Problem 1

Finding the MGF

$X \sim N(\mu, \sigma^2)$. Therefore, the density of $X$ is given by,

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right).$$

The moment generating function is then given by,

$$M_X(t) = \mathbb{E}[e^{tX}] = \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} \, dx$$

$$= \frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}((x-2\mu\sigma^2)(x+\sigma^2t+\mu)^2)} \, dx$$

$$= \frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}((x-2\mu\sigma^2+\mu)^2)} \, dx$$

$$= \exp\left(-\frac{1}{2\sigma^2}(\mu^2-(\mu+\sigma^2t)^2)\right) \frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}(x-\mu-\sigma^2t)^2} \, dx$$

$$= \exp\left(t\mu + \frac{t^2\sigma^2}{2}\right).$$

**$X$ is sub-Gaussian**

$X$ is sub-Gaussian with mean $\mathbb{E}[X] = \mu$ and variance $\text{Var}(X) = \sigma^2$ iff every $t$,

$$\mathbb{E}\left[e^{t(X-\mu)}\right] \leq e^{\sigma^2t^2/2}.$$

Now $\mathbb{E}[-X] = \mu$ and $\text{Var}(-X) = \sigma^2$. Hence, $-X$ is sub-Gaussian iff for every $t$,

$$\mathbb{E}\left[e^{t(-X+\mu)}\right] \leq e^{\sigma^2t^2/2}.$$

Now,

$$X \text{ is sub-Gaussian} \iff \mathbb{E}\left[e^{t(X-\mu)}\right] \leq e^{\sigma^2t^2/2} \text{ for all } t \iff \mathbb{E}\left[e^{t(X-\mu)}\right] \leq e^{\sigma^2(-t)^2/2} \text{ for all } t \iff \mathbb{E}\left[e^{(-t)(X-\mu)}\right] \leq e^{\sigma^2t^2/2} \text{ for all } t \iff \mathbb{E}\left[e^{t(-X+\mu)}\right] \leq e^{\sigma^2t^2/2} \text{ for all } t \iff -X \text{ is sub-Gaussian.}$$

The third iff holds because if something holds for all $t$, it should also hold for all $-t$. 


Problem 2

Need to show that:
\[ \inf_{k=0,1,2,...} \frac{\mathbb{E}[X^k]}{\mu^k} \leq \inf_{t\geq 0} \frac{\mathbb{E}[\exp(tX)]}{\exp(t\mu)}. \]

We will use the hint that if \( c \leq \frac{a_i}{b_i} \) for all \( i \) then
\[
c \leq \frac{\sum_{i=1}^{\infty} a_i}{\sum_{i=1}^{\infty} b_i}.
\]

In this case let us set \( c = \inf_{k=0,1,2,...} \frac{\mathbb{E}[X^k]}{\mu^k} \), then
\[
c = \inf_{k=0,1,2,...} \frac{\mathbb{E}[X^k]}{\mu^k} \leq \frac{\mathbb{E}[X^k]}{\mu^k} \quad \forall \ k.
\]

Now we fix a \( t \geq 0 \) and let
\[
a_k = \frac{t^k \mathbb{E}[X^k]}{k!} \quad \text{and} \quad b_k = \frac{t^k \mu^k}{k!}.
\]

Then
\[
a_k = \frac{t^k \mathbb{E}[X^k]}{k!} = \frac{\mathbb{E}[X^k]}{\mu^k} \geq \frac{c}{k} \quad \forall \ k.
\]

Hence using the hint,
\[
c \leq \frac{\sum_{k=0}^{\infty} a_k}{\sum_{k=0}^{\infty} b_k} = \frac{\sum_{k=0}^{\infty} \frac{t^k \mathbb{E}[X^k]}{k!}}{\sum_{k=0}^{\infty} \frac{t^k \mu^k}{k!}} = \frac{\mathbb{E}[\exp(tX)]}{\exp(t\mu)}.
\]

Therefore for all fixed \( t \geq 0 \),
\[
c = \inf_{k=0,1,2,...} \frac{\mathbb{E}[X^k]}{\mu^k} \leq \frac{\mathbb{E}[\exp(tX)]}{\exp(t\mu)}.
\]

Hence,
\[
\inf_{k=0,1,2,...} \frac{\mathbb{E}[X^k]}{\mu^k} \leq \inf_{t\geq 0} \frac{\mathbb{E}[\exp(tX)]}{\exp(t\mu)}.
\]

