Square root law

We assume that $X_1, \cdots, X_n \overset{iid}{\sim} F$, $\mathbb{E} X_i = \mu$ exists. The centre limit theorem implies that $\sqrt{n}(\overline{X}_n - \mu)$ convergents to a Gaussian element in distribution, i.e.

$$\mathbb{P}(\sqrt{n}(\overline{X}_n - \mu) \geq t) \to 1 - \Phi(t),$$

where Φ is the law of the Gaussian variable ($\sigma^2 = \mathrm{Var}\, X_i < \infty$)

$\Rightarrow$

$$\mathbb{P}(\sqrt{n}|\overline{X}_n - \mu| \geq t) \to 1 - \Phi(t) + \Phi(-t)$$

$\Rightarrow$

$$\overline{X}_n - \mu = O_p\left(\frac{1}{\sqrt{n}}\right)$$

(or we could use:

$$\mathbb{P}(|\overline{X}_n - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2}$$

$\Rightarrow$

$$\mathbb{P}(\sqrt{n}|\overline{X}_n - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}$$)

If X is complicated (e.g. High dimensional), the rate of concentration would be smaller and even diverging. These features are hard to be captured by the classical techniques, e.g. Markov inequality, characteristic function.

Chernoff inequality

Notation and assumption:

① $X_1, \cdots, X_n$ are independent.

② $\mathbb{E} X_i = \mu_i$ exists.

③ $\psi_{X_i}(\lambda) = \log \mathbb{E}\, e^{\lambda(X_i - \mu_i)}$

We have:

$$\mathbb{P}(X - \mu > t) \le \frac{\mathbb{E}\, e^{\lambda(X-\mu)}}{e^{\lambda t}}, \quad \lambda \ge 0,\ t \ge 0$$

$\Rightarrow$

$$\mathbb{P}(X-\mu > t) \le e^{\inf_{\lambda \ge 0}(\psi_X(\lambda) - \lambda t)},$$

which implies that we could bound $\psi(\lambda)$ to achieve the upper bound of tail probability.

Example

If $X \sim \mathcal{N}(\mu, \sigma^2)$, we have

$$\psi_X(\lambda) = \frac{\sigma^2 \lambda^2}{2}$$

$$\Rightarrow \mathbb{P}(X-\mu \geq t) \leq e^{\inf_{\lambda \geq 0}\left(\frac{\sigma^2 \lambda^2}{2} - \lambda t\right)}$$

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

where σ^2 could be viewed as the dispersion parameter.

Example

We call X the Rademacher variable if $\mathbb{P}(X=1) = \mathbb{P}(X=-1) = \frac{1}{2}$. Note that

$$
\begin{aligned}
\mathbb{E} e^{\lambda X} &= (e^{\lambda} + e^{-\lambda})/2 \\
&= \frac{1}{2}\left(\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} + \sum_{k=0}^{\infty} \frac{(-\lambda)^k}{k!}\right) \\
&= \sum_{k=0}^{\infty} \frac{\lambda^{2k}}{(2k)!} \le \sum_{k=0}^{\infty} \frac{(\lambda^2)^k}{2^k k!} \\
&= e^{\lambda^2/2}
\end{aligned}
$$

$$
\Rightarrow \quad \psi_X(\lambda) \le e^{\lambda^2/2}
$$

$$
\begin{aligned}
\Rightarrow \quad \psi_{\sum_{i=0}^{n} \alpha_i X_i}(\lambda) &= \mathbb{E} e^{\lambda \sum_{i=0}^{n} \alpha_i X_i} \\
&= \prod_{i=0}^{n} \mathbb{E} e^{\lambda \alpha_i X_i} \\
&\le \prod_{i=0}^{n} e^{\alpha_i^2 \lambda^2/2} \\
&= e^{\lambda^2 \|\alpha\|_2^2/2}
\end{aligned}
$$

$$
\Rightarrow \quad \mathbb{P}\left(\sum_{i=0}^{n} \alpha_i X_i > t\right) \le e^{-\frac{t^2}{2\|\alpha\|_2^2}}.
$$

Remark:

$$
\mathbb{P}\left(\left|\sum_{i=0}^{n} \alpha_i X_i\right| > t\right) \le 2e^{-t^2/2\|\alpha\|_2^2}.
$$

