Problem 1 [25 points: 5+10+10]

(i) Let $W = -X$ with mean $\nu = -\mu$, and note that if $X$ is sub-Gaussian then, if the inequality is true for $t \in \mathbb{R}$, then it is also true for $-t$. Thus

$$
\log \left( \mathbb{E} \left[ e^{t(W-\nu)} \right] \right) = \log \left( \mathbb{E} \left[ e^{-t(X-\mu)} \right] \right) \leq \frac{t^2 \sigma^2}{2},
$$

sub-Gaussianity of $X$

This is true $\forall t \in \mathbb{R}$, implying that $W$ is sub-Gaussian. The vice versa “$W$ sub-Gaussian $\Rightarrow X$ sub-Gaussian” can be proved with identical steps.

(ii) For any $c > 0$

$$
P(X - \mu \geq t) = P(c(X - \mu) \geq ct) = P\left(e^{c(X-\mu)} \geq e^{ct}\right)
$$

[Markov’s inequality] $\leq \mathbb{E}[e^{c(X-\mu)}] e^{-ct}$

[sub-Gaussianity of $X$] $\leq e^{\frac{c^2 \sigma^2}{2} - ct}$

Since this inequality holds for any $c > 0$, we can find $c^*$ that minimize the upper bound. Let $g(c) = \frac{c^2 \sigma^2}{2} - ct$. We have

- $\frac{\partial g}{\partial c} = c \sigma^2 - t = 0 \Rightarrow c^* = \frac{t}{\sigma^2}$
- $\frac{\partial^2 g}{\partial c^2} = \sigma^2 > 0 \Rightarrow$ strict convexity $\Rightarrow c^*$ is the global minimum

Thus the sharpest upper bound (in this Problem) is obtained by plugging $c^*$ into the formula above:

$$
P(X - \mu \geq t) \leq e^{-\frac{t^2}{2 \sigma^2}}$$
(iii) Let $\sigma^2_X, \sigma^2_Y$ be the constants of the sub-Gaussianity bounds for $X$ and $Y$, respectively. We have
\[
E \left[ e^{t(X+Y-\mu-\nu)} \right] = E \left[ e^{tX} \right] E \left[ e^{tY} \right] \quad [X, Y \text{ are independent}]
\leq e^{-\frac{t^2 \sigma_X^2}{2}} e^{-\frac{t^2 \sigma_Y^2}{2}} \quad \text{[sub-Gaussianity of } X, Y]\]
\[
eq e^{\frac{t^2(\sigma_X^2+\sigma_Y^2)}{2}}
\]
showing that $X+Y$ is sub-Gaussian with constant $\sigma^2 = \sigma_X^2 + \sigma_Y^2$.

**Problem 2 [20 points]**

Note that $Y_i \sim \text{Bernoulli}(\theta_n)$ with mean $\theta_n$ and variance $\theta_n(1-\theta_n)$, where $\theta_n$ is a non-increasing sequence with $\lim_{n \to \infty} \theta_n = \lim_{n \to \infty} \frac{1}{n^2} \int_{-1/n^2}^{1/n^2} p(x) dx = 0$ since the density $p$ of $X$ is bounded (so the distribution of $X$ cannot be a point mass on 0). Since $|Y_i| \leq 1$, then

- Hoeffding’s inequality
  \[
P(\bar{Y}_n - \theta_n > t) \leq P(|\bar{Y}_n - \theta_n| > t) \leq 2e^{-2nt^2}
  \]

- Bernstein’s inequality
  \[
P(\bar{Y}_n - \theta_n > t) \leq P(|\bar{Y}_n - \theta_n| > t) \leq 2\exp \left\{ -\frac{nt^2}{2\theta_n(1-\theta_n) + 2t/3} \right\}
  \]

Bernstein’s bound is tighter than Hoeffding’s if
\[
t < 3 \left( \frac{1}{4} - \theta_n(1-\theta_n) \right) = g(n)
\]  

