ISYE 8803 HW 1

William Yuan

Total: 95/100 (A).

Problem 1

(a) Proceeding from the definition of the MGF, we find that

$$
\begin{aligned}
M_X(\lambda)\exp(-\lambda u) &= \mathbb{E}\left[\exp(\lambda(X))\right]\exp(-\lambda u) \\
&= \left(\sum_{k=1}^{\infty}\frac{\lambda^k\mathbb{E}\left[X^k\right]}{k!}\right)\left(\sum_{j=1}^{\infty}\frac{\lambda^j u^j}{j!}\right)^{-1} \\
&= \left(\sum_{k=1}^{\infty}\frac{\mathbb{E}\left[X^k\right]}{u^k}\frac{\lambda^k u^k}{k!}\right)\left(\sum_{j=1}^{\infty}\frac{\lambda^j u^j}{j!}\right)^{-1} \\
&\geq \inf_{k\in\mathbb{Z}_+}\mathbb{E}\left[X^k\right]u^{-k}\left(\sum_{k=1}^{\infty}\frac{\lambda^k u^k}{k!}\right)\left(\sum_{j=1}^{\infty}\frac{\lambda^j u^j}{j!}\right)^{-1} \\
&= \inf_{k\in\mathbb{Z}_+}\mathbb{E}\left[X^k\right]u^{-k}
\end{aligned}
$$

⊕ Precisely

Since the above inequality holds regardless of the value of $\lambda > 0$, we know that

$$\inf_{\lambda>0} M_X(\lambda)\exp(-\lambda u) \geq \inf_{k\in\mathbb{Z}_+}\mathbb{E}\left[X^k\right]u^{-k}$$

(b) We begin by observing that

$$\begin{aligned}
\exp(-\lambda u) &\geq (\exp(-\lambda u) + \exp(\lambda u))^{-1} \\
&= \left(\sum_{j=1}^{\infty} \frac{\lambda^j(-u^j)}{j!} + \sum_{j'=1}^{\infty} \frac{\lambda^{j'} u^{j'}}{j'!}\right)^{-1} \\
&= \left(2\sum_{j=1}^{\infty} \frac{\lambda^{2j} u^{2j}}{(2j)!}\right)^{-1}
\end{aligned}$$

+

Furthermore, since the distribution of X is equal to that of $-X$, we know that all odd moments are zero:

$$\begin{aligned}
\mathbb{E}\left[X^{2k+1}\right] &= \frac{\mathbb{E}\left[X^{2k+1}\right] + \mathbb{E}\left[(-X)^{2k+1}\right]}{2} \\
&= \frac{\mathbb{E}\left[X^{2k+1}\right] - \mathbb{E}\left[X^{2k+1}\right]}{2} \\
&= 0
\end{aligned}$$

for all $k \in \mathbb{Z}_+$.

Putting the above together, we find that

$$\begin{aligned}
M_X(\lambda)\exp(-\lambda u) &\geq \mathbb{E}\left[\exp(\lambda(X))\right](\exp(-\lambda u) + \exp(\lambda u))^{-1} \\
&= \left(\sum_{k=1}^{\infty} \frac{\lambda^{2k}\mathbb{E}\left[X^{2k}\right]}{(2k)!}\right)\left(2\sum_{j=1}^{\infty} \frac{\lambda^{2j} u^{2j}}{(2j)!}\right)^{-1} \\
&= \left(\sum_{k=1}^{\infty} \frac{\mathbb{E}\left[X^{2k}\right]}{u^{2k}} \frac{\lambda^{2k} u^{2k}}{(2k)!}\right)\left(2\sum_{j=1}^{\infty} \frac{\lambda^{2j} u^{2j}}{(2j)!}\right)^{-1} \\
&\geq \inf_{k\in\mathbb{Z}_+} \mathbb{E}\left[X^{2k}\right] u^{-2k} \left(\sum_{k=1}^{\infty} \frac{\lambda^{2k} u^{2k}}{(2k)!}\right)\left(2\sum_{j=1}^{\infty} \frac{\lambda^{2j} u^{2j}}{(2j)!}\right)^{-1} \\
&= \frac{1}{2}\inf_{k\in\mathbb{Z}_+} \mathbb{E}\left[X^{2k}\right] u^{-2k}
\end{aligned}$$

Since the above inequality holds regardless of the value of $\lambda > 0$, we know that

$$\inf_{\lambda>0} M_X(\lambda)\exp(-\lambda u) \geq \frac{1}{2}\inf_{k\in\mathbb{Z}_+} \mathbb{E}\left[X^{2k}\right] u^{-2k}$$

+.

