0$, we have*

$$\mu(x \in S^{n-1} : d(x, A) \geq \varepsilon) \leq 2e^{-n\varepsilon^2/64}.$$

We have an equivalent form of the Lévy's Inequality above:

Theorem 2. *If $f : S^{n-1} \to \mathbb{R}$ is Lipschitz with constant L, i.e.*

$$|f(x) - f(y)| \leq L\|x - y\|_2.$$

Then, there exists $c \in \mathbb{R}$, such that, for any $\varepsilon > 0$,

$$\mu(x \in S^{n-1} : |f(x) - c| \geq \varepsilon) \leq 4e^{-\frac{n\varepsilon^2}{64L^2}}.$$

The Isoperimetric Problem: Given $a \in (0, 1)$, and $\varepsilon > 0$, we aim to find the Borel subsets $A \subseteq X$ with $\mu(A) = a$ such that $\mu(A_\varepsilon)$ is as small as possible, where

$$A_\varepsilon = \{x \in X : d(x, A) := \inf\{y \in A, d(x, y)\} < \varepsilon\}.$$

Definition 1. For $A \subset X$ Borel, the *boundary measure* is defined as

$$\mu^+(A) = \liminf_{\varepsilon \to 0} \frac{\mu(A_\varepsilon) - \mu(A)}{\varepsilon} = \liminf_{\varepsilon \to 0} \frac{1}{\varepsilon}\mu(A_\varepsilon \backslash A).$$

Definition 2. I_μ is called the *isoperimetric function* if it is the largest function such that

$$\forall A \subseteq X, \quad \mu^+(A) \geq I_\mu(\mu(A)).$$

Definition 3. Given $a \in [0, 1]$, $\varepsilon > 0$, define

$$\phi_a(\varepsilon) = \inf\{\mu(A_\varepsilon) : A \subseteq X, \mu(A) \geq a\}.$$

The concentration of measure problem is to find good lower bounds on $\phi_a(\varepsilon)$.

Lemma 1. *Let (X, d) be a metric space and μ is the Borel measure. If $f : X \to \mathbb{R}$ is Lipschitz with constant L, then, there exists some $M \in \mathbb{R}$, such that, for any $\varepsilon > 0$, we have*

$$\mu(x \in X : |f(x) - M| \geq \varepsilon) \leq 2\left(1 - \phi_{1/2}\left(\frac{\varepsilon}{L}\right)\right).$$

Proof. Let M be a median of f, i.e.

$$M = \inf\{t \in \mathbb{R} : \mu(x \in X : f(x) > t) \geq 1/2\}.$$

And let $A = \{x \in X : f(x) > M\}$. Then

$$x \in A_{\varepsilon/L} \Rightarrow \exists y, d(x, y) < \frac{\varepsilon}{L} \Rightarrow |f(x) - f(y)| \leq Ld(x, y) < \varepsilon,$$

and

$$y \in A \Rightarrow f(y) \geq M \Rightarrow f(x) > M - \varepsilon.$$

Hence $\{x \in X : f(x) \leq M - \varepsilon\} \subseteq X \backslash A_{\varepsilon/L}$ and $\mu(A) \geq 1/2$. Therefore, by the definition of $\phi_{1/2}$, we have

$$\mu(x \in X : f(x) \leq M - \varepsilon) \leq 1 - \phi_{1/2}\left(\frac{\varepsilon}{L}\right).$$

Now, let $B = \{x \in X : f(x) \leq M\}$. Then

$$\mu(x \in X : f(x) \geq M + \varepsilon) \leq 1 - \phi_{1/2}\left(\frac{\varepsilon}{L}\right).$$

Hence, we have

$$\mu(x \in X : |f(x) - M| \geq \varepsilon) \leq 2\left(1 - \phi_{1/2}\left(\frac{\varepsilon}{L}\right)\right).$$

□

Now, we will discuss isoperimetric problem for Gaussian measure. Here are just some notations we would use:

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad \Phi(x) = \int_{-\infty}^{x} \phi(s) ds.$$

$$\psi : [0, 1] \to [0, 1/\sqrt{2\pi}], \quad \psi(t) = \phi(\Phi^{-1}(t)).$$

Proposition 1. *(i)* $\psi(0) = \psi(1) = 0$.
(ii) $\psi(1-x) = \psi(x)$ *for all* $x \in [0,1]$.
(iii) $\psi''\psi = -1$.
(iv) $(\psi')^2$ *is a convex function on* $[0,1]$.

Proof. (i) $\psi(0) = \phi(\Phi^{-1}(0)) = \phi(-\infty) = 0$. Similarly, $\psi(1) = \phi(\infty) = 0$.
(ii) $\psi(1-x) = \phi(\Phi^{-1}(1-x)) = \phi(-\Phi^{-1}(x)) = \phi(\Phi^{-1}(x)) = \psi(x)$.
(iii) Direct computations.
(iv)

$$\left((\psi')^2\right)'' = 2\psi'\psi'' = -2\frac{\psi'}{\psi} = (-2)\frac{\psi''\psi - (\psi')^2}{\psi^2} = 2\frac{1+(\psi')^2}{\psi^2} > 0.$$

$\square$

Let $\phi_n : \mathbb{R}^n \to \mathbb{R}$,

$$\phi_n(x) = \frac{1}{(2\pi)^{n/2}} e^{-\frac{\|x\|_2^2}{2}},$$

and define the n-dimensional Gaussian measure as

$$\gamma_n(A) = \int_A \phi_n(x) dx, \quad A \subseteq \mathbb{R}^n.$$

Note that γ_n is invariant under rotations, i.e., for any orthogonal $n \times n$ matrix U, and any $A \subseteq \mathbb{R}^n$, we have $\gamma_n(A) = \gamma_n(UA)$.

Theorem 3. *(Bobkov's Inequality)(B_n)*
If $f : \mathbb{R}^n \to [0,1]$ *is smooth (locally Lipschitz), then,*

$$\psi\left(\int_{\mathbb{R}^n} f d\gamma_n\right) \leq \int_{\mathbb{R}^n} \sqrt{\|\nabla f\|_2^2 + \psi(f)^2} d\gamma_n.$$

Before we start to prove Bobkov's Inequality, let us state another result and use Bobkov's Inequality to prove it.

Theorem 4. *(Borell-Sudakov-Tsirelson) For any Borel* $A \subseteq \mathbb{R}^n$, *we have*

$$\gamma_n(A_\varepsilon) \geq \Phi(\Phi^{-1}(\gamma_n(A)) + \varepsilon).$$

Proof. Now, for $A \subseteq \mathbb{R}^n$, define

$$f(x) \begin{cases} 1 & x \in A \\ 1 - \frac{d(x,A)}{\varepsilon} & x \in A_\varepsilon \backslash A \\ 0 & x \in \mathbb{R}^n \backslash A_\varepsilon \end{cases}$$

Then, f is Lipschitz with constant $1/\varepsilon$ and $\|\nabla f\|_2 \leq 1/\varepsilon$ a.e. Now, also use the Bobkov's Inequality, we get

$$\begin{aligned} \psi(\gamma_n(A)) &\leq \psi\left(\int_{\mathbb{R}^n} f d\gamma_n\right) \\ &\leq \int_{\mathbb{R}^n} \sqrt{\|\nabla f\|_2^2 + \psi(f)^2} d\gamma_n \\ &\leq \int_{\mathbb{R}^n} \|\nabla f\|_2 d\gamma_n + \int_{\mathbb{R}^n} \psi(f) d\gamma_n \\ &\leq \frac{1}{\varepsilon}\gamma_n(A_\varepsilon \backslash A) + \gamma_n(A_\varepsilon \backslash A). \end{aligned}$$

Hence, we have

$$\gamma_n^+(A) = \liminf_{\varepsilon \to 0} \frac{1}{\varepsilon}\gamma_n(A_\varepsilon \backslash A) \geq \psi(\gamma_n(A)),$$

and thus $I_{\gamma_n} \geq \psi$. Define $h(\varepsilon) = \Phi^{-1}(\gamma_n(A_\varepsilon))$ and w.l.o.g. A is a finite union of balls. Then, we get

$$h'(\varepsilon) = \frac{\gamma_n^+(A_\varepsilon)}{\phi(\Phi^{-1}(\gamma_n(A_\varepsilon)))} = \frac{\gamma_n^+(A_\varepsilon)}{\psi(\gamma_n(A_\varepsilon))} \geq 1,$$

where I used the fact that

$$\frac{d}{d\varepsilon}\gamma_n(A_\varepsilon) = \lim_{h \to 0} \frac{\gamma_n(A_\varepsilon + h) - \gamma_n(A_\varepsilon)}{h} = \gamma_n^+(A_\varepsilon).$$

Now, since $h'(\varepsilon) \geq 1$ for all ϵ, we have $h(\varepsilon) \geq h(0) + \varepsilon$. Thus

$$\Phi^{-1}(\gamma_n(A_\varepsilon)) \geq \Phi^{-1}(\gamma_n(A)) + \varepsilon,$$

which implies that

$$\gamma_n(A_\varepsilon) \geq \Phi(\Phi^{-1}(\gamma_n(A)) + \varepsilon).$$

□

Remark 1. In general, if $I_\mu \geq g' \circ g^{-1}$, then

$$\mu(A_\varepsilon) \geq g(g^{-1}(\mu(A)) + \varepsilon).$$

Remark 2. The solution of the isoperimetric problem for γ_n will be half space. Let $a = \gamma_n(A)$ and there exists s such that $\Phi(s) = a$, i.e. $s = \Phi^{-1}(a)$. Now define

$$H = \{x \in \mathbb{R}^n : x_1 \leq s\} \quad \text{and} \quad H_\varepsilon = \{x \in \mathbb{R}^n : x_1 < s + \varepsilon\}.$$

Then, we have

$$\gamma_n(H) = \Phi(s) = a \quad \text{and} \quad \gamma_n(H_\varepsilon) = \Phi(s+\varepsilon) = \Phi(\Phi^{-1}(a) + \varepsilon).$$

Now, we will try to prove the Bobkov's Inequality. Let B_n denote the Bobkov's Inequality for $\mathbb{R}^n$, then, we have the following lemma.

Lemma 2. *$B_n \wedge B_m \Rightarrow B_{n+m}$.*

Proof. $f : \mathbb{R}^{n+m} \to [0,1]$ smooth. $\mathbb{R}^{n+m} = \mathbb{R}^n \times \mathbb{R}^m$. Write $f = f(x,y)$, $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$. For a fixed $x \in \mathbb{R}^n$, define

$$g(x) = \int_{\mathbb{R}^m} f(x,y)d\gamma_m(y).$$

Then, it is clear that

$$\nabla g(x) = \int_{\mathbb{R}^m} \nabla_x f d\gamma_m(y).$$

Also, notice that for any $u, v \geq 0$, we have

$$\left(\int u d\mu\right)^2 + \left(\int v d\mu\right)^2 \leq \left(\int \sqrt{u^2+v^2} d\mu\right)^2.$$

Therefore, we get

$$
\begin{aligned}
&\psi\left(\int_{\mathbb{R}^{n+m}} f(x,y)d\gamma_n(x)d\gamma_m(y)\right) = \psi\left(\int_{\mathbb{R}^n} g(x)d\gamma_n(x)\right)\\
&\leq \int_{\mathbb{R}^n}\left(\left\|\int_{\mathbb{R}^m}\nabla_x f(x,y)d\gamma_m(y)\right\|_2^2 + \psi\left(\int_{\mathbb{R}^m} f(x,y)d\gamma_m(y)\right)^2\right)^{1/2} d\gamma_n(x)\\
&\leq \int_{\mathbb{R}^n}\Bigg[\left(\int_{\mathbb{R}^m}\|\nabla_x f(x,y)\|_2 d\gamma_m(y)\right)^2\\
&+\left(\int_{\mathbb{R}^m}\left[\|\nabla_y f(x,y)\|_2^2 + \psi(f(x,y))^2\right]^{1/2} d\gamma_m(y)\right)^2\Bigg]^{1/2} d\gamma_n(x)\\
&\leq \int_{\mathbb{R}^m}\int_{\mathbb{R}^n}\left[\|\nabla_x f(x,y)\|_2^2 + \|\nabla_y f(x,y)\|_2^2 + \psi(f(x,y))^2\right]^{1/2} d\gamma_n(x)d\gamma_m(y)\\
&= \int_{\mathbb{R}^{n+m}}\sqrt{\|\nabla f\|_2^2 + \psi(f)^2 d\gamma_{n+m}}.
\end{aligned}
$$

□

Corollary 1. *To prove Bobkov's Inequality, it is enough to prove B_1.*

In order to prove B_1, we will introduce a discrete version of Bobkov's Inequality first, which will be useful later.

Let $\Omega_n = \{\pm 1\}^n$ and μ be the uniform probability measure on Ω_n. For $f : \Omega_n \to \mathbb{R}$, define

$$\partial_i f(x_1, x_2, \cdots, x_n) = \frac{1}{2}[f(x_1, \cdots, x_n) - f(x_1, \cdots, x_{i-1}, -x_i, x_{i+1}, \cdots, x_n)],$$

and $\nabla f = (\partial_1 f, \partial_2 f, \cdots, \partial_n f) \in \mathbb{R}^n$.

Theorem 5. *(Discrete Bobkov's Inequality)(DB_n)*

For any $f : \{\pm 1\}^n \to [0,1]$, we have

$$\psi\left(\int_{\{\pm 1\}^n} f d\mu\right) \leq \int_{\{\pm 1\}^n}\sqrt{\|\nabla f\|_2^2 + \psi(f)^2 d\mu}.$$

Lemma 3. *$DB_n \wedge DB_m \Rightarrow DB_{n+m}$.*

Proof. The proof is similar to the proof of $B_n \wedge B_m \Rightarrow B_{n+m}$. □

Lemma 4. *$DB_n, \forall n \Rightarrow B_1$.*

Proof. $f : \mathbb{R} \to [0,1]$ smooth and $\sup|f'|, \sup|f''| < \infty$. Define $F_n : \{\pm 1\}^n \to [0,1]$ as

$$F_n(x_1, \cdots, x_n) = f\left(\frac{x_1 + \cdots + x_n}{\sqrt{n}}\right).$$

Applying Mean Value Theorem, we get

$$\begin{aligned}
&\left| \|\nabla F_n(x_1, \cdots, x_n)\|_2^2 - f'\left(\frac{x_1 + \cdots + x_n}{\sqrt{n}}\right)^2 \right| \\
&= \left| \sum_{i=1}^n \left(\frac{f\left(\frac{x_1+\cdots+x_n}{\sqrt{n}}\right) - f\left(\frac{x_1+\cdots+x_{i-1}-x_i+\cdots+x_n}{\sqrt{n}}\right)}{2} \right)^2 - \frac{1}{n} f'\left(\frac{x_1 + \cdots + x_n}{\sqrt{n}}\right)^2 \right| \\
&\leq \sum_{i=1}^n \left| \frac{1}{n} f'\left(\frac{x_1 + \cdots + x_{i-1} + \xi_i + x_{i+1} + \cdots + x_n}{\sqrt{n}}\right)^2 - \frac{1}{n} f'\left(\frac{x_1 + \cdots + x_n}{\sqrt{n}}\right)^2 \right| \\
&\leq \frac{2}{n}\|f'\|_\infty \sum_{i=1}^n \left| f'\left(\frac{x_1 + \cdots + x_{i-1} + \xi_i + x_{i+1} + \cdots + x_n}{\sqrt{n}}\right) - f'\left(\frac{x_1 + \cdots + x_n}{\sqrt{n}}\right) \right| \\
&\leq \frac{2}{n}\|f'\|_\infty \sum_{i=1}^n \frac{2}{\sqrt{n}} |f''(\eta_i)| \leq \frac{4\|f'\|_\infty \|f''\|_\infty}{\sqrt{n}}.
\end{aligned}$$

Therefore, we have

$$\begin{aligned}
&\int_{\{\pm 1\}^n} \sqrt{\|\nabla F_n\|_2^2 + \psi(F_n)^2} d\mu \\
&\leq \int_{\{\pm 1\}^n} \sqrt{f'\left(\frac{x_1 + \cdots + x_n}{\sqrt{n}}\right)^2 + f\left(\frac{x_1 + \cdots + x_n}{\sqrt{n}}\right)^2} + O(\frac{1}{\sqrt{n}}) \\
&\longrightarrow \int_{\mathbb{R}} \sqrt{(f')^2 + \psi(f)^2} d\gamma_1 \quad \text{by CLT.}
\end{aligned}$$

Now, applying DB_n to F_n, we get

$$\begin{aligned}
&\int_{\{\pm 1\}^n} \sqrt{\|\nabla F_n\|_2^2 + \psi(F_n)^2} d\mu \geq \psi\left(\int_{\{\pm 1\}^n} F_n d\mu\right) \\
&= \psi\left(\int_{\{\pm 1\}^n} f\left(\frac{x_1 + \cdots + x_n}{\sqrt{n}}\right) d\mu\right) \longrightarrow \psi\left(\int_{\mathbb{R}} f d\gamma_1\right) \quad \text{by CLT.}
\end{aligned}$$

$\square$

Now, in order to prove the Bobkov's Inequality, it suffices to prove DB_1. For $f : \{\pm 1\} \to [0,1]$, $f(1) = a$, $f(-1) = b$, where $a, b \in [0,1]$, then it suffices to prove the following result.

