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

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

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

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

and

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

(1) holds.

Note that

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

then (2) holds by symmetry. □

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

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

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

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

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

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

then one has that

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

and

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

Hence

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

□

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

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

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

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

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

Proof of Theorem 8. For any $\boldsymbol{u} = (u_1,\ldots,u_d)^\top \in \mathbb{R}^d$, we construct F by $F(\boldsymbol{u}) = m^{-1/2}\mathbf{X}\boldsymbol{u}$, 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 $\boldsymbol{u} \ne \mathbf{0}$,

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

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

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

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

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

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

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

for all pairs $\boldsymbol{u}_i\neq\boldsymbol{u}_j$ 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

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

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

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

and

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

(1) holds.

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

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

By taking

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

one has that

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

Then (2) can be obtained by symmetry. □

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

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

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 $|\lambda| < \alpha_*^{-1}$. □

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. □

Proof of Theorem 17. Let

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

and

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

Note that

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

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

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

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

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

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

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

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

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

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

and

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

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

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

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

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

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

□