Problem 2

For $t_1, t_2 \in \mathbb{R}$ and $\theta \in [0,1]$, we know that

$$
\begin{aligned}
\log(\mathbb{E}[\exp((\theta t_1 + (1-\theta)t_2)X)]) &= \log\Big(\mathbb{E}\left[(\exp(t_1 X))^{\theta}\,(\exp(t_2 x))^{1-\theta}\right]\Big) \\
&= \log\left(\sum_{x\in\mathcal{X}} P_X(x)\,(\exp(t_1 X))^{\theta}\,(\exp(t_2 x))^{1-\theta}\right) \\
&= \log\left(\sum_{x\in\mathcal{X}} (P_X(x)\exp(t_1 X))^{\theta}\,(P_X(x)\exp(t_2 x))^{1-\theta}\right) \\
&\le \log\left(\left(\sum_{x\in\mathcal{X}} P_X(x)\exp(t_1 X)\right)^{\theta}\left(\sum_{x\in\mathcal{X}} P_X(x)\exp(t_2 x)\right)^{1-\theta}\right) \\
&= \log\Big(\mathbb{E}\left[\exp(t_1 X)\right]^{\theta}\,\mathbb{E}\left[\exp(t_2 X)\right]^{1-\theta}\Big) \\
&= \theta\log(\mathbb{E}\left[\exp(t_1 X)\right]) + (1-\theta)\log(\mathbb{E}\left[\exp(t_2 X)\right])
\end{aligned}
$$

where we have applied Young's inequality and the monotonicity of the logarithm in the fourth line.

+

(It's a pleasure to grade!)

Problem 3

(1) **(a)** Integrating by parts using $u = 1/t$ and $dv = t\phi(t)$, we find that $v = -\phi(t)$ and $du = -1/t^2$, where we have used the fact that $\phi'(t) = -t\phi(t)$.

This then yields the following bound:

$$\begin{aligned}\int_u^\infty \phi(t)\,dt &= \left[-\frac{1}{t}\phi(t)\right]_u^\infty - \int_u^\infty \frac{1}{t^2}\phi(t)\,dt \\ &\geq \left[-\frac{1}{t}\phi(t)\right]_u^\infty \\ &= \frac{1}{u}\phi(u)\end{aligned}$$

+

where non-negativity of $\phi(t)/t^2$ was applied in the second line.

Integrating by parts using $u = 1/t^3$ and $dv = t\phi(t)$, we also find that $v = -\phi(t)$ and $du = -3/t^4$.

This then yields the following bound:

$$\begin{aligned}\int_u^\infty \phi(t)\,dt &= \left[-\frac{1}{t}\phi(t)\right]_u^\infty - \int_u^\infty \frac{1}{t^2}\phi(t)\,dt \\ &= \left[-\frac{1}{t}\phi(t)\right]_u^\infty - \left[\left[-\frac{1}{t^3}\phi(t)\right]_u^\infty - \int_u^\infty \frac{3}{t^4}\phi(t)\,dt\right] \\ &\leq \left[-\frac{1}{t}\phi(t)\right]_u^\infty - \left[-\frac{1}{t^3}\phi(t)\right]_u^\infty \\ &= \left(\frac{1}{u} - \frac{1}{u^3}\right)\phi(u)\end{aligned}$$

+

where non-negativity of $3\phi(t)/t^4$ was applied in the third line.

(b) Integrating by parts using $u = 3/t^5$ and $dv = t\phi(t)$, we also find that $v = -\phi(t)$ and $du = -15/t^6$.

