High-Dimensional Statistics A Non-Asymptotic Viewpoint (2019) by Martin J. Wainwright

Exercises solns by Chen Du (chendu2017@sjtu.edu.cn, chen1443@e.ntu.edu.sg)

January 11, 2023

1 Chapter 2

Exercise 2.1.

(a) X=0, w.p. 1

(b) X=0, w.p. 1

□

Exercise 2.2.

(a) obvious

(b) We start from the intermediate term:

$$\begin{aligned}
\mathbb{P}[Z \geq z] &= \int_z^\infty \phi(t)dt \\
&= \int_z^\infty -\frac{\phi'(t)}{t}dt && \text{by part (a)} \\
&= \frac{\phi(z)}{z} - \frac{\phi(z)}{z^3} + \int_z^\infty 3\frac{\phi(t)}{t^4}dt && (\geq LHS, \text{by integration by parts}) \\
&= \frac{\phi(z)}{z} - \frac{\phi(z)}{z^3} + \int_z^\infty -3\frac{\phi'(t)}{t^5}dt \\
&= \frac{\phi(z)}{z} - \frac{\phi(z)}{z^3} + 3\frac{\phi(z)}{z^5} - \int_z^\infty 5\frac{\phi(t)}{t^6}dt && (\leq RHS, \text{integration by parts})
\end{aligned}$$

□

Exercise 2.3. (Polynimial Markov sharper than Chernoff)
For a given $r > 0$,

$$\begin{aligned}
\mathbb{E}\left[e^{rX}\right] = \mathbb{E}\left[\sum_k \frac{(rX)^k}{k!}\right] = \sum_k \frac{(r\delta)^k}{k!}\mathbb{E}\left[\frac{X^k}{\delta^k}\right] \geq \sum_k \frac{(r\delta)^k}{k!}\left(\inf_k \mathbb{E}\left[\frac{X^k}{\delta^k}\right]\right) = e^{r\delta}\left(\inf_k \mathbb{E}\left[\frac{X^k}{\delta^k}\right]\right) \\
\iff \frac{\mathbb{E}\left[e^{rX}\right]}{e^{r\delta}} \geq \inf_k \mathbb{E}\left[\frac{X^k}{\delta^k}\right], \forall r \geq 0 \\
\iff \inf_{r>0} \frac{\mathbb{E}\left[e^{rX}\right]}{e^{r\delta}} \geq \inf_k \mathbb{E}\left[\frac{X^k}{\delta^k}\right]
\end{aligned}$$

□

Exercise 2.4. (Hoeffding Lemma)

(a) By $\frac{d^k\psi(r)}{dr^k} = \mathbb{E}\left[X^k e^{rX}\right]$, we have $\psi(0) = 0$ and $\psi'(0) = \mathbb{E}[X] = \mu$.

(b) Note that

$$
\begin{aligned}
\psi'(r) &= \frac{g'_X(r)}{g_X(r)}, \\
\psi''(r) &= \frac{g''_X(r)}{g_X(r)} - \left(\frac{g'_X(r)}{g_X(r)}\right)^2 \\
&= \frac{\int x^2 e^{rx} d\mathbb{P}}{g_X(r)} - \left(\frac{\int x e^{rx} d\mathbb{P}}{g_X(r)}\right)^2 \\
&= \int x^2 d\mathbb{Q} - \left(\int x d\mathbb{Q}\right)^2 && \text{(by defining } d\mathbb{Q} = \frac{e^{rx}}{g_X(r)} d\mathbb{P}\text{)} \\
&= \mathsf{Var}_{\mathbb{Q}}[X] \\
&\le \left(\frac{b-a}{2}\right)^2
\end{aligned}
$$

Easy to check that the newly defined measure $d\mathbb{Q}$ whose Radon-Nikodym derivative with respect to $d\mathbb{P}$ is $\frac{d\mathbb{Q}}{d\mathbb{P}} = \frac{e^{rx}}{g_X(r)}$ is indeed a probability measure as $\int d\mathbb{Q} = \int \frac{e^{rx}}{g_X(r)} d\mathbb{P} = 1$. The last inequality follows from the fact that for a bounded univariate random variable, assigning half probability mass to each end point achieves the maximum variance, which provides an upper bound on $\sup_{r\in\mathbb{R}} |\psi''(r)|$.

(c) We show that X is Sub-Gaussian by its definition,

$$
\begin{aligned}
\mathbb{E}\left[e^{r(X-\mu)}\right] &= \exp\left(\psi(r) - r\mu\right) \\
&= \exp\left(\psi(0) + \psi'(0) r + \frac{1}{2}\psi''(\epsilon) r^2 - r\mu\right) && \text{(by Taylor expansion, where } \epsilon \in (0, r)\text{)} \\
&\le \exp\left(\frac{r^2 (\frac{b-a}{2})^2}{2}\right).
\end{aligned}
$$

The last inequality confirms that X belong to Sub-Gaussian class with parameter $\sigma = \frac{b-a}{2}$.

□

Exercise 2.5.

(a) By $\mathbb{E}[1 + rX] \le \mathbb{E}\left[e^{rX}\right] \le e^{\frac{r^2\sigma^2}{2} + r\mu}, \forall r \in \mathbb{R}$, we have

$$
\mathbb{E}[X] \le \frac{e^{\frac{r^2\sigma^2}{2} + r\mu} - 1}{r}, \forall r > 0.
$$

Taking limit on the right-hand side, and letting $r \searrow 0$, we have $\mathbb{E}[X] \le \mu$. Since for $-X$, the above analysis holds as well by modifying the r in the numerator to $-r$, we then have $\mathbb{E}[-X] \le -\mu$. Combining two inequalities, we obtain $\mathbb{E}[X] = \mu$.