There exist $n^*$ such that $\theta_{n^*} - 1 \geq \frac{1}{2} \geq \theta_{n^*}$, so that for $n << n^*$ and $n \gg n^*$ we get $g(n) \approx \frac{3}{4}$ and for $n \approx n^*$ we have $g(n) \approx 0$. Thus the result depends on the sequence $\theta_n$: if we are interested in small deviations of $\bar{Y}_n$ from its mean $\theta_n$, then Bernstein’s bound is tighter than the Hoeffding’s one for sufficiently small or sufficiently large $n$. However, we can also see that $\theta_n(1-\theta_n) \leq \theta_n \leq \sup_x p(x) \int_{-1/n^2}^{1/n^2} dx = 2 \sup_x p(x)/n^2$, i.e. $\theta_n(1-\theta_n) = O(n^{-2})$, so that for large $n$ we have $\frac{nt^2}{2\theta_n(1-\theta_n) + 2t/3} \sim \frac{3}{2} nt$, which implies that for $t < 3/4$ and $n$ large, Bernstein’s inequality might be preferred to Hoeffding’s.
Problem 3 [40 points: 0+15+0+15+10+0]

(i) Since \( X_n/a_n = O_P(1) \) then \( \forall \delta > 0, \exists C > 0 \) such that, \( \mathbb{P}(|X_n/a_n| > C) < \delta \). Moreover, since \( Y_n/b_n = O(1) \), then \( \exists B > 0 \) such that for \( n \) sufficiently large, \( \left| \frac{Y_n}{b_n} \right| \leq B \). Thus for sufficiently large \( n \):

\[
P(|X_n Y_n|/|a_n b_n| > BC) \leq P(B|X_n/a_n| > BC)
= P(|X_n/a_n| > C)
< \delta
\]

so that for large \( n, \forall \delta > 0, \exists C^* = BC > 0 \) such that \( P(|X_n Y_n|/|a_n b_n| > C^*) < \delta \).

(ii) There exists constants \( C_X, C_Y \) such that \( \forall \delta > 0, \mathbb{P}(|X_n/a_n| > C) < \delta \) and \( \mathbb{P}(|Y_n/b_n| > C) < \delta \). Note that

\[
|X_n + Y_n|/\max\{a_n, b_n\} > C_X + C_Y \Rightarrow |X_n/\max\{a_n, b_n\}| > C_X \text{ or } |Y_n/\max\{a_n, b_n\}| > C_Y
\]

\[
\Rightarrow |X_n/a_n| > C_X \text{ or } |Y_n/b_n| > C_Y
\]

since \( \max\{a_n, b_n\} \geq a_n, b_n \). Thus, for large \( n \), by the union bound we have that:

\[
\mathbb{P}(\frac{|X_n + Y_n|}{\max\{a_n, b_n\}} > C_X + C_Y) \leq
\leq \mathbb{P}(\frac{|X_n/\max\{a_n, b_n\}| > C_X} + \mathbb{P}(\frac{|Y_n/\max\{a_n, b_n\}| > C_Y) \leq
\leq \mathbb{P}(\frac{|X_n/a_n| > C_X} + \mathbb{P}(\frac{|Y_n/b_n| > C_Y) < 2\delta
\]

Since \( \delta \) was arbitrary this proves the claim.

(iii) The claim is false. Let \( X_n = a_n = 1 \) and \( Y_n = n, b_n = n^2 \). Then we have that \( X_n Y_n = n \neq o_P(1) = o_P(a_n) \).
(iv) $X_n = o_P(a_n), Y_n = o_P(b_n)$, that is $\forall \varepsilon, \delta > 0$ there exists $n_x^*, n_y^*$ such that if $n > n_x^*, n_y^*$ then

$$\mathbb{P}(|X_n/a_n| > \varepsilon) < \delta$$

and the claim is proved.

