Solution 2

Euler Cat

July 17, 2023

1 Solutions to exercises in chapter 2

1.1 2.1.4

The second inequality is trivial from equation Proposition 2.1.2, so we only work on the first one.

Note that:

$$\frac{1}{\sqrt{2\pi}}\int_t^\infty g^2 * e^{-\frac{g^2}{2}} dg \tag{1}$$

$$= \frac{1}{\sqrt{2\pi}}\int_t^\infty g * e^{-\frac{g^2}{2}} d(\frac{g^2}{2}) \tag{2}$$

$$= \frac{1}{\sqrt{2\pi}}\int_t^\infty g * d(-e^{\frac{-g^2}{2}}) \tag{3}$$

$$= \frac{1}{\sqrt{2\pi}}[t * e^{-\frac{t^2}{2}} + \int_t^\infty e^{-\frac{g^2}{2}} dg] \tag{4}$$

$$= \frac{1}{\sqrt{2\pi}} * t * e^{-\frac{t^2}{2}} + P\{g > t\} \tag{5}$$

Where from (3) to (4), we use the integration by parts.

1.2 2.2.7

Before entering into the main branch of proof, we need to first reply on a conclusion called Hoeffding's Lemma which says for each bounded random variable $Z \in [m, M]$ we have $E[exp(\lambda(Z - E[Z]))] \leq exp(\frac{\lambda^2(M-m)^2}{2})$. For a proof, please refer to the lemma 5 in the link.

Now we go back to the main proof by following a standard pipeline of moment generating method.

$$P\{\sum_{i=1}^{N}(X_i - EX_i) \geq t\} \tag{6}$$

$$= P\{\lambda * \sum_{i=1}^{N}(X_i - EX_i) \geq \lambda * t\} \tag{7}$$

$$= P\{exp(\lambda * \sum_{i=1}^{N}(X_i - EX_i)) \geq exp(\lambda * t)\} \tag{8}$$

$$\leq exp(-\lambda * t) * \prod_{i=1}^{N} E[exp(\lambda * (X_i - EX_i))] \tag{9}$$

$$\leq exp(-\lambda * t) * \prod_{i=1}^{N} exp(\frac{\lambda^2(M_i - m_i)^2}{2}) \tag{10}$$

$$= exp(-\lambda * t) * exp(\frac{\lambda^2 * \sum_{i=1}^{i=N}(M_i - m_i)^2}{2}) \tag{11}$$

Where in step (10) we apply the previously mentioned Hoffeding Lemma. Now we just need to optimize with respect to λ we can have

$$P\{\sum_{i=1}^{N}(X_i - EX_i) \geq t\}$$
$$\leq exp(-\frac{t^2}{2 * \sum_{i=1}^{N}(M_i - m_i)^2})$$

Note that our constant here is not optimal.

1.3 2.2.8

By denoting X_i as the indicator function for wrong answer in ith time we mean $X_i = 1$ if it is correct and $X_i = 0$ if it is wrong.

So the wrong probability is as following:

$$
\begin{align}
& P(\sum_{i=1}^{N} X_i \leq \frac{N}{2}) \tag{12} \\
& = P(\sum_{i=1}^{N} (X_i - EX_i) \leq \frac{N}{2} - N(\delta + \frac{1}{2})) \tag{13} \\
& = P(\sum_{i=1}^{N} (X_i - EX_i) \leq -N\delta) \tag{14} \\
& \leq exp(-\frac{N^2\delta^2}{2 * N}) \tag{15} \\
& = exp(-2 * N * \delta^2) \tag{16} \\
& \tag{17}
\end{align}
$$

Where in (17) we use the Hoeffding inequality for general bounded random variable.

Let $exp(-2 * N * \delta^2) \leq \epsilon$ will let us get $N \geq \frac{1}{2}\delta^{-2}ln(\epsilon^{-1})$

1.4 2.2.9

1.4.1

$$P(|\frac{1}{N}\sum_{i=1}^{N} X_i - \mu| > \epsilon) \leq \frac{Var(\frac{1}{N}\sum_{i=1}^{N} X_i)}{\epsilon^2} \tag{18}$$

$$= \frac{Var(X)}{N * \epsilon^2} \tag{19}$$

And the last term should be less than $\frac{1}{4}$, so we can have

$$N \geq \frac{4 * Var(X)}{\epsilon^2} = O(\frac{Var(X)}{\epsilon^2})$$

1.4.2

The proof comes from link.

