

Concentration Inequalities

Jiuzhou Miao

March 5, 2025

Proof of Proposition 1. We first prove (1). One has that the moment generating function of X is

$$\mathbb{E}(e^{\lambda X}) = e^{\lambda\mu + \lambda^2\sigma^2/2},$$

implies that the moment generating function of $X - \mu$ is $\varphi(\lambda) = e^{\lambda^2\sigma^2/2}$. Then

$$\inf_{\lambda \in [0, b]} \{-\lambda t + \log \varphi(\lambda)\} = \inf_{\lambda \in [0, b]} \{-\lambda t + \lambda^2\sigma^2/2\} = -\frac{t^2}{2\sigma^2}.$$

By Chernoff's inequality, one has that for all $t > 0$,

$$\log \mathbb{P}(X - \mu \geq t) \leq \inf_{\lambda \in [0, b]} \{-\lambda t + \log \varphi(\lambda)\} = -\frac{t^2}{2\sigma^2}$$

and

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

(1) holds.

Note that

$$\mathbb{P}(|X - \mu| \geq t) = \mathbb{P}(X - \mu \geq t) + \mathbb{P}(X - \mu \leq -t),$$

then (2) holds by symmetry. \square

Proofs of Proposition 5. I will provide the proofs for both cases that $\sigma = b - a$ and $\sigma = (b - a)/2$. We firstly prove that X is sub-Gaussian with parameter $\sigma = b - a$. Let X' be an independent copy of X and ε be a Rademacher random variable independent with X and X' . In the following, the notation \mathbb{E}_\star means that take expectation with respect to the random variables placed in the position of \star . Then one has that

$$\begin{aligned} \mathbb{E}(e^{\lambda X}) &= \mathbb{E}_X \left[e^{\lambda \{X - \mathbb{E}_{X'}(X')\}} \right] \\ &= \mathbb{E}_X \left[e^{\mathbb{E}_{X'} \{ \lambda(X - X') \}} \right] \\ &\leq \mathbb{E}_{X, X'} \{ e^{\lambda(X - X')} \} \\ &= \mathbb{E}_{X, X'} \left[\mathbb{E}_\varepsilon \{ e^{\lambda \varepsilon(X - X')} \} \right] \\ &\leq \mathbb{E}_{X, X'} \{ e^{\lambda^2(X - X')^2/2} \} \\ &\leq e^{\lambda^2(b-a)^2/2}, \end{aligned}$$

where the first " \leq " by using Jensen's inequality, the second " \leq " by using ε is sub-Gaussian with parameter $\sigma = 1$, and the last " \leq " by using $|X - X'| \leq b - a$.

Then we prove that X is sub-Gaussian with parameter $\sigma = (b - a)/2$. Denote $\psi(\lambda) = \log \mathbb{E}(e^{\lambda X})$. One can easily obtain that

$$\begin{aligned}\psi^{(1)}(\lambda) &= \frac{\mathbb{E}(X e^{\lambda X})}{\mathbb{E}(e^{\lambda X})}, \\ \psi^{(2)}(\lambda) &= \frac{\mathbb{E}(X^2 e^{\lambda X}) \mathbb{E}(e^{\lambda X}) - \{\mathbb{E}(X e^{\lambda X})\}^2}{\{\mathbb{E}(e^{\lambda X})\}^2} \\ &= \frac{\mathbb{E}(X^2 e^{\lambda X})}{\mathbb{E}(e^{\lambda X})} - \{\psi^{(1)}(\lambda)\}^2,\end{aligned}$$

and $\psi(0) = \psi^{(1)}(0) = 0$. Define another probability measure \mathbb{Q} which satisfies that