(b) Note that $e^{\frac{r^2\sigma^2}{2} + r\mu} \ge \mathbb{E}\left[e^{rX}\right] = \mathbb{E}\left[\sum_{k=0}^{\infty} \frac{(rX)^k}{k!}\right] \ge \mathbb{E}\left[1 + rX + \frac{1}{2}(rX)^2\right]$ holds for all $r > 0$ if the third-moment of X is positive and holds for all $r < 0$ otherwise. WLOG, we assume $\mathbb{E}\left[X^3\right] \ge 0$, then we know that

$$
\mathbb{E}\left[X^2\right] \le 2\frac{e^{\frac{r^2\sigma^2}{2} + r\mu} - 1 - r\mu}{r^2}, \forall r > 0
$$

Let $r \searrow 0$, we have $\mathbb{E}\left[X^2\right] \le \mu^2 + \sigma^2$, which implies $\mathsf{Var}[X] \le \sigma^2$.

(c) We provide a counterexample to disprove the statement. Consider a Bernoulli distribution with success probability $p = \frac{1}{4}$, whose moment generating function is $g(r) = 1 - p + pe^r$ and its variance is $\mathsf{Var}[X] = 0.1875$. To ensure the inequality $g(r) \leq e^{\frac{r^2\sigma^2}{2}+r\mu}$ holds for all $r \in \mathbb{R}$, the parameter σ^2 is required to be bigger than $\sup_{r\in\mathbb{R}} 2\frac{\psi(r)-rp}{r^2}$. However when $r = 1$, $2\frac{\psi(r)-rp}{r^2} \approx 0.2147$ is larger than $\mathsf{Var}[X]$. Therefore, the parameter σ^2 should be larger than $\mathsf{Var}[X]$. [Conjecture: for an uni-variate random variable that is symmetric around its center, the parameter σ^2 is indeed its variance]

□

Exercise 2.6.

Since $X_i^2 \geq 0, a.s.$, we can use one-sided Bernstein's inequality (2.23) to derive the bound:

$$\mathbb{P}\left[Z - \mathbb{E}\left[Z\right] \leq -\sigma^2\delta\right] \leq \exp\left(-\frac{n\delta^2\sigma^4}{\frac{2}{n}\sum_{i=1}^n \mathbb{E}\left[X_i^4\right]}\right).$$

Then, to obtain the desired bound, it suffices to show that $\mathbb{E}\left[X_i^4\right] \leq 8\sigma^4$:

$$\begin{aligned}
\mathbb{E}\left[X_i^4\right] &= \int_0^\infty \mathbb{P}\left[X_i^4 \geq t\right] dt \\
&= \int_0^\infty \mathbb{P}\left[X_i^4 \geq u^4\right] du^4 && (\text{let } t = u^4) \\
&\leq 4\int_0^\infty u^3 \cdot 2e^{-\frac{u^2}{2\sigma^2}} du && (\text{by Sub-Gaussian's concentration}) \\
&= 4 \cdot (2\sigma^2)^2 \int_0^\infty e^{-\frac{u^2}{2\sigma^2}} \left(\frac{u^2}{2\sigma^2}\right)^{2-1} d\frac{u^2}{2\sigma^2} \\
&= 16\sigma^4\Gamma(2) \\
&= 16\sigma^4 && (\text{worse by a constant 2})
\end{aligned}$$

□

Exercise 2.7. (Bennett's inequality)

(a) Since $|X| \leq b,\ a.s.$, we know that

$$\mathbb{E}\left[\sum_{k=2}^\infty (\lambda X)^k\right] = \mathbb{E}\left[(\lambda X)^2 \sum_{k=2}^\infty (\lambda X)^{k-2}\right] \leq \mathbb{E}\left[(\lambda X)^2\right] \sum_{k=2}^\infty (\lambda b)^{k-2},\ \forall\lambda \geq 0$$

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

$$\begin{aligned}
\ln\left(\mathbb{E}\left[e^{\lambda X}\right]\right) &= \ln\left(\mathbb{E}\left[1 + \lambda X + \sum_{k=2}^\infty \frac{(\lambda X)^k}{k!}\right]\right) && \forall\lambda > 0 \\
&= \ln\left(1 + \mathbb{E}\left[\sum_{k=2}^\infty \frac{(\lambda X)^k}{k!}\right]\right) \\
&\leq \mathbb{E}\left[\sum_{k=2}^\infty \frac{(\lambda X)^k}{k!}\right] \\
&\leq \mathbb{E}\left[(\lambda X)^2\right] \sum_{k=2}^\infty \frac{(\lambda b)^{k-2}}{k!} \\
&= \lambda^2\sigma^2 \frac{e^{\lambda b} - 1 - \lambda b}{\lambda^2 b^2}
\end{aligned}$$

Now, we consider another random variable denoted as $Y = -X$, the above analysis still holds for all $\lambda \geq 0$. This implies that, for random variable X with any negative λ, the desired inequality holds, which completes the proof.