Suppose that we have k estimates $(\mu_1, ..., \mu_k)$ and each of these estimates is $\frac{3}{4}$-correct. Let us consider the median $\bar{\mu}$ and $X_i = 1(|\mu_i - \mu| > \epsilon)$. As a result of this, we can know that

$$P(|\bar{\mu} - \mu| > \epsilon) \tag{20}$$

$$= P(\sum_{i=1}^{k} X_i > \frac{k}{2}) \tag{21}$$

$$= P(\sum_{i=1}^{k} (X_i - \frac{1}{4}) > \frac{k}{4}) \tag{22}$$

$$<= exp(-\frac{\frac{k^2}{8}}{k}) \tag{23}$$

$$= exp(-\frac{k}{8}) \tag{24}$$

Where in (21) we use the fact that if median surpasses μ then half of them must also be larger than μ and in (22) we use the fact that X_i is an indicator function that has $\frac{1}{4}$ probability and in (23) we use the Hoeffding inequality for bounded random variables with the interval [0,1].

So we only need to let $exp(-\frac{k}{8}) \leq \delta$ then we can have $k \geq \frac{ln(\frac{1}{\delta})}{8}$.

Note that for each of the estimation μ_i, we need to sample at least $0(\frac{\delta^2}{\epsilon^2})$ times. So combined together, we can have $O(ln(\frac{1}{\delta}) * \frac{\delta^2}{\epsilon^2})$ can guarantee the accuracy.

1.5 2.2.10

1.5.1

Note that

$$\begin{aligned}
& Eexp(-tX_i) && (25)\\
& = \int_0^\infty exp(-tx) * p(x)dx && (26)\\
& \leq \frac{1}{t}\int_0^\infty -de^{-tx} && (27)\\
& = \frac{1}{t} && (28)
\end{aligned}$$

1.5.2

Note that

$$\begin{aligned}
& P(\sum_{i=1}^{N} X_i \leq \epsilon * N) && (29)\\
& = P(\sum_{i=1}^{N} \frac{-X_i}{\epsilon} \geq -N) && (30)\\
& = P(e^{\lambda * \sum_{i=1}^{N} \frac{-X_i}{\epsilon}} \geq e^{-\lambda * N}) && (31)\\
& \leq e^{\lambda * N} * \prod_{i=1}^{N} Ee^{-\frac{\lambda * X_i}{\epsilon}} && (32)\\
& \leq e^{\lambda * N} * (\frac{\epsilon}{\lambda})^N && (33)\\
& \leq (\frac{e^\lambda}{\lambda})^N * \epsilon^N && (34)
\end{aligned}$$

By basic calculus, we know that function $f(\lambda) = \frac{e^\lambda}{\lambda}$ achieves minimum value when $\lambda = 1$. So the final result is $(e\epsilon)^N$.

1.6 2.3.2

Note that

$$
\begin{aligned}
&P(S_N \leq t) && (35)\\
&= P(-S_N \geq -t) && (36)\\
&= P(-\lambda * S_N \geq -\lambda * t) && (37)\\
&= P(e^{-\lambda * S_N} \geq e^{-\lambda * t}) && (38)\\
&\leq e^{\lambda * t} * \prod_{i=1}^{N} E e^{-\lambda * X_i} && (39)\\
&= e^{\lambda * t} * \prod_{i=1}^{N} [e^{-\lambda} * p + (1 - p_i)] && (40)\\
&\leq e^{\lambda * t} * e^{(e^{-\lambda} - 1)\mu} && (41)
\end{aligned}
$$

Set $\lambda = ln(\frac{\mu}{t})$ we can have the desired result.

1.7 2.3.3

From theorem 2.3.1, we know that

$$P(S_N \geq t) \leq e^{-\mu}(\frac{e\mu}{t})^t \tag{42}$$

On the other hand, from theorem 1.3.4, we can know that when N approaches infinity, $\mu = ES_N$ will converge to a number λ and S_N will approach $Pois(\lambda)$ in distribution. So (42) will become

$$P(X \geq t) \leq e^{-\lambda} * (\frac{e\lambda}{t})^t \tag{43}$$

Note that the condition $t > \mu$ will become $t > \lambda$.

1.8 2.3.5

The proof is borrowed from link.

Note that

$$P(S_N \geq (1+\delta)*\mu) \leq e^{-\mu}*(\frac{e}{1+\delta})^{(1+\delta)\mu} \tag{44}$$

So in order to let the desired inequality happen, we should have

$$e^{-\mu}*(\frac{e}{1+\delta})^{(1+\delta)\mu} \leq e^{-c\mu\delta^2} \tag{45}$$

Taking log operation on both sizes, we can have

$$-\mu+(1+\delta)\mu(1-ln(1+\delta)) \leq -c\mu\delta^2 \tag{46}$$
$$-1+(1+\delta)(1-ln(1+\delta)) \leq -c\delta^2 \tag{47}$$
$$\delta+c\delta^2 \leq (1+\delta)ln(1+\delta) \tag{48}$$

So our mission is to prove the equation (48). More specifically, we need to take care of the following function:

$$f(x)=\frac{(1+x)ln(1+x)-x}{x^2} \tag{49}$$

By calculus, we know that the function f is a decreasing function in interval $(0,1]$. So we can select any c such that $0<c\leq 2ln2-1$, which will meet our requirement.