Lemma 5. *(DB_1) For any $a, b \in [0,1]$, we have*

$$\psi\left(\frac{a+b}{2}\right) \leq \frac{1}{2}\left(\sqrt{\left(\frac{a-b}{2}\right)^2 + \psi(a)^2} + \sqrt{\left(\frac{a-b}{2}\right)^2 + \psi(b)^2}\right).$$

Proof. Fix $c \in [0,1]$, $g(x) = \psi(c+x)^2 + x^2$, where

$$x \in \Delta(c) := [-\min\{c, 1-c\}, \min\{c, 1-c\}].$$

Choosing $c = \frac{a+b}{2}$ and $x = \frac{a-b}{2}$, we need to prove that

$$\sqrt{g(0)} \leq \frac{\sqrt{g(x)} + \sqrt{g(-x)}}{2}.$$

Now, after squaring twice, it is equivalent to prove that

$$16g(0)^2 + (g(x) - g(-x))^2 \leq 8g(0)(g(x) + g(-x)).$$

Set $h(x) := g(x) - g(0) = \psi(c+x)^2 + x^2 - \psi(c)^2$. Our goal is to prove that for fixed $c \in [0,1]$, then,

$$(h(x) - h(-x))^2 \leq 8\psi(c)^2(h(x) + h(-x)) \quad \forall x \in \Delta(c).$$

Next, notice that the function $R(x) := h(x) + h(-x) - 2\psi'(c)x^2$ is convex on $\Delta(c)$, because

$$R''(x) = \psi\left[\frac{\psi'(c+x)^2 + \psi'(c-x)^2}{2} - \psi'(c)^2\right] \geq 0,$$

by Jensen's Inequality. Now, R is even, convex and $R(0) = 0$, which implies that $R(x) \geq R(0) = 0$, i.e.

$$h(x) + h(-x) \geq 2\psi'(c)^2x^2.$$

Hence, it suffices to prove that

$$16\psi'(c)^2\psi(c)^2x^2 \geq (h(x) - h(-x))^2,$$

or, equivalently,

$$4\psi(c)|\psi'(c)| \geq \left|\frac{h(x)-h(-x)}{x}\right| = \left|\frac{\psi(c+x)^2-\psi(c-x)^2}{c}\right|.$$

Since by Proposition 1, ψ is symmetric around $1/2$, w.l.o.g., $c \in [0,1/2], x \geq 0, 0 \leq x < c \leq 1/2$ and $\psi(c+x) \geq \psi(c-x)$. So we need to prove

$$4\psi(c)\psi'(c) \geq \frac{\psi(c+x)^2-\psi(c-x)^2}{x}.$$

Set $u(x) := \psi(c+x)^2 - \psi(c-x)^2$. Then, using $\psi''\psi = -1$ from Proposition 1, we get

$$u''(x) = 2(\psi'(c+x)^2 - \psi'(c-x)^2) \leq 0,$$

since by Proposition 1, $(\psi')^2$ is convex and symmetric about $1/2$. Now u is concave and nonnegative, thus $u(x) \leq u(0)$. □

We now completed the proof of DB_1, and hence Bobkov's Inequality.

Corollary 2. *If $f : \mathbb{R}^n \to \mathbb{R}$ is Lipschitz with constant L, then $\exists M \in \mathbb{R}$, such that*

$$\gamma_n\left(x \in \mathbb{R}^n : |f(x) - M| \geq \varepsilon\right) \leq Ke^{-C\varepsilon^2/(2L^2)},$$

where K is a universal constant.

Proof. For $A \subseteq \mathbb{R}^n$ such that $\alpha := \gamma_n(A) \geq 1/2$, then, by Theorem 4,

$$\begin{aligned}\gamma_n(A_\varepsilon) &\geq \Phi(\Phi^{-1}(\alpha)+\varepsilon) = 1 - \frac{1}{\sqrt{2\pi}}\int_{\Phi^{-1}(\alpha)+\varepsilon}^{\infty} e^{-x^2/2}dx \\ &\geq 1 - \frac{1}{\sqrt{2\pi}}\int_{\varepsilon}^{\infty} e^{-x^2/2}dx \geq 1 - \frac{K}{2}e^{-C\varepsilon^2/2},\end{aligned}$$

for some universal constants K and C. Now let M be the median of f and define

$$A = \{x \in \mathbb{R}^n : f(x) > M\} \quad \text{and} \quad B = \{x \in \mathbb{R}^n : f(x) \leq M\}.$$

Then, we have $\gamma_n(A) = \gamma_n(B) = 1/2$, and

$$\begin{aligned}&\gamma_n\left(x \in \mathbb{R}^n : |f(x) - M| \geq \varepsilon\right) \\ &\leq \gamma_n\left(x \in \mathbb{R}^n : f(x) \geq M + \varepsilon\right) + \gamma_n\left(x \in \mathbb{R}^n : f(x) \leq M - \varepsilon\right) \\ &\leq \gamma_n(\mathbb{R}^n \backslash A_{\varepsilon/L}) + \gamma_n(\mathbb{R}^n \backslash B_{\varepsilon/L}) \leq Ke^{-C\varepsilon^2/(2L^2)}.\end{aligned}$$

□

Proposition 2. *(X, d) is a metric space and μ is the Borel probability measure. $f : X \to \mathbb{R}$. The followings are equivalent:*

(i) $\exists K_1, \delta_1$, $\exists A \in \mathbb{R}$, such that $\forall C > 0$, $\mathbb{P}(|f - A| \geq C) \leq K_1 e^{-\delta_1 C^2}$.

(ii) $\exists K_2, \delta_2$, such that $\forall C > 0$, $\mathbb{P}(|f - \tilde{f}| \geq C) \leq K_2 e^{-\delta_2 C^2}$, where $f, \tilde{f}$ are independent identical distributions.

(iii) $\exists K_3, \delta_3$, such that $\forall C > 0$, $\mathbb{P}(|f - \mathbb{E}f| \geq C) \leq K_3 e^{-\delta_3 C^2}$.

(iv) $\exists K_4, \delta_4$, such that $\forall C > 0$, $\mathbb{P}(|f - M_f| \geq C) \leq K_4 e^{-\delta_4 C^2}$, where M_f is the median of f.

Moreover, if we have

$$K_1 \leq K_4 \leq 2K_3 \leq 2K_2 \leq 4K_1 \quad \textit{and} \quad \delta_1 \geq \delta_4 \geq \frac{\delta_3}{4} \geq \frac{\delta_2}{8} \geq \frac{\delta_1}{32},$$

then, the inequalities in (i)-(iv) can hold simultanenously.

Proof. (i)$\Rightarrow$(ii).

$$\begin{aligned}\mathbb{P}(|f - \tilde{f}| \geq C) &\leq \mathbb{P}\left(|f - A| \geq \frac{C}{2}\right) + \mathbb{P}\left(|\tilde{f} - A| \geq \frac{C}{2}\right) \\ &\leq 2K_1 e^{-\delta_1 \frac{C^2}{4}}.\end{aligned}$$

(ii)$\Rightarrow$(iii). For $\lambda > 0$,

$$\begin{aligned}\mathbb{E}\left(e^{\lambda^2 |f - \tilde{f}|^2}\right) &= \int_0^\infty 2\lambda^2 C e^{\lambda^2 C^2} \mathbb{P}(|f - \tilde{f}| \geq C) dC \\ &\leq \int_0^\infty 2\lambda^2 C e^{\lambda^2 C^2} K_2 e^{-\delta_2 C^2} dC.\end{aligned}$$

Choose $\lambda = \sqrt{\frac{\delta_2}{2}}$ and $\mathbb{E}e^{\frac{\delta_2}{2}|f - \tilde{f}|^2} \leq K_2$. But $\phi(t) := e^{\delta_2^2 t^2/2}$ is convex. Hence

$$K_2 \geq \mathbb{E}_{f,\tilde{f}} e^{\frac{\delta_2}{2}|f - \tilde{f}|^2} \geq \mathbb{E}_f e^{\frac{\delta_2}{2}|f - \mathbb{E}f|^2}.$$

Hence, using Markov's Inequality, we have

$$\begin{aligned}\mathbb{P}(|f - \mathbb{E}f| \geq C) &= \mathbb{P}\left(e^{\frac{\delta_2}{2}|f - \mathbb{E}f|^2} \geq e^{\frac{\delta_2}{2}C^2}\right) \\ &\leq e^{-\frac{\delta_2}{2}C^2} \mathbb{E}e^{\frac{\delta_2}{2}|f - \mathbb{E}f|^2} \leq K_2 e^{-\frac{\delta_2}{2}C^2}.\end{aligned}$$

(iii)$\Rightarrow$(iv). Let $C_0 = \sqrt{(\log 2K_3)/\delta_3}$. Then,

$$\mathbb{P}(|f - \mathbb{E}f| \geq C_0) \leq \frac{1}{2} = K_3 e^{-\delta_3 C_0^2}.$$

We get $\mathbb{P}(f \geq C_0 + \mathbb{E}f), \mathbb{P}(f \leq C_0 - \mathbb{E}f) \leq 1/2$, hence,

$$\mathbb{E}f - C_0 \leq M_f \leq \mathbb{E}f + C_0.$$

So, if $C \geq 2C_0$, then

$$\begin{aligned}\mathbb{P}(|f - M_f| \geq C) &\leq \mathbb{P}(|f - \mathbb{E}f| \geq C - C_0) \\ &\leq K_3 e^{-\delta_3 (C - C_0)^2} \leq K_3 e^{-\frac{\delta_3}{4} C^2}.\end{aligned}$$

If $C \leq 2C_0$, then

$$e^{-\frac{\delta_3}{4} C^2} \geq e^{-\frac{\delta_3}{4} C_0^2} = \frac{1}{2K_3}.$$

Therefore,

$$\mathbb{P}(|f - M_f| \geq C_0) \leq 1 \leq 2K_3 e^{-\delta_3 C^2/4}.$$

(iv)$\Rightarrow$(i). Trivial. □

Remark 3. We can also formulate and prove a version for general tails.

Now, let us discuss the case of the Euclidean sphere:

$$S^{n-1} := \{x \in \mathbb{R}^n : \|x\|_2 = 1\}.$$

Let μ be the normalized surface area measure on S^{n-1} and let d be the geodesic metric on S^{n-1}.

Theorem 6. *(Lévy's Theorem) Balls (spherical caps) are the solutions of the isoperimetric problems on S^{n-1}. i.e. if we have $A \subseteq S^{n-1}$, $x \in S^{n-1}$, $r > 0$ such that*

$$\mu(A) = \mu(y \in S^{n-1} : d(x, y) \leq r\}.$$

Then $\mu(A_\varepsilon) \geq \mu(B(x, r + \varepsilon))$.

Theorem 7. *$f : S^{n-1} \to \mathbb{R}$, Lipschitz with constant L. Then,*

$$\mu\left(\left|f(x) - \int_{S^{n-1}} f d\mu\right| \geq \varepsilon\right) \leq K e^{-\frac{Cn\varepsilon^2}{L^2}}.$$

Theorem 8. *(Maurey-Pisier) $F : \mathbb{R}^n \to \mathbb{R}$, Lipschitz with constant L. $\gamma_n \sim (g_1, \ldots, g_n)$, where $g_1, g_2, \ldots, g_n$ are i.i.d. standard Gaussian random variables. Then,*

$$\mathbb{P}(|F(g_1, \ldots, g_n) - \mathbb{E}F(g_1, \ldots, g_n)| \geq t) \leq 2e^{-\frac{2t^2}{\pi^2 L^2}}.$$

Proof. By approximation, we may assume that F is continuously differentiable. Let

$$G = (g_1, \dots, g_n) \quad \text{and} \quad H = (h_1, \dots, h_n)$$

be independent Gaussians. For $\theta \in [0, \frac{\pi}{2}]$, define $G_\theta = G \sin\theta + H \cos\theta \in \mathbb{R}^n$. Then, $\frac{d}{d\theta} G_\theta = G\cos\theta - H \sin\theta$. Since

$$\begin{pmatrix} G_\theta \\ \frac{d}{d\theta} G_\theta \end{pmatrix} = \begin{pmatrix} \sin\theta & \cos\theta \\ \cos\theta & -\sin\theta \end{pmatrix} \begin{pmatrix} G \\ H \end{pmatrix}$$

and we know the orthogonal transformation of Gaussian is again Gaussian, thus, $(G_\theta, \frac{d}{d\theta} G_\theta)$ has the same distribution as (G, H) (We call this Key Fact). Let $\phi : \mathbb{R} \to \mathbb{R}$ be any convex function. By Jensen's Inequality and Key Fact,

$$\begin{aligned}
&\mathbb{E}_G(\phi(F(G) - \mathbb{E}_H F(H))) \leq \mathbb{E}_{G,H} \phi(F(G) - F(H)) \\
&= \mathbb{E}\left[\phi\left(\int_0^{\pi/2} \frac{d}{d\theta} F(G_\theta) d\theta\right)\right] = \mathbb{E}\left[\phi\left(\int_0^{\pi/2} \langle \nabla F(G_\theta), \frac{d}{d\theta} G_\theta \rangle d\theta\right)\right] \\
&\leq \frac{2}{\pi} \mathbb{E} \int_0^{\pi/2} \phi\left(\frac{\pi}{2} \langle \nabla F(G_\theta), \frac{d}{d\theta} G_\theta \rangle d\theta\right) = \mathbb{E}\phi\left(\frac{\pi}{2} \langle \nabla F(G), H \rangle d\theta\right).
\end{aligned}$$

Therefore, we have

$$\mathbb{E}(\phi(F(G) - \mathbb{E}F(G))) \leq \mathbb{E}\phi\left(\frac{\pi}{2} \sum_{i=1}^n \frac{\partial F}{\partial x_i}(G) h_i\right).$$

For $\lambda \in \mathbb{R}$, $\phi(t) = e^{\lambda t}$, we have

$$\begin{aligned}
\mathbb{E}\left(e^{\lambda \frac{\pi}{2} \sum_{i=1}^n \frac{\partial F}{\partial x_i}(G) h_i}\right) &= \mathbb{E}_G \prod_{i=1}^n \mathbb{E}_H e^{\lambda \frac{\pi}{2} \frac{\partial F}{\partial x_i}(G) h_i} \\
&= \mathbb{E}_G e^{\frac{1}{2} \frac{\lambda^2 \pi^2}{4} \sum_{i=1}^n \left(\frac{\partial F}{\partial x_i}(G)\right)^2} \leq e^{\frac{\lambda^2 \pi^2 L^2}{8}}.
\end{aligned}$$

Thus, for any $\lambda \in \mathbb{R}$,

$$\mathbb{E} e^{\lambda(F(G) - \mathbb{E}F(G))} \leq e^{\frac{\lambda^2 \pi^2 L^2}{8}}.$$

Define $X := F(G) - \mathbb{E}F(G)$. Then, for any $\lambda \in \mathbb{R}$, $\mathbb{E}e^{\lambda x} \leq e^{\pi^2 L^2 \lambda^2/8}$. Applying Markov's Inequality, we get

$$\mathbb{P}(X \geq t) = \mathbb{P}(e^{\lambda X} \geq e^{\lambda t}) \leq e^{-\lambda t} \mathbb{E} e^{\lambda X} \leq e^{-\lambda t + \frac{\pi^2 L^2 \lambda^2}{8}}.$$

Now, let $\psi(\lambda) = -\lambda t + \frac{\pi^2 L^2 \lambda^2}{8}$. Minimizing ψ over λ, we get our desired result. $\square$

Theorem 9. *(Lévy's Isoperimetric Theorem) Let μ be the normalized surface measure on S^{n-1} with Euclidean metric. $A \subseteq S^{n-1}$ is Borel and $C \subseteq S^{n-1}$ is a cap (ball in the Euclidean metric) with $\mu(C) = \mu(A)$. Then, for any $\varepsilon > 0$, $\mu(A_\varepsilon) \geq \mu(C_\varepsilon)$.*

Definition 4. For $A, B \subseteq \mathbb{R}^n$, define the *Minkowski sum* of A and B as

$$A + B := \{a + b : a \in A, b \in B\}.$$

Theorem 10. *(Brunn-Minkowski Inequality) If $A, B \subseteq \mathbb{R}^n$ are compact and nonempty, then*

$$Vol(A + B)^{\frac{1}{n}} \geq Vol(A)^{\frac{1}{n}} + Vol(B)^{\frac{1}{n}},$$

where the equality holds if and only if $A = x + rB$ for some $x \in \mathbb{R}^n$, $r > 0$, up to measure zero.