(v) This proposition is false. We provide two counterexamples.

- Take $X_n = 1/n^2, Y_n = 1/n^4, a_n = 1/n$ and $b_n = 1/n^2$. Then

$$Y_n = o_P(b_n) \implies Y_n = O_P(b_n)$$

Now note that

$$\left| \frac{X_n b_n}{Y_n a_n} \right| = n$$

Which goes to infinity with probability 1.

- Take $Y_n = X_n^2$ and $b_n = a_n^2$. Then

$$Y_n = o_P(b_n) \implies Y_n = O_P(b_n)$$

Now note that $X_n/Y_n = 1/X_n \neq o_P(1/a_n)$ since for $\varepsilon > 0$

$$\mathbb{P}(|a_n/X_n| > \varepsilon) = \mathbb{P}(1/\varepsilon > |X_n/a_n|) \to 1$$

(vi) Case $X_n = O_P(a_n), Y_n = O_P(b_n)$ \implies $X_n Y_n = O_P(a_n b_n)$:

We need to show $\forall \varepsilon > 0, \lim_{n \to \infty} \mathbb{P}\left(\left| \frac{X_n Y_n}{a_n b_n} \right| > \varepsilon \right) = 0$, i.e. $\forall \varepsilon, \delta > 0, \exists N \in \mathbb{N} \text{ s.t. } \forall n \geq N,

$$\mathbb{P}\left(\left| \frac{X_n Y_n}{a_n b_n} \right| > \varepsilon \right) < \delta.$$ 

Let $\varepsilon, \delta > 0$ be fixed. Since $X_n = O_P(a_n)$, then $\exists C > 0$ such that for large $n$, say $n \geq n_x^*$,

$$\mathbb{P}\left(\left| \frac{X_n}{a_n} \right| > C \right) < \frac{\delta}{2}. $$

Moreover $Y_n = o_P(b_n)$, so that $\exists n_y^*$ s.t. $\forall n \geq n_y^*$, $\mathbb{P}\left(\left| \frac{Y_n}{b_n} \right| > \frac{\varepsilon}{C} \right) < \frac{\delta}{2}$. 

4
Since \( \frac{X_n Y_n}{a_n b_n} > \varepsilon \Rightarrow \frac{X_n}{a_n} > C \) or \( \frac{Y_n}{b_n} > \varepsilon \), then \( \forall n \geq N = \max\{n_X^*, n_Y^*\} \),

\[
 P \left( \frac{X_n Y_n}{a_n b_n} > \varepsilon \right) \leq P \left( \left\{ \frac{X_n}{a_n} > C \right\} \cup \left\{ \frac{Y_n}{b_n} > \varepsilon \right\} \right)
\leq P \left( \frac{X_n}{a_n} > C \right) + P \left( \frac{Y_n}{b_n} > \varepsilon \right)
< \frac{\delta}{2} + \frac{\delta}{2} = \delta
\]

Note that we can find such \( N \) for any pair \( \varepsilon, \delta > 0 \), where \( \delta \) can be chosen arbitrarily close to 0. Therefore \( \forall \varepsilon > 0, \lim_{n \to \infty} P \left( \frac{X_n Y_n}{a_n b_n} > \varepsilon \right) = 0 \), i.e. \( X_n Y_n = o_p(a_n b_n) \).

**Problem 4 [15 points]**

In the case where \( F \) is continuous and also strictly continuous, to show \( F(X) \sim \text{Uniform}(0, 1) \) we can just exploit the invertibility of \( F \) to obtain that for \( p \in (0, 1) \)

\[
P(F(X) \leq p) = P(X \leq F^{-1}(p)) = p
\]

(2)

where \( F^{-1} \) is the inverse of \( F \) and so the quantile function.