Example

If $|Y_i| \le M_i$ and $\mathbb{E} Y_i = 0$,

$\mathbb{E} e^{\lambda Y_i} = \mathbb{E} e^{\lambda(Y_i - \mathbb{E} Y_i')}$, where Y' is the independent copy of Y, and

$$
\begin{aligned}
&\mathbb{E} e^{\lambda(Y_i - \mathbb{E} Y_i')} \\
\overset{\text{Jesen}}{\le}\ & \mathbb{E}_{Y_i} \mathbb{E}_{Y_i'} e^{\lambda(Y_i - Y_i')} \\
=\ & \mathbb{E}_{Y_i} \mathbb{E}_{Y_i'} \mathbb{E}_X e^{\lambda X (Y_i - Y_i')},
\end{aligned}
$$

where X is the Rademacher variable and $X \perp (Y_i, Y_i')$, and the last "=" is achieved since:

$$Y_i - Y_i' =_d Y_i' - Y_i.$$

Then

$$
\begin{aligned}
\mathbb{E}_X e^{\lambda X (Y_i - Y_i')} &\le e^{\lambda^2 (Y_i - Y_i')^2 / 2} \\
&\le e^{\lambda^2 (2M_i)^2 / 2},
\end{aligned}
$$

$$\Rightarrow \mathbb{P}\left(\left|\sum_{i=1}^{n} Y_i\right| > t\right) \le 2e^{-\frac{t^2}{2\sum_{i=1}^{n}(2M_i)^2}}.$$

The above procedure shows that the dispersion parameter of variable valued in the interval with length L is L. We could use another approach to get a tighter dispersion.

For $X \in [a, b]$, note that

$$\psi_X(0) = 0,$$

$$\psi_X'(0) = \frac{d \log \mathbb{E} e^{\lambda(X-\mu)}}{d\lambda} \Big|_{\lambda=0}$$

$$= \frac{\mathbb{E}(X-\mu)e^{\lambda(X-\mu)}}{\mathbb{E} e^{\lambda(X-\mu)}} \Big|_{\lambda=0}$$

$$= 0,$$

$$\psi_X''(\lambda) = \frac{\mathbb{E}(X-\mu)^2 e^{\lambda(X-\mu)}}{\mathbb{E} e^{\lambda(X-\mu)}} - \frac{(\mathbb{E}(X-\mu)e^{\lambda(X-\mu)})^2}{(\mathbb{E} e^{\lambda(X-\mu)})^2}$$

$= \operatorname{Var} Y_\lambda$, where

$$p_{Y_\lambda}(y) = \frac{e^{\lambda(y-\mu)}}{\mathbb{E} e^{\lambda(X-\mu)}} \cdot p_X(y), \quad Y_\lambda \in [a, b]$$

Notice that

$$\operatorname{Var} Y_\lambda = \mathbb{E}(Y_\lambda - \mathbb{E} Y_\lambda)^2$$
$$\leq \mathbb{E}\left(Y_\lambda - \frac{a+b}{2}\right)^2$$
$$\leq (b-a)^2/4,$$

By Taylor expansion,

$$\psi_X(\lambda) = \frac{\lambda^2}{2} \psi''(\lambda_0)$$
$$= \frac{\lambda^2}{2} \operatorname{Var} Y_{\lambda_0}$$
$$\leq \frac{\lambda^2}{2} \frac{(b-a)^2}{4},$$

which indicates the dispersion parameter of X is at most $\frac{b-a}{2}$.

Sub-gaussian property

Random variable is sub-gaussian if $\exists\, \sigma > 0$ s.t.

$$\psi_X(\lambda) \le \frac{\sigma^2 \lambda^2}{2}, \quad \forall \lambda \in \mathbb{R}$$

where σ is called the sub-gaussian parameter.

Hoefding Inequality

$X_1, \dots, X_n$ are independent with sub-gaussian parameter σ_i, then

$$\mathbb{P}\left(\left|\sum_{i=1}^{n} (X_i - \mu_i)\right| \ge t\right) \le 2e^{-\frac{t^2}{2\sum_{i=1}^{n} \sigma_i^2}},$$

for $t \ge 0$.