1.9 2.3.6

Just taking the limits on both sides of the equation in exercise 2.3.5 with the help of law of large numbers. Besides, we need to make the replacement $\delta = \frac{t}{\mu}$, which will give us the desired results.

1.10 2.3.8

The proof comes from the link. The basic idea still lie in the form of moment generating function.

Note that $M_X(t) = E[e^{tX}] = e^{\lambda(e^t-1)}$ for Possion distribution

As a result of this:

$$
\begin{aligned}
lim_\lambda M_{(X-\lambda)/\sqrt{\lambda}}(t) &= lim_\lambda E[e^{t*\frac{X-\lambda}{\sqrt{\lambda}}}] && (50)\\
&= lim_\lambda e^{-t*\lambda} * Ee^{\frac{t*X}{\sqrt{\lambda}}} && (51)\\
&= lim_\lambda e^{-t*\sqrt{\lambda}} * e^{\lambda*(e^{\frac{t}{\sqrt{\lambda}}}-1)} && (52)\\
&= lim_\lambda e^{-t*\sqrt{\lambda}+\lambda*(\frac{t}{\sqrt{\lambda}}+\frac{t^2}{2*\lambda}+\frac{t^3}{6*\lambda^{\frac{3}{2}}}+...)} && (53)\\
&= e^{\frac{t^2}{2}+f(\frac{1}{\lambda})} && (54)\\
&= e^{\frac{t^2}{2}} && (55)
\end{aligned}
$$

Where in equation (53) we use the Taylor expansion of exp function. And in (54) we note that $\frac{1}{\lambda}$ approaches zero when λ goes into infinity.

1.11 2.4.2

Following the same notations in proposition 2.4.1, we can have:

$$P(d_i \geq C * logn) \leq e^{-d}(\frac{e * d}{C * logn})^{C*logn} \tag{56}$$

Since n here can be quite large, we can have some n such that $d < logn$. Then

$$P(d_i \geq C * logn) \leq e^{-d}(\frac{e}{C})^{C*logn} \tag{57}$$

Then we just need to select one C such that $\alpha = \frac{e}{C} < 1$ then we can get that

$$P(d_i \geq C * logn) \leq \alpha^{C*logn} \tag{58}$$

Making a union bound with respect to all nodes, we can have that

$$P(\exists i, d_i \geq C * logn) \leq n * \alpha^{C*logn} \tag{59}$$

If we make C large enough, then we will have the above probability small enough, which means with high probability the degree of each node in the graph will be controlled by $logn$.

1.12 2.4.3

Similar to the previous exercise, we have

$$
\begin{align}
&P(d_i \geq C * \frac{logn}{loglogn}) \tag{60}\\
&\leq e^{-d} * (\frac{e * d * loglogn}{C * logn})^{C * \frac{logn}{loglogn}} \tag{61}
\end{align}
$$

Since d is a fixed number, we can have $d \leq \sqrt{logn}$ when n is large enough. Note that $loglogn \leq B * logn$ when n is large enough, so we can have

$$
\begin{align}
&P(d_i \geq C * \frac{logn}{loglogn}) \tag{62}\\
&\leq e^{-d} * (\frac{e * B}{C})^{C * \frac{logn}{loglogn}} \tag{63}
\end{align}
$$

Make C large enough, we can have $\alpha = \frac{e*B}{C} < 1$. So,

$$
\begin{align}
&P(d_i \geq C * \frac{logn}{loglogn}) \tag{64}\\
&\leq \alpha^{C * \frac{logn}{loglogn}} \tag{65}
\end{align}
$$

Making C large enough again and union bound, we then have with high probability degree of each node is bounded by $\frac{logn}{loglogn}$.

1.13 2.4.4

The proof comes from the link and link which replies on some arguments with respect to independence among nodes.

First all of, note that by definition we have

$$d = (n-1) * p \tag{66}$$

Since

$$d = o(logn) \tag{67}$$

Then

$$p = o(\frac{logn}{n}) \tag{68}$$

Now suppose we select a group of nodes V' of size $n^{\frac{1}{3}}$ which is a subset of V. So the expected number of edges existed among nodes in V' is $n^{\frac{2}{3}} * p = o(\frac{logn}{n^{1/3}})$ which approaches zero when n goes into infinity. So with high probability, we can assume V' has no edges among its nodes.