(b) We prove the inequality after taking logarithm on both sides:

$$\begin{aligned}
\ln\left(\mathbb{P}\left[\sum_i^n X_i \geq n\delta\right]\right) &\leq -n\lambda\delta + \sum_{i=1}^n \psi_{X_i}(\lambda) \\
&\leq -n\lambda\delta + \sum_{i=1}^n \sigma_i^2\lambda^2\left(\frac{e^{\lambda b}-1-\lambda b}{\lambda^2 b^2}\right) && \text{by part (a)} \\
&= -n\lambda\delta + n\sigma^2\lambda^2\left(\frac{e^{\lambda b}-1-\lambda b}{\lambda^2 b^2}\right) \\
&= -\frac{n\sigma^2}{\delta^2}\left[\frac{b\delta}{\sigma^2}\lambda b - e^{\lambda b} + 1 + \lambda b\right] \\
&= -\frac{n\sigma^2}{b^2} h\left(\frac{b\delta}{\sigma^2}\right) && \text{by setting } \lambda : \lambda b = \ln\left(1+\frac{b\delta}{\sigma^2}\right)
\end{aligned}$$

(c) We compare the inequalities for $\sum_i^n X_i$:

- **Eq.(2.22b) One-sided Bernstein's inequality**: $\mathbb{P}\left[\sum_i^n (X_i-\mu_i) \geq n\delta\right] \leq \exp\left(-\frac{n\delta^2}{2(\frac{1}{n}\sum_i^n \sigma_i^2 + \frac{b\delta}{3})}\right)$
- **Eq.(2.62) Bennett concentration inequality**: $\mathbb{P}\left[\sum_i^n X_i \geq n\delta\right] \leq \exp\left(-\frac{n\sigma^2}{b^2} h\left(\frac{b\delta}{\sigma^2}\right)\right)$

  It suffices to show that

$$\begin{aligned}
&\frac{n\sigma^2}{b^2} h\left(\frac{b\delta}{\sigma^2}\right) \geq \frac{n\delta^2}{2(\sigma^2+\frac{b\delta}{3})} \\
\Longleftrightarrow &\frac{n\sigma^2}{b^2} h\left(\frac{b\delta}{\sigma^2}\right) \geq \frac{n\sigma^2}{b^2}\frac{\left(\frac{b\delta}{\sigma^2}\right)^2}{2(1+\frac{b\delta}{3\sigma^2})} \\
\Longleftrightarrow &h\left(\frac{b\delta}{\sigma^2}\right) \geq g\left(\frac{b\delta}{\sigma^2}\right),
\end{aligned}$$

  where $h(u) = (1+u)\ln(1+u) - u$ and $g(u) = \frac{u^2}{2(1+\frac{u}{3})}$. We notice that $h(0) = g(0) = 0$, $h'(0) = g'(0) = 0$, but

$$h''(u) = \frac{1}{1+u} \geq \frac{216}{(6+2u)^3} = g''(u), \forall u \geq 0.$$

  Elementary calculus results in $h(u) \geq g(u), \forall u \geq 0$. Therefore, Bennett's inequality is at least as sharp as Bernstein's inequality.

$\square$

Exercise 2.8.

(a) **I can only prove a special case when** $C = 1$. We first note that the given inequality (2.63) can be separated into two parts according to t's range.

$$\mathbb{P}[Z \geq t] \leq e^{-\frac{t^2}{2(v^2+bt)}} = e^{-\frac{t^2}{2b(\frac{v^2}{b}+t)}} \leq \begin{cases} \exp\left(-\frac{t}{4b}\right) & , \quad \text{when } t \geq \frac{v^2}{b} \\ \exp\left(-\frac{t^2}{4v^2}\right) & , \quad \text{when } t < \frac{v^2}{b} \end{cases}$$

By layer-cake representation, we have

$$\begin{aligned}
\mathbb{E}[Z] &= \int_0^\infty \mathbb{P}[Z \geq t]\, dt \\
&\leq \int_0^{\frac{v^2}{b}} e^{-\frac{t^2}{4v^2}}\, dt + \int_{\frac{v^2}{b}}^\infty e^{-\frac{t}{4b}}\, dt \\
&\leq \int_0^\infty e^{-\frac{t^2}{4v^2}}\, dt + \int_0^\infty e^{-\frac{t}{4b}}\, dt \\
&= v\sqrt{\pi} + 4b
\end{aligned}$$

(b) Using part (a), it suffices to show that

$$\mathbb{P}\left[\frac{1}{n}\sum_{i}^{n} X_i \geq t\right] \leq e^{-\frac{t^2}{2\left(\frac{\sigma^2}{n}+\frac{b}{n}t\right)}}$$

Since X_i satisfies Bernstein's conditions, by Eq. (2.17a),

$$\mathbb{E}\left[e^{\lambda X}\right] \leq e^{\frac{\lambda^2\sigma^2/2}{1-\lambda b}}, \forall \lambda \in (0, 1/b).$$

Then, we are able to calculate the probability:

$$\begin{aligned}
\ln\left(\mathbb{P}\left[\frac{1}{n}\sum_{i}^{n} X_i \geq t\right]\right) &\leq -\lambda n t + \sum_{i}^{n} \frac{\lambda^2\sigma^2/2}{(1-\lambda b)} && \forall \lambda \in (0, 1/b) \\
&= -\frac{nt^2}{2(\sigma^2+bt)} && \text{by setting } \lambda^* = \frac{t}{\sigma^2+bt}
\end{aligned}$$

□

Exercise 2.9.