Now, let $f: \mathbb{R}^n \to \mathbb{R}$, we introduce two typical method to bound

$$\mathbb{P}(|f(X^{(n)}) - \mathbb{E}f(X^{(n)})| \geq t),$$

where $X^{(n)} = (X_1, \cdots, X_n)$

(Bound differences inequality)

Suppose that

$$|f(\cdots, x_k, \cdots) - f(\cdots, x_k', \cdots)| \leq L_k,$$

then

$$\mathbb{P}(|f(X^{(n)}) - \mathbb{E}f(X^{(n)})| \geq t) \leq 2e^{-\frac{2t^2}{\sum_{k=1}^n L_k^2}}$$

Pf: Let

$$D_k = \mathbb{E}(f(X^{(n)}) | X^{(k)}) - \mathbb{E}(f(X^{(n)}) | X^{(k-1)})$$

$\Rightarrow$

$$f(X^{(n)}) - \mathbb{E}f(X^{(n)}) = \sum_{k=1}^n D_k$$

Let $\mathcal{F}_k = \sigma(X_1, \cdots, X_k)$,

Then

$$\mathbb{E}\, D_k \mid \mathcal{F}_{k-1} = 0$$

and

$$\mathbb{E}\, e^{\lambda \sum_{k=1}^{n} D_k}$$
$$= \mathbb{E}\; \mathbb{E}\, e^{\lambda \sum_{k=1}^{n} D_k} \mid \mathcal{F}_{n-1}$$
$$= \mathbb{E}\, e^{\lambda \sum_{k=1}^{n-1} D_k} (\mathbb{E}\; e^{\lambda D_n} \mid \mathcal{F}_{n-1}).$$

It's sufficent to show

$$\mathbb{E}\, e^{\lambda D_k} \mid \mathcal{F}_{k-1} \leq e^{\frac{\lambda^2}{2}\left(\frac{L_k}{2}\right)^2},$$

let

$$D_k(x) \mid \mathcal{F}_{k-1}$$
$$= \mathbb{E}\; f(X^{(n)}) \mid X^{(k)}) - \mathbb{E} f(X^{(n)}) \mid (X^{(k-1)}, x)$$
$$A_k = \sup_x\; D_k(x) \mid \mathcal{F}_{k-1},$$
$$B_k = \inf_x\; D_k(x) \mid \mathcal{F}_{k-1}$$

$$\Rightarrow \; A_k - B_k$$
$$= \sup_x \mathbb{E} f(X^{(n)}) \mid (X^{(k-1)}, x) - \inf_x \mathbb{E} f(X^{(n)}) \mid (X^{(k-1)}, x)$$
$$\leq \sup_{x,y} \mathbb{E}_{X_{k+1,\dots,n}} \mid f(X^{(k-1)}, x, X_{k+1}, \cdots, X_n) -$$

$$f(X^{(k-1)}, y, X_{k+1}, \cdots, X_n)|$$
$$\leq L_k. \qquad \square$$

Example

Let $g: \mathbb{R}^2 \to \mathbb{R}$, $|g| \leq b$

$$U(X_1, \cdots X_n) = \frac{1}{C_n^2} \sum_{j<k} g(X_j, X_k),$$

then (U statistics)

$$|U(X_1, \cdots, x, \cdots, X_n) - U(X_1, \cdots, y, \cdots, X_n)|$$
$$\leq \frac{1}{C_n^2} (n-1) \cdot 2b = \frac{4b}{n}, \text{ then}$$

$$\mathbb{P}(|U - \mathbb{E}U| \geq t) \leq 2e^{-\frac{nt^2}{8b^2}}.$$

Sub-exponential tail bound

A random variable X is sub-exponential if $\exists\,(\sigma, \alpha)$ s.t.

$$\psi_X(\lambda) \le \frac{\lambda^2\sigma^2}{2}, \text{ for}$$

$$|\lambda| \le \frac{1}{\alpha}.$$

Example

If $Z \sim N(0,1)$, $X = Z^2$

$$\mathbb{E}\, e^{\lambda(X-1)}$$

$$= \int \frac{1}{\sqrt{2\pi}} e^{\lambda(Z^2-1)} e^{-Z^2/2}\, dz$$

$$= e^{-\lambda} \int \frac{1}{\sqrt{2\pi}} e^{-(\frac{1}{2}-\lambda)z^2}\, dz$$

$$= e^{-\lambda} / \sqrt{1-2\lambda}, \quad \lambda \le \frac{1}{2}$$

$$= \frac{e^{-\lambda}}{e^{-\lambda-2\lambda^2+o(\lambda^2)}}$$

(use Taylor to $\log\sqrt{1-2\lambda}$)

$$\leq e^{2\lambda^2} = e^{4\lambda^2/2}, |\lambda| < \frac{1}{4}.$$

then X is sub-exponential with parameter $(2, 4)$

Theorem

$$\mathbb{P}(X-\mu \geq t) \leq \begin{cases} e^{-\frac{t^2}{2\sigma^2}}, & 0 \leq t \leq \frac{\sigma^2}{\alpha} \\ e^{-\frac{t}{2\alpha}}, & t > \sigma^2/\alpha \end{cases}$$