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

then one has that

$$\begin{aligned}\psi^{(1)}(\lambda) &= \int X \frac{e^{\lambda X}}{\mathbb{E}(e^{\lambda X})} d\mathbb{P} = \int X d\mathbb{Q} = \mathbb{E}_{X \sim \mathbb{Q}}(X), \\ \psi^{(2)}(\lambda) &= \int X^2 \frac{e^{\lambda X}}{\mathbb{E}(e^{\lambda X})} d\mathbb{P} - \{\psi^{(1)}(\lambda)\}^2 = \int X^2 d\mathbb{Q} - \{\psi^{(1)}(\lambda)\}^2 = \text{Var}_{X \sim \mathbb{Q}}(X),\end{aligned}$$

and

$$\sup_{\lambda \in \mathbb{R}} |\psi^{(2)}(\lambda)| \leq (b - a)^2/4.$$

Hence

$$\mathbb{E}(e^{\lambda X}) = e^{\psi(\lambda)} = \exp \left\{ \int_0^\lambda \int_0^\mu \psi^{(2)}(t) dt d\mu \right\} \leq \exp \left\{ 2^{-1} \lambda^2 \sup_{\lambda \in \mathbb{R}} |\psi^{(2)}(\lambda)| \right\} \leq e^{\lambda^2 (b-a)^2/8}.$$

□

Proof of Proposition 7. By using the sub-Exponential properties of X_i , one has that

$$\mathbb{E}\{e^{\lambda \sum_{i=1}^n (X_i - \mu_i)}\} = \prod_{i=1}^n \mathbb{E}\{e^{\lambda (X_i - \mu_i)}\} \leq e^{\frac{\lambda^2 \sum_{i=1}^n \nu_i^2}{2}} = e^{\frac{\nu_*^2 \lambda^2}{2}},$$

for $|\lambda| < \min_{1 \leq i \leq n} \alpha_i^{-1} = \alpha_*^{-1}$. Then by Chernoff's inequality, one has that

$$\begin{aligned}\log \mathbb{P}\left\{\sum_{i=1}^n (X_i - \mu_i) \geq t\right\} &\leq \inf_{\lambda \in [0, \alpha_*^{-1}]} \{-\lambda t + 2^{-1} \nu_*^2 \lambda^2\} \\ &\leq \begin{cases} -\frac{t^2}{2\nu_*^2} & , 0 < t \leq \frac{\nu_*^2}{\alpha_*} \\ -\frac{t}{2\alpha_*} & , t > \frac{\nu_*^2}{\alpha_*} \end{cases}.\end{aligned}$$

Then the two-sided inequality can be obtained by symmetry. □

Proof of Theorem 8. For any $\mathbf{u} = (u_1, \dots, u_d)^\top \in \mathbb{R}^d$, we construct F by $F(\mathbf{u}) = m^{-1/2} \mathbf{X} \mathbf{u}$, where $\mathbf{X} = (X_{jk})_{m \times d}$ with $\{X_{jk}\}_{j,k=1}^{m,d}$ are independent $\mathcal{N}(0, 1)$ random variables. Then one has that for any $\mathbf{u} \neq \mathbf{0}$,

$$\frac{\|F(\mathbf{u})\|_2^2}{\|\mathbf{u}\|_2^2} = \frac{\|\mathbf{X} \mathbf{u}\|_2^2}{m \|\mathbf{u}\|_2^2} = m^{-1} \sum_{j=1}^m \underbrace{\frac{\left(\sum_{k=1}^d X_{jk} u_k\right)^2}{\sum_{k=1}^d u_k^2}}_{\text{Denote by } Y_j}$$

and $\{Y_j\}_{j=1}^m$ are independent Chi-square random variables with 1 degree of freedom. By using Proposition 7,

$$\mathbb{P}\left\{\left|\frac{\|F(\mathbf{u})\|_2^2}{\|\mathbf{u}\|_2^2} - 1\right| \geq \delta\right\} = \mathbb{P}\left\{\left|\sum_{j=1}^m (Y_j - 1)\right| \geq m\delta\right\} \leq 2e^{-\frac{m\delta^2}{8}}.$$

For $\mathbf{u}_i \neq \mathbf{u}_j$, one has that

$$\frac{\|F(\mathbf{u}_i) - F(\mathbf{u}_j)\|_2^2}{\|\mathbf{u}_i - \mathbf{u}_j\|_2^2} = \frac{\|\mathbf{X}(\mathbf{u}_i - \mathbf{u}_j)\|_2^2}{m\|\mathbf{u}_i - \mathbf{u}_j\|_2^2} = \frac{\|F(\mathbf{u}_i - \mathbf{u}_j)\|_2^2}{\|\mathbf{u}_i - \mathbf{u}_j\|_2^2}.$$