Proving the hint:

We have for all \( i \),
\[
c \leq \frac{a_i}{b_i} \implies cb_i \leq a_i.
\]

Summing up on both sides we get,
\[
\sum_{i=1}^{\infty} b_i \leq \sum_{i=1}^{\infty} a_i \implies c \leq \frac{\sum_{i=1}^{\infty} a_i}{\sum_{i=1}^{\infty} b_i}.
\]
Problem 3

To show:

\[
\frac{1}{n-1} \sum_{i=1}^{n} (X_i - \hat{\mu})^2 = \frac{1}{n} \sum g(X_j, X_k).
\]

\[
\frac{1}{n} \sum_{j<k} g(X_j, X_k) = \frac{1}{2\binom{n}{2}} \sum_{j<k} g(X_j, X_k)
\]

\[
= \frac{1}{4\binom{n}{2}} \sum (X_j - X_k)^2
\]

\[
= \frac{1}{4\binom{n}{2}} \sum (X_j^2 - 2X_jX_k + X_k^2)
\]

\[
= \frac{1}{4\binom{n}{2}} \left( \sum_{j=1}^{n} (n-1)X_j^2 - \sum_{j=1}^{n} \sum_{k=1}^{n} 2X_jX_k + 2 \sum_{j=1}^{n} X_j^2 + \sum_{k=1}^{n} (n-1)X_k^2 \right)
\]

\[
= \frac{1}{4\binom{n}{2}} \left( 2n \sum_{j=1}^{n} X_j^2 - 2n^2\overline{X}^2 \right)
\]

\[
= \frac{4n}{4n(n-1)} \left( \sum_{j=1}^{n} X_j^2 - n\overline{X}^2 \right)
\]

\[
= \frac{1}{n-1} \sum_{j=1}^{n} (X_j - \overline{X})^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \hat{\mu})^2.
\]

Problem 4

Tightness of Markov’s inequality:

We need to find a random variable \( X \geq 0 \), such that

\[
P(X \geq t) = \frac{\mathbb{E}[X]}{t}.
\]

Once we fix a \( t \geq 0 \), consider a random variable \( X \) which takes the value \( t \) with probability \( p \) and 0 with probability \( 1-p \). Then \( P(X = 0) = 1-p \) and \( P(X = t) = p \). Hence,

\[
\mathbb{E}[X] = tP(X = t) = tp \quad \text{and} \quad P(X \geq t) = P(X = t) = p.
\]

Therefore,

\[
P(X \geq t) = P(X = t) = p = \frac{\mathbb{E}[X]}{t}.
\]

Alternative Solution. In the previous solution you could have also fixed the value of \( p \). For example if you take \( p = 1 \), \( P(X = t) = 1 \). Then \( \mathbb{E}[X] = tP(X = t) = t \). Hence,

\[
P(X \geq t) = P(X = t) = 1 = \frac{\mathbb{E}[X]}{t}.
\]
Tightness of Chebyshev’s inequality:

In this case, we need to find a random variable $X$, such that

$$P(|X - \mathbb{E}[X]| \geq t) = \frac{Var(X)}{t^2}. \tag{1}$$

We can get the above inequality by plugging in $k = t/\sigma$, where $\sigma^2 = Var(X)$.

Once we fix a $t \geq 0$, consider a random variable $X$ which takes the value $t$ with probability $p/2$, $-t$ with probability $p/2$ and 0 with probability $(1 - p)$. That is, $P(X = t) = \frac{p}{2}$, $P(X = -t) = \frac{p}{2}$ and $P(X = 0) = (1 - p)$. Then,

$$\mathbb{E}[X] = tP(X = t) - tP(X = -t) = \frac{tp}{2} - \frac{tp}{2} = 0,$$

and the variance is given by,

$$Var(X) = E[X^2] = t^2P(X = t) + (-t)^2P(X = t) = pt^2.$$

Therefore,

$$P(|X - \mathbb{E}[X]| \geq t) = P(|X| \geq t) = P(|X| = t) = p = \frac{Var(X)}{t^2}. \tag{2}$$

