Exercise 2.1

(a) Let X be 1 almost surely, $t=1$, Then $P\{X \geq t\} = 1 = \frac{\boldsymbol{E}X}{t}$.

(b) Let X satisfy $P\{Y=1\} = P\{Y=-1\} = \frac{1}{2}$, $t=1$, thus $\boldsymbol{E}Y = 0$ and $Var(Y) = 1$, $P\{|Y - \boldsymbol{E}Y| \geq 1\} = 1 = \frac{Var(Y)}{t^2}$.

Exercise 2.2

(a) $\phi'(z) = -z\phi(z)$

(b) Use (a) and integration by parts for mutiple times:

$$\begin{aligned}
P\{Z \geq z\} &= \int_z^{+\infty} -\frac{1}{z}\phi'(z)dz = -\frac{1}{z}\phi'(z) - \int_z^{+\infty} \frac{\phi(z)}{z}dz \\
&= \frac{1}{z}\phi(z) + \int_z^{+\infty} \frac{\phi'(z)}{z^3}dz = \frac{\phi(z)}{z} + \frac{\phi(z)}{z^3} + \int_z^{+\infty} \frac{3\phi(z)}{z^4}dz \\
&= \phi(z)(\frac{1}{z} - \frac{1}{z^3}) - \int_z^{+\infty} \frac{3\phi'(z)}{z^5}dz \\
&= \phi(z)(\frac{1}{z} - \frac{1}{z^3} + \frac{3}{z^5}) - \int_z^{+\infty} \frac{15\phi(z)}{z^7}dz
\end{aligned}$$

Exercise 2.3

$$\begin{aligned}
\frac{\boldsymbol{E}e^{\lambda X}}{e^{\lambda\delta}} &= \frac{\boldsymbol{E}[\sum_{k=0}^{+\infty} \frac{(\lambda X)^k}{k!}]}{\sum_{k=0}^{+\infty} \frac{(\lambda\delta)^k}{k!}} = \frac{\sum_{k=0}^{+\infty} \frac{(\lambda\delta)^k}{k!} \frac{\boldsymbol{E}X^k}{\delta^k}}{\sum_{k=0}^{+\infty} \frac{(\lambda\delta)^k}{k!}} \\
&\geq \frac{\sum_{k=0}^{+\infty} \frac{(\lambda\delta)^k}{k!} \inf_k \frac{\boldsymbol{E}X^k}{\delta^k}}{\sum_{k=0}^{+\infty} \frac{(\lambda\delta)^k}{k!}} = \inf_k \frac{\boldsymbol{E}X^k}{\delta^k}
\end{aligned}$$

take infimum for λ on both sides, we get

$$\inf_k \frac{\boldsymbol{E}X^k}{\delta^k} \leq \inf_{\lambda > 0} \frac{\boldsymbol{E}e^{\lambda X}}{e^{\lambda\delta}}.$$

Exercise 2.4

(a) $\psi(0) = 0$, $\psi'(0) = \frac{\boldsymbol{E}[Xe^{\lambda X}]}{\boldsymbol{E}e^{\lambda X}}|_{\lambda=0} = \mu$.

(b) $\psi''(\lambda) = \frac{\boldsymbol{E}[X^2 e^{\lambda X}]}{\boldsymbol{E}[e^{\lambda X}]} - \frac{\boldsymbol{E}^2[Xe^{\lambda X}]}{\boldsymbol{E}^2[e^{\lambda X}]}$.

(c) Observe first that if a random variable Y ranges over $[a,b]$, then its variance can be bounded:

$$Var(Y) = Var(Y - \frac{a+b}{2}) \leq \boldsymbol{E}[|Y - \frac{a+b}{2}|^2] \leq \frac{(b-a)^2}{4}.$$

Now let P denote the distribution of X and let P_λ be the probability distribution with density $x \to e^{-\psi(\lambda)}e^{\lambda x}$ with respect to P. Then

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

where Y is a random variable with distribution P_λ relative to P. The sub-Gaussian property follows by Taylor's theorem, for some $\theta \in [0, \lambda]$,

$$\psi(\lambda) = \psi(0) + \psi'(0)\lambda + \psi''(\theta)\frac{\lambda^2}{2} \leq \frac{(b-a)^2}{8}.$$

Exercise 2.5

(a) Using Jensen's Inequality, we have

$$e^{\lambda \boldsymbol{E}X} \leq \boldsymbol{E}e^{\lambda X} \leq e^{\frac{\lambda^2\sigma^2}{2}+\lambda\mu}, \forall \lambda \in \mathbb{R} \Rightarrow \frac{\sigma^2}{2}\lambda^2 + (\mu - \boldsymbol{E}X)\lambda \geq 0, \forall \lambda \in \mathbb{R} \Rightarrow \boldsymbol{E}X = \mu$$