(a) This is a straightforward conclusion by applying Chernoff bound:

$$\begin{aligned}
\ln\left(\mathbb{P}\left[Z_n \leq \delta n\right]\right) &= \ln\left(\mathbb{P}\left[-Z_n \geq -n\delta\right]\right) \\
&\leq n \ln\left((1-\alpha+\alpha e^{-\lambda})e^{\lambda\delta}\right), \ \forall \lambda > 0 && \text{by i.i.d and Chernoff bound} \\
&= n\left[(1-\delta)\ln\left(\frac{1-\alpha}{1-\delta}\right) + \delta \ln\left(\frac{\alpha}{\delta}\right)\right] && \text{by setting } \lambda^* = \ln\left(\frac{\alpha}{\delta}\right) - \ln\left(\frac{1-\alpha}{1-\delta}\right) \\
&= -nD_{KL}(\delta\|\alpha)
\end{aligned}$$

The minimizer λ^* we used in the second to the last equality is positive as $\delta < \alpha \in (0, 1/2]$

(b)
- **Hoeffding**: $\mathbb{P}\left[\sum_i X_i - n\alpha \leq n(\delta-\alpha)\right] \leq \exp\left(-\frac{n^2(\delta-\alpha)^2}{2\sum_i \alpha(1-\alpha)}\right)$
- **Inequality in part (a)**: $\mathbb{P}\left[\sum_i X_i \leq n\delta\right] \leq \exp\left(-nD_{KL}(\delta\|\alpha)\right)$

It suffices to show

$$\frac{n(\delta-\alpha)^2}{2\alpha(1-\alpha)} \leq nD_{KL}(\delta\|\alpha),$$

Denote the l.h.s as a function $g_\alpha(\delta)$ and the r.h.s as another function $h_\alpha(\delta)$. Then we have $g_\alpha(\alpha) = h_\alpha(\alpha) = 0$. The first-order derivatives are

$$\begin{cases} g'_\alpha(\delta) = \frac{\delta-\alpha}{\alpha(1-\alpha)} < 0 & \forall \delta \in (0, \alpha) \\ h'_\alpha(\delta) = \ln\left(\frac{\delta}{\alpha}\right) - \ln\left(\frac{1-\delta}{1-\alpha}\right) < 0 & \forall \delta \in (0, \alpha) \end{cases}.$$

Moreover,

$$g''_\alpha(\delta) = \frac{1}{\delta(1-\delta)} \geq \frac{1}{\alpha(1-\alpha)} = h''_\alpha(\delta), \ \forall \delta \in (0, \alpha).$$

These results mean that both functions are decreasing and have value 0 at point $\delta = \alpha$, but $g_\alpha(\cdot)$ decreases faster than $h_\alpha(\cdot)$. Therefore, $g(\delta) \geq h(\delta), \forall \delta \in (0, \alpha)$. (A same argument as that in Exercise 2.7)

□

Exercise 2.10.

(a) We have relationships $m = \tilde{\delta} n \le \delta n$. Since

$$\mathbb{P}[Y_n \le \delta n] = \mathbb{P}\left[Z_n \le \tilde{\delta} n\right] \ge \mathbb{P}\left[Z_n = \tilde{\delta} n\right] = \binom{n}{m} \alpha^m (1-\alpha)^{n-m}$$

Taking logarithm on both sides, then dividing by n, and finally replacing $\frac{m}{n}$ with $\tilde{\delta}$, we get the desired inequality.

(b) Let $Z_n \sim \mathsf{Bin}(\mathsf{n}, \tilde{\delta})$. With the fact that Y_n are most likely to be $m = \tilde{\delta} n$, we have the following relationship:

$$\mathbb{P}\left[Y_n = \tilde{\delta} n\right] \ge \mathbb{P}[Y_n = l], \ \forall l \in \{0, 1, 2, \dots, n\}.$$

Following the same operations in part (a), we can rewrite the inequality as:

$$\begin{aligned}
\tfrac{1}{n} \ln \tbinom{n}{m} + \tilde{\delta} \ln^{\alpha} + (1-\tilde{\delta}) \ln(1-\alpha) \quad & \ge \frac{1}{n} \ln(\mathbb{P}[Y_n = l]) \\
\iff \qquad \tfrac{1}{n} \ln \tbinom{n}{m} \quad & \ge -\tilde{\delta} \ln^{\alpha} - (1-\tilde{\delta}) \ln(1-\alpha) + \frac{1}{n} \ln(\mathbb{P}[Y_n = l]) \\
& \ge -\tilde{\delta} \ln^{\tilde{\delta}} - (1-\tilde{\delta}) \ln\left(1-\tilde{\delta}\right) + \frac{1}{n} \ln(\mathbb{P}[Y_n = l]).
\end{aligned}$$

The last inequality holds as the function is increasing in α and $\tilde{\delta} \le \alpha$. Then it remains to show

$$\exists l \in \{0, 1, \dots, n\}, \ s.t., \ \mathbb{P}[Y_n = l] \ge \frac{1}{1+n}.$$

There indeed exists such an l since Y_n has only $1+n$ outcomes, and obviously at least one outcome should have a probability no smaller than $\frac{1}{1+n}$.

(c) Substituting the inequality in part (b) into the r.h.s of the inequality in part (a), and rearranging, we have:

$$\begin{aligned}
\tfrac{1}{n} \ln(\mathbb{P}[Z_n \le \delta n]) \quad & \ge -D_{KL}(\tilde{\delta} \| \alpha) - \frac{\ln(n+1)}{n} \\
\iff \qquad \mathbb{P}[Z_n \ge \delta n] \quad & \ge \frac{1}{1+n} \exp(-n D_{KL}(\tilde{\delta} \| \alpha))
\end{aligned}$$

Combining the result here and the result in Exercise 2.9 part (a), we bound Z_n's lower tail bound from both above and below. The bounds differ only in a constant factor.

$$\frac{1}{1+n} e^{-n D_{KL}(\tilde{\delta} \| \alpha)} \le \mathbb{P}\left[Z_n \le \tilde{\delta} n\right] \le e^{-n D_{KL}(\tilde{\delta} \| \alpha)}$$

Exercise 2.11.