In the more general case where the cdf \( F \) is assumed to be only continuous, but not necessarily strictly increasing (e.g. if \( F \) is flat on some interval \([a, b]\)), then \( F \) is not strictly increasing but just non-decreasing), we need more steps. We start showing that if \( F \) is continuous, then \( \forall p \in (0, 1) \) we have \( F(x) \geq p \Leftrightarrow x \geq F^{-1}(p) \), where \( F^{-1}(p) = \inf \{ t : F(t) \geq p \} \) is the generalized inverse of \( F \) and the quantile function.

- “\( F(x) \geq p \Rightarrow x \geq F^{-1}(p) \)”. Note that \( F^{-1}(F(x)) \leq x \) and that \( F^{-1}(p) \) is strictly increasing in \( p \in (0, 1) \). Thus

\[
F(x) \geq p \Leftrightarrow F^{-1}(F(x)) \geq F^{-1}(p) \Rightarrow x \geq F^{-1}(p)
\]

- “\( F(x) \geq p \Leftrightarrow x \geq F^{-1}(p) \)”. Note that \( F(F^{-1}(p)) = P(X \leq \inf \{ t : F(t) \geq p \}) \geq p, \forall p(0, 1) \). Thus

\[
x \geq F^{-1}(p) \Rightarrow F(x) \geq F(F^{-1}(p)) \geq p
\]

Now we can complete the proof

\[
P(F(X) > p) = P(X > F^{-1}(p)) = p
\]

where we used the fact that \( F^{-1} \) is strictly monotone and \( X \) is continuous.
**Problem 5 [0 points]**

For all \( A, B \)

\[
P(g(X) \in A, h(Y) \in B) = P(X \in g^{-1}(A), Y \in h^{-1}(B))
\]

\[
= P(X \in g^{-1}(A))P(Y \in h^{-1}(B))
\]

\[
= P(g(X) \in A)P(h(X) \in B)
\]

where \( g^{-1}(A) = \{ x \in \mathcal{X} : g(x) \in A \} \) and \( h^{-1}(B) = \{ y \in \mathcal{Y} : h(y) \in B \} \).

**Problem 6 [BONUS 20 points]**

(a) From hint,

\[
L(X_1, \ldots, X_i, \ldots, X_n) - L(X_1, \ldots, \hat{X}_i, \ldots, X_n)
\]

\[
\leq \left[ L(\{X_1, \ldots, X_n\} \setminus \{X_i\}) + 2 \min_{j > i} \|X_i - X_j\| \right] - L(\{X_1, \ldots, X_n\} \setminus \{X_i\})
\]

\[
= 2 \min_{i < j \leq n} \|X_i - X_j\|,
\]

and by symmetry, \( L(X_1, \ldots, X_i, \ldots, X_n) - L(X_1, \ldots, \hat{X}_i, \ldots, X_n) \geq -2 \min_{i < j \leq n} \|\hat{X}_i - X_j\| \).

Therefore

\[
Y_i = \mathbb{E} \left[ L(X_1, \ldots, X_i, \ldots, X_n) - L(X_1, \ldots, \hat{X}_i, \ldots, X_n) | X_1, \ldots, X_i \right]
\]

\[
\leq 2 \mathbb{E} \left[ \min_{i < j \leq n} \|X_i - X_j\| | X_1, \ldots, X_i \right]
\]

\[
= 2 \mathbb{E} \left[ \min_{i < j \leq n} \|X_i - X_j\| | X_i \right] = 2g_{n-i}(X_i)
\]

and

\[
Y_i = \mathbb{E} \left[ L(X_1, \ldots, X_i, \ldots, X_n) - L(X_1, \ldots, \hat{X}_i, \ldots, X_n) | X_1, \ldots, X_i \right]
\]

\[
\geq -2 \mathbb{E} \left[ \min_{i < j \leq n} \|\hat{X}_i - X_j\| | X_1, \ldots, X_i \right]
\]

\[
= -2 \mathbb{E} \left[ \min_{i < j \leq n} \|\hat{X}_i - X_j\| \right]
\]

\[
= -2 \mathbb{E} \left[ \min_{i < j \leq n} \|\hat{X}_i - X_j\| | \hat{X}_i \right] = -2 \mathbb{E} \left[ g_{n-i}(\hat{X}_i) \right].
\]