(b) Let $h(\lambda) = e^{\frac{\lambda^2\sigma^2}{2}+\lambda\mu} - \boldsymbol{E}e^{\lambda X}$, then $h(0) = h'(0) = 0$, sine $h(\lambda) \geq 0$, thus

$$h''(0) = \left(e^{\frac{\lambda^2\sigma^2}{2}+\lambda\mu}[\sigma^2 + (\sigma^2\lambda + \mu)^2] - \boldsymbol{E}X^2 e^{\lambda X}\right)|_{\lambda=0} \geq 0.$$

we get $\boldsymbol{E}[X^2] \leq \sigma^2 + \mu^2$, thus $Var(X) \leq \sigma^2$.

(c) Disprove. An asymmetric Bernoulli is a counter-example. And intuitively, in the bounded case, $\sigma^2_{min} >= \frac{(b-a)^2}{4}$, as long as $\exists \lambda_0 \in \mathbb{R}, \psi''(\lambda_0) = Var_{\lambda_0}(Y) = \frac{(b-a)^2}{4}$, while $Var(X)$ can still smaller than $\frac{(b-a)^2}{4}$.

Exercise 2.6

From the special case of Bennett's Inequality (2.23), we get

$$P(Z_n \leq \boldsymbol{E}[Z_n] - \sigma^2\delta) = P(\sum_{i=1}^n X_i^2 \leq E[\sum_{i=1}^n X_i^2] - n\sigma^2\delta) \leq \exp(-\frac{n\sigma^4\delta^2}{\frac{2}{n}\sum_{i=1}^n \boldsymbol{E}X_i^4})$$

A important property of sub-Gaussian variables is the growth of their moments (Boucheron 2013), if $X \in \mathcal{G}(v)$, then $\forall q \geq 1, \boldsymbol{E}[X^{2q}] \leq 2q!(2v)^q \leq q!(4v)^q$, thus here we have $\boldsymbol{E}X_i^4 \leq 16\sigma^2$, and the original objective inequality follows based on that.

Exercise 2.7

(a) Since $\phi(x) = \frac{e^x-x-1}{x^2}$ is an increasing on $\mathbb{R}$, we have

$$\begin{aligned}\log \boldsymbol{E}[e^{\lambda X_i}] &\leq \boldsymbol{E}(e^{\lambda X_i} - \lambda X_i - 1) = \boldsymbol{E}[X_i^2 \frac{(e^{\lambda X_i} - \lambda X_i - 1)}{X_i^2}] \\ &\leq \boldsymbol{E}[X_i^2]\frac{(e^{\lambda b} - \lambda b - 1)}{b^2} = \sigma_i^2\lambda^2[\frac{(e^{\lambda b} - \lambda b - 1)}{(\lambda b)^2}]\end{aligned}$$

(b) We can calculate the Cramér transformation of X_i using the bound for $\psi(\lambda) = \log \boldsymbol{E}[e^{\lambda X_i}]$ given by (a), which is

$$\psi^*(t) = \sup_{\lambda \geq 0}(\lambda t - \psi(t)) = -\frac{\sigma_i^2}{b^2}h(\frac{bt}{\sigma_i^2})$$

where $h(t) = (1+t)\log(1+t) - t, t > 0$. And $\log \boldsymbol{E}[e^{\lambda \sum_{i=1}^n X_i}] \leq n\sigma^2\lambda^2[\frac{(e^{\lambda b}-\lambda b-1)}{(\lambda b)^2}]$, thus

$$P(\sum_{i=1}^n X_i \geq n\delta) \leq \exp(-\frac{n\sigma^2}{b^2}h(\frac{b\delta}{\sigma^2}))$$

(c) It can be proved that $h(u) \geq \frac{u^2}{2(1+\frac{u}{3})}$, then

$$\exp(-\frac{n\sigma^2}{b^2}h(\frac{b\delta}{\sigma^2})) \leq \exp(-\frac{\delta^2}{2(\sigma^2 + \frac{b\delta}{3})})$$

which is a Bernstein type upper bound.

Exercise 2.8

(a)(*) gg :(, where can be the term $logC$ from?