Now, we can have that (let $k = n^{\frac{1}{3}}$)

$$P(\forall i \in V', d_i \neq 10d) \tag{69}$$

$$= (1 - C_{n-k}^{10d} * p^{10d} * (1-p)^{n-k-10d})^k \tag{70}$$

Now recall the results from Exercise 0.0.5 which says that

$$C_n^m \leq (\frac{en}{m})^m \tag{71}$$

We can know that

$$C_{n-k}^{10d} * p^{10d} * (1-p)^{n-k-10d} \tag{72}$$

$$\geq (\frac{e * (n-k) * p}{10d})^{10d} * (1-p)^n \tag{73}$$

$$\geq (\frac{e}{20})^{10 * \epsilon * logn} * (1 - \frac{\epsilon * logn}{n})^n \tag{74}$$

Where in the last step, we let n large enough such that $n - k > \frac{n-1}{2}$. By using the inequality $1 - x > e^{-2x}, 0 < x < \frac{1}{2}$ we can have

$$(\frac{e}{20})^{10*\epsilon*logn} * (1 - \frac{\epsilon * logn}{n})^n \geq e^{-10*\epsilon*log(\frac{20}{e})*logn} * e^{-2*\epsilon*logn} = n^{-C*\epsilon} \tag{75}$$

Where C here is some constant number greater than zero. As a result of this, we can have

$$P(\forall i \in V', d_i \neq 10d) \leq (1 - n^{-C*\epsilon})^{n^{\frac{1}{3}-C*\epsilon+C*\epsilon}} \leq e^{-n^{\frac{1}{3}-C*\epsilon}} \tag{76}$$

Making ϵ small enough such that $\frac{1}{3} - C*\epsilon > 0$ we can get our results by letting n approaches ∞.

1.14 2.4.5(TODO)

Following the same notations as above, we extract a subset V' from V such that with high probability there is no edge among nodes of V'.

So we have(let $k = n^{\frac{1}{3}}$)

$$P(\exists i \in V', d_i \geq \frac{logn}{loglogn}) \tag{77}$$

$$= 1 - P(\forall i \in V', d_i < \frac{logn}{loglogn}) \tag{78}$$

$$= 1 - P(d_i < \frac{logn}{loglogn})^k \tag{79}$$

$$= 1 - (1 - P(d_i \geq \frac{logn}{loglogn}))^k \tag{80}$$

$$\tag{81}$$

Now for the tail distribution in (80), we should apply Chernoff inequality in which case we have

$$P(\forall i \in V', d_i < \frac{logn}{loglogn}) = (1 - P(d_i \geq \frac{logn}{loglogn}))^k \tag{82}$$

$$\leq (1 - P(d_i = \frac{logn}{loglogn}))^k \tag{83}$$

Let $d' = \frac{logn}{loglogn}$, then

$$C_{n-k}^{d'} * p^{d'} * (1-p)^{n-k-d'} \tag{84}$$

$$\geq (\frac{(n-k)*p}{d'})^{d'} * (1-p)^n \tag{85}$$

Since $(n-1)*p = d = O(1)$, we have $p = O(\frac{1}{n})$. Then, let n be enough large such that $n - k > \frac{n-1}{2} * \frac{2}{d}$.

$$(85) \geq (\frac{1}{d'})^{d'} * (1 - \frac{C}{n})^n \tag{86}$$

$$\geq (\frac{1}{d'})^{d'} * e^{-2*C} \tag{87}$$

1.15 2.5.1

Making the replacement $t = \frac{x^2}{2}$ will lead to the result.

The second part of the proof comes from link. The main idea here is to utilize the log-convexity of gamma function:

$$\Gamma(\frac{p+1}{2}) \leq \sqrt{\Gamma(1) * \Gamma(p)} \tag{88}$$

$$= \sqrt{\Gamma(p)} \tag{89}$$

$$\leq \sqrt{\Gamma(p+1)} \tag{90}$$

$$\leq \sqrt{3 * (\frac{3 * p}{5})^p} \tag{91}$$

Where in the last step we bound the gamma function by the conclusions in the link.

1.16 2.5.4

By Jensen inequality, we can have

$$exp(E\lambda * X) \leq exp(K^2 * \lambda^2) \tag{92}$$

Note that λ can be less than or greater than zero, so $EX = 0$ is necessary.

1.17 2.5.5

1.17.1

Note that

$$Ee^{(\lambda^2 * X^2)} \tag{93}$$
$$= \frac{1}{\sqrt{2 * \pi}} * \int e^{(\lambda - \frac{1}{2}) * x^2} dx \tag{94}$$

To guarantee a convergence of the above integral, we can know that $\lambda < \frac{1}{2}$, which is just a bounded neighbour of zero.