(a) We use layer-cake representation and do integration on two different regions.

$$\begin{aligned}
\mathbb{E}[Z_n] &= \int_0^\infty \mathbb{P}\left[\max_i |X_i| \ge t\right] dt \\
&= \int_0^{u_0} \mathbb{P}\left[\max_i |X_i| \ge t\right] dt + \int_{u_0}^\infty \mathbb{P}\left[\max_i |X_i| \ge t\right] dt \\
&\le u_0 + 2n \int_{u_0}^\infty \mathbb{P}[X_i \ge t]\, dt && \text{by union bound and } X \text{ is symmetric} \\
&\le u_0 + 2n \int_{u_0}^\infty e^{-\frac{t^2}{2\sigma^2}} dt && \text{by Sub-Gaussian tail bound} \\
&= u_0 + 2n \cdot \sqrt{2\pi\sigma^2} \cdot \mathbb{P}\left[\mathsf{N}(0, \sigma^2) \ge u_0\right] \\
&\le u_0 + 2n \cdot \sqrt{2\pi\sigma^2} \cdot \sqrt{\frac{2}{\pi}} \frac{\sigma}{u_0} e^{-u_0^2/(2\sigma^2)} && \text{by the given tail bound} \\
&= \sqrt{2\sigma^2 \ln^n} + \frac{4\sigma}{\sqrt{2 \ln^n}} && \text{by setting } u_0 = \sqrt{2\sigma^2 \ln^n} \text{ and simplifying}
\end{aligned}$$

(b)

(c)

$\square$

Exercise 2.12.

(a) Note that, for any $\lambda > 0$:

$$\begin{aligned}
\mathbb{E}\left[\lambda \max_i X_i\right] &= \ln\left(\exp\left(\mathbb{E}\left[\lambda \cdot \max_i X_i\right]\right)\right) \\
&\leq \ln\left(\mathbb{E}\left[\exp\left(\lambda \cdot \max_i X_i\right)\right]\right) && \text{by Jensen's inequality} \\
&\leq \ln\left(\sum_i \cdot \mathbb{E}\left[\exp(\lambda X_i)\right]\right) \\
&\leq \ln n + \frac{\lambda^2\sigma^2}{2} && X_i \in SubG
\end{aligned}$$

Therefore, the expectation is upper bounded:

$$\begin{aligned}
\mathbb{E}\left[\max_i X_i\right] &= \frac{\mathbb{E}\left[\lambda \max_i X_i\right]}{\lambda} \\
&\leq \frac{\ln n}{\lambda} + \frac{\lambda\sigma^2}{2} \\
&= \sqrt{2\sigma^2 \ln n} && \text{by setting } \lambda = \sqrt{\frac{2\ln n}{\sigma^2}}
\end{aligned}$$

(b) The first inequality follows by noticing that $Z = \max_i |X_i| = \max_i\{X_1, X_2, \ldots, X_n, -X_1, -X_2, \ldots, -X_n\}$. The second inequality is obviously true for $n \geq 2$.

$\square$

Exercise 2.13.

(a) $\mathbb{E}\left[e^{\lambda(X_1+X_2)}\right] = \mathbb{E}\left[e^{\lambda X_1}\right]\mathbb{E}\left[e^{\lambda X_2}\right] \leq e^{\frac{\lambda^2\sigma_1^2}{2}} e^{\frac{\lambda^2\sigma_2^2}{2}} = e^{\frac{\lambda^2(\sigma_1^2+\sigma_2^2)}{2}}$, where the first equality holds due to independence.

(b) We use Hölder's inequality to do the proof:

$$\begin{aligned}
\mathbb{E}\left[e^{\lambda(X_1+X_2)}\right] &= \mathbb{E}\left[e^{\lambda X_1} e^{\lambda X_2}\right] \\
&\leq \left(\mathbb{E}\left[e^{p\lambda X_1}\right]\right)^{\frac{1}{p}} \left(\mathbb{E}\left[e^{q\lambda X_2}\right]\right)^{\frac{1}{q}} && \text{by Hölder's inequality, where } \frac{1}{p} + \frac{1}{q} = \\
&\leq \sqrt{\exp\left(\frac{(2\lambda)^2\sigma_1^2}{2}\right)\exp\left(\frac{(2\lambda)^2\sigma_2^2}{2}\right)} && X_1, X_2 \in SG \text{ and by setting } p = q = 2 \\
&= \exp\left(\frac{\lambda^2(2\sigma_1^2 + 2\sigma_1^2)}{2}\right)
\end{aligned}$$

(c) We can get the desired result by setting $p = \frac{\sigma_1+\sigma_2}{\sigma_1}, q = \frac{\sigma_1+\sigma_2}{\sigma_2}$ in part (b).

(d) By conditional expectation, we have

$$\begin{aligned}
\mathbb{E}\left[e^{\lambda X_1 X_2}\right] &= \mathbb{E}_{X_2}\left[\mathbb{E}_{X_1}\left[e^{\lambda X_1 X_2} | X_2\right]\right] \\
&\leq \mathbb{E}_{X_2}\left[e^{\frac{\lambda^2 X_2^2 \sigma_1^2}{2}}\right] && \text{conditional on } X_2 \text{ it is a SubG} \\
&\leq \frac{1}{\sqrt{1 - \lambda^2\sigma_1^2\sigma_2^2}},\ \forall \lambda \in [0, 1) && \text{by Theorem 2.6 (4)} \\
&\leq e^{\lambda^2\sigma_1^2\sigma_2^2},\ \forall \lambda^2\sigma_1^2\sigma_2^2 \leq 1/2
\end{aligned}$$

The last inequality is true since $\sqrt{1-x}\cdot e^x \geq 1,\ \forall x \in (0, 1/2)$. Reorganizing the final term will match the definition of Sub-Exponential variable with required parameters.