Proof. It is enough to prove the case when A, B are disjoint unions of axis parallel boxes. Let A, B be disjoint unions of finitely many axis parallel boxes and K be the total number of boxes. Let us do induction on K.

For $K = 2$, $A = \prod_{i=1}^n I_i$, $B = \prod_{i=1}^n J_i$, where $I_i, J_i \subseteq \mathbb{R}$ are intervals. Let $a_i = \mathrm{Vol}(I_i)$ and $b_i = \mathrm{Vol}(J_i)$, then $\mathrm{Vol}(A + B) = \prod_{i=1}^n (a_i + b_i)$. So, we want to show that $\prod_{i=1}^n (a_i + b_i)^{1/n} \geq \prod_{i=1}^n a_i^{1/n} + \prod_{i=1}^n b_i^{1/n}$. This follows from AM-GM since

$$\left(\prod_{i=1}^n \frac{a_i}{a_i + b_i}\right)^{1/n} + \left(\prod_{i=1}^n \frac{b_i}{a_i + b_i}\right)^{1/n} \leq \frac{1}{n}\sum_{i=1}^n \frac{a_i}{a_i + b_i} + \frac{1}{n}\sum_{i=1}^n \frac{b_i}{a_i + b_i} = 1.$$

Next, let's do the induction step. Let H be an axis parallel hyperplane such that on both sides of H there is an entire box from A. (Such a hyperplane H always exists.) Let H^+, H^- be the two sides of the hyperplane. Define

$$\tilde{A} = A \cap H^+ \quad \text{and} \quad \hat{A} = A \cap H^-.$$

$\mathrm{Vol}(\tilde{A})/\mathrm{Vol}(A) = \lambda \in [0, 1]$ and translate B perpendicular to H such that if $\tilde{B} = B \cap H^+$ and $\hat{B} = B \cap H^-$, then $\mathrm{Vol}(\tilde{B})/\mathrm{Vol}(B) = \lambda$. For $\tilde{A}, \tilde{B}, \hat{A}, \hat{B}$, the total number of boxes is smaller than K. By induction hypothesis, we have

$$\mathrm{Vol}(\tilde{A} + \tilde{B})^{\frac{1}{n}} \geq \mathrm{Vol}(\tilde{A})^{\frac{1}{n}} + \mathrm{Vol}(\tilde{B})^{\frac{1}{n}}, \quad \mathrm{Vol}(\hat{A} + \hat{B})^{\frac{1}{n}} \geq \mathrm{Vol}(\hat{A})^{\frac{1}{n}} + \mathrm{Vol}(\hat{B})^{\frac{1}{n}}.$$

Now, $(\tilde{A}+\tilde{B}) \cup (\hat{A}+\hat{B}) \subseteq A+B$, where $\tilde{A}+\tilde{B} \in H^+$ and $\hat{A}+\hat{B} \in H^-$ are disjoint. Hence, we get

$$\begin{aligned}
&\mathrm{Vol}(A+B)^{1/n} \geq \left[\mathrm{Vol}(\tilde{A}+\tilde{B}) + \mathrm{Vol}(\hat{A}+\hat{B})\right]^{1/n} \\
&\geq \left[\left(\mathrm{Vol}(\tilde{A})^{1/n} + \mathrm{Vol}(\tilde{B})^{1/n}\right)^n + \left(\mathrm{Vol}(\hat{A})^{1/n} + \mathrm{Vol}(\hat{B})^{1/n}\right)^n\right]^{1/n} \\
&= \left[\lambda \left(\mathrm{Vol}(A)^{1/n} + \mathrm{Vol}(B)^{1/n}\right)^n + (1-\lambda)\left(\mathrm{Vol}(A)^{1/n} + \mathrm{Vol}(B)^{1/n}\right)^n\right]^{1/n} \\
&= \mathrm{Vol}(A)^{1/n} + \mathrm{Vol}(B)^{1/n}.
\end{aligned}$$

□

Corollary 3. *(Isoperimetric Theorem for Lebesgue Measure) Let* $B_2^n := \{x \in \mathbb{R}^n : \|x\|_2 \leq 1\}$. *$A \subseteq \mathbb{R}^n$ is Borel and $r > 0$ such that $Vol(rB_2^n)$=$Vol(A)$, then*

$$Vol(A_\varepsilon) \geq Vol((r+\varepsilon)B_2^n).$$

Proof. $A_\varepsilon = A + \varepsilon B_2^n$. Applying Brunn-Minkowski Inequality, we get

$$\begin{aligned}
\mathrm{Vol}(A_\varepsilon) &\geq \left(\mathrm{Vol}(A)^{\frac{1}{n}} + \mathrm{Vol}(\varepsilon B_2^n)^{\frac{1}{n}}\right)^n = \left(\mathrm{Vol}(rB_2^n)^{\frac{1}{n}} + \mathrm{Vol}(\varepsilon B_2^n)^{\frac{1}{n}}\right)^n \\
&= \left((r+\varepsilon)\mathrm{Vol}(B_2^n)^{\frac{1}{n}}\right)^n = \mathrm{Vol}((r+\varepsilon)B_2^n).
\end{aligned}$$

□

Lemma 6. *The following two inequalities imply each other.*

(i) $\forall A, B \neq \emptyset$ *and* $A, B \subseteq \mathbb{R}^n$ *compact,*

$$Vol(A+B)^{\frac{1}{n}} \geq Vol(A)^{\frac{1}{n}} + Vol(B)^{\frac{1}{n}}.$$

(ii) $\forall A, B \subseteq \mathbb{R}^n$ *compact and* $\forall \lambda \in [0,1]$,

$$Vol(\lambda A + (1-\lambda)B) \geq Vol(A)^\lambda\, Vol(B)^{1-\lambda}.$$

Proof. (i)⇒(ii). By (i) and concavity of log, we have

$$\begin{aligned}
\frac{1}{n}\log(\mathrm{Vol}(\lambda A + (1-\lambda)B)) &\geq \log(\lambda \mathrm{Vol}(A)^{\frac{1}{n}} + (1-\lambda)\mathrm{Vol}(B)^{\frac{1}{n}}) \\
&\geq \frac{\lambda}{n}\log \mathrm{Vol}(A) + \frac{1-\lambda}{n}\log \mathrm{Vol}(B).
\end{aligned}$$

(ii)⇒(i). Define

$$\tilde{A} = \frac{1}{\mathrm{Vol}(A)^{1/n}} A, \quad \tilde{B} = \frac{1}{\mathrm{Vol}(B)^{1/n}} B \quad \text{and} \quad \lambda = \frac{\mathrm{Vol}(A)^{1/n}}{\mathrm{Vol}(A)^{1/n} + \mathrm{Vol}(B)^{1/n}}.$$

Then, by (ii), we get

$$\lambda \tilde{A} + (1-\lambda)\tilde{B} = \mathrm{Vol}\left(\frac{A+B}{\mathrm{Vol}(A)^{1/n} + \mathrm{Vol}(B)^{1/n}} \right) \geq \mathrm{Vol}(\tilde{A})^{\lambda} \mathrm{Vol}(\tilde{B})^{1-\lambda} = 1,$$

which implies that

$$\mathrm{Vol}(A+B)^{\frac{1}{n}} \geq \mathrm{Vol}(A)^{\frac{1}{n}} + \mathrm{Vol}(B)^{\frac{1}{n}}.$$

□

Definition 5. A measure μ on $\mathbb{R}^n$ is *log-concave* if $\forall \lambda \in [0,1]$, $\forall A, B \subseteq \mathbb{R}^n$ compact,

$$\mu(\lambda A + (1-\lambda)B) \geq \mu(A)^{\lambda} \mu(B)^{1-\lambda}.$$

Example 1. If $d\mu = e^{-f(x)}dx$, where $f(x)$ is convex, then, μ is log-concave.

Lemma 7. *(Borell's Lemma) Let μ be log-concave probability measure on $\mathbb{R}^n$ and $K \subseteq \mathbb{R}^n$ closed, convex and centrally symmetric (i.e. $x \in K \Leftrightarrow -x \in K$) and $\mu(K) = a$. Then $\forall r \geq 1$,*

$$1 - \mu(rK) \leq a \left(\frac{1-a}{a} \right)^{\frac{r+1}{2}}.$$

Proof. For $r \geq 1$ and any $x \in K, y \notin rK$, if

$$\frac{r-1}{r+1} x + \frac{2}{r+1} y = z \in K,$$

then, since K is convex,

$$y = \frac{r+1}{2r}(rz) + \frac{r-1}{2r}(-rx) \in rK.$$

Contradiction. Hence,

$$\frac{r-1}{r+1} x + \frac{2}{r+1} y \notin K.$$

Therefore, we have

$$\frac{2}{r+1}(\mathbb{R}^n \backslash (rK)) + \frac{r-1}{r+1}K \subseteq (\mathbb{R}^n \backslash K).$$

Hence, we get

$$\begin{aligned} 1-a = 1-\mu(K) &\geq \mu\left(\frac{2}{r+1}(\mathbb{R}^n \backslash (rK)) + \frac{r-1}{r+1}K\right) \\ &\geq (1-\mu(rK))^{\frac{2}{r+1}}\mu(K)^{\frac{r-1}{r+1}} = (1-\mu(rK))^{\frac{2}{r+1}}a^{\frac{r-1}{r+1}}. \end{aligned}$$

□

Theorem 11. *(Prékopa-Leindler Inequality) Let $m, f, g : \mathbb{R}^n \to [0, \infty)$ be measurable functions and $\lambda \in [0, 1]$. Assume that for any $x, y \in \mathbb{R}$,*

$$m(\lambda x + (1-\lambda)y) \geq f(x)^\lambda g(y)^{1-\lambda}.$$

Then, we have

$$\int_{\mathbb{R}^n} m dx \geq \left(\int_{\mathbb{R}^n} f dx\right)^\lambda \left(\int_{\mathbb{R}^n} g dx\right)^{1-\lambda}.$$

Remark 4. For $A, B \subseteq \mathbb{R}^n$ compact, let $f = 1_A$, $g = 1_B$ and $m = 1_{\lambda A+(1-\lambda)B}$. Then it is easy to check that

$$1_{\lambda A+(1-\lambda)B} \geq 1_A^\lambda 1_B^{1-\lambda},$$

since it is trivial if $x \notin A$ or $y \notin B$. If $x \in A$ and $y \in B$, then $\lambda x + (1-\lambda)y \in \lambda A + (1-\lambda)B$. Therefore, by Prékopa-Leindler Inequality and Lemma 6, we proved the Brunn-Kinkowski Inequality

$$\mathrm{Vol}(\lambda A + (1-\lambda)B) \geq \mathrm{Vol}(A)^\lambda \mathrm{Vol}(B)^{1-\lambda}.$$

Proof of Prékopa-Leindler Inequality. We do induction on n. For $n = 1$, we will first prove Brunn-Minkowski in dimension 1. $A, B \subseteq \mathbb{R}$, compact, measurable. For any $\varepsilon > 0$, there exists $t \in A \cap B \subseteq \mathbb{R}$ and a translation of A such that

$$\mathrm{Vol}(A \cap [t, \infty)) \geq \mathrm{Vol}(A) - \varepsilon > 0, \mathrm{Vol}(B \cap (-\infty, t]) \geq \mathrm{Vol}(B) - \varepsilon > 0.$$

W.l.o.g., we can let $t = 0$, then, we get

$$\mathrm{Vol}(A') \geq \mathrm{Vol}(A) - \varepsilon, \mathrm{Vol}(B') \geq \mathrm{Vol}(B) - \varepsilon,$$

where $A' = A \cap [0, \infty)$ and $B' = B \cap (-\infty, t]$ and we have

$$A' \cup B' \subseteq A' + B',$$

where A' and B' are disjoint. Therefore, we get

$$\mathrm{Vol}(A + B) = \mathrm{Vol}(A' + B') \geq \mathrm{Vol}(A') + \mathrm{Vol}(B') \geq \mathrm{Vol}(A) + \mathrm{Vol}(B) - 2\varepsilon.$$

It is true for any $\varepsilon > 0$, thus $\mathrm{Vol}(A + B) \geq$Vol(A)+Vol(B). Now, w.l.o.g., assume f and g to be bounded and $\|f\|_\infty = \|g\|_\infty = 1$ since otherwise we can consider

$$\frac{f}{\|f\|_\infty}, \quad \frac{g}{\|g\|_\infty} \quad \text{and} \quad \frac{m}{\|f\|_\infty^\lambda \|g\|_\infty^{1-\lambda}}.$$

Now, for $t \in (0, 1)$, define

$$A = \{x \in \mathbb{R} : f(x) \geq t\}, B = \{x \in \mathbb{R} : g(x) \geq t\}, C = \{x \in \mathbb{R} : m(x) \geq t\}.$$

Since $A, B \neq \emptyset$, $t < 1$, $\|f\|_\infty = \|g\|_\infty = 1$, we have

$$\mathrm{Vol}(\lambda A + (1 - \lambda)B) \geq \mathrm{Vol}(A) + \mathrm{Vol}(B).$$

Also, $\lambda A + (1 - \lambda)B \subseteq C$, since for $x \in A$, $y \in B$, we have

$$m(\lambda x + (1 - \lambda)y) \geq f(x)^\lambda g(x)^{1-\lambda} \geq t^\lambda t^{1-\lambda} = t.$$

Hence, we get

$$\mathrm{Vol}(m \geq t) \geq \lambda \mathrm{Vol}(f \geq t) + (1 - \lambda)\mathrm{Vol}(g \geq t).$$