Hence

\[
|Y_i| \leq 2 \max \left\{ g_{n-i}(X_i), \mathbb{E} \left[ g_{n-i}(\hat{X}_i) \right] \right\}.
\]
(b) From \( g_m(x) \leq \sqrt{\frac{\pi}{m}} \) and (a), \( |Y_i| \leq 2\sqrt{\frac{\pi}{n-i}} \) for \( 1 \leq i \leq n-1 \), and we have \( |Y_n| \leq 2\sqrt{2} \).

By applying McDiarmid inequality to \( g(Y_1, \cdots, Y_n) = \sum_{i=1}^{n} Y_i \) then \( |g(Y_1, \cdots, Y_i, \cdots, Y_n) - g(Y_1, \cdots, \hat{Y}_i, \cdots, Y_n)| = |Y_i - \hat{Y}_i| \leq 4\sqrt{\frac{\pi}{n-i}} \) for \( 1 \leq i \leq n-1 \) and \( |g(Y_1, \cdots, Y_n) - g(Y_1, \cdots, \hat{Y}_n)| \leq 4\sqrt{2} \), so

\[
\mathbb{P}(|L(X_1, \cdots, X_n) - \mathbb{E}[L(X_1, \cdots, X_n)]| > \varepsilon) \leq 2 \exp\left(-\frac{2\varepsilon^2}{(4\sqrt{2})^2 + \sum_{i=1}^{n-1} \left(4\sqrt{\frac{\pi}{n-i}}\right)^2}\right)
\]

\[
\leq 2 \exp\left(-\frac{2\varepsilon^2}{16 + 8\pi + 8\pi \sum_{i=2}^{n-1} \int_{i-1}^{i} \frac{1}{x} dx}\right)
\]

\[
\leq 2 \exp\left(-\frac{\varepsilon^2}{16 + 8\pi + 8\pi \log n}\right).
\]

Actually, \( Y_i \)'s are not independent, so McDiarmid inequality is not directly applicable. Instead, you can apply Azuma-Hoeffding's inequality in this case, which gives essentially the same result as applying McDiarmid inequality without worrying about \( Y_i \)'s independence.

**Proofs of hints**

- **Showing** \( L(\{x_1, \cdots, x_n\} \setminus \{x_i\}) \leq L(x_1, \cdots, x_n) \leq L(\{x_1, \cdots, x_n\} \setminus \{x_i\}) + 2 \min_{i < j \leq n} \|x_i - x_j\|\)

Let the shortest tour of \( \{x_1, \cdots, x_n\} \) be \( T = \cdots x_j x_i x_k \cdots \). If we remove \( x_i \) from this tour and make \( T' = \cdots x_j x_k \cdots \), then length of \( T' \) is shorter than length of \( T \). Then \( T' \) is one possible tour of \( \{x_1, \cdots, x_n\} \setminus \{x_i\} \), so

\[
L(\{x_1, \cdots, x_n\} \setminus \{x_i\}) \leq (\text{length of } T') \leq (\text{length of } T) = L(x_1, \cdots, x_n).
\]