This then yields the following bound:

$$\begin{aligned}\int_u^\infty \phi(t)\,dt &= \left[-\frac{1}{t}\phi(t)\right]_u^\infty - \int_u^\infty \frac{1}{t^2}\phi(t)\,dt \\ &= \left[-\frac{1}{t}\phi(t)\right]_u^\infty - \left[\left[-\frac{1}{t^3}\phi(t)\right]_u^\infty - \int_u^\infty \frac{3}{t^4}\phi(t)\,dt\right] \\ &= \left[-\frac{1}{t}\phi(t)\right]_u^\infty - \left[\left[-\frac{1}{t^3}\phi(t)\right]_u^\infty - \left[\left[-\frac{3}{t^5}\phi(t)\right]_u^\infty - \int_u^\infty \phi(t)\frac{15}{t^6}\,dt\right]\right] \\ &\geq \left[-\frac{1}{t}\phi(t)\right]_u^\infty - \left[-\frac{1}{t^3}\phi(t)\right]_u^\infty + \left[-\frac{3}{t^5}\phi(t)\right]_u^\infty \\ &= \left(\frac{1}{u} - \frac{1}{u^3} + \frac{3}{u^5}\right)\phi(u)\end{aligned}$$

+

where non-negativity of $15\phi(t)/t^6$ was applied in the fourth line.

(2) Changing variables according to $x = t - u$, we find that

$$\begin{aligned}\frac{1}{2} - \Phi(u) &= \frac{1}{2} - \int_u^\infty \phi(t)\, dt \\ &= \frac{1}{2} - \int_0^\infty \phi(x+u)\, dx\end{aligned}$$

Next, suppose we define $F(u) = \frac{1}{2} - \Phi(u)$.

We therefore find that

$$\begin{aligned}F'(u) &= \int_0^\infty (x+u)\phi(x+u)\, dx \\ &= [\phi(x+u)]_0^\infty \\ &= \phi(u)\end{aligned}$$

We then find that

$$\begin{aligned}F(u) &= \int_0^u F'(x)\, dx \\ &= \int_0^u \phi(x)\, dx \\ &= \int_0^u \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx \\ &= \int_0^u \frac{1}{\sqrt{2\pi}} \sum_{k=0}^\infty \left(\frac{1}{k!}\right)\left(-\frac{x^2}{2}\right)^k dx \\ &= \frac{1}{\sqrt{2\pi}} \sum_{k=0}^\infty \int_0^u \left(\frac{1}{k!}\right)\left(-\frac{x^2}{2}\right)^k dx \\ &= \frac{1}{\sqrt{2\pi}} \sum_{k=0}^\infty \left[\left(\frac{1}{k!}\right)\left(\frac{(-1)^k x^{2k+1}}{2^k(2k+1)}\right)\right]_0^u \\ &= \frac{1}{\sqrt{2\pi}} \sum_{k=0}^\infty \frac{(-1)^k x^{2k+1}}{2^k k!(2k+1)}\end{aligned}$$

$\oplus$ Good Job!

Problem 4

(a) We know that

$$\begin{aligned}
\mathbb{E}[X] &= \mathbb{E}[X\mathbb{1}[X < (1-t)\mathbb{E}[X]]] + \mathbb{E}[X\mathbb{1}[X \geq (1-t)\mathbb{E}[X]]] \\
&\leq (1-t)\mathbb{E}[X] + \mathbb{E}[X\mathbb{1}[X \geq (1-t)\mathbb{E}[X]]] \\
&\leq (1-t)\mathbb{E}[X] + (\mathbb{E}[X^2])^{1/2}(\mathbb{E}[(\mathbb{1}[X \geq (1-t)\mathbb{E}[X]])^2])^{1/2} \\
&= (1-t)\mathbb{E}[X] + (\mathbb{E}[X^2])^{1/2}(\mathbb{P}[X \geq (1-t)\mathbb{E}[X]])^{1/2}
\end{aligned}$$

where in the last line we have applied the Cauchy Schwarz inequality. Rearranging yields

$$\mathbb{P}[X \geq (1-t)\mathbb{E}[X]] \geq t^2 \frac{(\mathbb{E}[X])^2}{\mathbb{E}[X^2]}$$