Now, integrate w.r.t. t, we get

$$\begin{aligned}
\int_{\mathbb{R}} m = \int_0^\infty \mathrm{Vol}(m \geq t)dt &\geq \int_0^1 \mathrm{Vol}(m \geq t)dt \\
&\geq \lambda \int_0^1 \mathrm{Vol}(f \geq t)dt + (1 - \lambda) \int_0^1 \mathrm{Vol}(g \geq t)dt \\
&= \lambda \int f + (1 - \lambda) \int g \geq \left(\int f \right)^\lambda \left(\int g \right)^{1-\lambda},
\end{aligned}$$

where the last inequality above is obtained by AM-GM. Now, let's do induction on dimension n. Suppose $f, g, m : \mathbb{R}^n \to [0, \infty)$, $\mathbb{R}^n = \mathbb{R} \times \mathbb{R}^{n-1}$, $t \in \mathbb{R}$, $s \in \mathbb{R}^{n-1}$, and let

$$f_t(s) = f(t, s), g_t(s) = g(t, s), m_t(s) = m(t, s).$$

Fix $t_0, t_1 \in \mathbb{R}$, define $t = \lambda t_0 + (1 - \lambda)t_1$ for every $u, v \in \mathbb{R}^n$,

$$\begin{aligned} m_t(\lambda u + (1 - \lambda)v) &= m(t, \lambda u + (1 - \lambda)v) = m(\lambda(t_0, u) + (1 - \lambda)(t_1, v)) \\ &\geq f(t_0, u)^\lambda g(t_1, v)^{1-\lambda} = f_{t_0}(u)^\lambda g_{t_1}(v)^{1-\lambda}, \end{aligned}$$

where f_{t_0}, g_{t_1}, m_t satisfy the induction hypothesis at $n - 1$. Thus,

$$\int_{\mathbb{R}^{n-1}} m_t(s)ds \geq \left(\int_{\mathbb{R}^{n-1}} f_{t_0}(s)ds\right)^\lambda \left(\int_{\mathbb{R}^{n-1}} g_{t_1}(s)ds\right)^{1-\lambda}$$

Define

$$\bar{f}(t) = \int_{\mathbb{R}^{n-1}} f_t(s)ds, \bar{g}(t) = \int_{\mathbb{R}^{n-1}} g_t(s)ds, \bar{m}(t) = \int_{\mathbb{R}^{n-1}} m_t(s)ds.$$

For all $t_0, t_1 \in \mathbb{R}$,

$$\bar{m}(\lambda t_0 + (1 - \lambda)t_1) \geq \bar{f}(t_0)^\lambda \bar{g}(t_1)^{1-\lambda}.$$

$\bar{m}, \bar{f}, \bar{g}$ also satisfy the induction hypothesis at dimension 1. Thus,

$$\int_{\mathbb{R}^n} m = \int_{\mathbb{R}} \bar{m} \geq \left(\int_{\mathbb{R}} \bar{f}\right)^\lambda \left(\int_{\mathbb{R}} \bar{g}\right)^{1-\lambda} \geq \left(\int_{\mathbb{R}^n} f\right)^\lambda \left(\int_{\mathbb{R}^n} g\right)^{1-\lambda}.$$

□

Now, we're going to state and prove Gromov-Milman Theorem. Before we do that, let's give the basic settings first. Let $\|\cdot\|$ be a norm on $\mathbb{R}^n$.

$$B_{\|\cdot\|} = K = \{x \in \mathbb{R}^n : \|x\| \leq 1\}$$

is the unit ball. Then, it is clear that K is convex and centrally symmetric. Conversely, any convex centrally symmetric body is a unit ball of a norm.

Definition 6. A norm $\|\cdot\|$ is called *uniformly convex* if $\forall \varepsilon > 0$, $\exists \delta = \delta(\varepsilon)$, such that

$$\forall x, y \in B_{\|\cdot\|}, \quad \|x - y\| \geq \varepsilon, \quad \text{then} \quad \left\|\frac{x + y}{2}\right\| \leq 1 - \delta.$$

Remark 5. On $\mathbb{R}^n$, uniformly convexity is equivelant to strictly convexity plus compactness.

Definition 7. The *modulus of uniform convexity* of $\|\cdot\|$ is

$$\delta_{\|\cdot\|}(\varepsilon) := \inf\left\{1 - \left\|\frac{x+y}{2}\right\| : \|x\|, \|y\| \le 1, \|x-y\| \ge \varepsilon\right\}.$$

Now, if $\|x\|_2, \|y\|_2 \le 1$ and $\|x-y\|_2 = \varepsilon$. Then

$$\|x+y\|_2^2 + \|x-y\|_2^2 = 2\|x\|_2^2 + 2\|y\|_2^2 \le 4,$$

which implies that $\|x+y\|_2^2 + \varepsilon^2 \le 4$. Hence

$$\left\|\frac{x+y}{2}\right\|_2 \le \sqrt{1 - \frac{\varepsilon^2}{4}} \sim 1 - \frac{\varepsilon^2}{8},$$

and

$$\delta_{\|\cdot\|_2}(\varepsilon) = 1 - \sqrt{1 - \varepsilon^2/4} \sim \frac{\varepsilon^2}{8}.$$

For every norm $\|\cdot\|$, $\delta_{\|\cdot\|}(\varepsilon) \le 1 - \sqrt{1 - \varepsilon^2/4}$.

$$\delta_{\|\cdot\|_p}(\varepsilon) \sim \begin{cases} \frac{p-1}{8}\varepsilon^2 + o(\varepsilon^2) & 1 \le p < 2 \\ C_p\varepsilon^p + o(\varepsilon^p) & p > 2 \end{cases}$$

Let $\|\cdot\|$ be a norm on $\mathbb{R}^n$. $\delta(\varepsilon) = \delta_{\|\cdot\|}(\varepsilon)$. $K = \{x \in \mathbb{R}^n : \|x\| \le 1\}$. ν is a Borel probability measure such that

$$\nu(A) := \frac{\mathrm{Vol}(A \cap K)}{\mathrm{Vol}(K)}.$$

μ is a Borel probability measure on $\partial K = S$ (cone measure),

$$\forall A \subseteq S, \quad \mu(A) := \frac{\mathrm{Vol}([0,1]A)}{\mathrm{Vol}(K)} := \frac{\mathrm{Vol}(\cup_{0\le t\le 1}\{ta : a \in A\})}{\mathrm{Vol}(K)}.$$

Theorem 12. *(Gromov-Milman Theorem)*

(i) $\forall A \subseteq K$, *the unit ball, and* $\forall \varepsilon > 0$, *we have*

$$\nu(A_\varepsilon) \ge 1 - \frac{1}{\nu(A)} e^{-2\delta(\varepsilon)n},$$

where $A_\varepsilon = \{x \in \mathbb{R}^n, x \in K : \exists y \in A, \|x - y\| \leq \varepsilon\}$.

(ii) $\forall A \subseteq S := \partial K$, *we have*

$$\mu(A_\varepsilon) \geq 1 - \frac{4}{\mu(A)} e^{-2\delta(\varepsilon/2)n},$$

where $A_\varepsilon = \{x \in \mathbb{R}^n, x \in S : \exists y \in A, \|x - y\| \leq \varepsilon\}$.

Proof. (i) For $A \subseteq K$, $B = K \backslash A_\varepsilon$, if $x \in A$, $y \in B$ and $\|x - y\| \geq \varepsilon$, then, by definition of $\delta(\varepsilon)$, we have $\|(x + y)/2\| \leq 1 - \delta(\varepsilon)$, which implies that $(A + B)/2 \subseteq (1 - \delta(\varepsilon))K$. Hence, by Brunn-Minkowski Inequality, we get

$$\nu(A)^{1/2}\nu(B)^{1/2} \leq \nu\left(\frac{A + B}{2}\right) \leq (1 - \delta(\varepsilon))^n \leq e^{-n\delta(\varepsilon)}.$$

Hence, we have

$$1 - \nu(A_\varepsilon) = \nu(B) \leq \frac{1}{\nu(A)} e^{-2n\delta(\varepsilon)}.$$

(ii) $A \subseteq S$, $\tilde{A} = [1/2, 1]A$, $B = S \backslash A_\varepsilon$ and $\tilde{B} = [1/2, 1]B$. For $x \in \tilde{A}$, $y \in \tilde{B}$, it is not hard to show that $\|x - y\| \geq \varepsilon/2$. Then we have $\|(x+y)/2\| \leq 1 - \delta(\varepsilon/2)$, which implies that $(\tilde{A} + \tilde{B})/2 \subseteq (1 - \delta(\varepsilon/2))K$. Hence, by Brunn-Minkowski Inequality, we get

$$\nu(\tilde{A})^{1/2}\nu(\tilde{B})^{1/2} \leq \nu\left(\frac{\tilde{A} + \tilde{B}}{2}\right) \leq (1 - \delta(\varepsilon/2))^n \leq e^{-n\delta(\varepsilon/2)}.$$

Notice that

$$[0, 1]A \backslash \tilde{A} = K \backslash [1/2, 1]A = [0, 1/2]A = \frac{1}{2}[0, 1]A.$$

Hence,

$$\nu([0, 1]A) \geq \nu(\tilde{A}) \geq \left(1 - \frac{1}{2^n}\right) \nu([0, 1]A).$$

$\square$

Now, we will discuss *Martingale Method* in concentration of measure. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space.

Lemma 8. *Let $\{\emptyset, \Omega\} = \mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \cdots \subseteq \mathcal{F}_n = \mathcal{F}$ be a filtration. And $f_0 = \mathbb{E}f_0, f_1, f_2, \ldots, f_n = f$ is a Martingale w.r.t. this filtration. For $1 \leq i \leq n$, define the Martingale difference as $d_i = f_i - f_{i-1}$. Then, we have*

$$\mathbb{P}(|f - \mathbb{E}f| \geq t) \leq 2e^{-\frac{t^2}{4\sum_{i=1}^n \|d_i\|_\infty^2}}.$$

Proof. Notice that $\forall x \in \mathbb{R}$, $e^x \leq x + e^{x^2}$. Therefore, $\forall \lambda \in \mathbb{R}$,

$$e^{\lambda d_i} \leq \lambda d_i + e^{\lambda^2 d_i^2}.$$

Hence, we get

$$\mathbb{E}[e^{\lambda d_i}|\mathcal{F}_{i-1}] \leq \lambda \mathbb{E}[d_i|\mathcal{F}_{i-1}] + \mathbb{E}[e^{\lambda^2 d_i^2}|\mathcal{F}_{i-1}] = \mathbb{E}[e^{\lambda^2 d_i^2}|\mathcal{F}_{i-1}] \leq e^{\lambda^2\|d_i\|_\infty^2}.$$

Therefore, we get

$$\begin{aligned}
\mathbb{E}e^{\lambda(f-\mathbb{E}f)} &= \mathbb{E}\left[\prod_{i=1}^n e^{\lambda d_i}\right] = \mathbb{E}\left[\mathbb{E}\left(\prod_{i=1}^n e^{\lambda d_i}|\mathcal{F}_{n-1}\right)\right] \\
&= \mathbb{E}\left[\prod_{i=1}^{n-1} e^{\lambda d_i}\mathbb{E}(e^{\lambda d_n}|\mathcal{F}_{n-1})\right] \leq e^{\lambda\|d_i\|_\infty^2}\mathbb{E}\left(\prod_{i=1}^{n-1} e^{\lambda d_i}\right) \\
&\leq \cdots \leq e^{\lambda^2\sum_{i=1}^n \|d_i\|_\infty^2}.
\end{aligned}$$

Thus, by Markov's Inequality, $\forall \lambda > 0$, we have

$$\mathbb{P}(f - \mathbb{E}f \geq t) \leq e^{-\lambda t}\mathbb{E}e^{\lambda(f-\mathbb{E}f)} \leq e^{-\lambda t + \lambda^2\sum_{i=1}^n \|d_i\|_\infty^2}.$$

Now, minimizing the RHS in λ, and consider $\mathbb{P}(f - \mathbb{E}f \leq -t)$ similarly, we will get our desired result. $\square$

Remark 6. Indeed, a stronger inequality is known. For $M = \max_{1\leq i\leq n}\|d_i\|_\infty$ and $\sigma^2 = \|\sum_{i=1}^n \mathbb{E}(d^2|\mathcal{F}_{i-1})\|_\infty$, we have

$$\mathbb{P}(|f - \mathbb{E}f| \geq t) \leq 2e^{-\frac{Ct}{M}\sinh^{-1}(tM/\sigma^2)}.$$

Definition 8. Let S_n be the permutation group of $\{1, 2, \ldots, n\}$. We define a metric d on S_n as

$$d(\pi, \tau) := \frac{1}{n}\{1 \leq i \leq n : \pi(i) \neq \tau(i)\}.$$

Define the uniform measure μ on S_n as

$$\forall A \subseteq S_n, \quad \mu(A) = \frac{|A|}{n!}.$$

Theorem 13. *(Maurey) If $f : S_n \to \mathbb{R}$ is Lipschitz with constant L, then,*

$$\mu(\pi \in S_n : |f(\pi) - \mathbb{E}f| \geq t) \leq 2e^{-\frac{t^2 n}{16L^2}}.$$

Proof. Fix $k \in \{1, 2, \ldots, n\}$, for any distinct indices $i_i, \ldots, i_k \in \{1, 2, \ldots, n\}$, define

$$A_{i_1,\ldots,i_k} := \{\pi \in S_n : \pi(1) = i_1, \ldots, \pi(k) = i_k\},$$

which forms a partition of S_n. Let $\mathcal{F}_k = \sigma(A_{i_1,\ldots,i_k} : i_1, \ldots, i_k)$. Take $i_1, \ldots, i_k$ distinct and $A = A_{i_1,\ldots,i_k}$ an atom of $\mathcal{F}_k$ and let

$$B = A_{i_1,\ldots,i_k,r} \in \mathcal{F}_{k+1}, \quad C = A_{i_1,\ldots,i_k,s} \in \mathcal{F}_{k+1}.$$

B, C are atoms of $\mathcal{F}_{k+1}$. If $\pi \in B$, $|B| = |C|$, then $(r, s)\pi \in C$. Thus, we get

$$\begin{aligned}\mathbb{E}(f|\mathcal{F}_{k+1})|_B - \mathbb{E}(f|\mathcal{F}_{k+1})|_C &= \frac{1}{|B|}\sum_{\pi\in B} f(\pi) - \frac{1}{|C|}\sum_{\pi \in C} f(\pi) \\ &= \frac{1}{|B|}\sum_{\pi\in B}(f(\pi) - f((r,s)\pi)).\end{aligned}$$

Therefore, we have

$$\left|\mathbb{E}(f|\mathcal{F}_{k+1})|_B - \mathbb{E}(f|\mathcal{F}_{k+1})|_C\right| \leq \frac{1}{|B|}\sum_{\pi\in B} Ld(\pi, (r,s)\pi) \leq \frac{2L}{n}.$$

Now, let $f_k = \mathbb{E}(f|\mathcal{F}_k)$ and $d_{k+1} = f_{k+1} - f_k$. If $B \in \mathcal{F}_{k+1}$ is an atom, then there exists a unique $A \in \mathcal{F}_k$ which is an atom such that $B \subseteq A$. Let $N = \#\{C \subseteq A : \text{ atoms of } \mathcal{F}_{k+1}\}$. Then, we have

$$\begin{aligned}\left|d_{k+1}|_B\right| &= \left|\mathbb{E}(f|\mathcal{F}_{k+1})|_B - \frac{1}{N}\sum_{C\subseteq A\text{: atoms of }\mathcal{F}_{k+1}} \mathbb{E}(f|\mathcal{F}_{k+1})|_C\right| \\ &\leq \frac{1}{N}\sum_{C\subseteq A\text{: atoms of }\mathcal{F}_{k+1}} \left|\mathbb{E}(f|\mathcal{F}_{k+1})|_B - \mathbb{E}(f|\mathcal{F}_{k+1})|_C\right| \leq \frac{2L}{n}.\end{aligned}$$

By Lemma 8, we have

$$\mu(\pi \in S_n : |f(\pi) - \mathbb{E}f| \geq t) \leq 2e^{-\frac{t^2}{4n\cdot 4L^2/n^2}} = 2e^{-\frac{t^2 n}{16L^2}}.$$

$\square$

Now, let G be a finite group and μ is the uniform measure on it.

$$G = G_0 \supseteq G_1 \supseteq G_2 \supseteq \cdots \supseteq G_n = \{e\}.$$

d is an invariant metric on G and the quotient metric on G/G_i is given by

$$\rho(G_i x, G_i y) = \min\{d(a,b) : a \in G_i x, b \in G_i y\} = d(xy^{-1}, G_i).$$

We can check that since d is an invariant measure, ρ is also a metric. Let $a_i = \operatorname{diam}(G_i/G_{i+1})$, then the "length" of $G_0 \supseteq \cdots \supseteq G_n$ is $l = \sqrt{\sum_{i=1}^n a_i^2}$.