1.17.2

Note that(let C be an arbitrary number)

$$P(|X| > \sqrt{C}) \tag{95}$$
$$= P(|X|^2 > C) \tag{96}$$
$$= P(\lambda^2 * |X|^2 > \lambda^2 * C) \tag{97}$$
$$\leq e^{-\lambda^2 * C} * E[\lambda^2 * |X|^2] \tag{98}$$
$$\leq e^{-\lambda^2 * C} * e^{(K * \lambda^2)} \tag{99}$$
$$= e^{(K * \lambda^2 - \lambda^2 * C)} \tag{100}$$

So we just need to let $C = K + \epsilon$, then

$$(100) = e^{-\epsilon * \lambda^2} \tag{101}$$

Since λ can be arbitrary large, we can have that $|X| \leq \sqrt{C + \epsilon}$ almost everywhere.

1.18 2.5.7

Note that $\frac{C*X^2}{C*t^2} = \frac{X^2}{t^2}$ so $||C*X||_{\phi_2} = |C| * ||X||_{\phi_2}$. And the only way that can let $t = 0$ is that $X = 0$ almost everywhere. So the remained thing is to prove $||X+Y||_{\phi_2} \le ||X||_{\phi_2} + ||Y||_{\phi_2}$.

Note that

$$Ee^{\frac{(X+Y)^2}{(||X||+||Y||)^2}} \tag{102}$$

$$= Ee^{(\frac{X+Y}{|X|+|Y|})^2} \tag{103}$$

$$= Ee^{(\frac{|X|}{|X|+|Y|}*\frac{X}{|X|}+\frac{|Y|}{|X|+|Y|}*\frac{Y}{|Y|})^2} \tag{104}$$

Note that the function $f(x) = e^{x^2}$ is a convex function, so

$$(104) \le E[\frac{|X|}{|X|+|Y|} * e^{(\frac{X}{|X|})^2} + \frac{|Y|}{|X|+|Y|} * e^{(\frac{Y}{|Y|})^2}] \tag{105}$$

$$\le 2 * \frac{|X|}{|X|+|Y|} + 2 * \frac{|Y|}{|X|+|Y|} \tag{106}$$

$$= 2 \tag{107}$$

Where in (106) we use the assumption $||X||_{\phi_2} \le 2, ||Y||_{\phi_2} \le 2$

1.19 2.5.9

1.19.1 Poisson

The proof comes from link. Without loss of generality we assume that t here is an integer number.

Note that

$$P(X \geq t) \geq P(X = t) \tag{108}$$
$$= \frac{e^{-\lambda} * \lambda^t}{t!} \tag{109}$$
$$\approx \frac{e^{-\lambda} * \lambda^t}{\sqrt{2 * \pi * t}} * (\frac{e}{t})^t \tag{110}$$
$$= \frac{e^{-\lambda}}{\sqrt{2 * \pi * t}} * (\frac{e * \lambda}{t})^t \tag{111}$$
$$= \frac{e^{-\lambda}}{\sqrt{2 * \pi * t}} * e^{t + t * ln\lambda - t * lnt} \tag{112}$$

Clearly, the power term here can not match up with $\frac{t^2}{2}$, so it can not be a sub-gaussian random variable.

1.19.2 Exponential

Note that

$$P(X \geq t) = e^{-\lambda * t} \tag{113}$$

Then, if the RHS of the above equation is smaller than $e^{-\frac{t^2}{2}}$ then λ should be arbitrary large, which is absurd.

1.19.3 Pareto

By definition, the Pareto distribution is $P(X \geq t) = (\frac{C}{t})^{\alpha}, t \geq C$ and $P(X \geq t) = 0, t \leq C$. So we focus on the part $t \geq C$.

Note that $(\frac{C}{t})^{\alpha} = e^{\alpha * (lnC - lnt)}$ and $lnt < \frac{t^2}{2}$.

1.19.4 Cauchy

It is well known that Cauchy distribution does not have mean and variance but on the other hand sub-gaussian distribution ineed have finite variance.