(b) $\boldsymbol{E}[e^{\lambda \frac{\sum_{i=1}^n X_i}{n}}] \leq \exp(\frac{\lambda^2\sigma^2/n}{2(1-b|\lambda|/n)})$, which implies $P(|\frac{1}{n}\sum_{i=1}^n X_i| \geq t) \leq 2\exp(-\frac{nt^2}{2(\sigma+bt)})$, and we can set $C = 2, v = \frac{\sigma}{\sqrt{n}}, b = \frac{b}{n}$ in (2.63), Then the inequality follows the result of (a).

Exercise 2.9

(a)(*) Let $m = \lfloor n\delta \rfloor$, then

$$P(Z_n \le \delta n) = \sum_{k=0}^{m} \tbinom{n}{k}\alpha^k(1-\alpha)^{n-k} \le (1+n\delta)\tbinom{n}{n\delta}\alpha^{n\delta}(1-\alpha)^{n(1-\delta)}$$
$$\le (\frac{\alpha}{\delta})^{n\delta}(\frac{1-\alpha}{1-\delta})^{n(1-\delta)} \le e^{-nD(\delta\|\alpha)}$$

(b) The Hoeffding bound is $e^{-2n(\alpha-\delta)^2}$. So it suffices to prove that $D(\delta \parallel \alpha) \ge 2(\delta-\alpha)^2$.

Exercise 2.10

(a) It can be derived from $P(Z_n \le \delta n) \ge P(Z_n = m) = \tbinom{n}{m}\alpha^m(1-\alpha)^{n-m}$.

(b) It suffices to show that $\tbinom{n}{m}(n+1)\tilde{\delta}^m(1-\tilde{\delta})^{n-m} \ge 1$, which follows from $\sum_{l=0}^{n}\tbinom{n}{l}\tilde{\delta}^l(1-\tilde{\delta})^{n-l} = 1$ and the hint.

(c) (*)Based on previous relations, $P(Z_n \le \delta n) \ge \frac{1}{n+1}e^{-nD(\tilde{\delta}\|\alpha)}$. However, since $\tilde{\delta} \le \delta$, $D(\tilde{\delta} \parallel \alpha) \ge D(\delta \parallel \alpha)$, which means bounding it furthur to $\frac{1}{n+1}e^{-nD(\delta\|\alpha)}$ is not feasible.

Exercise 2.11

gg, no idea of how to split the bound with 2 terms.

Exercise 2.12

(a) Using the convexity of exponential function and Jensen's inequality,

$$e^{\lambda \boldsymbol{E}[\max\limits_{i=1,\dots,n} X_i]} \le \boldsymbol{E}e^{\lambda \max\limits_{i=1,\dots,n} X_i} = \boldsymbol{E}\max_{i=1,\dots,n} e^{\lambda X_i} \le \sum_{i=1}^{n} \boldsymbol{E}e^{\lambda X_i} \le ne^{\lambda^2\sigma^2}$$
$$\Rightarrow \boldsymbol{E}[\max_{i=1,\dots,n} X_i] \le \frac{\log n}{\lambda} + \frac{\sigma^2\lambda}{2}$$

Take infimum for λ on right side, which yields

$$\boldsymbol{E}\max_{i=1,\dots,n} \le \sqrt{2\sigma^2\log n}$$

(b) $Z = \max\limits_{i=1,\dots,n}|X_i| = \max\{X_1,\dots,X_n,-X_1,\dots,-X_n\}$. Notice that no independence assumptions are needed, (a) can be applied to Z, thus

$$\boldsymbol{E}Z \le \sqrt{2\sigma^2\log(2n)} \le 2\sqrt{\sigma^2\log n}$$

Exercise 2.13

(a) $\boldsymbol{E}[e^{\lambda(X_1+X_2)}] \le \boldsymbol{E}(e^{\lambda X_1})\boldsymbol{E}(e^{\lambda X_2}) \le e^{\frac{\lambda^2(\sigma_1^2+\sigma_2^2)}{2}}$

(b) By Cauchy-Schwarz inequality, $\boldsymbol{E}[e^{\lambda(X_1+X_2)}] \le (\boldsymbol{E}e^{2\lambda X_1}\boldsymbol{E}e^{2\lambda X_2})^{\frac{1}{2}} \le e^{\lambda^2(\sigma_1^2+\sigma_2^2)}$