Hence, one has that the probability of the opposite event of

$$(1 - \delta)\|\mathbf{u}_i - \mathbf{u}_j\|_2^2 \leq \|F(\mathbf{u}_i) - F(\mathbf{u}_j)\|_2^2 \leq (1 + \delta)\|\mathbf{u}_i - \mathbf{u}_j\|_2^2,$$

for all pairs $\mathbf{u}_i \neq \mathbf{u}_j$ can be bounded by

$$2\binom{N}{2}e^{-\frac{m\delta^2}{8}} \leq N^2e^{-\frac{m\delta^2}{8}}.$$

□

Proof of Theorem 10. By using Taylor's expansion, one has that

$$\begin{aligned} \left|\mathbb{E}\{e^{\lambda(X_i - \mu_i)}\} - 1 - 2^{-1}\lambda^2\sigma_i^2\right| &\leq \sum_{k=3}^{\infty} \frac{|\lambda|^k}{k!} \left|\mathbb{E}\{(X_i - \mu_i)^k\}\right| \\ &\leq \sum_{k=3}^{\infty} \frac{|\lambda|^k}{k!} 2^{-1}k!\sigma_i^2b^{k-2} \\ &= 2^{-1}\lambda^2\sigma_i^2 \sum_{k=3}^{\infty} (|\lambda|b)^{k-2} \\ &= \frac{2^{-1}\lambda^2\sigma_i^2|\lambda|b}{1 - |\lambda|b} \end{aligned}$$

for $|\lambda| < 1/b$, then one has that

$$\mathbb{E}\{e^{\lambda(X_i - \mu_i)}\} \leq 1 + 2^{-1}\lambda^2\sigma_i^2 + \frac{2^{-1}\lambda^2\sigma_i^2|\lambda|b}{1 - |\lambda|b} = 1 + \frac{2^{-1}\lambda^2\sigma_i^2}{1 - |\lambda|b} \leq \exp\left\{\frac{2^{-1}\lambda^2\sigma_i^2}{1 - |\lambda|b}\right\}$$

and

$$\mathbb{E}\{e^{\lambda\sum_{i=1}^n(X_i - \mu_i)}\} \leq \exp\left\{\frac{2^{-1}\lambda^2\sum_{i=1}^n\sigma_i^2}{1 - |\lambda|b}\right\},$$

(1) holds.

By using Chernoff's inequality, one has that for all $t > 0$,

$$\mathbb{P}\left\{\sum_{i=1}^n (X_i - \mu_i) \geq t\right\} \leq e^{-\lambda t} \mathbb{E}\{e^{\lambda\sum_{i=1}^n(X_i - \mu_i)}\} \leq \exp\left\{-\lambda t + \frac{2^{-1}\lambda^2\sum_{i=1}^n\sigma_i^2}{1 - |\lambda|b}\right\}.$$

By taking

$$\lambda = \frac{t}{bt + \sum_{i=1}^n\sigma_i^2} \in [0, b^{-1}),$$

one has that

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

Then (2) can be obtained by symmetry. \square

Proof of Theorem 14. We only prove (1). By tower property, one has that for $|\lambda| < \alpha_n^{-1}$

$$\begin{aligned}\mathbb{E}(e^{\lambda \sum_{k=1}^n D_k}) &= \mathbb{E}\{\mathbb{E}(e^{\lambda \sum_{k=1}^n D_k} | \mathcal{A}_{n-1})\} \\ &= \mathbb{E}\{e^{\lambda \sum_{k=1}^{n-1} D_k} \mathbb{E}(e^{\lambda D_n} | \mathcal{A}_{n-1})\} \\ &\leq e^{\nu_n^2 \lambda^2 / 2} \mathbb{E}(e^{\lambda \sum_{k=1}^{n-1} D_k}).\end{aligned}$$

Iterating this procedure yields the bound

$$\mathbb{E}(e^{\lambda \sum_{k=1}^n D_k}) \leq e^{\nu_*^2 \lambda^2 / 2}$$

for all $|\lambda| < \alpha_*^{-1}$. \square

Proof of Theorem 15. We only need to prove that $\mathbb{E}(e^{\lambda D_k} | \mathcal{A}_{k-1}) \leq e^{\lambda^2 (b_k - a_k)^2 / 8}$ almost surely. This argument is similar to the proof of Proposition 5. \square