1.20 2.5.10

1.20.1

The proof comes from link. By using the layer cake representation of integral, we can have

$$
\begin{aligned}
Emax\frac{|X_i|}{\sqrt{1+logi}} &= \int_0^\infty P(\frac{max|X_i|}{\sqrt{1+logi}} \geq t)dt && (114)\\
&= \int_0^{t_0} dt + \int_{t_0}^\infty dt && (115)\\
&\leq t_0 + \sum_i \int_{t_0}^\infty P(|X_i| \geq t * \sqrt{1+logi})dt && (116)\\
&\leq t_0 + \int_{t_0}^\infty \sum_i 2 * e^{-\frac{t^2*(1+logi)}{||X_i||_{\phi_2}^2}} dt && (117)\\
&\leq t_0 + \int_{t_0}^\infty \sum_i 2 * e^{-\frac{t^2*(1+logi)}{K^2}} dt && (118)\\
&\leq t_0 + \sum_i 2 * \int_{t_0}^\infty e^{-\frac{t^2}{K^2}} * \frac{1}{i}^{\frac{t_0^2}{K^2}} dt && (119)\\
&= t_0 + \int_{t_0}^\infty e^{-\frac{t^2}{K}} dt * \sum_i 2 * \frac{1}{i}^{\frac{t_0^2}{K^2}} && (120)
\end{aligned}
$$

Let $t_0 = 2 * K$, then we can have

$$
Emax\frac{|X_i|}{\sqrt{1+logi}} \leq C * K \qquad (121)
$$

1.20.2

By the previous inequality, and note that

$$
\begin{aligned}
Emax\frac{|X_i|}{\sqrt{1+logi}} &\geq E\frac{max|X_i|}{max\sqrt{1+logi}} && (122)\\
&= \frac{Emax|X_i|}{\sqrt{1+logN}} && (123)\\
&\geq \frac{Emax|X_i|}{\sqrt{3*logN}} && (124)\\
& && (125)
\end{aligned}
$$

$$
E[max|X_i|] \leq C * K * \sqrt{logN} \qquad (126)
$$

1.21 2.5.11

The proof comes from link.

Note that

$$
\begin{aligned}
Emax|X_i| &= \int P(max|X_i| > t)dt && (127)\\
&= \int (1 - P(|X_i| \leq t)^N)dt && (128)\\
&= \int (1 - (1 - P(|X_i| > t))^N)dt && (129)
\end{aligned}
$$

Also note that,

$$
\begin{aligned}
P(|X| > t) &= \frac{2}{\sqrt{2 * \pi}} * \int_0^\infty e^{-\frac{|t+x|^2}{2}} dx && (130)\\
&\geq \frac{2}{\sqrt{2 * \pi}} * \int_0^1 e^{-\frac{|t+x|^2}{2}} dx && (131)\\
&\geq \frac{2}{\sqrt{2 * \pi}} * e^{-\frac{|1+t|^2}{2}} && (132)\\
&\geq \frac{2}{\sqrt{2 * \pi}} * C * e^{-t^2} && (133)
\end{aligned}
$$

Where in the last step, we determine the constant C such that

$$(t + 1)^2 \leq 2 * t^2 - lnC^2 \quad (134)$$

Which is equivalent to say

$$lnC^2 \leq t^2 - 2 * t - 1, x > 0 \quad (135)$$

So $lnC^2 \leq -2$ which means $0 < C < e^{-1}$. Now, go back to (129).

$$
\begin{aligned}
Emax|X_i| &\geq \int (1 - (1 - c * e^{-t^2})^N)dt && (136)\\
& && (137)
\end{aligned}
$$

To induce the $\sqrt{logN}$ term into our formulation, we make the replacement $t = \sqrt{logN} * u$ which gives us

$$
\begin{aligned}
Emax|X_i| &\geq \sqrt{logN} * \int (1 - (1 - \frac{c}{N^{u^2}})^N)du && (138)\\
&\geq \sqrt{logN} * \int (1 - \alpha^{\frac{1}{u^2}})du && (139)\\
& && (140)
\end{aligned}
$$

Here $\alpha = e^{-c} < 1$. So by the numerical inequality $e^{-2*x} < 1 - x$ when x is small enough(or u is large enough), we can know that

$$\alpha^{\frac{1}{u^2}} < 1 - \frac{C}{u^2} \tag{141}$$

For some constant $C > 0$. As a result of this, we can know that

$$Emax|X_i| \geq \sqrt{logN} * \int_M \frac{C}{u^2} du \tag{142}$$

And the conclusion follows.

1.22 2.6.4

Note that $m_i \leq X_i \leq M_i, EX_i$ implies $m_i \leq 0 \leq M_i$. As a result of this, $||X||_{\phi_2} \leq C * ||X||_\infty \leq C * |M_i - m_i|$ and apply theorem 2.6.2 will suffice to get our conclusion

In this exercise, I do not utilize theorem 2.6.3.

1.23 2.6.5

Since $p \geq 2$, the first inequality can be proved by using the inclusion between L_2 and L_P space.