Theorem 14. *(Schechtman) If $f : G \to \mathbb{R}$ is Lipschitz with constant L w.r.t. metric d, then,*

$$\mu(|f - \mathbb{E}f| \geq t) \leq 2e^{-\frac{t^2}{4L^2}}.$$

Remark 7. The case $G = S_n$ gives us the Maurey Theorem.

Theorem 15. *(Talagrand) Let $(\Omega, \mathcal{F}, \mu)$ be a probability space. For any $n \in \mathbb{N}$, define the probability space $(\Omega^n, \mathcal{F}^{\otimes n}, \mu^{\otimes n})$ and let $\mathbb{P}_n = \mu^{\otimes n}$. Define the metric d as*

$$d((x_1, x_2, \ldots, x_n), (y_1, y_2, \ldots, y_n)) = \#\{1 \leq i \leq n : x_i \neq y_i\}.$$

Then, $\forall A \subseteq \Omega^n$, we have

$$\mathbb{E}_{\mathbb{P}_n} e^{td(x,A)} \leq \frac{1}{\mathbb{P}_n(A)} \left(\frac{1}{2} + \frac{e^t + e^{-t}}{4} \right)^n \leq \frac{1}{\mathbb{P}_n(A)} e^{t^2 n/4}.$$

Remark 8. $\forall f : \Omega^n \to \mathbb{R}$ Lipschitz with constant $1/n$,

$$\mathbb{P}_n(x \in \Omega^n : |f(x) - \mathbb{E}_{\mathbb{P}_n} f| \geq t) \leq 4e^{-\frac{t^2 n}{16}}.$$

We need to following Lemma to prove Talagrand's Theorem.

Lemma 9. *If $g : \Omega \to [0,1]$ is measurable, then,*

$$\int_\Omega \min\left\{e^t, \frac{1}{g(\omega)}\right\} d\mu(\omega) \cdot \int_\Omega g(\omega) d\mu(\omega) \leq \frac{1}{2} + \frac{e^t + e^{-t}}{4}.$$

Proof of Talagrand's Theorem. We do induction on n and use the Lemma above in our proof. For $n = 1$, we have

$$\begin{aligned}\int e^{td(x,A)}d\mu(x) &= \int_\Omega (1_A + e^t 1_{\Omega\setminus A}d\mu = \int_\Omega \min\left\{e^t, \frac{1}{1_A}\right\} d\mu \\ &\le \frac{1}{\int_\Omega 1_A d\mu} \cdot \left(\frac{1}{2} + \frac{e^t + e^{-t}}{4}\right) = \frac{1}{\mu(A)}\left(\frac{1}{2} + \frac{e^t + e^{-t}}{4}\right).\end{aligned}$$

Assuming the case n is true, let's prove that the case $n+1$ holds as well. For $A \subseteq \Omega^{n+1} = \Omega^n \times \Omega$ and $\omega \in \Omega$, define

$$A(\omega) := \{x \in \Omega^n : (x, \omega) \in A\}, \quad B = \bigcup_{\omega\in\Omega} A(\omega).$$

It is not hard to see that

(i) If $A(\omega) \neq \emptyset$, then $d(A, (x, \omega)) = d(A(\omega), x)$.

(ii) $d(A, (x, \omega)) \le 1 + d(B, x)$.

Now, we get

$$\begin{aligned}&\int_{\Omega^{n+1}} e^{td(A,x)}d\mathbb{P}_{n+1}(x) = \int_\Omega \int_{\Omega^n} e^{td(A,(x,\omega))}d\mathbb{P}_n(x)d\mu(\omega) \\ &\le \int_\Omega \left(\int_{\Omega^n} \min\{e^{td(A(\omega),x)}, e^t \cdot e^{td(B,x)}\}d\mathbb{P}_n(x)\right) d\mu(\omega) \\ &\le \int_\Omega \min\left\{\int_{\Omega^n} e^{td(A(\omega),x)}d\mathbb{P}_n(x), e^t \int_{\Omega^n} e^{td(B,x)}d\mathbb{P}_n(x)\right\} d\mu(\omega) \\ &\le \int_\Omega \min\left\{\frac{1}{\mathbb{P}_n(A(\omega))}\left(\frac{1}{2} + \frac{e^t + e^{-t}}{4}\right)^n, \frac{e^t}{\mathbb{P}_n(B)}\left(\frac{1}{2} + \frac{e^t + e^{-t}}{4}\right)^n\right\} d\mu(\omega) \\ &= \left(\frac{1}{2} + \frac{e^t + e^{-t}}{4}\right)^n \frac{1}{\mathbb{P}_n(B)} \int_\Omega \min\left\{e^t, \frac{1}{\mathbb{P}(A(\omega))/\mathbb{P}_n(B)}\right\} d\mu(\omega) \\ &= \left(\frac{1}{2} + \frac{e^t + e^{-t}}{4}\right)^{n+1} \frac{1}{\mathbb{P}_n(B) \cdot \frac{\mathbb{P}_{n+1}(A)}{\mathbb{P}_n(B)}} = \frac{1}{\mathbb{P}_{n+1}(A)}\left(\frac{1}{2} + \frac{e^t + e^{-t}}{4}\right)^{n+1}.\end{aligned}$$

□

Let $X_1, X_2, \dots, X_n$ be normed spaces, with norms $\|\cdot\|_{X_i}$, $1 \le i \le n$. For $\Omega_i \subseteq X_i$, let μ_i be probability measures on Ω_i such that $\operatorname{diam}(\Omega_i) \le 1$ for all i. Define

$$(\Omega, \mathbb{P}_n) = (\Omega_1 \times \cdots \times \Omega_n, \mu_1 \times \cdots \times \mu_n),$$

and

$$d((x_1, y_2, \cdots, x_n), (y_1, y_2, \cdots, y_n)) = \left(\sum_{i=1}^{n} \|x_i - y_i\|_{X_i}^2 \right)^{1/2}.$$

Also, define conv(Ω) as the convex hull of Ω. Under these assumptions, we have the following Theorem.

Theorem 16. *(Concentration for Convex Functions) If $g : conv(\Omega) \to \mathbb{R}$ is a convex function and is Lipschitz with constant 1, then,*

$$\mathbb{P}_n(|g - M| \geq t) \leq 4e^{-t^2/4},$$

where M is a median of g w.r.t. $\mathbb{P}_n$.

Proof. Let $A = \{x \in \Omega : g(x) \leq M\}$, then, $\mathbb{P}_n(A) \geq 1/2$. We claim that

$$\{x \in \Omega : g(x) \geq M + t\} \subseteq \{x \in \Omega : d(x, \mathrm{conv}(A)) \geq t\}.$$

Because if $d(x, \mathrm{conv}(A)) < t$, then $\exists y_1, \ldots, y_k \in A$, $\lambda_i \geq 0$, $\sum_{i=1}^{k} \lambda_i = 1$ such that

$$d\left(x, \sum_{i=1}^{k} \lambda_i y_i\right) < t.$$

Now, g is Lipschitz with constant 1, so we have

$$g(x) \leq t + g\left(\sum_{i=1}^{k} \lambda_i y_i\right) \leq t + \sum_{i=1}^{k} \lambda_i g(y_i) \leq t + \sum_{i=1}^{k} \lambda i M = t + M.$$

By (ii) of Talagrand's Convex Hull Theorem which I will state and prove later, we get

$$\mathbb{P}_n(g(x) \geq M + t) \leq \frac{1}{\mathbb{P}_n(A)} e^{-t^2/4} \leq 2e^{-t^2/4}.$$

Now, define $B = \{x \in \Omega : g(x) \leq M - t\}$, similarly, we have

$$\{x : g(x) \geq M\} \subseteq \{x : d(x, \mathrm{conv}(B)) \geq t\},$$

and

$$\frac{1}{2} \leq \mathbb{P}_n(g \geq M) \leq \frac{1}{\mathbb{P}_n(B)} e^{-t^2/4},$$

which implies that $\mathbb{P}_n(B) \leq 2e^{-t^2/4}$. Therefore, we conclude that

$$\mathbb{P}_n(|g - M| \geq t) \leq 4e^{-t^2/4}.$$

□

Remark 9. We need the convexity assumption on g to get dimension independent concentration.

Example 2. $\Omega_i = \{0,1\} \subseteq [0,1]$ and $\Omega = \{0,1\}^n$. d is the Euclidean metric and here n is even. Define

$$A = \left\{ (x_1, x_2, \ldots, x_n) \in \{0,1\}^n : \sum_{i=1}^n x_i \le \frac{n}{2} \right\},$$

and $g : \{0,1\}^n \to \mathbb{R}$ such that

$$g(x) = d(x, A).$$

Then, g is Lipschitz with constant 1. Take $(x_1, \ldots, x_n) \in \{0,1\}^n$ such that $\sum_{i=1}^n (2x_i - 1) \ge \rho\sqrt{n}$ for some $\rho > 0$. Then, for every $y \in A$,

$$\sum_{i=1}^n (x_i - y_i)^2 \ge \sum_{i=1}^n (x_i - y_i) \ge \frac{n + \rho\sqrt{n}}{2} - \frac{n}{2} = \frac{\rho}{2}\sqrt{n},$$

which implies that $\|x - y\|_2 \ge \sqrt{\rho/2}n^{1/4}$. Since it is true for every $y \in A$, we get

$$g(x) = d(x, A) \ge \sqrt{\frac{\rho}{2}}n^{1/4}.$$

Since

$$\left\{ x \in \{0,1\}^n : g(x) \ge \sqrt{\frac{\rho}{2}}n^{1/4} \right\} \supseteq \left\{ x : \sum_{i=1}^n (2x_i - 1) \ge \rho\sqrt{n} \right\},$$

we have

$$\mathbb{P}_n\left(g(x) \ge \sqrt{\frac{\rho}{2}}n^{1/4} \right) \ge \mathbb{P}_n\left(\sum_{i=1}^n (2x_i - 1) \ge \rho\sqrt{n} \right) \to C(\rho) \text{ as } n \to \infty.$$

Therefore, there is no dimension-free concentration.

Theorem 17. *(Talagrand's Convex Hull Theorem)*
$\forall A \subseteq \Omega$, *define* $f(x, A) = d(x, conv(A))$, *then*
(i)

$$\int_\Omega e^{\frac{1}{4}f(x,A)^2} d\mathbb{P}_n(X) \le \frac{1}{\mathbb{P}_n(A)}.$$

(ii)

$$\mathbb{P}_n(f(x, A) \ge t) \le \frac{1}{\mathbb{P}_n(A)} e^{-t^2/4}.$$

We will need the following Numerical Lemma to prove Talagrand's Convex Hull Theorem.

Lemma 10. *(Numerical Lemma)* $\forall r \in [0,1]$, $\exists \lambda = \lambda(r) \in [0,1]$, *such that*

$$r^{-\lambda} e^{\frac{(1-\lambda)^2}{4}} \leq 2 - r.$$

Proof. Set $f(\lambda) = r^{-\lambda} \exp[(1-\lambda)^2/4]$. Then,

$$f'(\lambda) = \left(\log\left(\frac{1}{r}\right) - \frac{1-\lambda}{2}\right) e^{-\frac{(1-\lambda)^2}{4}} r^{-\lambda} = 0,$$

when $\lambda = 1 + 2\log r < 1$. So the best λ for our Lemma is

$$\lambda = \max\{0, 1 + 2\log r\}.$$

(Case 1: $\lambda = 0$). Then, $1 + 2\log r \leq 0$, so $r \leq e^{-1/2}$. It is easy to check that

$$e^{1/4} \leq 2 - r \leq 2 - e^{-1/2}.$$

(Case 2: $\lambda > 0$). Then, $r > e^{-1/2}$. We need to have

$$2 - r \geq r^{-1-2\log r} e^{(\log r)^2},$$

i.e. $\log(2-r) \geq -(1 + 2\log r)\log r + (\log r)^2 = -\log r - (\log r)^2$. So, we need to show that

$$f(r) := \log(2-r) + (\log r)^2 + \log r \geq 0, \quad 0 \leq r \leq 1.$$

Notice that

$$f'(r) = \frac{-1}{2-r} + \frac{2\log r}{r} + \frac{1}{r},$$

and $f(1) = 0$. So, it suffices to check $f'(r) \leq 0$ for $0 \leq r \leq 1$. $\square$

Proof of Talagrand's Convex Hull Theorem. (ii) can be obtained from (i) by applying Markov's Inequality. Therefore, we only need to prove (i). We do induction on n. For $n = 1$, if $x \in A$, then $f(x, A) = 0$ and if $x \notin A$, then $f(x, A) \leq 1$. Therefore

$$\int_{\Omega_i} e^{\frac{1}{4} f(x,A)^2} d\mu_i(x) \leq \mu_i(A) + e^{1/4}(1 - \mu_i(A)).$$

Denoting $\lambda = \mu_i(A)$, we need $\lambda + (1-\lambda)e^{1/4} \leq 1/\lambda$, i.e.,

$$\psi(\lambda) := \lambda^2 + \lambda(1-\lambda)e^{1/4} \leq 1.$$

This is true since $\psi(1) = 1$ and

$$\begin{aligned}\psi'(\lambda) &= 2\lambda + (1-2\lambda)e^{1/4} \geq 2\lambda(1-e^{1/4}) + e^{1/4}\\ &\geq 2(1-e^{1/4}) + e^{1/4} = 2 - e^{1/4} \geq 0.\end{aligned}$$

Now, assuming the case n and we will try to show the case $n+1$. For $A \subseteq \Omega \times \Omega_{n+1}$ and for every $\omega \in \Omega_{n+1}$, let

$$A(\omega) = \{y \in \Omega : (y,\omega) \in A\}, \quad B = \bigcup_{\omega \in \Omega_{n+1}} A(\omega).$$

Claim: Fix $z = (y,\omega) \in \Omega \times \Omega_{n+1}$, then, for every $\lambda \in [0,1]$,

$$f(z,A)^2 \leq \lambda f(y, A(\omega))^2 + (1-\lambda)f(y,B)^2 + (1-\lambda)^2.$$

Let's prove the Claim first. Take $x \in \mathrm{conv}(A(\omega))$, for which

$$f(y,A(\omega))^2 = \|y-x\|^2 = \sum_{i=1}^n \|y_i - x_i\|_{X_i}^2.$$

Take $u \in \mathrm{conv}(B)$, for which

$$f(y,B) = \|y-u\|^2 = \sum_{i=1}^n \|y_i - u_i\|_{X_i}^2.$$

For any $(x,\omega) \in \mathrm{conv}(A)$, there exist some $v_1, \ldots, v_m \in A(\omega)$, $\lambda_i \geq 0$, $\sum_{i=1}^m \lambda_i =$, such that $x = \sum_{i=1}^m \lambda_i v_i$.

$$(x,\omega) = \sum_{i=1}^m \lambda_i (v_i, \omega) \in \mathrm{conv}(A).$$

$\exists \omega' \in \mathrm{conv}(\Omega_{n+1})$, $(u,\omega') \in \mathrm{conv}(A)$. $u = \sum_{i=1}^m v_i$, $v_i \in B$, $\exists \omega_i \in \Omega_{n+1}$, $\omega' = \sum_{i=1}^m \lambda_i \omega_i \in \mathrm{conv}(\Omega_{n+1})$, $(v_i, \omega_i) \in A$. Now, we have

$$\begin{aligned}f(z,A)^2 &\leq d(z, \lambda(z,\omega) + (1-\lambda)(u,\omega'))^2\\ &= d(y, \lambda x + (1-\lambda)u)^2 + \|\omega - \lambda\omega - (1-\lambda)\omega'\|_{X_{n+1}}^2\\ &\leq (\lambda\|y-x\| + (1-\lambda)\|y-u\|)^2 + (1-\lambda)^2\|\omega-\omega'\|_{X_{n+1}}\\ &\leq \lambda|y-x\|^2 + (1-\lambda)\|y-u\|^2 + (1-\lambda)^2\\ &= \lambda f(y,A(\omega))^2 + (1-\lambda)f(y,B)^2 + (1-\lambda)^2,\end{aligned}$$

which proves the Claim.