Alternative Solution. In the previous solution you could have also fixed the value of $p$. For example if you take $p = 1$, $P(X = t) = 0.5$ and $P(X = -t) = 0.5$. Then,

$$\mathbb{E}[X] = tP(X = t) - tP(X = -t) = \frac{t}{2} - \frac{t}{2} = 0,$$

and the variance is given by,

$$Var(X) = E[X^2] = t^2P(X = t) + (-t)^2P(X = t) = t^2.$$

Therefore,

$$P(|X - \mathbb{E}[X]| \geq t) = P(|X| \geq t) = P(|X| = t) = 1 = \frac{Var(X)}{t^2}. \tag{3}$$

Problem 5

(a) We know that $a \leq X_i \leq b$ for all $i$. Hence $a \leq \mu \leq b$.

$$\sigma^2 = Var(X_i) = E[(X_i - \mu)^2]$$

$$= E[(X_i - \mu)^2 \mathbb{1}_{X_i > \mu}] + E[(X_i - \mu)^2 \mathbb{1}_{X_i \leq \mu}]$$

$$\leq (b - \mu)^2 P(X_i > \mu) + (\mu - a)^2 P(X_i \leq \mu),$$

where equality occurs if and only if $X_i$ takes only values $a$ and $b$. Then $P(X_i > \mu) = P(X_i = b)$ and $P(X_i \leq \mu) = P(X_i = a)$. In that case, $\mu = aP(X_i = a) + bP(X_i = b)$. For simplicity of notation, let $P(X_i = a) = p$, and $P(X_i = b) = (1 - p)$. So,

$$\sigma^2 \leq (b - ap - b(1-p))^2(1-p) + (ap + b(1-p) - a)^2 p$$

$$= (b - a)^2 p^2 (1 - p) + (b - a)^2 (1 - p)^2 p$$

$$= (b - a)^2 p(1 - p) \leq \frac{(b - a)^2}{4}. \tag{4}$$
**Alternative Solution.** We know that \( a \leq X_i \leq b \) for all \( i \). Hence \( a \leq E[X_i] \leq b \). Therefore, \( |X_i - E[X_i]| \leq b - a \). Therefore,

\[
(X_i - E[X_i])^2 \leq (b - a)^2.
\]

Hence,

\[
\sigma^2 = \text{Var}(X_i) = E[(X_i - E[X_i])^2] \leq (b - a)^2.
\]

(b) We consider the regime where \( 0 \leq t \leq (b - a) \). We consider this because when \( t > (b - a) \),

\[
P(|X - \mu| \geq t) = 0,
\]

since \( a \leq X \leq b \), \( a \leq \mu \leq b \) and therefore \( |X - \mu| \leq (b - a) \).

Since \( \sigma^2 \leq (b - a)^2 \) and \( t \leq (b - a) \), under this regime the Bernstein bound can be bounded above by,

\[
2 \exp\left(-\frac{nt^2}{2(\sigma^2 + (b-a)t)}\right) \leq 2 \exp\left(-\frac{nt^2}{2((b-a)^2 + (b-a)^2)}\right)
= 2 \exp\left(-\frac{nt^2}{4(b-a)^2}\right),
\]

which ignoring constants is similar to the Hoeffding’s bound.

(c) For Bernstein’s bound to be better/tighter than Hoeffding’s bound, we need

\[
2 \exp\left(-\frac{nt^2}{2(\sigma^2 + (b-a)t)}\right) \leq 2 \exp\left(-\frac{nt^2}{(b-a)^2}\right)
\]

\[
\iff \frac{nt^2}{2(\sigma^2 + (b-a)t)} \geq \frac{nt^2}{(b-a)^2}
\]

\[
\iff \frac{2(\sigma^2 + (b-a)t)}{nt^2} \leq \frac{(b-a)^2}{nt^2}
\]

\[
\iff \sigma^2 \leq \frac{(b-a)^2 - 2(b-a)t}{2} = \frac{(b-a)(b-a-2t)}{2},
\]

for any \( t \geq 0 \). Therefore, if \( \sigma^2 \leq \frac{(b-a)(b-a-2t)}{2} \) for a fixed \( t \geq 0 \), then Bernstein’s bound is better/tighter than Hoeffding’s bound.

**Comment.** Notice that if \( (b-a) \geq t \geq \frac{(b-a)(b-a-2t)}{2} \), \( \frac{(b-a)(b-a-2t)}{2} \leq 0 \). Then Hoeffding’s bound is always better than Bernstein’s for any \( \sigma^2 \).