Proof of Theorem 17. Let

$$D_k = \mathbb{E}\{f(\mathbf{X}) | X_1, \dots, X_{k-1}, X_k\} - \mathbb{E}\{f(\mathbf{X}) | X_1, \dots, X_{k-1}\},$$

and

$$\begin{aligned}A_k &= \inf_x \mathbb{E}\{f(\mathbf{X}) | X_1, \dots, X_{k-1}, X_k = x\} - \mathbb{E}\{f(\mathbf{X}) | X_1, \dots, X_{k-1}\}, \\ B_k &= \sup_x \mathbb{E}\{f(\mathbf{X}) | X_1, \dots, X_{k-1}, X_k = x\} - \mathbb{E}\{f(\mathbf{X}) | X_1, \dots, X_{k-1}\}.\end{aligned}$$

Note that

$$\begin{aligned}D_k - A_k &= \mathbb{E}\{f(\mathbf{X}) | X_1, \dots, X_{k-1}, X_k\} - \inf_x \mathbb{E}\{f(\mathbf{X}) | X_1, \dots, X_{k-1}, X_k = x\} \geq 0, \\ B_k - D_k &= \sup_x \mathbb{E}\{f(\mathbf{X}) | X_1, \dots, X_{k-1}, X_k = x\} - \mathbb{E}\{f(\mathbf{X}) | X_1, \dots, X_{k-1}, X_k\} \geq 0,\end{aligned}$$

Then one has that $D_k \geq A_k$ and $D_k \leq B_k$ almost surely. Then we will show that $B_k - A_k \leq L_k$ almost surely. By using the independence of $\sigma(X_{k+1}^n)$ and $\sigma(X_1^k)$, one has that

$$\begin{aligned}B_k - A_k &= \sup_x \mathbb{E}\{f(\mathbf{X}) | X_1, \dots, X_{k-1}, X_k = x\} - \inf_x \mathbb{E}\{f(\mathbf{X}) | X_1, \dots, X_{k-1}, X_k = x\} \\ &= \sup_{x, x'} \left| \mathbb{E}\{f(\mathbf{X}) | X_1, \dots, X_{k-1}, X_k = x\} - \mathbb{E}\{f(\mathbf{X}) | X_1, \dots, X_{k-1}, X_k = x'\} \right| \\ &= \sup_{x, x'} \left| \mathbb{E}_{X_{k+1}^n} \{f(X_1^{k-1}, x, X_{k+1}^n)\} - \mathbb{E}_{X_{k+1}^n} \{f(X_1^{k-1}, x', X_{k+1}^n)\} \right| \\ &\leq L_k\end{aligned}$$

almost surely. Then the inequality can be obtained by using Theorem 15. \square

Lemma 1. Suppose that $f : \mathbb{R}_n \rightarrow \mathbb{R}$ is differentiable. Then for any convex function $\phi : \mathbb{R} \rightarrow \mathbb{R}$, one has that

$$\mathbb{E} \left\{ \phi \left(f(\mathbf{X}) - \mathbb{E}\{f(\mathbf{X})\} \right) \right\} \leq \mathbb{E} \left\{ \phi \left(\frac{\pi}{2} \langle \nabla f(\mathbf{X}), \mathbf{Y} \rangle \right) \right\},$$

where \mathbf{X} and \mathbf{Y} are independent $\mathcal{N}(\mathbf{0}, \mathbf{I}_n)$ random vectors.

Proof. By symmetry and Jensen's inequality, one has that

$$\begin{aligned} \mathbb{E} \left\{ \phi \left(f(\mathbf{X}) - \mathbb{E}\{f(\mathbf{X})\} \right) \right\} &= \mathbb{E}_{\mathbf{X}} \left\{ \phi \left(f(\mathbf{X}) - \mathbb{E}_{\mathbf{Y}}\{f(\mathbf{Y})\} \right) \right\} \\ &= \mathbb{E}_{\mathbf{X}} \left\{ \phi \left(\mathbb{E}_{\mathbf{Y}}\{f(\mathbf{X}) - f(\mathbf{Y})\} \right) \right\} \\ &\leq \mathbb{E}_{\mathbf{X}, \mathbf{Y}} \left\{ \phi(f(\mathbf{X}) - f(\mathbf{Y})) \right\} \\ &= \mathbb{E} \left\{ \phi(f(\mathbf{X}) - f(\mathbf{Y})) \right\}. \end{aligned}$$