Fix $\omega \in \Omega_{n+1}$ and $\lambda \in [0,1]$. We get

$$\begin{aligned}
&\int_\Omega e^{\frac{1}{4}f((y,\omega),A)^2} d\mathbb{P}_n(y) \le \int_\Omega e^{\frac{1}{4}(\lambda f(y,A(\omega))^2+(1-\lambda)f(y,B)^2+(1-\lambda)^2)} d\mathbb{P}_n(y) \\
&= e^{\frac{(1-\lambda)^2}{4}} \int_\Omega \left(e^{\frac{1}{4}f(y,A(\omega))^2}\right)^\lambda \left(e^{\frac{f(y,B)^2}{4}}\right)^{1-\lambda} d\mathbb{P}_n \\
&\le e^{\frac{(1-\lambda)^2}{4}} \left(\int_\Omega e^{\frac{1}{4}f(y,A(\omega))^2} d\mathbb{P}_n(y)\right)^\lambda \left(\int_\Omega e^{\frac{f(y,B)^2}{4}} d\mathbb{P}_n(y)\right)^{1-\lambda} \\
&\le e^{\frac{(1-\lambda)^2}{4}} \frac{1}{\mathbb{P}_n(A(\omega))^\lambda} \cdot \frac{1}{\mathbb{P}_n(B)^{1-\lambda}} = \mathbb{P}_n(B)^{-1} \left(\frac{\mathbb{P}_n(A(\omega))}{\mathbb{P}_n(B)}\right)^{-\lambda} e^{(1-\lambda)^2/4}.
\end{aligned}$$

Now, apply Numerical Lemma with $r = \mathbb{P}_n(A(\omega))/\mathbb{P}_n(B) \le 1$, we can find some $\lambda = \lambda(\omega)$, such that

$$\int_\Omega e^{\frac{1}{4}f((y,\omega),A)^2} d\mathbb{P}_n(y) \le \mathbb{P}_n(B)^{-1} \left(2 - \frac{\mathbb{P}_n(A(\omega))}{\mathbb{P}_n(B)}\right).$$

Finally, integrating over $\omega \in \Omega_{n+1}$, we get

$$\begin{aligned}
&\int_{\Omega\times\Omega_{n+1}} e^{\frac{1}{4}f(z,A)^2} d\mathbb{P}_{n+1}(z) \le \frac{1}{\mathbb{P}_n(B)} \left(2 - \frac{\mathbb{P}_{n+1}(A)}{\mathbb{P}_n(B)}\right) \\
&= \frac{1}{\mathbb{P}_{n+1}(A)} \cdot \frac{\mathbb{P}_{n+1}(A)}{\mathbb{P}_n(B)} \left(2 - \frac{\mathbb{P}_{n+1}(A)}{\mathbb{P}_n(B)}\right) \le \frac{1}{\mathbb{P}_{n+1}(A)}.
\end{aligned}$$

□

Next, we introduce the Khintchine Inequality, is a theorem from probability, and is also frequently used in analysis. Heuristically, it says that if we pick n real numbers $a_1, \dots, a_n$, and add them together each multiplied by a random sign ± 1, then the expected value of the modulus, or the modulus it will be closest to on average, will be not too far off from $\sqrt{|a_1|^2 + \cdots + |a_n|^2}$

Theorem 18. *(Khintchine's Inequality) Let ξ_i, be i.i.d. random variables such that $\mathbb{P}(\xi_i = 1) = \mathbb{P}(\xi_i = -1) = 1/2$. Then, for all $1 \le p < \infty$, $\exists A_p, B_p$ such that for all n, $\forall a_1, \dots, a_n \mathbb{R}$, we have*

$$A_p \left(\sum_{i=1}^n a_i^2\right)^{1/2} \le \left(\mathbb{E}_{\xi\in\{\pm1\}^n} \left|\sum_{i=1}^n \xi_i a_i\right|^p\right)^{1/p} \le B_p \left(\sum_{i=1}^n a_i^2\right)^{1/2},$$

where $B_p = O(\sqrt{p})$.

Proof. Notice first that $(\sum_{i=1}^n a_i^2)^{1/2} = \sqrt{\mathbb{E}|\sum_{i=1}^n a_i\xi_i|^2}$. Let $X = \sum_{i=1}^n a_i\xi_i$. By the simple fact that $\frac{1}{2}(e^t + e^{-t}) \le e^{t^2/2}$, we get

$$\mathbb{E}e^{\lambda X} = \prod_{i=1}^n \frac{e^{\lambda a_i} + e^{-\lambda a_i}}{2} \le \prod_{i=1}^n e^{\lambda^2 a_i^2/2} = e^{\frac{\lambda^2}{2}\sum_{i=1}^n a_i^2}.$$

W.l.o.g., we assume $\sum_{i=1}^n a_i^2 = 1$. Thus, $\mathbb{E}e^{\lambda X} \le e^{\lambda^2/2}$. Now by the usual argument, i.e., applying Markov's Inequality and optimizing over λ, we get

$$\mathbb{P}(|X| \ge t) \le 2e^{-t^2/2}.$$

Therefore, we have

$$\mathbb{E}|X|^p = \int_0^\infty pt^{p-1}\mathbb{P}(|X| \ge t)dt \le \int_0^\infty 2pt^{p-1}e^{-t^2/2}dt \le (Cp)^{p/2}.$$

Hence, for $B_p = O(\sqrt{p})$, we have

$$\left(\mathbb{E}_{\xi\in\{\pm1\}^n}\left|\sum_{i=1}^n \xi_i a_i\right|^p\right)^{1/p} \le B_p\left(\sum_{i=1}^n a_i^2\right)^{1/2}.$$

Now, if $p \ge 2$, we obtain lower bound with $A_p = 1$ by Jensen's Inequality. If $1 \le p < 2$, define $\theta \in [0,1]$ by $\frac{1}{2} = \frac{\theta}{p} + \frac{1-\theta}{4}$. Applying Hölder's Inequality, we get

$$\left(\sum_{i=1}^n a_i^2\right)^{1/2} = \|X\|_2 \le \|X\|_p^\theta \|X\|_4^{1-\theta} \le \|X\|_p^\theta B_4^{1-\theta}\|X\|_2^{1-\theta}.$$

Thus, $\|X\|_2 \le B_4^{\frac{1-\theta}{\theta}}\|X\|_p$. $\square$

Theorem 19. *(Kahane's Inequality) Let ξ_i, be i.i.d. random variables such that $\mathbb{P}(\xi_i = 1) = \mathbb{P}(\xi_i = -1) = 1/2$. Then, for all $1 \le p < \infty$, $\exists A_p, B_p = O(\sqrt{p})$ such that for any normed space X, any $n \in \mathbb{N}$, any $x_1, \ldots, x_n \in X$, we have*

$$A_p\mathbb{E}\|\sum_{i=1}^n \xi_i x_i\| \le \left(\mathbb{E}\|\sum_{i=1}^n \xi_i x_i\|^p\right)^{1/p} \le B_p\mathbb{E}\|\sum_{i=1}^n \xi_i x_i\|.$$

Proof. We only prove the upper bound here. $\Omega_i = \{-1, 1\}$ with uniform measure. $f : \mathbb{R}^n \to \mathbb{R}$,

$$f(a_1, \ldots, a_n) = \|\sum_{i=1}^n a_i x_i\|,$$

and f is convex. Define

$$\sigma = \sup_{x^* \in X^*, \|x^*\|=1} \left(\sum_{i=1}^n x^*(x_i)^2 \right)^{1/2}.$$

Claim: $\sigma = \|f\|_{\text{Lip}}$.

Now, let's prove the Claim. Notice that

$$\begin{aligned}
\|f\|_{\text{Lip}} = \sup_{a \in S^{n-1}} \|\sum_{i=1}^n a_i x_i\| &= \sup_{a \in S^{n-1}} \sup_{\|x^*\|=1} x^* \left(\sum_{i=1}^n a_i x_i \right) \\
= \sup_{\|x^*\|=1} \sup_{a \in S^{n-1}} \sum_{i=1}^n a_i x^*(x_i) &= \sup_{\|x^*\|=1} \left(\sum_{i=1}^n x^*(x_i)^2 \right)^{1/2} = \sigma.
\end{aligned}$$

Let M be a median of f on $\{\pm 1\}^n$. Theorem 16 implies that

$$\mathbb{P}(|f - M| \geq t) \leq 4e^{-\frac{t^2}{16\sigma^2}}.$$

Hence, we get

$$\begin{aligned}
&\mathbb{E}\|\sum_{i=1}^n \xi_i x_i\|^p = \int_0^\infty p t^{p-1} \mathbb{P}(f \geq t) dt \\
&\leq \int_0^{2M} p t^{p-1} dt + \int_{2M}^\infty p t^{p-1} e^{-\frac{(t-M)^2}{16\sigma^2}} dt \leq (CM)^p + (C\sigma\sqrt{p})^p.
\end{aligned}$$

□

Corollary 4. $(\mathbb{E}_\xi \|\sum_{i=1}^n \xi_i x_i\|^p)^{1/p} \leq C(M + \sigma\sqrt{p})$, *where* M *is a median of* $\|\sum_{i=1}^n \xi_i x_i\|$ *and* $\sigma = \sup_{\|x^*\|=1} (\sum_{i=1}^n x^*(x_i)^2)^{1/2}$.

Next, we will discuss spectral methods. Let $(\Omega, \mathbb{F}, \mathbb{P})$ be a probability space. A is a set of measurable function. $\mathcal{E} : A \to \mathbb{R}_+$, where $\mathcal{E}$ is the energy functional. A key example is (X, d, μ) where d is a metric, μ a Borel probability measure, for $f : X \to \mathbb{R}$, define the "abstract length of the gradient of f at $x \in X$" as

$$|\nabla f|(x) = \limsup_{y \to x} \frac{|f(y) - f(x)|}{d(y,x)} \leq \|f\|_{\text{Lip}},$$

if f is Lipschitz. Let $A = \{f : X \to \mathbb{R} \text{ Lipschitz}\}$ and

$$\mathcal{E}(f) = \int_X |\nabla f|(x)^2 d\mu(x).$$

Definition 9. (i) $f : X \to \mathbb{R}$,

$$\sigma_\mu^2(f) := \mathbb{E}(f - \mathbb{E}f)^2 = \mathbb{E}f^2 - (\mathbb{E}f)^2.$$

(ii) $f : X \to \mathbb{R}_+$,

$$\text{Ent}_\mu(f) := \mathbb{E}(f \log f) - (\mathbb{E}f)\log(\mathbb{E}f).$$

Definition 10. The pair $(A, \mathcal{E})$ satisfies the *Poincaré Inequality with constant C* if

$$\forall f \in A, \quad \sigma^2(f) \leq C\mathcal{E}(f).$$

Definition 11. The pair $(A, \mathcal{E})$ satisfies the *log-Sobolev Inquality with constant C* if

$$\forall f \in A, \quad \text{Ent}(f^2) \leq C\mathcal{E}(f).$$

Example 3. Here's an example of Poincaré Inequality. If (X, g) is Riemannian manifold and here

$$\mathcal{E}(g) = \int |\nabla g|^2 d\mu,$$

where μ is the normalized Riemannian volume. Then we have Poincaré Inequality with constant $C = 1/\lambda_1$, where λ_1 is the first eigenvalue of Laplacian.

Proposition 3. *(i) $\sigma^2(\lambda f) = \lambda^2\sigma^2(f)$.*

(ii) Ent$(\lambda f) = \lambda$Ent(f), for $\lambda \geq 0, f \geq 0$.

(iii) Ent$(f) \geq 0$.

Proof. (i) $\sigma^2(\lambda f) = \mathbb{E}(\lambda f - \mathbb{E}\lambda f)^2 = \lambda^2\sigma^2(f)$.

(ii) We do the following computations.

$$\begin{aligned}
&\text{Ent}(\lambda f) = \mathbb{E}[\lambda f \log(\lambda f)] - \mathbb{E}(\lambda f)\log \mathbb{E}\lambda f \\
&= \lambda\mathbb{E}(f \log f) + \lambda \log \lambda \mathbb{E}f - \lambda\mathbb{E}f \log \mathbb{E}f - \lambda \log \lambda \mathbb{E}f = \lambda\text{Ent}(f).
\end{aligned}$$

(iii) $\psi(x) := x \log x$ is convex. So, Jensen's Inequality implies that

$$\mathbb{E}\psi(f) - \psi(\mathbb{E}f) = \mathrm{Ent}(f) \geq 0.$$

$\square$

Lemma 11. *Given $\Phi : \mathbb{R} \to \mathbb{R}$ a C^1 function, (X, d, μ) a probability metric space and $f : X \to \mathbb{R}$ Lipschitz, then, we have*

$$\mathcal{E}(\Phi \circ f) \leq \|f\|_{Lip}^2 \int_X |\Phi'(f(x))|^2 d\mu(x).$$

Proof. For any $x \in X$, we have

$$\begin{aligned}|\nabla \Phi \circ f|(x) &= \limsup_{y \to x} \frac{|\Phi(f(y)) - \Phi(f(x))|}{d(y,x)} \\ &= \limsup_{y \to x} \frac{|\Phi(f(y)) - \Phi(f(x))|}{|f(y) - f(x)|} \cdot \frac{|f(y) - f(x)|}{d(y,x)} \leq \|f\|_{\mathrm{Lip}} |\Phi'(f(x))|.\end{aligned}$$

Hence, we get

$$\mathcal{E}(\Phi \circ f) = \int_X |\nabla \Phi \circ f|^2 d\mu \leq \|f\|_{\mathrm{Lip}}^2 \int_X |\Phi'(f(x))|^2 d\mu(x).$$

$\square$

Theorem 20. *(Gromov-Milman)* $(\Omega, \mathbb{P}, d)$ *is a metric probability space.* $A =$ *{bounded Lipschitz functions}. And $(A, \mathcal{E})$ satisfies the Poincaré Inequality with constant C. Then, for any $|lambda| \leq 2/\sqrt{C}$, and any $f : \Omega \to \mathbb{R}$ Lipschitz with constant 1, we have*

(i) $\mathbb{E}[e^{\lambda(f - \mathbb{E}f)}] \leq 240/(4 - C\lambda^2)$.