Pf: $\psi_X(\lambda) \leq \frac{\lambda^2\sigma^2}{2}$

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

$$\forall |\lambda| \leq \frac{1}{\alpha},$$

the optimal λ is: t/σ^2,

if $t \leq \sigma^2/\alpha \Rightarrow t/\sigma^2 \leq \frac{1}{\alpha}$

hence

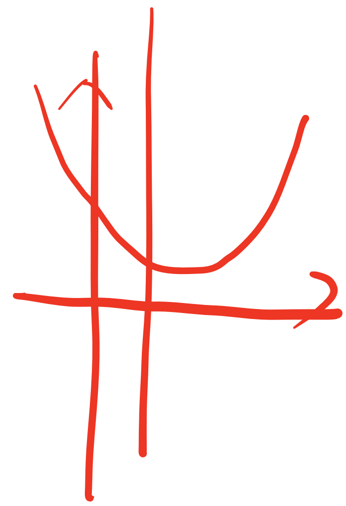
$$\mathbb{P}(X-\mu \geq t) \leq e^{-\frac{t^2}{2\sigma^2}}$$

If $t \geq \sigma^2/\alpha$,

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

$$\leq e^{\frac{t}{2\alpha}-\frac{t}{\alpha}}$$

$$\leq e^{-\frac{t}{2\alpha}} \qquad \square$$



Example

Let $X_1, \dots, X_n \sim \chi^2(1)$, then $\left(\sum_{k=1}^{n} X_i - n\right)$ is sub-exponential with parameters $(\sqrt{4n}, 4)$

$\Rightarrow$

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{k=1}^{n} X_i - 1\right| \geq t\right) \leq 2e^{-\frac{n^2t^2}{8n}} = 2e^{-\frac{nt^2}{8}},$$

if $0 < tn < \frac{4n}{4} \iff 0 < t < 1$

Example Randon projection

Consider $\{u_1, \cdots, u_n\}$ and a linear map:

$$F: u \mapsto Xu/\sqrt{m},$$

where $u \in \mathbb{R}^d$ and $X \in \mathbb{R}^{m\times d}$, $m \leq d$. Moreover, let $X = \begin{pmatrix} x_1^T \\ \vdots \\ x_m^T \end{pmatrix}$, which elements $\sim N(0,1)$, then

$$\frac{\|F(\Delta u)\|_2^2}{\|\Delta u\|_2^2} = \frac{1}{m}\sum_{i=1}^{m}\left\langle x_i, \frac{\Delta u}{\|\Delta u\|_2}\right\rangle^2$$

$$\Rightarrow \mathbb{P}\left(\left|\frac{\|F(\Delta u)\|_2^2}{\|\Delta u\|_2^2} - 1\right| \geq \varepsilon\right) \leq 2e^{-\frac{m\varepsilon^2}{8}}$$

$$\Rightarrow \mathbb{P}\left(\left|\frac{\|F(u_i - u_j)\|_2^2}{\|u_i - u_j\|_2^2} - 1\right| \geq \varepsilon, \forall i \neq j \leq N\right)$$

$$\leq 2C_N^2 e^{-\frac{m\varepsilon^2}{8}} \leq N^2 e^{-\frac{m\varepsilon^2}{8}},$$

which implied the projected dimension m should not be too small.

Example Bernstein-type bound

If we assume a moment constraint:

$$\left| \mathbb{E}(X-\mu)^k \right| \le \frac{1}{2} k!\, \sigma^2 b^{k-2},$$

where $\sigma^2 = \operatorname{Var} X$, then

$$\mathbb{E} e^{\lambda(X-\mu)}$$

$$= 1 + \frac{\lambda^2\sigma^2}{2} + \sum_{k\ge 3} \frac{\lambda^k \mathbb{E}(X-\mu)^k}{k!}$$

$$\le 1 + \lambda^2\sigma^2/2 \left(1 + \sum_{k\ge 3} (|\lambda| b)^{k-2}\right)$$

$$= 1 + \frac{\lambda^2\sigma^2}{2} \frac{1}{1-|\lambda| b} \le e^{\frac{\lambda^2\sigma^2}{2}\frac{1}{1-|\lambda| b}},$$

$|\lambda| < \frac{1}{b}$.