For the second inequality, we need layer cake trick again:

$$E|\sum_{i=1}^{N} a_i * X_i)|^p = \int_0^\infty p * t^{p-1} * P(|\sum_{i=1}^{N} a_i * X_i| \geq t)dt \tag{143}$$

$$\leq \int_0^\infty p * t^{p-1} * 2 * e^{-\frac{ct^2}{K^2||a||_2^2}} dt \tag{144}$$

$$= 2p * \int_0^\infty t^{p-1} * e^{-\frac{ct^2}{K^2||a||_2^2}} dt \tag{145}$$

$$\tag{146}$$

Note that

$$\int_0^\infty t^{p-1} * e^{-\frac{ct^2}{K^2||a||_2^2}} \tag{147}$$

$$= (k * ||a||_2 * c)^p \int_0^\infty y^{\frac{p}{2}-1} e^{-y} dy \tag{148}$$

$$= (k * ||a||_2 * c)^p * \Gamma(\frac{p}{2}) \tag{149}$$

Where in the second-to-last inequality, we use the replacement $y = \frac{ct^2}{K^2 * ||a||_2^2}$.

Finally, we just need to take the pth root on both sides and note that $\Gamma(\frac{p}{2})^{\frac{1}{p}} \leq \sqrt{p}$(actually, this is a conclusion from exercise 2.5.1).

1.24 2.6.6

The second inequality comes from the inclusion relationship $L_1 \subset L_2$.

For the first inequality, we note that

$$||Z||_1 \leq ||Z||_1^{\frac{1}{4}} * ||Z||_3^{\frac{3}{4}} \tag{150}$$

This can be proved by splitting $Z^2 = Z^{\frac{1}{2}} * Z^{\frac{3}{2}}$ and Cauchy-Schwarz inequality.

Now, let $Z = \sum a_i * X_i$ and apply exercise 2.6.5, we can have

$$||Z||_3^{\frac{3}{4}} \leq C * K * (\sum_{i=1}^{N} a_i^2)^{\frac{3}{8}} \tag{151}$$

So

$$||\sum_{i=1}^{N} a_i * X_i||_{L_1} \geq c(K) * (\sum_{i=1}^{N} a_i^2)^{\frac{1}{2}} \tag{152}$$

1.25 2.6.7

The same as before, we can know that $||\sum_{i=1}^{N} a_i * X_i||_{L_p} \leq (\sum_{i=1}^{N} a_i^2)^{\frac{1}{2}}$ due to the inclusion relationship.

For the first inequality, note that $0 < p < 2$ so $4 - p > 2$. We need to split 2 into two parts $\frac{p}{2}$ and $\frac{4-p}{2}$ and apply Cauchy-Schwarz inequality:

$$\int |Z|^2 = \int |Z|^{\frac{p}{2}} * |Z|^{\frac{4-p}{2}} \tag{153}$$

$$\leq (\int |Z|^p)^{\frac{1}{2}} * (\int |Z|^{4-p})^{\frac{1}{2}} \tag{154}$$

So

$$(\int |Z|^p)^{\frac{1}{2}} \geq \frac{\int |Z|^2}{(\int |Z|^{4-p})^{\frac{1}{2}}} \tag{155}$$

$$\geq \frac{\sum_{i=1}^{N} a_i^2}{(C * K * \sqrt{4-p} * (\sum a_i^2)^{\frac{1}{2}})^{\frac{4-p}{2}}} \tag{156}$$

Then

$$(\int |Z|^p)^{\frac{1}{p}} \geq \frac{(\sum_{i=1}^{N} a_i^2)^{\frac{2}{p}}}{(C * K * \sqrt{4-p} * (\sum a_i^2)^{\frac{1}{2}})^{\frac{4-p}{p}}} \tag{157}$$

$$= C(K, \sqrt{p}) * (\sum_{i=1}^{N} a_i^2)^{\frac{1}{2}} \tag{158}$$

Where $C(K, \sqrt{p})$ here is a constant related to K and $\sqrt{p}$.

1.26 2.6.9

The proof comes from link. The basic idea is to give some counter-example.

Consider a random variable that takes value a and $-a$ with probability p and $1-p$ respectively. So we can get its expectation of second-order moment generating function

$$Ee^{|X|^2} = p * e^{a^2} + (1-p) * e^{a^2} = e^(a^2). \tag{159}$$

Let $a = \sqrt{log2}$, then we can have $Ee^{|X|^2} = 2$. Since e^x is monotone, we know that $||X||_{\phi_2} = 1$. So we only need to prove $Ee^{\frac{(X-EX)^2}{1}} > 2$.