+

(b) Using the same approach as in (a), we find that

$$\begin{aligned}
\mathbb{E}[X] &= \mathbb{E}[X\mathbb{1}[X < (1-t)\mathbb{E}[X]]] + \mathbb{E}[X\mathbb{1}[X \geq (1-t)\mathbb{E}[X]]] \\
&\leq \mathbb{E}[(1-t)\mathbb{E}[X]\mathbb{1}[X < (1-t)\mathbb{E}[X]]] + \mathbb{E}[X\mathbb{1}[X \geq (1-t)\mathbb{E}[X]]] \\
&= \mathbb{E}[(1-t)\mathbb{E}[X](1-\mathbb{1}[X \geq (1-t)\mathbb{E}[X]])] + \mathbb{E}[X\mathbb{1}[X \geq (1-t)\mathbb{E}[X]]] \\
&= (1-t)\mathbb{E}[X] + \mathbb{E}[(X-(1-t)\mathbb{E}[X])\mathbb{1}[X \geq (1-t)\mathbb{E}[X]]] \\
&\leq (1-t)\mathbb{E}[X] + (\mathbb{E}[(X-(1-t)\mathbb{E}[X])^2])^{1/2}(\mathbb{E}[(\mathbb{1}[X \geq (1-t)\mathbb{E}[X]])^2])^{1/2} \\
&= (1-t)\mathbb{E}[X] + (\mathbb{E}[(X-(1-t)\mathbb{E}[X])^2])^{1/2}(\mathbb{P}[X \geq (1-t)\mathbb{E}[X]])^{1/2}
\end{aligned}$$

From here we note that

$$\begin{aligned}
\mathbb{E}[(X-(1-t)\mathbb{E}[X])^2] &= \mathbb{E}[((X-\mathbb{E}[X]) - t\mathbb{E}[X])^2] \\
&= \mathbb{E}[(X-\mathbb{E}[X])^2] + 2\mathbb{E}[X-\mathbb{E}[X]]t\mathbb{E}[X] + t^2(\mathbb{E}[X])^2 \\
&= \mathrm{Var}(X) + t^2(\mathbb{E}[X])^2
\end{aligned}$$

Rearranging then yields

$$\mathbb{P}[X \geq (1-t)\mathbb{E}[X]] \geq t^2 \frac{(\mathbb{E}[X])^2}{t^2(\mathbb{E}[X])^2 + \mathrm{Var}(X)}$$

Bravo! +

(c) We know that

$$\begin{aligned}
\mathbb{E}[X] &= \mathbb{E}[X\mathbb{1}[X < (1-t)\mathbb{E}[X]]] + \mathbb{E}[X\mathbb{1}[X \geq (1-t)\mathbb{E}[X]]] \\
&\leq (1-t)\mathbb{E}[X] + \mathbb{E}[X\mathbb{1}[X \geq (1-t)\mathbb{E}[X]]] \\
&\leq (1-t)\mathbb{E}[X] + (\mathbb{E}[|X|^p])^{1/p}(\mathbb{E}[(\mathbb{1}[X \geq (1-t)\mathbb{E}[X]])^{p/(p-1)}])^{(p-1)/p} \\
&= (1-t)\mathbb{E}[X] + (\mathbb{E}[|X|^p])^{1/p}(\mathbb{P}[X \geq (1-t)\mathbb{E}[X]])^{(p-1)/p}
\end{aligned}$$

where in the last line we have applied the Hölder's inequality for some $p \geq 1$. Rearranging yields

+

$$\mathbb{P}[X \geq (1-t)\mathbb{E}[X]] \geq \left(t^p \frac{\mathbb{E}[X]^p}{\mathbb{E}[|X|^p]}\right)^{1/(p-1)}$$

Problem 5

(a) Using a polar change of coordinates, we find that

$$
\begin{aligned}
M_2(t) &= \mathbb{E}\left[\exp\big(t(Z_1^2 + Z_2^2)\big)\right] \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{1}{2\pi}\exp\left(-\left(\frac{z_1^2+z_2^2}{2}\right)\right)\exp\big(t(z_1^2+z_2^2)\big)\,dz_1\,dz_2 \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{1}{2\pi}\exp\left(\left(t-\frac{1}{2}\right)(z_1^2+z_2^2)\right)dz_1\,dz_2 \\
+ \quad &= \int_{0}^{2\pi}\int_{0}^{\infty} \frac{1}{2\pi}\exp\left(\left(t-\frac{1}{2}\right)r^2\right)r\,dr\,d\theta \quad \text{correct} \\
&= \int_{0}^{2\pi}\left[\frac{1}{2\pi}\exp\left(\left(t-\frac{1}{2}\right)r^2\right)\left(\frac{1}{2t-1}\right)\right]_0^{\infty} d\theta \\
&= \frac{1}{1-2t}\int_0^{2\pi}\frac{1}{2\pi}d\theta \\
&= \frac{1}{1-2t} \quad +
\end{aligned}
$$

Note that this also shows that $M_1(t) = \frac{1}{\sqrt{1-2t}}$ and $M_d(t) = (1-2t)^{-d/2}$.

where the second to last line follows since $t < 1/2$. Furthermore, if $t \geq 1/2$, we note that the inner integral over r diverges and that $M_2(t) = \infty$ in this case.