Let $\mathbf{Z}(\theta) = \{Z_1(\theta), \dots, Z_n(\theta)\}^\top$, where $Z_k(\theta) = X_k \sin \theta + Y_k \cos \theta$. One has that

$$\begin{aligned} f(\mathbf{X}) - f(\mathbf{Y}) &= f(\mathbf{Z}(\pi/2)) - f(\mathbf{Z}(0)) \\ &= \int_0^{\pi/2} \frac{df(\mathbf{Z}(\theta))}{d\theta} d\theta \\ &= \int_0^{\pi/2} \sum_{k=1}^n \frac{\partial f(\mathbf{Z}(\theta))}{\partial x_k} \frac{dZ_k(\theta)}{d\theta} d\theta \\ &= \int_0^{\pi/2} \langle \nabla f(\mathbf{Z}(\theta)), \nabla \mathbf{Z}(\theta) \rangle d\theta \\ &= \frac{2}{\pi} \int_0^{\pi/2} \frac{\pi}{2} \langle \nabla f(\mathbf{Z}(\theta)), \nabla \mathbf{Z}(\theta) \rangle d\theta, \end{aligned}$$

and

$$\begin{aligned} \mathbb{E} \left\{ \phi \left(f(\mathbf{X}) - \mathbb{E}\{f(\mathbf{X})\} \right) \right\} &\leq \mathbb{E} \left\{ \phi(f(\mathbf{X}) - f(\mathbf{Y})) \right\} \\ &= \mathbb{E} \left\{ \phi \left(\frac{2}{\pi} \int_0^{\pi/2} \frac{\pi}{2} \langle \nabla f(\mathbf{Z}(\theta)), \nabla \mathbf{Z}(\theta) \rangle d\theta \right) \right\} \\ &\leq \frac{2}{\pi} \int_0^{\pi/2} \mathbb{E} \left\{ \phi \left(\frac{\pi}{2} \langle \nabla f(\mathbf{Z}(\theta)), \nabla \mathbf{Z}(\theta) \rangle \right) \right\} d\theta. \end{aligned}$$

Note that for $\theta \in [0, \pi/2]$, $(Z_k(\theta), Z_k^{(1)}(\theta))^\top$ is jointly Gaussian with mean vector $(0, 0)^\top$ and covariance matrix \mathbf{I}_2 , then one has that

$$\begin{aligned} \frac{2}{\pi} \int_0^{\pi/2} \mathbb{E} \left\{ \phi \left(\frac{\pi}{2} \langle \nabla f(\mathbf{Z}(\theta)), \nabla \mathbf{Z}(\theta) \rangle \right) \right\} d\theta &= \frac{2}{\pi} \int_0^{\pi/2} \mathbb{E} \left\{ \phi \left(\frac{\pi}{2} \langle \nabla f(\tilde{\mathbf{X}}), \tilde{\mathbf{Y}} \rangle \right) \right\} d\theta \\ &= \mathbb{E} \left\{ \phi \left(\frac{\pi}{2} \langle \nabla f(\tilde{\mathbf{X}}), \tilde{\mathbf{Y}} \rangle \right) \right\}, \end{aligned}$$

where $\tilde{\mathbf{X}}$ and $\tilde{\mathbf{Y}}$ are independent $\mathcal{N}(\mathbf{0}, \mathbf{I}_n)$ random vectors. □

Proof of Theorem 19. In Lemma 1, take $\phi(t) = e^{\lambda t}$. One has that

$$\begin{aligned}
\mathbb{E} \left[\exp \left\{ \lambda \left[f(\mathbf{X}) - \mathbb{E}\{f(\mathbf{X})\} \right] \right\} \right] &\leq \mathbb{E}_{\mathbf{X}, \mathbf{Y}} \left[\exp \left\{ \frac{\pi \lambda}{2} \langle \nabla f(\mathbf{X}), \mathbf{Y} \rangle \right\} \right] \\
&\leq \mathbb{E}_{\mathbf{X}} \left[\exp \left\{ \frac{\lambda^2 \pi^2}{8} \|\nabla f(\mathbf{X})\|_2^2 \right\} \right] \\
&\leq \exp \{ \lambda^2 \pi^2 L^2 / 8 \}.
\end{aligned}$$

□