(c) By Jensen's inequality, $\boldsymbol{E}[\exp\{\lambda(X_1+X_2)\}] = \boldsymbol{E}[\exp\{\frac{\sigma_1}{\sigma_1+\sigma_2}(\frac{\sigma_1+\sigma_2}{\sigma_1}\lambda X_1)+\frac{\sigma_2}{\sigma_1+\sigma_2}(\frac{\sigma_1+\sigma_2}{\sigma_2}\lambda X_2)\}] \le \frac{\sigma_1}{\sigma_1+\sigma_2}\exp\{(\sigma_1+\sigma_2)^2\lambda^2/2\} + \frac{\sigma_2}{\sigma_1+\sigma_2}\exp\{(\sigma_1+\sigma_2)^2\lambda^2/2\} = \exp\{(\sigma_1+\sigma_2)^2\lambda^2/2\}$

Exercise 2.14

(a) $Var(X) = \int_0^{+\infty} P(|X-\boldsymbol{E}X|^2 \ge t)dt \le \int_0^{+\infty} c_1e^{-c_2t^2}dt = \frac{c_1}{c_2}$.

(b) For example, when $P(X=0) = P(X=1) = \frac{1}{2}$, any $m \in (0,1)$ can be a median.

(c) By Jensen's inequality,

$$\boldsymbol{E}[\exp\{\lambda(X_1+X_2)\}] = \boldsymbol{E}[\exp\{\frac{\sigma_1}{\sigma_1+\sigma_2}(\frac{\sigma_1+\sigma_2}{\sigma_1}\lambda X_1) + \frac{\sigma_2}{\sigma_1+\sigma_2}(\frac{\sigma_1+\sigma_2}{\sigma_2}\lambda X_2)\}]$$
$$\le \frac{\sigma_1}{\sigma_1+\sigma_2}e^{(\sigma_1+\sigma_2)^2\lambda^2/2} + \frac{\sigma_2}{\sigma_1+\sigma_2}e^{(\sigma_1+\sigma_2)^2\lambda^2/2}$$
$$= e^{(\sigma_1+\sigma_2)^2\lambda^2/2}$$

Exercise 2.15

Let $g(x) = \|\hat{f}_n - f\|_1$, then g satisfies a bounded differences property,

$$|g(x) - g(x^{(k)})| = |\|\hat{f}_n - f\|_1 - \|\hat{f}_n^{(k)} - f\|_1| \leq \|\hat{f}_n - \hat{f}_n^{(k)}\|_1$$
$$= \int_{-\infty}^{+\infty} \frac{1}{nh}|K(\frac{x - x_k}{h}) - K(\frac{x - x_k^{'}}{h})|dx \leq \frac{2}{n}$$

By Corollary 2.21 (Bounded differences inequality),

$$P(|\|\hat{f}_n - f\|_1) \geq \boldsymbol{E}\|\hat{f}_n - f\|_1 + \delta \leq e^{-\frac{n\delta^2}{2}}$$

Exercise 2.16

(a) S_n satisfies a bounded differences property,

$$|S_n - S_n^{(k)}| \leq \|x_k - x_k^{'}\|_H \leq 2b_k$$

By Corollary 2.21 (Bounded differences inequality),

$$P(|S_n - \boldsymbol{E}S_n| \geq n\delta) \leq 2e^{-\frac{n\delta^2}{2b^2}}$$

(b) As $\{X_i\}$ are independent, thus

$$na = \sqrt{\sum_{i=1}^{n} \boldsymbol{E}[\|X_i\|_{\mathbb{H}}^2]} \geq \boldsymbol{E}[\|\sum_{i=1}^{n} X_i\|_{\mathbb{H}}] = \boldsymbol{E}S_n$$

then $P(S_n \geq na + n\delta) \leq P(S_n - \boldsymbol{E}S_n \geq n\delta) \leq e^{-\frac{n\delta^2}{2b^2}}$.

Exercise 2.17

$Q \in \mathcal{S}_+^{n\times n} \Rightarrow Q = A^T\Lambda A$, then $Z = X^TQX = (AX)^T\Lambda AX = Y^T\Lambda Y$, where A is orthogonal and $Y = AX \sim \mathcal{N}(0, I)$. Observe that $P(Z \geq tr(Q) + \sigma t) = P(\sum \lambda_i Y_i^2 \geq \sum \lambda_i + \sigma t)$, since $\lambda_i Y_i^2$ is sub-Exponential with $(2\lambda_i, 4\lambda_i)$, thus $\sum \lambda_i Y_i^2$ is sub-Exponential with $(2\sqrt{\sum \lambda_i^2}, 4\max_i \lambda_i)$. Therefore,

$$|P(\sum \lambda_i Y_i^2 - \sum \lambda_i| \geq t) \leq 2\exp\{-\min(\frac{t^2}{8\sum \lambda_i^2}, \frac{t}{8\max_i \lambda_i})\} = 2\exp\{-\min(\frac{t^2}{8\|Q\|_F^2}, \frac{t}{8\|Q\|_2})\}.$$

Exercise 2.18

(a) From the definition of $\|X\|_{\psi_q}$, we have $\boldsymbol{E}e^{\frac{|X|^2}{\|X\|_{\psi_q}}} \leq 2$. On the other side, we also have $\boldsymbol{E}e^{\frac{|X|^2}{\|X\|_{\psi_q}}} \geq P(|X| > t)\exp(\frac{t^q}{\|X\|_{\psi_q}})$. Thus $P(|X| > t) \leq 2\exp(-\frac{t^q}{\|X\|_{\psi_q}})$.