Next, we note that because all Z_i are iid, we know that for $t < 1/2$ that

$$
\begin{aligned}
M_{2d}(t) &= \mathbb{E}\left[\exp\left(t\sum_{i=1}^{2d} Z_i^2\right)\right] \\
&= \prod_{i=1}^{d}\mathbb{E}\left[\exp\big(t(Z_{2i-1}^2 + Z_{2i}^2)\big)\right] \\
&= M_2(t)^d \\
&= \frac{1}{(1-2t)^d}
\end{aligned}
$$

(b) Using Chernoff's method, we find that

$$
\begin{aligned}
\mathbb{P}[X > x] &\leq \inf_{t<\frac{1}{2}} \exp(-tx)M_{2d}(t) \\
&= \inf_{t<\frac{1}{2}} \frac{\exp(-tx)}{(1-2t)^d} \\
&= \inf_{t<\frac{1}{2}} \exp\left(\log\left(\frac{\exp(-tx)}{(1-2t)^d}\right)\right) \\
&= \inf_{t<\frac{1}{2}} \exp(-tx - d\log(1-2t))
\end{aligned}
$$

From here we note that

$$
\begin{aligned}
f(t) &= -tx - d\log(1-2t) \\
f'(t) &= -x + 2d\frac{1}{1-2t} \\
f''(t) &= 2d\frac{2}{(1-2t)^2} > 0
\end{aligned}
$$

From here we observe that $f''(t) > 0$ implies that $f(t)$ is convex for $t < 1/2$, and hence the optimal solution is given by the first order condition for optimality:

$$
\begin{aligned}
-x(1-2t^*) + 2d = 0 \\
t^* = \frac{1}{2} - \frac{d}{x}
\end{aligned}
$$

Since the exponential function is monotoni~~c~~e, we know that t^* being the optimizer for $f(t)$ implies that it is the optimizer for $\exp(f(t))$.

Substituting the optimal value of t into the Chernoff bound, we find that

$$
\mathbb{P}[X > x] \le \exp\left(d\log\left(\frac{x}{2d}\right) - \frac{x-2d}{2}\right) \boxed{\frac{2d}{x}}
$$

⊕

(iii) ⊖ (I recommend doing it if you haven't yet. it's just algebra, but a useful run.)

Problem 6

(a) Since $Y = sX - \mu \sim \mathcal{N}(\mu(s-1), s^2 I_d)$, we know that the risk is given by

$$\begin{aligned}
\text{Risk}_\mu[sX] &= \mathbb{E}_\mu\left[\|sX - \mu\|^2\right] \\
&= \mathbb{E}_\mu\left[\sum_{i=1}^{d} Y_i^2\right] \\
&= \mathbb{E}_\mu\left[\sum_{i=1}^{d}(s^2 + \mu_i^2(s-1)^2)\right] \\
&= ds^2 + (s-1)^2\mathbb{E}_\mu\left[\|\mu\|^2\right]
\end{aligned}$$

From here we note that if $s < 0$, $(s-1)^2 > 1$ and $s^2 > 0$. We then see that the estimator for $s = 0$ satisfies

$$\begin{aligned}
\text{Risk}_\mu[0] &= \mathbb{E}_\mu\left[\|\mu\|^2\right] \\
&< ds^2 + (s-1)^2\mathbb{E}_\mu\left[\|\mu\|^2\right] \\
&= \text{Risk}_\mu[sX]
\end{aligned}$$

Since the above inequality is strict, we know that the estimator when $s = 0$ dominates any estimator such that $s < 0$. +

From here we note that if $s > 1$, $(s-1)^2 > 0$ and $s^2 > 1$. We then see that the estimator for $s = 1$ satisfies

$$\begin{aligned}
\text{Risk}_\mu[X] &= d \\
&< ds^2 + (s-1)^2\mathbb{E}_\mu\left[\|\mu\|^2\right] \\
&= \text{Risk}_\mu[sX]
\end{aligned}$$

Since the above inequality is strict, we know that the estimator when $s = 1$ dominates any estimator such that $s >$ ~~0~~. 1 (typo) +

(b) From (a), we know that the risk is given by

$$\text{Risk}_\mu[sX] = ds^2 + (s-1)^2\mathbb{E}_\mu\left[\|\mu\|^2\right]$$

Differentiating with respect to s, we see that the first order condition for optimality is

$$\begin{aligned}
2ds + 2(s^* - 1)\mathbb{E}_\mu\left[\|\mu\|^2\right] &= 0 \\
\frac{\mathbb{E}_\mu\left[\|\mu\|^2\right]}{d + \mathbb{E}_\mu\left[\|\mu\|^2\right]} &= s^* \\
\frac{\|\mu\|^2}{d + \|\mu\|^2} &= s^*
\end{aligned}$$

where the expectation over μ is removed since we are considering a deterministic choice of μ.