□

Exercise 2.14.

(a) We use layer-cake representation:

$$\begin{aligned}
\mathsf{Var}[X] &= \mathbb{E}\left[(X-\mathbb{E}[X])^2\right] \\
&= \int_0^\infty \mathbb{P}\left[(X-\mathbb{E}[X])^2 \geq t\right] dt \\
&\leq \int_0^\infty c_1 e^{-c_2 t} dt \\
&= \frac{c_1}{c_2}
\end{aligned}$$

(b) Bernoulli distribution with success probability 1/2. Then, both 0 and 1 are the medians.

(c) WLOG, we assume $\delta = \mu - m \leq 0$, where μ is the mean and m median. We are interested in the upper bound of l.h.s in Eq (2.69):

$$\begin{aligned}
\mathbb{P}[|X-m| \geq t] &\leq \mathbb{P}[|(X-\mu)| + |(\mu - m)| \geq t] \\
&= \mathbb{P}[|X-\mu| \geq t - \delta], \quad \forall t \geq 0
\end{aligned}$$

We separately check two regions:

- $t \in [2\delta, \infty)$: In this range, we have $\frac{1}{2}t \leq t - \delta$. Thus, we get another looser upper bound

  $$\mathbb{P}[|X-m| \geq t] \leq \mathbb{P}\left[|X-\mu| \geq \frac{1}{2}t\right] \leq c_1 e^{-\frac{c_2}{4}t^2}, \quad \forall t \in [2\delta, \infty)$$

  The last inequality follows from the given inequality Eq (2.68). The most-right-side term shows that it suffices to let $\beta \geq 1, \alpha = \frac{1}{4}$, then we are fine with that

  $$\mathbb{P}[|X-m| \geq t] \leq \beta c_1 e^{-\alpha c_2 t^2}, \quad t \in [2\delta, \infty).$$

- .

□

Exercise 2.15.

Let the random vector $X = (X_1, X_2, \ldots, X_n)$. We define a new function $g(X) = \left\|\widehat{f_n^X} - f\right\|_1$. Easy to check that the function g satisfies the bounded difference property Eq. (2.32), that is, if we change one element in X to X', we should have:

$$\begin{aligned}
|g(X) - g(X')| &= \left| \left\|\widehat{f_n^X} - f\right\|_1 - \left\|\widehat{f_n^{X'}} - f\right\|_1 \right| \\
&\leq \left\|\widehat{f_n^X} - \widehat{f_n^{X'}}\right\|_1 \\
&= \frac{1}{n}\int_{-\infty}^{\infty} \left| \frac{1}{h}K\left(\frac{x-X}{h}\right) - \frac{1}{h}K\left(\frac{x-X'}{h}\right)\right| dx \\
&\leq \frac{2}{n}
\end{aligned}$$

Both inequalities hold due to triangular inequality. By Corollary 2.21, we have

$$\mathbb{P}[g(X) - \mathbb{E}[g(X)] \geq \delta] \leq e^{-\frac{n\delta^2}{2}},$$

which is better than the desired result by a constant factor 4 in the exponent.

□

Exercise 2.16.

(a) Notice that S_n satisfies the bounded difference property (2.32) with parameter $2b_i$, then according to Corollary 2.21, we have,

$$\mathbb{P}\left[|S_n - \mathbb{E}\left[S_n\right]| \geq n\delta\right] \leq 2e^{-\frac{n\delta^2}{2b^2}},$$

which is better than the desired bounded by a constant. This solution follows the same idea as in Exercise 2.15.

(b) By rearranging the required inequality, we notice that the desired inequality is actually:

$$\mathbb{P}\left[S_n - \sqrt{\sum_i \mathbb{E}\left[\|X_i\|_H^2\right]} \geq n\delta\right] \leq e^{-\frac{n\delta^2}{8b^2}}.$$

By inserting $-\mathbb{E}\left[S_n\right] + \mathbb{E}\left[S_n\right]$ into the probability argument's r.h.s, it suffices to show

$$\mathbb{E}\left[S_n\right] \leq \sqrt{\sum_i \mathbb{E}\left[\|X_i\|_H^2\right]}.$$

This is true since

$$\sqrt{\sum_i \mathbb{E}\left[\|X_i\|_H^2\right]} \geq \sqrt{\mathbb{E}\left[\left\|\sum_i X_i\right\|_H^2\right]} = \sqrt{\mathbb{E}\left[S_n^2\right]} \geq \mathbb{E}\left[S_n\right]$$

□

Exercise 2.17.

We first note that $\mathbb{E}\left[Z\right] = tr(Q)$ as the X_i are standard normal variables. Since Q is a semi-definite positive matrix, it can be decomposed into $Q = P^\top \Lambda P$, where $\Lambda = Diag(\lambda_1, \ldots, \lambda_n)$ and P is a unitary matrix ($PP^{-1} = I$) with orthogonal and unit-length columns. Because P also satisfies $P^\top = P^{-1}$, it is a rotation matrix. And by rotation invariance of Gaussian variables, Px is also a standard Gaussian vector. Therefore,

$$Z = \left\langle \Lambda, Px \cdot (Px)^\top \right\rangle = \sum_i \lambda_i \cdot \chi^2(1).$$

Since $\chi^2(1) \sim SubExp(2, 4)$, and by the definition of Sub-Exponential variable, we know that $\lambda_i \cdot \chi^2(1) \sim SubExp(2\lambda_i, 4\lambda_i)$. As Z is the sum over some sub-exponential variables, $Z \sim (2\sqrt{\sum_i (\lambda_i)^2}, 4\max_i\{\lambda_i\})$. By sub-exponential's tail bound (Proposition 2.9), we have

$$\mathbb{P}\left[Z - tr(Q) \geq \sigma t\right] \leq \exp\left(-\min\left(\frac{\sigma^2 t^2}{8\sum_i \lambda_i^2}, \frac{\sigma t}{8\max_i\{\lambda_i\}}\right)\right).$$

The desired bound follows by SDP matrix's facts that $\max_i\{\lambda_i\} = \sup_{\|u\|_2 \leq 1} \|Qu\|_2 = \|Q\|_2$ and $\sum_i \lambda_i^2 = \|Q\|_F^2$.