$(\lambda > 0)$

$$\Rightarrow \mathbb{P}(|X-\mu| \ge t) \le 2e^{\frac{\lambda^2\sigma^2}{2}\frac{1}{1-\lambda b} - \lambda t},$$

Let $\frac{t(1-\lambda b)}{\sigma^2} = \lambda$

$$\Rightarrow \lambda = \frac{t}{bt+\sigma^2} < \frac{1}{b}$$

$$\Rightarrow \mathbb{P}(|X-\mu| \ge t) \le 2e^{-\frac{t^2}{2(\sigma^2+bt)}}.$$

A geometric perspective on concentration

Notation

① $(\mathcal{X}, \rho)$ is a general metrix space.

② The law of $X \in \mathcal{X}$: P.

③ $f: \mathcal{X} \longmapsto \mathbb{R}$, s.t.

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

④ $m_f = \text{median}(f(X))$

We want to give a tail bound to $\mathbb{P}(|f(X)-m_f| \geq \varepsilon)$. Claim that

$$\{f(X) \leq m_f + \varepsilon\} \supset A^{\frac{\varepsilon}{L}},$$

where $A := \{f(X) \leq m_f\}$ and

$$A^{\frac{\varepsilon}{L}} = A + \bar{B}(0; \tfrac{\varepsilon}{L}).$$

Pf: $\forall x \in A^{\frac{\varepsilon}{2}}$, $\exists y \in A$ s.t.

$$\rho(x,y) \leq \frac{\varepsilon}{2}$$

$$\Rightarrow \quad |f(x)-f(y)| \leq \varepsilon$$

$$\Rightarrow \quad f(x) \leq \varepsilon + m_f \qquad \square$$

Then

$$\mathbb{P}(f(x)-m_f \geq \varepsilon)$$

$$\leq 1-\mathbb{P}(A^{\frac{\varepsilon}{2}}),$$ where

$\mathbb{P}(A)=\frac{1}{2}$. Let

$$\alpha_p(\varepsilon) = \sup_{\{A;\mathbb{P}(A)\geq\frac{1}{2}\}} (1-\mathbb{P}(A^{\varepsilon}))$$

then

$$\mathbb{P}(f(x)-m_f \geq \varepsilon) \leq \alpha_p(\tfrac{\varepsilon}{2})$$

Similarly,

$$\mathbb{P}(m_f - f(x) \geq \varepsilon) \leq \alpha_p(\tfrac{\varepsilon}{2})$$

Then

$$\mathbb{P}(|f(X)-m_f| \geq \varepsilon) \leq 2\alpha_p(\frac{\varepsilon}{L})$$

Example 5 (Sphere with radius 1)

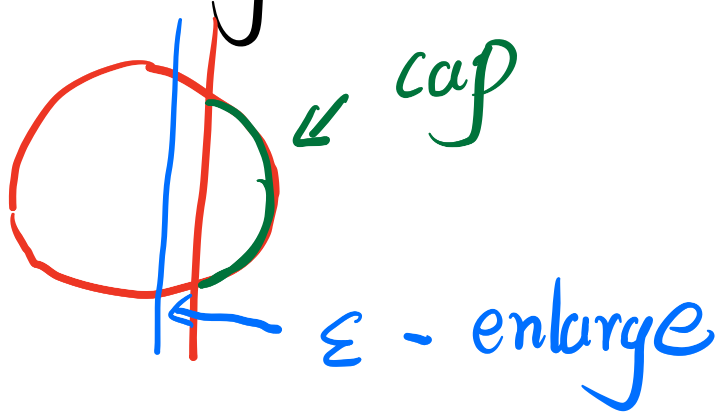
If $X \sim \mathrm{Unif}(S^{n-1})$, m_X is the median of X. $\rho(x,y) = \|x-y\|_2$.

(Isoperimetric inequality on the sphere)

For $\forall t>0$, $\forall A \subset S^{n-1}$ ($P = \mathrm{Unif}(S^{n-1})$) s.t.

$$P(A) = \alpha,$$

the spherical caps minimize $P(A^\varepsilon)$.