Furthermore, we find that the second derivative is given by $2d + 2\mathbb{E}_\mu\left[\|\mu\|^2\right] > 0$, implying that the risk function is convex and that $s^* \in [0,1]$ is indeed the optimal minimizer.

+ (It's just a 1d quadratic but ok ...)

(c) We first note that $\hat{\mu}^* = s^* X$ is not a "proper" estimator since it requires knowledge of the true value of μ, which in practice is never known.

Instead, we consider estimator $(1 - d/\|X\|^2)X$, where we note that for a given value of μ,

$$\begin{aligned}
\mathbb{E}_X\left[\|X\|^2\right] &= \mathbb{E}_X\left[\sum_{i=1}^{d} X_i^2\right] \\
&= \sum_{i=1}^{d}(1 + \mu_i^2) \\
&= d + \|\mu\|^2
\end{aligned}$$

Hence,

$$\begin{aligned}
\left(1 - \frac{d}{\mathbb{E}_X\left[\|X\|^2\right]}\right)X &= \left(1 - \frac{d}{d + \|\mu\|^2}\right)X \\
&= \frac{\|\mu\|^2}{d + \|\mu\|^2}X \\
&= \hat{\mu}^*
\end{aligned}$$

If we assume that $\|X\|^2 \approx \mathbb{E}_X\left[\|X\|^2\right]$, we would expect the estimator $(1 - d/\|X\|^2)X$ to be reasonably close to the optimal estimator $\hat{\mu}^*$.

Yep, that's the idea.

(d) We first note that the risk function can be written according to

$$\begin{aligned}
\text{Risk}_\mu(\hat{\mu}^\delta) &= \mathbb{E}_\mu\left[\left\|\hat{\mu}^\delta - \mu\right\|^2\right] \\
&= \mathbb{E}_\mu\left[\left\|\left(1 - \frac{\delta}{\|X\|^2}\right)X - \mu\right\|^2\right] \\
&= \mathbb{E}_\mu\left[\left\|(X - \mu) - \frac{\delta}{\|X\|^2}X\right\|^2\right] \\
&= \mathbb{E}_\mu\left[\|X - \mu\|^2 - 2(X - \mu)^T\frac{\delta}{\|X\|^2}X + \left\|\frac{\delta}{\|X\|^2}X\right\|^2\right] \\
&= \mathbb{E}_\mu\left[\|X - \mu\|^2 - 2\sum_{i=1}^{d}\left((X_i - \mu_i)\frac{\delta}{\|X\|^2}X_i\right) + \frac{\delta^2}{\|X\|^2}\right]
\end{aligned}$$

Applying Stein's lemma to $g(X) = \delta/\|X\|^2 X_i$ for all $i \in [d]$, we find that

$$
\begin{aligned}
\mathbb{E}_\mu\left[\sum_{i=1}^d (X_i - \mu_i)\frac{\delta}{\|X\|^2}X_i\right] &= \mathbb{E}_\mu\left[\sum_{i=1}^d \frac{\partial}{\partial X_i}\left\{\frac{\delta}{\|X\|^2}X_i\right\}\right] \\
&= \mathbb{E}_\mu\left[\sum_{i=1}^d \delta\left(\frac{\|X\|^2 - 2X_i^2}{\|X\|^4}\right)\right] \\
&= \delta(d-2)\mathbb{E}_\mu\left[\frac{1}{\|X\|^2}\right]
\end{aligned}
$$