□

Exercise 2.18.

(a) By the definition of function ψ and the fact that $\|X\|_{\psi_q} \leq +\infty$, we have that

$$\mathbb{E}\left[\psi(|X| \,/\, \|\psi_q\| \, X^q)\right] \leq 1 \Longleftrightarrow \mathbb{E}\left[\exp\left(\frac{|X|^q}{\|X\|_{\psi_q}^q}\right)\right] \leq 2$$

Then we look at the probability to be bounded. With Chernoff bound, we can show

$$\begin{aligned}
\mathbb{P}\left[|X|>t\right] &= \mathbb{P}\left[\exp\left(\frac{\lambda^q |X|^q}{t^q}\right) > \exp\left(\lambda^q\right)\right], \quad \forall \lambda > 0 \\
&\leq \frac{\mathbb{E}\left[\frac{|X|^q}{\|X\|_{\psi_q}^q}\right]}{\exp\left(t^q / \|X\|_{\psi_q}^q\right)}, \quad \text{let } \lambda = t/\|X\|_{\psi_q} \\
&\leq 2\exp\left(\|X\|_{\psi_q}^{-q} \cdot t^q\right).
\end{aligned}$$

(b) This can be proved by contradiction. Suppose that $\|\psi_q\|$ is infinite, that means, by taking $\psi_q(u) = \exp(u^q) - 1$,

$$\mathbb{E}\left[\exp\left(|X|/t\right)\right] > 2, \quad \forall t > 0.$$

By layer cake representation, and the given inequality (2.73), we can show that

$$\begin{aligned}
LHS &= \int_0^\infty \mathbb{P}\left[\exp\left(|X|/t\right) \geq u\right] du && \forall t > 0 \\
&= \int_0^1 \mathbb{P}\left[|X| \geq t\ln(u)\right] du + \int_0^\infty \mathbb{P}\left[|X| \geq tu\right] du && \forall t > 0 \\
&\leq 1 + \int_0^\infty c_1 \exp\left(-c_2 t^q u^q\right) du && \forall t > 0, \text{by the given ineq.}
\end{aligned}$$

Notice that, for any given $\delta > 0$, there must exist $t > 0$ such that the second term is less than δ, we get a contradiction with $LHS > 2$.

□

Exercise 2.19.

$$\mathbb{E}\left[\max_i |X_i|\right] \leq \sigma\mathbb{E}\left[\sum_i |X_i|/\sigma\right] = \sigma\mathbb{E}\left[\psi^{-1}\left(\psi\left(\sum_i |X_i|/\sigma\right)\right)\right] \leq \sigma\psi^{-1}\left(\mathbb{E}\left[\psi\left(\sum_i |X_i|/\sigma\right)\right]\right),$$

the last inequality comes from the concavity of ψ^{-1} (since ψ is strictly convex). Then we focus the inner expectation part:

$$\mathbb{E}\left[\psi\left(\sum_i |X_i|/\sigma\right)\right] \leq \sum_i \mathbb{E}\left[\psi\left(|X_i|/\sigma\right)\right] \leq n.$$

□

Exercise 2.20.

The desired inequality is a direct result by applying Chernoff bound and the given Rosenthal's inequality:

$$\mathbb{P}\left[\left|\frac{1}{n}\sum_i X_i\right| \geq \delta\right] \leq \frac{\mathbb{E}\left[\left(\sum_i X_i\right)^{2m}\right]}{(n\delta)^{2m}} = \frac{1}{(\sqrt{n}\delta)^{2m}} \cdot \frac{\mathbb{E}\left[\left(\sum_i X_i\right)^{2m}\right]}{n^m} \leq \frac{R_m}{(\sqrt{n}\delta)^{2m}}\left\{\frac{\sum_i \mathbb{E}\left[X_i^{2m}\right]}{n^m} + \left(\frac{\sum_i \mathbb{E}\left[X_i^2\right]}{n}\right)^m\right\}.$$

Then it remains to show the expression in the parentheses is constant. The second part $\sum_i \mathbb{E}\left[X_i^2\right]/n$ is obviously constant. We show that the first part is also a constant:

$$\frac{\sum_i \mathbb{E}\left[X_i^{2m}\right]}{n^m} = \frac{\sum_i \left[\left(\mathbb{E}\left[X_i^{2m}\right]\right)^{1/(2m)}\right]^{2m}}{n^m} \leq \frac{\sum_i C_m^{2m}}{n^m} \leq \frac{\left(\sum_i C_m^2\right)^m}{n^m} = C_m^{2m}$$

□

Exercise 2.21.

(a) $\Pr\left[d(X) \le \delta\right] = \Pr\left[\min_{j\in[N]} \left\|X - z^j\right\|_1 \le n\delta\right] \le \sum_{j\in[N]} \Pr\left[\left\|X - z^j\right\|_1 \le n\delta\right] \le N \cdot e^{-n\cdot D_{KL}(\delta\|1/2)}$, where the last inequality is from Exercise 2.9. This inequality holds only when $\delta < 1/2$ (this assumption is not mentioned in this exercise) since $X_i - z_i^j$ is a $Ber(1/2)$. Additionally, by assumption that $N \le 2^{nR}$, we can further control above terms:

$$N \cdot e^{-n\cdot D_{KL}(\delta\|1/2)} \le e^{n(R - D_{KL}(\delta\|1/2))} \to 0, \quad \text{as } n \to \infty.$$

The last convergence comes from another assumption that $R < D_{KL}(\delta\|1/2)$.