(ii) $\forall t > 0$, *we have the exponential concentration*

$$\mathbb{P}(|f - \mathbb{E}f| > t) \leq 240 e^{-\sqrt{\frac{2}{C}} \cdot t}.$$

Proof. (i) Using the Poincaré Inequality with the function $e^{\frac{\lambda}{2} f}$ and applying Lemma 11 with $\Phi(x) = e^x$, we get

$$\mathbb{E}e^{\lambda f} - (\mathbb{E}e^{\frac{\lambda}{2} f})^2 = \sigma^2(e^{\frac{\lambda}{2} f}) \leq C\mathcal{E}(e^{\frac{\lambda}{2} f}) \leq C \frac{\lambda^2}{4} \mathbb{E}e^{\lambda f}.$$

Hence, we get

$$\mathbb{E}e^{\lambda f} \leq \frac{1}{1-\frac{C\lambda^2}{4}}\left(\mathbb{E}e^{\frac{\lambda}{2}f}\right)^2 \leq \frac{1}{1-\frac{C\lambda^2}{4}}\left(\frac{1}{1-\frac{C\lambda^2}{16}}\right)^2\left(\mathbb{E}e^{\frac{\lambda}{4}f}\right)^4$$
$$\leq \cdots \leq \left(\prod_{k=1}^{m}\left(\frac{1}{1-\frac{C\lambda^2}{4^k}}\right)^{2^{k-1}}\right)\cdot\left(\mathbb{E}e^{\frac{\lambda}{2^m}f}\right)^{2^m}.$$

And notice that

$$\left(\mathbb{E}e^{\frac{\lambda}{2^m}f}\right)^{2^m} = \left(\mathbb{E}\left(1+\frac{\lambda}{2^m}f+o\left(\frac{1}{2^m}\right)\right)\right)^{2^m} \to e^{\lambda\mathbb{E}f},$$

as $m \to \infty$ and

$$\leq \cdots \leq \left(\prod_{k=1}^{\infty}\left(\frac{1}{1-\frac{C\lambda^2}{4^k}}\right)^{2^{k-1}}\right) \leq \frac{240}{4-C\lambda^2}.$$

Therefore, we have

$$\mathbb{E}[e^{\lambda(f-\mathbb{E}f)}] \leq \frac{240}{4-C\lambda^2}.$$

(ii) By (i), for any $\lambda > 0$, we have

$$\mathbb{P}(|f-\mathbb{E}f|>t) \leq \left(\mathbb{E}e^{\lambda|f-\mathbb{E}f|}\right)e^{-\lambda t} \leq \frac{240}{4-C\lambda^2}e^{-\lambda t}.$$

So taking $\lambda = \sqrt{2/C}$, we get (ii). □

Remark 10. Same holds for f to be Lipschitz (not necessarily bounded). For example, we can truncate f by considering

$$f^M := \begin{cases} f & |f| \leq M \\ M & f > M \\ -M & f < -M \end{cases}$$

Theorem 21. *(Herbst's Theorem)* $(\Omega, \mathbb{P}, d)$ *is a metric probability space and* $A = \{$*bounded Lipschitz functions*$\}$ *and* $(A, \mathcal{E})$ *satisfies log-Sobolev Inequality with constant* C*. Then for every function* $f : \Omega \to \mathbb{R}$ *Lipschitz with constant* 1*, we have for any* $\lambda \in \mathbb{R}$,

$$\mathbb{E}e^{\lambda(f-\mathbb{E}f)} \leq e^{C\lambda^2/4}, \quad \mathbb{P}(|f-\mathbb{E}f|>t) \leq 2e^{-t^2/C}.$$

Proof. Applying log-Sobolev Inequality to the function $e^{\lambda f/2}$ and applying Lemma 11 with $\Phi(x) = e^x$, we get

$$\begin{aligned}\mathrm{Ent}((e^{\frac{\lambda f}{2}})^2) = \mathrm{Ent}(e^{\lambda f}) &= \mathbb{E}e^{\lambda f}\lambda f - (\mathbb{E}e^{\lambda f})\log(\mathbb{E}e^{\lambda f}) \\ &\le C\mathcal{E}(e^{\lambda f/2}) \le C\frac{\lambda^2}{4}\mathbb{E}(e^{\lambda f}).\end{aligned}$$

Therefore, we have

$$\mathbb{E}(e^{\lambda f}\lambda f) - \mathbb{E}(e^{\lambda f})\log(\mathbb{E}e^{\lambda f}) \le C\frac{\lambda^2}{4}\mathbb{E}(e^{\lambda f}).$$

Define $h(\lambda) = \mathbb{E}(e^{\lambda f})$. Then, $h'(\lambda) = \mathbb{E}(fe^{\lambda f})$ and

$$\lambda h'(\lambda) - h(\lambda)\log h(\lambda) \le C\frac{\lambda^2}{4}h(\lambda),$$

which implies $(\frac{1}{\lambda}\log h(\lambda))' \le C/4$. Hence

$$\lim_{\lambda\to 0}\frac{1}{\lambda}\log h(\lambda) = \lim_{\lambda\to 0}\frac{h'(\lambda)}{h(\lambda)} = \lim_{\lambda\to 0}\frac{\mathbb{E}e^{\lambda f}f}{\mathbb{E}^{\lambda f}} = \mathbb{E}f.$$

Hence, we have

$$\frac{1}{\lambda}\log h(\lambda) \le \mathbb{E}f + \frac{C}{4}\lambda.$$

Thus,

$$\mathbb{E}e^{\lambda f} = h(\lambda) \le e^{\lambda\mathbb{E}f}e^{\frac{C}{4}\lambda^2} \quad\Rightarrow\quad \mathbb{E}e^{\lambda(f-\mathbb{E}f)} \le e^{\frac{C}{4}\lambda^2}.$$

□

Remark 11. Same holds for f to be Lipschitz (not necessarily bounded) by truncation.

Now we consider the product spaces and Poincaré Inequality and log-Sobolev Inequality tensorize.

$$\Omega = \Omega_1 \times \cdots \times \Omega_n, \quad \mu = \mu_1 \times \cdots \times \mu_n.$$

$f : \Omega \to \mathbb{R}$. Fix i and $x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n \in \prod_{j\neq i}\Omega_j$.

$$f_i^{x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_n} : \Omega_i \to \mathbb{R}, \quad f_i^{x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_n} = f(x_1, \ldots, x_n).$$

Proposition 4.

$$Ent_\mu(f) \le \sum_{i=1}^n \mathbb{E}_\mu \left(Ent_{\mu_i}(f_i^{x_1,\dots,x_{i-1},x_{i+1},\dots,x_n}) \right).$$

Remark 12. We also have

$$\sigma^2_\mu(f) \le \sum_{i=1}^n \mathbb{E}_\mu \left(\sigma^2_{\mu_i}(f_i^{x_1,\dots,x_{i-1},x_{i+1},\dots,x_n}) \right).$$

By Proposition 4, we get the following Corollary.

Corollary 5. *If each Ω_i satisfies log-Sobolev Inequality with constant C, A_i with energy $\mathcal{E}_i$ and*

$$A = \{f : \Omega \to \mathbb{R} : f_i \in A_i\}, \quad \mathcal{E}(f) = \sum_{i=1}^n \mathbb{E}_\mu \mathcal{E}_i(f_i),$$

then, $(\Omega, \mathcal{E}, A)$ satisfies log-Sobolev Inequality with the same constant C.

Before we prove Proposition 4, we recall the famous Young's Inequality.

Lemma 12. *(Young's Inequality) $p : \mathbb{R}_+ \to \mathbb{R}_+$ increasing and $p(0) = 0$, $q = p^{-1}$, i.e. q is the inverse function of p. Then*

$$\forall u, v \in \mathbb{R}_+, \quad uv \le \int_0^u p(s)ds + \int_0^v q(t)dt.$$

Now, we prove the following Lemma which will be used to prove Proposition 4.

Lemma 13.

$$\forall f \in \Omega \to \mathbb{R}_+, \quad Ent(f) = \sup\{\mathbb{E}(fg) : \mathbb{E}e^g \le 1\}.$$

Proof. Let $p(t) = \log(t+1)$ and $q(t) = e^t - 1$. Then, by Young's Inequality,

$$uv \le (u+1)\log(u+1) - u + e^v - 1 - v.$$

Now, let $x = u + 1$ and $y = v$, we get

$$\forall x, y \ge 0, \quad xy \le x \log x - x + e^y.$$

Hence, $fg \le f \log f - f + e^g$. Normalize $\mathbb{E}f = 1$ and assume $\mathbb{E}e^g \le 1$, we get

$$\mathbb{E}fg \le \mathbb{E}f \log f = \mathrm{Ent} f.$$

Taking $g = \log f$, we finish the proof. □

Proof of Proposition 4. $f : \Omega_1 \times \cdots \times \Omega_n := \Omega \to \mathbb{R}_+$. $g : \Omega \to \mathbb{R}$ such that $\mathbb{E}e^g \le 1$.

$$g^i(x_1, \ldots, x_n) = \log \left(\frac{\int e^{g(x_1,\ldots,x_n)} d\mu_1(x_1) \cdots d\mu_{i-1}(x_{i-1})}{\int e^{g(x_1,\ldots,x_n)} d\mu_1(x_1) \cdots d\mu_i(x_i)} \right)$$

depends only on $(x_i, \ldots, x_n)$. We have

$$\sum g^i = \log \frac{e^g}{\mathbb{E}e^g} \ge g \quad \Rightarrow \quad g \le \sum g^i,$$

and $\mathbb{E}_{\mu_i} e^{(g^i)_i} = 1$. Hence,

$$\mathbb{E}(fg) \le \sum_i \mathbb{E} f g^i = \sum_i \mathbb{E}_\mu (\mathbb{E}_{\mu_i} f_i (g^i)_i) \le \sum_i \mathbb{E}_\mu \mathrm{Ent}(f_i).$$

Now, by Lemma 13, we complete the proof. □

A key example for Poincaré's Inequality is for μ the measure on $\mathbb{R}$, such that $d\mu(t) = \frac{1}{2} e^{-|t|} dt$. We have the following Lemma.

Lemma 14. *μ is a measure on $\mathbb{R}$ such that $d\mu(t) = \frac{1}{2} e^{-|t|} dt$. Then, for every Lipschitz function $f : \mathbb{R} \to \mathbb{R}$, we have*

$$\sigma_\mu^2(f) \le 4 \int_{\mathbb{R}} (f')^2 d\mu.$$

Proof. Standard approximation tells us that it is enough to deal with $f \in C^1$ bounded. Now, for any $\phi : \mathbb{R} \to \mathbb{R}$ smooth, we have

$$\begin{aligned} \mathbb{E}_\mu(\phi) &= \frac{1}{2} \int_{-\infty}^{\infty} \phi(t) e^{-|t|} dt \\ &= \phi(0) + \frac{1}{2} \int_{-\infty}^{\infty} \phi'(t) \mathrm{sgn}(t) e^{-|t|} dt, \end{aligned}$$

by integration by parts. If we choose $\phi(t) = (f(t) - f(0))^2$, then we get

$$\begin{aligned} \mathbb{E}_\mu(f(t) - f(0))^2 &= 2 \int_{-\infty}^{\infty} (f(t) - f(0)) f'(t) \mathrm{sgn}(t) \frac{e^{-|t|}}{2} dt \\ &\le 2 \left(\mathbb{E}_\mu (f - f(0))^2 \right)^{1/2} \left(\mathbb{E}_\mu (f')^2 \right)^{1/2}. \end{aligned}$$

But we know that $\mathbb{E}_\mu(f(t) - \mathbb{E}_\mu(f))^2 \le \mathbb{E}_\mu(f(t) - f(0))^2$, hence

$$\mathbb{E}_\mu(f - \mathbb{E}_\mu f)^2 \le 4 \mathbb{E}_\mu (f')^2.$$

□

Corollary 6. *(Tensorization) For $x \in \mathbb{R}^n$, we have exponential measure on $\mathbb{R}^n$ as*

$$d\mu(x) = \frac{1}{2^n} e^{-\|x\|_1} dx.$$

Then, for any $f : \mathbb{R}^n \to \mathbb{R}$ Lipschitz, we have

$$\sigma_\mu(f) \leq 4 \int_{\mathbb{R}^n} \|\nabla f\|_2^2 d\mu.$$

Furthermore, by Gromov-Milman, if $f : \mathbb{R}^n \to \mathbb{R}$ is 1-Lipschitz, i.e.

$$|f(x) - f(y)| \leq \|x - y\|_2,$$

then, $\mu(|f - \mathbb{E}_\mu f| \geq t) \leq Ce^{-Ct}$.

Lemma 15. *(X, d) is a metric space. $A \subseteq X$ and $f : A \to \mathbb{R}$ is 1-Lipschitz, i.e.*

$$|f(x) - f(y)| \leq d(x, y) \quad \forall x, y \in A,$$

then, $\exists F : X \to \mathbb{R}$, such that $F|_A = f$ and $\|F\|_{Lip} \leq 1$.

Proof. Define for $x \in X$,

$$F(x) = \inf\{f(y) + d(x, y) : y \in A\}.$$

Then, for $x \in A$, $F(x) = f(x)$, since $\forall y \in A$, $f(y) + d(x, y) \geq f(x)$. Now, for any $x_1, x_2 \in X$, and for any $\varepsilon > 0$, there exists $y_1 \in A$, such that

$$f(y_1) + d(y_1, x_1) \leq F(x_1) + \varepsilon.$$

Hence, we have

$$\begin{aligned} F(x_2) &\leq f(y_1) + d(y_1, x_2) \leq F(x_1) + \varepsilon + d(y_1, x_2) - d(y_1, x_1) \\ &\leq F(x_1) + d(x_1, x_2) + \varepsilon. \end{aligned}$$

Since this is true for any $\varepsilon > 0$, we have $F(x_2) \leq F(x_1) + d(x_1, x_2)$. Similarly, we have $F(x_2) \geq F(x_1) - d(x_1, x_2)$. □

Now, let us state and prove a result for concentration on the sphere of $(\mathbb{R}^n, \|\cdot\|_1)$.

Theorem 22. *(Schechtman-Zinn) $B_1^n := \{x \in \mathbb{R}^n : \|x\|_1 = \sum_{i=1}^n |x_i| \leq 1\}$. λ is the normalized surface area measure on ∂B_1^n. If $f : \partial B_1^n \to \mathbb{R}$ satisfies $|f(x) - f(y)| \leq \|x - y\|_2$, then,*

$$\lambda(|f - \mathbb{E}_\lambda f| \geq t) \leq Ce^{-Cnt}.$$

Proof. $f : \partial B_1^n \to \mathbb{R}$ is 1-Lipschitz w.r.t. L_2 norm. So by Lemma 15, we can extend f to a 1-Lipschitz function F on $\mathbb{R}^n$. Next, consider $\mathbb{R}^n$ with exponential measure

$$d\mu(x) = \frac{1}{2^n} e^{-\sum_{i=1}^n |x_i|} dx, \quad S = \sum_{i=1}^n |x_i|.$$

Then, the random variable $(x_1, \ldots, x_n)/S$ is uniformly distributed on ∂B_1^n and the random variable $(x_1, \ldots, x_n)/S$ and S are independent. Applying the Tensorization Corollary above, we complete our proof. $\square$

Theorem 23. *(Gross' Theorem) Let γ_n be the standard Gaussian measure on $\mathbb{R}^n$, then, $\forall f : \mathbb{R}^n \to \mathbb{R}$, C^1, we have*

$$Ent_{\gamma_n}(f^2) \leq 2 \int_{\mathbb{R}^n} \|\nabla f\| - 2^2 d\gamma_n.$$

Using Herbst's Theorem, this implies concentration inequality for the Gaussian measure on $\mathbb{R}^n$.

We need Two Point log-Sobolev Inequality to prove Gross' Theorem.