Under this inequality, consider the hemisphere:

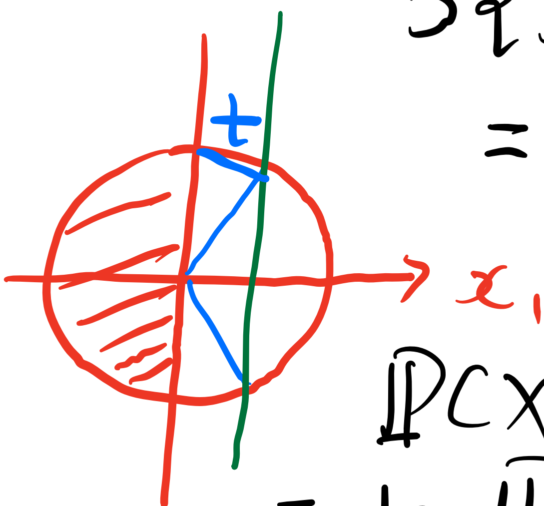
$$H := \{ x \in S^{n-1} ;\ x_1 \le 0 \},$$

we have $P(H) = \frac{1}{2}$, hence,

$$\alpha_P(t) \le 1 - P(H^t),$$

where

$$H^t = \{ x \in S^{n-1} ;\ \rho(y, H) \le t \}$$
$$= \{ x \in S^{n-1} ;\ x_1 \le t\sqrt{1 - \tfrac{t^2}{4}} \}$$
$$\supset \{ x \in S^{n-1} ;\ x_1 \le \tfrac{t}{\sqrt{2}} \}$$
$$=: \tilde{H}_t, \text{ then}$$



$$\mathbb{P}(X \in H^t) \ge \mathbb{P}(X \in \tilde{H}^t)$$
$$= 1 - \mathbb{P}(x_1 \ge \tfrac{t}{\sqrt{2}})$$

Now, let $Z \sim N(0, I_n)$, note that

$$\frac{Z}{\|Z\|_2} \sim \mathrm{Unif}(S^{n-1}),$$

then

$$\mathbb{P}(X_1 \geq \varepsilon), \quad \varepsilon \leq 1$$

$$= \mathbb{P}(Z_1 \geq \|Z\|_2 \varepsilon)$$

$$\leq \mathbb{P}(Z_1 \geq \|Z\|_2 \varepsilon \mid \|Z\|_2 \geq C\sqrt{n}) +$$

$$\mathbb{P}(\|Z\|_2 < C\sqrt{n}) \quad (C<1)$$

$$\leq \mathbb{P}(Z_1 \geq C\sqrt{n}\varepsilon) + \mathbb{P}\left(1 - \frac{\|Z\|_2^2}{n} > 1 - C^2\right)$$

$$\leq e^{-\frac{C^2 n \varepsilon^2}{8}} + e^{-\frac{n(1-C^2)^2}{8}},$$

we chose C s.t.

$$1 \leq (1-C^2)^2 / C^2$$

$$\Rightarrow e^{-C^2 n \varepsilon^2/8} \geq e^{-n(1-C^2)^2/8}$$

$$\Rightarrow \mathbb{P}(X_1 \geq \varepsilon) \leq 2e^{-cn\varepsilon^2}$$

$$\Rightarrow \mathbb{P}\left(X_1 \geq \frac{t}{\sqrt{2}}\right) \leq 2e^{-cnt^2}$$

$$\Rightarrow \mathbb{P}(|f(X) - m_f| > t) \lesssim e^{-\frac{cnt^2}{L^2}}$$

1

Lemma

If $\mathbb{P}(|X-m_X|\geq t)\leq c_1 e^{-c_2 t^2}$,

then

$$\mathbb{P}(|X-\mathbb{E}X|\geq t)\leq c_3 e^{-\frac{c_2}{4}t^2}$$

Pf: Note that

$$(\mathbb{E}X-m_X)^2\leq \mathbb{E}|X-m_X|^2$$

$$=\int_0^{+\infty}\mathbb{P}(|X-m_X|\geq t)dt$$

$$\leq c_1/c_2$$

and $\{|X-\mathbb{E}X|\geq t\}$

$$\Rightarrow \{|X-m_X|\geq t-\Delta\},$$

where $\Delta=|m_X-\mathbb{E}X|$, if

① $t\geq 2\sqrt{c_1/c_2}$

$$\mathbb{P}(|X-\mathbb{E}X|\geq t)\leq c_1 e^{-\frac{c_2}{4}t^2}$$

② $t<2\sqrt{c_1/c_2}$,

$$e^{-c_2t^2/4}\geq e^{-c_1},$$

chose a c_3 s.t. $c_3e^{-c_1}>1$. $\square$

Remark: This Lemma implies that

$$\mathbb{P}(|f(X)-\mathbb{E}f(X)|>t) \lesssim e^{-\frac{cnt}{4L^2}},$$

based on ①.