(b) (i) Define a new r.v. $\mathbb{1}\{V \ge 1\}$. Note that $V = \sum_{j\in[N]} V^j \in \mathbb{N}$, so $V \cdot \mathbb{1}\{V \ge 1\} = V$, a.s. Then, by Cauchy's inequality, we have

$$\begin{aligned}
\mathbb{E}^2[V] &= \mathbb{E}^2\left[V \cdot \mathbb{1}\{V \ge 1\}\right] && (V \cdot \mathbb{1}\{V \ge 1\} = V, a.s.)\\
&\le \mathbb{E}\left[V^2\right] \cdot \mathbb{E}\left[\mathbb{1}^2\{V \ge 1\}\right] && \text{(by Cauchy's inequality)}\\
&= \mathbb{E}\left[V^2\right] \cdot \Pr\left[V \ge 1\right].\\
\Longrightarrow \Pr\left[V \ge 1\right] &\ge \frac{\mathbb{E}^2[V]}{\mathbb{E}\left[V^2\right]}.
\end{aligned}$$

(i) TODO

Exercise 2.22.

(a) Denote $f(\boldsymbol{\theta}, \boldsymbol{x}) = \exp\left(\frac{1}{\sqrt{d}} \sum_{j\ne k} \theta_{jk} x_j x_k\right)$. Notice that $f(\boldsymbol{\theta}, \boldsymbol{x})$ is an exponential-like function and is convex in $\boldsymbol{\theta}$. Since $F_d(\boldsymbol{\theta}) := \ln\left(\sum_{\boldsymbol{x}} f(\boldsymbol{\theta}, \boldsymbol{x})\right)$, by vector composition of convex functions, we know F_d is convex in $\boldsymbol{\theta}$.

(b) By mean-value theorem, there exists $\alpha \in [0, 1]$ such that $F_d(\boldsymbol{\theta} + \boldsymbol{\delta}) - F_d(\boldsymbol{\theta}) = \nabla F(\boldsymbol{\theta} + \alpha \cdot \boldsymbol{\delta})^\top \boldsymbol{\delta}$. The latter is upper bounded by $\left\|\nabla F(\boldsymbol{\theta} + \alpha \cdot \boldsymbol{\delta})\right\|_2 \left\|\boldsymbol{\delta}\right\|_2$. Now, it remains to control the gradient's norm. Given any vector $\boldsymbol{\theta}$, we have:

$$\left\|\nabla F_d(\boldsymbol{\theta})\right\|_2 = \sqrt{\sum_{j\ne k} \frac{1}{d} \left(\frac{\sum_{\boldsymbol{x}} x_j x_k f(\boldsymbol{\theta}, \boldsymbol{x})}{\sum_{\boldsymbol{x}} f(\boldsymbol{\theta}, \boldsymbol{x})}\right)^2} \le \sqrt{\sum_{j\ne k} \frac{1}{d}} \le \sqrt{d},$$

where the first inequality is because $x_j x_k \le 1$. Replace back, we obtain $F_d(\boldsymbol{\theta}+\boldsymbol{\delta}) - F_d(\boldsymbol{\theta}) \le \sqrt{d}\left\|\boldsymbol{\delta}\right\|_2, \forall \boldsymbol{\delta}$, which is equivalent to desired statement.

(c) From Jensen's inequality, we have

$$\mathbb{E}\left[F_d(\boldsymbol{\theta})\right] := \mathbb{E}\left[\ln\left(\sum_{\boldsymbol{x}} f(\boldsymbol{\theta}; \boldsymbol{x})\right)\right] \le \ln\left(\mathbb{E}\left[\sum_{\boldsymbol{x}} f(\boldsymbol{\theta}, \boldsymbol{x})\right]\right),$$

where the last term can be further bounded:

$$\begin{aligned}
\ln\left(\mathbb{E}\left[\sum_{\boldsymbol{x}} f(\boldsymbol{\theta}, \boldsymbol{x})\right]\right) &= \ln\left(\sum_{\boldsymbol{x}} \mathbb{E}_{\boldsymbol{\theta}}\left[\exp\left(\sum_{j\ne k} \theta_{jk} \cdot \frac{x_j x_k}{\sqrt{d}}\right)\right]\right)\\
&\le \ln\left(\sum_{\boldsymbol{x}} \exp\left(\frac{\sum_{j\ne k} x_j^2 x_k^2 \beta^2}{2d}\right)\right)\\
&= d\ln 2 + \frac{d\beta^2}{4}.
\end{aligned}$$

The inequality above is by MGF of Gaussian variable and by combining $\binom{n}{2}$ such variables. The last line is due to $x_j^2 x_k^2 = 1$. Therefore,

$$\begin{aligned}\Pr\left[\frac{F_d(\boldsymbol{\theta})}{d} - (\ln 2 + \frac{\beta^2}{4}) \geq t\right] &\leq \Pr\left[\frac{F_d(\boldsymbol{\theta})}{d} - \mathbb{E}\left[F_d(\boldsymbol{\theta})\right] \geq t\right] \\ &\leq 2\exp\left(-\beta d t^2/2\right), \forall t > 0,\end{aligned}$$

where the last inequality is by Theorem 2.26. We can apply the theorem as F_d is shown to be Lip-cts in part (b).

□