(b)(*) Let $Y = e^{\frac{|X|^q}{t^q}}$, then

$$\boldsymbol{E}Y = \int_0^{+\infty} P(Y > s)ds = 1 + \int_1^{+\infty} P(|X| > t\ln^{\frac{1}{q}} s)ds$$
$$\leq 1 + \int_1^{+\infty} c_1\exp(-c_2t^q \ln s)ds = \frac{c_1 s^{1-c_2t^q}}{1 - c_2t^q}|_1^{+\infty} + 1$$

$\boldsymbol{E}Y < +\infty$ if $t > (\frac{1}{c_2})^{\frac{1}{q}}$, thus $\|X\|_{\psi_q}$ is finite.

Exercise 2.19

Based on the fact that ψ is strictly increasing convex and the definition of Orlicz norm σ,

$$
\begin{aligned}
\psi\left(\frac{\boldsymbol{E}[\max_{i=1,\dots,n}|X_i|]}{\sigma}\right) &\le \boldsymbol{E}\left[\psi\left(\frac{\max_{i=1,\dots,n}|X_i|}{\sigma}\right)\right] = \boldsymbol{E}\left[\max_{i=1,\dots,n}\psi\left(\frac{|X_i|}{\sigma}\right)\right] \\
&\le \boldsymbol{E}\left[\sum_{i=0}^{n}\psi\left(\frac{|X_i|}{\sigma}\right)\right] \le n \\
&\Rightarrow \boldsymbol{E}\left[\max_{i=1,\dots,n}|X_i|\right] \le \sigma\psi^{-1}(n)
\end{aligned}
$$

Exercise 2.20

Using Markov inequality and Rosenthal's inequality on $(\sum_{i=0}^{n} X_i)^{2m}$,

$$
\begin{aligned}
P(|\sum_{i=0}^{n} X_i| \ge n\delta) &\le \frac{\boldsymbol{E}\left[(\sum_{i=1}^{n} X_i)^{2m}\right]}{(n\delta)^{2m}} \le \frac{R_m\left\{\sum_{i=1}^{n}\boldsymbol{E}\left[X_i^{2m}\right] + (\sum_{i=1}^{n}\boldsymbol{E}\left[X_i^2\right])^m\right\}}{(n\delta)^{2m}} \\
&\le \frac{R_m\left\{nC_m^{2m} + (\sum_{i=1}^{n}\boldsymbol{E}\left[X_i^2\right])^m\right\}}{(n\delta)^{2m}} \\
&\le B_m\left(\frac{1}{\sqrt{n}\delta}\right)^{2m}
\end{aligned}
$$

where B_m is a constant depending on C_m and m.

Exercise 2.21

(a) From exercise 2.9, we have $P(V_j = 1) = P(\rho_H(X, z^j)) \le e^{-nD(\delta\|1/2)}$, thus $P(V \ge 1) = 1 - \prod_{i=1}^{N} P(V_i = 0) \le 1 - (1 - e^{-nD(\delta\|1/2)})^N$, which goes to zero as n goes to infinity.

(b)

(i) Let $P(V \ge 1) := p$, it suffices to show that $p\boldsymbol{E}V^2 \ge (EV)^2$. Obeserve that the left and right side equal to $p^2\boldsymbol{E}[V_+^2], p^2(\boldsymbol{E}V_+)^2$ respectively, then the inequality follows by $Var(V_+) \ge 0$.

(ii) Let $P(V_j = 1) := p$, then the tail bounds from exercise 2.10 show $p \ge \frac{1}{n}e^{-D(\delta\|1/2)}$. By (i), $\boldsymbol{E}V = np, \boldsymbol{E}V^2 = np + n(n-1)p^2$, $P(V \ge 1) \ge \frac{np}{1+(n-1)p}$, thus $P(V \ge 1) \to 1$ as $n \to \infty$.

Exercise 2.22

(a) The convexity can be proved using Hölder's inequality.

(b)(c) gg