Note that:

$$EX = a * p - a * (1-p) = a * (2 * p - 1) \tag{160}$$

So $X - EX$ takes value $2 * a * (1-p)$ and $-2 * a * p$ with probability p and $1-p$ respectively. So

$$\begin{aligned} Ee^{|X-EX|^2} &= e^{4*a^2*(1-p)^2} * p + e^{4*a^2*p^2} * (1-p) & (161)\\ &= 2^{4*(1-p)^2} * p + 2^{4*p^2} * (1-p) & (162) \end{aligned}$$

Let $p = \frac{1}{4}$, then we can get

$$Ee^{|X-EX|^2} = \frac{7}{4} * 2^{\frac{1}{4}} > 2 \tag{163}$$

1.27 2.7.2(TODO)

TO DO

1.28 2.7.3(TODO)

TO DO

1.29 2.7.4

We argue by contradiction. If it can be extended to all $|\lambda| \leq \frac{1}{K_3}$. Then by Jensen's inequality, we can have

$$\lambda * E|X| \leq K_3 * \lambda \tag{164}$$

So $E|X| = 0$. Clearly, this means $X = 0$ almost everywhere. That tells us in general λ can not be extended to negative domain.

1.30 2.7.10

$$||X - EX||_{\phi_1} \leq ||X||_{\phi_1} + ||EX||_{\phi_1} \tag{165}$$
$$\leq ||X||_{\phi_1} + C * ||X||_{\phi_1} \tag{166}$$

Where the last step comes from

$$exp(\frac{E|X|}{||X||_{\phi_2}}) \leq Eexp^{\frac{|X|}{||X||_{\phi_2}}} \leq 2 \tag{167}$$

And

$$E|X| \leq ln2 * ||X||_{\phi_2} \tag{168}$$

As well as $||EX||_{\phi_2} \leq 2 * |EX| \leq 2 * E|X|$.

1.31 2.7.11

Just replace the e^{x^2} with ϕ in the proof of exercise 2.5.7.

1.32 2.8.5

Using the inequality from hints, we can have

$$
\begin{aligned}
Ee^{\lambda * X} &\leq 1 + \frac{\lambda^2 * EX^2/2}{1 - \frac{|\lambda * X|}{3}} && (169)\\
&\leq 1 + \frac{\lambda^2 * EX^2/2}{1 - \frac{\lambda * K}{3}} && (170)\\
&\leq e^{\frac{\lambda^2 * EX^2/2}{1 - \frac{\lambda * K}{3}}} && (171)\\
&= e^{(g(\lambda) * EX^2)} && (172)
\end{aligned}
$$

Note that we use the inequality $1 + x \leq e^x, x > 0$. So we need to guarantee that $|\lambda| < \frac{3}{K}$.

1.33 2.8.6

Note that

$$
\begin{aligned}
P(\sum_{i=1}^{N} X_i \geq t) &= P(\lambda * \sum_{i=1}^{N} X_i \geq \lambda * t) && (173)\\
&= P(e^{\lambda * \sum_{i=1}^{N} X_i} \geq e^{\lambda * t}) && (174)\\
&\leq \frac{\prod_{i=1}^{N} e^{\lambda * X_i}}{e^{\lambda * t}} && (175)\\
&\leq \frac{\prod_{i=1}^{N} e^{g(\lambda) * EX_i^2}}{e^{\lambda * t}} && (176)\\
&= \frac{e^{g(\lambda) * \sigma^2}}{e^{\lambda * t}} && (177)\\
&= e^{g(\lambda) * \sigma^2 - \lambda * t} && (178)\\
&= e^{-\lambda * t + \frac{\frac{\sigma^2}{2} * \lambda^2}{1 - |\lambda| * K/3}} && (179)
\end{aligned}
$$

Now, we recall a lemma(lemma 14 in [1] or lemma 12 in [2])

Theorem 1 *Let C,b denote two positive real number, $t > 0$. Then*

$$inf_{\beta \in [0,1/b)}(-\beta * t + \frac{C * \beta^2}{1 - b * \beta}) \leq -\frac{t^2}{2 * (2 * C + b * t)} \tag{180}$$

So (note that RHS of (173) actually requires $\lambda \geq 0$, so we can actually remove the $||$) we can have

$$P(\sum_{i=1}^{N} X_i \geq t) \leq e^{-\frac{t^2}{2 * (\sigma^2 + K * t/3)}} \tag{181}$$

Then

$$P(|\sum_{i=1}^{N} X_i| \geq t) \leq 2 * e^{-\frac{t^2}{2 * (\sigma^2 + K * t/3)}} \tag{182}$$

References

[1] Maurer, Andreas, and Massimiliano Pontil. "Concentration inequalities under sub-Gaussian and sub-exponential conditions." Advances in Neural Information Processing Systems 34 (2021): 7588-7597.

[2] Maurer, Andreas. "Concentration inequalities for functions of independent variables." Random Structures and Algorithms 29.2 (2006): 121-138.