(Other choices of $g(X)$ are possible ...) +

To verify that Stein's lemma indeed holds, we check the integrability conditions for all $i \in [d]$ such that $d > 2$:

$$
\begin{aligned}
\mathbb{E}_\mu\left[\left|\frac{\partial}{\partial X_i}\left\{\frac{\delta}{\|X\|^2}X_i\right\}\right|\right] &= \delta(d-2)\mathbb{E}_\mu\left[\frac{1}{\|X\|^2}\right] \\
&= \delta(d-2)\frac{1}{d-2} = \delta < \infty \\
\mathbb{E}_\mu\left[\left|(X_i - \mu_i)\frac{\delta}{\|X\|^2}X_i\right|\right] &\le \delta\left(\mathbb{E}_\mu\left[\frac{X_i^2}{\|X\|^2}\right] + \mathbb{E}_\mu\left[\left|\frac{\mu_i X_i}{\|X\|^2}\right|\right]\right) \\
&\le \delta\left(\mathbb{E}_\mu\left[\frac{X_i^2}{\|X\|^2}\right] + \left(\mathbb{E}_\mu\left[\frac{X_i^2}{\|X\|^2}\right]\right)^{1/2}\left(\mathbb{E}_\mu\left[\frac{\mu_i^2}{\|X\|^2}\right]\right)^{1/2}\right) \\
&\le \delta\left(1 + \frac{\mu_i^2}{d-2}\right) < \infty
\end{aligned}
$$

where we have used the triangle inequality in the third line, the Cauchy-Schwarz inequality in the fourth line, and the bound $X_i/\|X\|^2 \le 1$ in the last line.

Putting the above together, we find that the risk function can be expressed as

$$
\begin{aligned}
\mathrm{Risk}_\mu(\hat{\mu}^\delta) &= \mathbb{E}_\mu\left[\|X - \mu\|^2 - 2\delta(d-2)\mathbb{E}_\mu\left[\frac{1}{\|X\|^2}\right] + \frac{\delta^2}{\|X\|^2}\right] \\
&= \mathbb{E}_\mu\left[\|X-\mu\|^2\right] - 2\delta(d-2)\mathbb{E}_\mu\left[\frac{1}{\|X\|^2}\right] + \delta^2\mathbb{E}_\mu\left[\frac{1}{\|X\|^2}\right]
\end{aligned}
$$

Differentiating the risk function with respect to δ, we find that the first order condition for optimality is given by

$$
\begin{aligned}
&0 = -2(d-2)\mathbb{E}_\mu\left[\frac{1}{\|X\|^2}\right] + 2\delta\mathbb{E}_\mu\left[\frac{1}{\|X\|^2}\right] \\
&\delta = d-2
\end{aligned}
$$

Finally, we note that the second derivative of the risk function is given by $2\mathbb{E}_\mu\left[\frac{1}{\|X\|^2}\right] = 2/(d-2) > 0$ for $d \ge 3$, implying that the risk function is indeed strongly convex and that the estimator associated with $\delta = d-2$ is indeed optimal. +

Great Job!

Problem 7

We first note that a circle can only intersect another congruent circle at at most 2 different points, and that every new intersection generates an edge. Hence, adding a circle to a set of $n-1$ circles can only add at most $2(n-1)$ new edges.

Furthermore, we know that the number of vertices contained in a planar graph formed by n circles is trivially at least that of a planar graph formed by $n-1$ circles, since adding a new circle cannot remove any vertices.

This gives rise to the following inequalities for $n \geq 3$

$$E_n \leq E_{n-1} + 2(n-1) \quad +$$
$$V_n \geq V_{n-1} \quad +$$

(didn't know this would suffice!)

Using Euler's formula for planar graphs, we then discover that

$$
\begin{aligned}
F_n &= E_n - V_n + 2\\
F_n &\leq E_{n-1} + 2(n-1) - V_{n-1} + 2\\
F_n &\leq F_{n-1} + 2(n-1)
\end{aligned}
$$

In general, we then find that

$$
\boxed{
\begin{aligned}
F_n &\leq F_3 + \sum_{i=4}^{n} 2(i-1)\\
&= 8 + 2\left(\frac{n(n-1)}{2} - \underbrace{6}_{3}\right)\\
&= n^2 - n \cancel{+4} + 2
\end{aligned}
}
$$

where we observe that $n^2 - n \cancel{+4} \overset{+2}{} < 2^n$ for $n \geq 4$, implying that all intersections of indices between sets are not possible as there cannot be 2^n faces in the corresponding planar graph.

(Conclusion still holds, $n^2 - n + 2 \,|_{n=4} = 14 < 16$.

$\oplus$

Shorter than my own solution – bravo!