Fix any \( j > i \), and let the shortest tour of \( \{x_1, \cdots, x_n\} \) be \( T'' = \cdots x_k x_j x_k \cdots \). If we add \( x_i \) between \( x_j, x_k \) from \( T'' \) and make \( T''' = \cdots x_k x_i x_j x_k \cdots \), then

\[
(\text{length of } T''') = (\text{length of } T'') - \|x_j - x_k\| + \|x_j - x_i\| + \|x_i - x_k\| \leq (\text{length of } T'') - \|x_j - x_k\| + \|x_j - x_i\| + \|x_i - x_j\| + \|x_j - x_k\| = L(\{x_1, \cdots, x_n\} \setminus \{x_i\}) + 2\|x_j - x_i\|.
\]
Then $T''$ is one possible tour of $\{x_1, \cdots, x_n\}$, so

$$L(x_1, \cdots, x_n) \leq (\text{length of } T'') \leq L(\{x_1, \cdots, x_n\} \setminus \{x_i\}) + 2\|x_i - x_j\|.$$ 

Since this holds for any $i < j \leq n$,

$$L(x_1, \cdots, x_n) \leq L(\{x_1, \cdots, x_n\} \setminus \{x_i\}) + 2 \min_{i < j \leq n} \|x_i - x_j\|.$$

- Showing $g_m(x) = \mathbb{E}\left( \min_{1 \leq i \leq m} \|x - X_i\| \right) \leq \sqrt{\frac{\pi}{m}}$

For any $0 \leq \lambda \leq \sqrt{2}$, $P\left( \min_{1 \leq i \leq m} \|x - X_i\| \geq \lambda \right)$ is maximized when $x$ is in one of corners, and in that case

$$P(\|x - X_1\| \geq \lambda) = \text{vol}(\{0, 1\}^2 \setminus B(0, \lambda)) \leq \text{vol}\left(\left[0, 1\right]^2 \setminus \left[0, \frac{\lambda}{2}\right]\right) = 1 - \frac{\lambda^2}{4}.$$ 

Hence

$$P\left( \min_{1 \leq i \leq m} \|x - X_i\| \geq \lambda \right) \leq \left(1 - \frac{\lambda^2}{4}\right)^m \leq \exp\left(-\frac{m}{4} \lambda^2\right),$$

and therefore,

$$\mathbb{E}\left( \min_{1 \leq i \leq m} \|x - X_i\| \right) = \int_0^{\sqrt{2}} P\left( \min_{1 \leq i \leq m} \|x - X_i\| \geq \lambda \right) d\lambda 
\leq \int_0^{\infty} \exp\left(-\frac{m}{4} \lambda^2\right) d\lambda 
= \sqrt{\frac{\pi}{m}}.$$

- Showing $L(X_1, \cdots, X_n) - \mathbb{E}[L(X_1, \cdots, X_n)] = \sum_{i=1}^n Y_i$

$$Y_i = \mathbb{E}\left[ L(X_1, \cdots, X_i, \cdots, X_n) - L(X_1, \cdots, \hat{X}_i, \cdots, X_n) \right| X_1, \cdots, X_i]$$

$$= \mathbb{E}\left[ L(X_1, \cdots, X_i, \cdots, X_n) \right| X_1, \cdots, X_i] - \mathbb{E}\left[ L(X_1, \cdots, \hat{X}_i, \cdots, X_n) \right| X_1, \cdots, X_i]$$

$$= \mathbb{E}\left[ L(X_1, \cdots, X_i, \cdots, X_n) \right| X_1, \cdots, X_i] - \mathbb{E}\left[ L(X_1, \cdots, \hat{X}_i, \cdots, X_n) \right| X_1, \cdots, X_i - 1]$$

$$= \mathbb{E}\left[ L(X_1, \cdots, X_i, \cdots, X_n) \right| X_1, \cdots, X_i] - \mathbb{E}\left[ L(X_1, \cdots, X_i, \cdots, X_n) \right| X_1, \cdots, X_i - 1]$$

hence

$$\sum_{i=1}^n Y_i = \mathbb{E}\left[ L(X_1, \cdots, X_n) \right| X_1, \cdots, X_n] - \mathbb{E}\left[ L(X_1, \cdots, X_n) \right] = L(X_1, \cdots, X_n) - \mathbb{E}\left[ L(X_1, \cdots, X_n) \right].$$