Lemma 16. *(Two Point log-Sobolev Inequality) $\Omega = \{0, 1\}$, $\mu(0) = \mu(1) = 1/2$. For $f : \{0, 1\} \to \mathbb{R}$, $|Df| = |f(1) - f(0)|$. We have*

$$Ent_\mu(f^2) \leq \frac{1}{2} \mathbb{E}_\mu |Df|^2.$$

Proof. W.l.o.g., we assume $\mathbb{E}_\mu(f^2) = 1$ and let $f(1) = a_1$, $f(0) = a_0$, then, $a_1^2 + a_0^2 = 2$. We want to prove that

$$\frac{1}{2}(a_1^2 \log a_1^2 + a_0^2 \log a_0^2) \leq \frac{1}{2}(a_1 - a_0)^2 = 1 - a_1 a_0.$$

Let $\lambda = a_0^2$, w.l.o.g., $0 \leq \lambda \leq 1$. Then, we want to prove that

$$\frac{1}{2}(\lambda \log \lambda + (2 - \lambda) \log(2 - \lambda)) \leq 1 - \sqrt{\lambda(2 - \lambda)}.$$

Hence, we want to show that

$$\psi(\lambda) := \lambda \log \lambda + (2-\lambda)\log(2-\lambda) + 2\sqrt{\lambda(2-\lambda)} \le 2 \quad \text{on } [0,1].$$

Now, $\psi(0) = 2\log 2$ and $\psi(1) = 2$.

$$\psi'(\lambda) = \log \frac{\lambda}{2-\lambda} + \sqrt{\frac{2-\lambda}{\lambda}} - \sqrt{\frac{\lambda}{2-\lambda}}.$$

Now, let $\mu = \lambda/(2-\lambda)$. It suffices to show that

$$g(\mu) := \log \mu + \frac{1}{\sqrt{\mu}} - \sqrt{\mu} > 0.$$

We know that $g(0) = \infty$, $g(1) = 0$ and

$$g'(\mu) = \frac{1}{\mu} - \frac{1}{2\mu^{3/2}} - \frac{1}{2\sqrt{\mu}} < 0.$$

Hence, we get our desired result. □

Proof of Gross' Theorem. The tensorization result for entropy tells us that it is enough to prove Gross' Theorem for the case $n = 1$. Indeed, 2 is the optimal constant in Gross' Theorem. Let $\gamma_1 = \gamma$ be the standard Gaussian measure on $\mathbb{R}$. We aim to prove that

$$\mathrm{Ent}_\gamma(f^2) \le 2 \int_{\mathbb{R}} (f')^2 d\gamma.$$

By Two Point log-Sobolev Inequality, we have $\mathrm{Ent}_\mu(f) \le \frac{1}{2}\mathbb{E}_\mu |Df|^2$. Now, for $f : \{0,1\}^n \to \mathbb{R}$, by tensorization for entropy, we get

$$\mathrm{Ent}_\mu(f^2) \le \frac{1}{2}\mathbb{E}_\mu \left(\sum_{i=1}^n |D_i f|^2 \right),$$

where $|D_i f| = |f(x_1, \ldots, x_{i-1}, 1, x_i, \ldots, x_n) - f(x_1, \ldots, x_{i-1}, 0, x_{i+1}, \ldots, x_n)|$. Now, for $\phi : \mathbb{R} \to \mathbb{R}$ smooth and compactly supported. Let

$$f(x_1, \ldots, x_n) = \phi\left(\frac{x_1 + \cdots + x_n - n/2}{\sqrt{n/4}} \right).$$

Then, applying Mean Value Theorem, we get

$$
\begin{aligned}
\mathrm{Ent}_\mu(f^2) &\leq \frac{1}{2}\mathbb{E}_\mu\left(\sum_{i=1}^n |D_i f|^2\right) \\
&= \frac{1}{2}\mathbb{E}_\mu \sum_{i=1}^n \left(\phi'\left(\frac{x_1 + \cdots x_{i-1} + y_i + x_{i+1} + \cdots + x_n - n/2}{\sqrt{n/4}}\right)\sqrt{\frac{4}{n}}\right)^2 \\
&= \frac{2}{n}\mathbb{E}_\mu \sum_{i=1}^n \left(\phi'\left(\frac{x_1 + \cdots + x_n - n/2}{\sqrt{n/4}}\right) + o(1)\right)^2 \\
&= 2\mathbb{E}_\mu \frac{1}{n}\sum_{i=1}^n \phi'\left(\frac{x_1 + \cdots + x_n - n/2}{\sqrt{n/4}}\right)^2 + o(1) \to 2\mathbb{E}_\gamma|\phi'|^2 \text{ as } n \to \infty.
\end{aligned}
$$

Finally, $\mathrm{Ent}_\mu(f^2) \to \mathrm{Ent}_\gamma(\phi^2)$ as $n \to \infty$ and hence we complete the proof. $\square$

Theorem 24. *(Bobkov-Ledoux-Maurey-Talagrand) Let ν be the symmetric exponential measure on $\mathbb{R}$, i.e., with density $\frac{1}{2}e^{-|t|}$. $\nu_n = \nu^{\otimes n}$ is the exponential measure on $\mathbb{R}^n$. $f : \mathbb{R}^n \to \mathbb{R}$ such that*

$$\forall x, y \in \mathbb{R}^n, \quad |f(x) - f(y)| \leq \beta\|x - y\|_1, \quad |f(x) - f(y)| \leq \alpha\|x - y\|_2.$$

Then, we have

$$\nu_n(|f - \mathbb{E}f| \geq r) \leq Ce^{-C\min\{\frac{r}{\beta}, \frac{r^2}{\alpha^2}\}}.$$

Lemma 17. *$\forall 0 \leq c < 1$, if $f : \mathbb{R} \to \mathbb{R}$ is C^1 such that $|f'| \leq c < 1$, then, we have*

$$Ent_\nu(e^f) \leq \frac{2}{1-c}\mathbb{E}_\nu((f')^2 e^f).$$

Proof. W.l.o.g., assume $f(0) = 0$. Then, we have

$$
\begin{aligned}
&\mathrm{Ent}_\nu(e^f) = \mathbb{E}_\nu(fe^f) - (\mathbb{E}_\nu e^f)\log(\mathbb{E}_\nu e^f) \\
&\leq \mathbb{E}_\nu(fe^f) - \mathbb{E}_\nu(e^f - 1) = \mathbb{E}_\nu(e^f f - e^f + 1),
\end{aligned}
$$

where I used the fact that $u - 1 \leq u \log u$ for all $u > 0$. Now, for $\phi : \mathbb{R} \to \mathbb{R}$, C^1, with bounded derivative, then, we have

$$\mathbb{E}_\nu(\phi) = \phi(0) + \int_{\mathbb{R}} \mathrm{sgn}(t)\phi'(t)d\nu(t).$$

$$
\begin{aligned}
&\mathrm{Ent}_\nu(e^f) \le \mathbb{E}_\nu(fe^f - e^f + 1) \\
&= \int_{\mathbb{R}} \mathrm{sgn}(t) \left(f(t) f'(t) e^{f(t)} + f'(t) e^{f(t)} - f'(t) e^{f(t)} \right) d\nu(t) \\
&= \int_{\mathbb{R}} \mathrm{sgn}(t) f(t) f'(t) e^{f(t)} d\nu(t) \\
&\le \left(\int_{\mathbb{R}} f^2 e^f d\nu \right)^{1/2} \left(\int_{\mathbb{R}} (f')^2 e^f d\nu \right)^{1/2}.
\end{aligned}
$$

Also, notice that

$$
\begin{aligned}
\int_{\mathbb{R}} f^2 e^f d\nu &= \int_{\mathbb{R}} \mathrm{sgn}(t) \left(2f(t) f'(t) e^{f(t)} + f(t)^2 f'(t) e^{f(t)} \right) d\nu(t) \\
&\le 2 \int_{\mathbb{R}} f f' e^f + c \int_{\mathbb{R}} f^2 e^f d\nu.
\end{aligned}
$$

Hence, we have

$$
\begin{aligned}
\int_{\mathbb{R}} f^2 e^f d\nu &\le \frac{2}{1-c} \int_{\mathbb{R}} f f' e^f d\nu \\
&\le \frac{2}{1-c} \left(\int_{\mathbb{R}} f^2 e^f d\nu \right)^{1/2} \left(\int_{\mathbb{R}} (f')^2 e^f d\nu \right)^{1/2},
\end{aligned}
$$

which implies that

$$
\int_{\mathbb{R}} f^2 e^f d\nu \le \left(\frac{2}{1-c} \right)^2 \int_{\mathbb{R}} (f')^2 e^f d\nu.
$$

Therefore, we conclude that

$$
\mathrm{Ent}_\nu(e^f) \le \frac{2}{1-c} \mathbb{E}_\nu((f')^2 e^f).
$$

□

Lemma 18. *(Tensorization) $f : \mathbb{R}^n \to \mathbb{R}$ is smooth and $\|\nabla f\|_\infty \le 1$. Then for any $|\lambda| < 1$, we have*

$$
Ent_{\nu^n}(e^{\lambda f}) \le \frac{2\lambda^2}{1-\lambda} \int_{\mathbb{R}^n} \|\nabla f\|_2^2 e^{\lambda f} d\nu_n.
$$

Proof. By the tensorization result for entropy (Proposition 4) and the previous Lemma, we get

$$\begin{aligned}
\mathrm{Ent}_{\nu^n}(e^{\lambda f}) &\leq \mathbb{E}_{\nu^n} \sum_{i=1}^{n} \mathrm{Ent}_{\nu}(e^{\lambda f_i}) \\
&\leq \frac{2}{1-\lambda} \mathbb{E}_{\nu^n} \sum_{i=1}^{n} \mathbb{E}_{\nu} \left[(\lambda f_i')^2 e^{\lambda f_i} \right] = \frac{2\lambda^2}{1-\lambda} \mathbb{E}_{\nu^n} \left[\sum_{i=1}^{n} \left(\frac{\partial f}{\partial x_i} \right)^2 e^{\lambda f} \right].
\end{aligned}$$

□

Proof of Theorem 24. $f : \mathbb{R}^n \to \mathbb{R}$. W.l.o.g., we can assume that f is smooth and compactly supported and

$$\|\nabla f\|_2 \leq \alpha, \quad \max_{1 \leq i \leq n} \left| \frac{\partial f}{\partial x_i} \right| \leq \beta.$$

Assume first that $\beta = 1$ and $\|\nabla f\|_\infty$. By Tensorization Lemma, if $|\lambda| \leq 1/2$, then,

$$\begin{aligned}
\mathrm{Ent}_{\nu^n}(e^{\lambda f}) &\leq 4\lambda^4 \int_{\mathbb{R}^n} \|\nabla f\|_2^2 e^{\lambda f} d\nu_n \\
&\leq 4\lambda^2 \alpha^2 \int_{\mathbb{R}^n} e^{\lambda f} d\nu_n.
\end{aligned}$$

Now, we do the similar arguments as in the proof of Herbst's Theorem. Let $h(\lambda) = \int_{\mathbb{R}^n} e^{\lambda f} d\nu_n$ and $K(\lambda) = \frac{1}{\lambda} \log h(\lambda)$. $K(0) = \mathbb{E}_{\nu^n} f$.

$$K'(\lambda) = \frac{-1}{\lambda^2} \log h(\lambda) + \frac{h'(\lambda)}{\lambda h(\lambda)} = \frac{1}{\lambda^2 h(\lambda)} \mathrm{Ent}_{\nu^n}(e^{\lambda f}).$$

Hence, we have

$$\lambda^2 h(\lambda) \geq \frac{1}{4\alpha^2} \mathrm{Ent}_{\nu^n}(e^{\lambda f}).$$

$\lambda > 0$, $K'(\lambda) \leq 4\alpha^2$, $|\lambda| \leq 1/2$, $K(\lambda) - K(0) \leq 4\alpha^2 \lambda$, $\lambda(K(\lambda) - \mathbb{E}f) \leq 4\alpha^2 \lambda^2$. Thus, $\mathbb{E}e^{\lambda(f - \mathbb{E}f)} \leq e^{4\alpha^2 \lambda^2}$. Hence, for any $0 < \lambda < 1/2$, applying Markov's Inequality, we get

$$\nu_n(f - \mathbb{E}f > r) \leq 3^{4\alpha^4 \lambda^2 - \lambda r},$$

with global minimum of $4\alpha^2 \lambda^2 - \lambda r$ attained at $\lambda = r/(8\alpha^2)$.

(Case 1). If $\frac{r}{8\alpha^2} \leq \frac{1}{2}$, then take $\lambda = \frac{r}{8\alpha^2}$, then,

$$\nu_n(f - \mathbb{E}f > r) \leq e^{-\frac{r^2}{16\alpha^2}}.$$

(Case 2). If $\frac{r}{8\alpha^2} > \frac{1}{2}$, then take $\lambda = \frac{1}{2}$, then

$$\nu_n(f - \mathbb{E}f > r) \leq e^{\alpha^2 - \frac{r}{2}} \leq e^{-\frac{r}{4}}.$$

Hence, we have

$$\nu_n(|f - \mathbb{E}f| > r) \leq 2e^{-\frac{1}{4}\min\{r, \frac{r^2}{4\alpha^2}\}}.$$

Now, we rescale back to the original case by applying to f/β, we get

$$\nu_n(|f - \mathbb{E}f| > r) = \nu_n\left(\left|\frac{f}{\beta} - \mathbb{E}\frac{f}{\beta}\right| \geq \frac{r}{\beta}\right) \leq 2e^{-\frac{1}{4}\min\{\frac{r}{\beta}, \frac{r^2}{4\alpha^2}\}}.$$

□

At the end of our notes, let us discuss Stein's Method for concentration inequalities.

Definition 12. X, X' are *exchangeable* if $(X, X') \sim (X', X)$.

We assume X and X' to be exchangeable and F to be skew-symmetric, i.e.,

$$F(X, X') = -F(X', X).$$

We also define

(i) $\mathbb{E}(F(X, X')|X) = f(X)$.

(ii) $V(X) = \frac{1}{2}\mathbb{E}(|f(X) - f(X')| \cdot |F(X, X')||X)$.

We will see that bounds on $V(X)$ will imply bounds on $\mathbb{P}(|f(X)| > t)$.

Lemma 19.

$$\mathbb{E}(h(X)f(X)) = \frac{1}{2}\mathbb{E}((h(X) - h(X'))F(X, X')).$$

Proof. Since we have

$$\mathbb{E}(h(X)F(X, X')) = \mathbb{E}[\mathbb{E}(h(X)F(X, X')|X)] = \mathbb{E}(fh),$$

therefore, we get

$$\mathbb{E}(h(X)F(X, X')) = \mathbb{E}(h(X')F(X', X)) = -\mathbb{E}(h(X')F(X, X')) = \mathbb{E}(fh).$$

□

Remark 13. $\mathbb{E}f^2 \leq \mathbb{E}(V(X))$.

Theorem 25. *Assume, for any θ, $\mathbb{E}[e^{\theta f(X)}F(X,X')] < \infty$ and $|V(X)| < C$ a.s. for some constant $C > 0$. Then, we have*

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

Proof. Define $m(\theta) = \mathbb{E}[e^{\theta f(X)}]$. Then, by the Lemma above, we get

$$m'(\theta) = \mathbb{E}[e^{\theta f(X)}f(X)] = \frac{1}{2}\mathbb{E}[(e^{\theta f(X)} - e^{\theta f(X')})F(X,X')].$$

Notice the fact that $|(e^x - e^y)/(x-y)| \leq \frac{1}{2}(e^x + e^y)$, since

$$\left|\frac{e^x - e^y}{x - y}\right| = \int_0^1 e^{tx+(1-t)y}dt \leq \int_0^1 te^x + (1-t)e^y dt = \frac{1}{2}(e^x + e^y).$$

Then, we have

$$\begin{aligned}
m'(\theta) &\leq \frac{1}{2}\mathbb{E}\left(\frac{|e^{\theta f(X)} - e^{\theta f(X')}|}{|\theta f(X) - \theta f(X')|} \cdot |\theta f(X) - \theta f(X')| \cdot |F(X,X')|\right) \\
&\leq \frac{|\theta|}{4}\mathbb{E}\left|(e^{\theta f(X)} + e^{\theta f(X')})(f(X) - f(X'))F(X,X')\right| \\
&\leq \frac{|\theta|}{4}\left[2\mathbb{E}((e^{\theta f(X)}V(X)) + 2\mathbb{E}((e^{\theta f(X')}V(X'))\right] \\
&= |\theta|\mathbb{E}(e^{\theta f(X)}V(X)) \leq C|\theta|\mathbb{E}(e^{\theta f(X)}) = C|\theta|m(\theta).
\end{aligned}$$

Therefore, $m(\theta) \leq e^{C\theta^2/2}$. Hence,

$$\mathbb{P}(|f(X)| \geq t) \leq 2e^{-\theta t + C\theta^2/2}$$

Now, minimizing the RHS in θ by taking $\theta = t/C$, we get our desired result. □

Theorem 26. *(Chatterjee) Suppose for any θ, $\mathbb{E}[e^{\theta f(X)}F(X,X')] < \infty$ and $|V(X)| \leq Bf(X) + C$ a.s. Then, we have*

$$\mathbb{P}(|f(X)| \geq t) \leq 2e^{-\frac{t^2}{2C+2Bt}}.$$