Problem 1

Let X_i be a Bernoulli random variable which takes value 1 when the algorithm predicts *incorrectly* in the *ith* attempt. So, for $i \in [N], \delta > 0$, we have:

$$P(X_i = 1) = \frac{1}{2} + \delta$$
$$P(X_i = 0) = \frac{1}{2} - \delta$$

Let Y be the random variable which tracks the total number of incorrect answers given by the algorithm. So, $Y = \sum_{i=1}^{N} X_i$. As majority vote is used to predict the final answer, the answer is incorrect if $Y > N/2$.

$$\begin{aligned}
P(Y > N/2) &= P(\lambda Y > \lambda N/2) = P(e^{\lambda Y} > e^{\lambda N/2}) && \lambda > 0 \\
&\leq e^{-\lambda N/2} \cdot \mathbb{E}\left[e^{\lambda Y}\right] && \text{Markov's Inequality} \\
&= e^{-\lambda N/2} \cdot \mathbb{E}\left[e^{\lambda \sum_{i=1}^{N} X_i}\right] \\
&= e^{-\lambda N/2} \cdot \prod_{i=1}^{N} \mathbb{E}\left[e^{\lambda X_i}\right] && X_i\text{'s are independent} \\
&= e^{-\lambda N/2} \cdot \left(\mathbb{E}\left[e^{\lambda X_1}\right]\right)^N && X_i\text{'s are identically distributed} \\
&= e^{-\lambda N/2} \cdot \left(e^{\lambda}\left(\frac{1}{2} - \delta\right) + \left(\frac{1}{2} + \delta\right)\right)^N && \text{MGF of a Bernoulli r.v.} \\
&= \left(e^{\lambda/2}\left(\frac{1}{2} - \delta\right) + e^{-\lambda/2}\left(\frac{1}{2} + \delta\right)\right)^N = \left(\sqrt{1 - 4\delta^2}\right)^N
\end{aligned}$$

The last step is obtained by finding the tightest bound. For tightest bound, we need the smallest possible value of RHS. As both terms inside the brackets is non-negative, we can apply AM-GM inequality.

$$\left(e^{\lambda/2}\left(\frac{1}{2} - \delta\right) + e^{-\lambda/2}\left(\frac{1}{2} + \delta\right)\right) \geq 2 \cdot \sqrt{e^{\lambda/2}\left(\frac{1}{2} - \delta\right) \cdot e^{-\lambda/2}\left(\frac{1}{2} + \delta\right)} = \sqrt{1 - 4\delta^2}$$

which is obtained for $\lambda = \ln(\frac{1+2\delta}{1-2\delta}) > 0$. So, we get:

$$P(Y > N/2) \leq (1 - 4\delta^2)^{N/2} \leq e^{-2N\delta^2} \qquad \text{As } 1 + x \leq e^x, \forall x \in \mathbb{R}$$

To obtain a correct answer with probability $1-\epsilon$, i.e., $P(Y < N/2) \geq 1-\epsilon$, we need $P(Y > N/2) \leq \epsilon$. To ensure $P(Y > N/2) \leq \epsilon$ for any $\epsilon \in (0, 1)$, it is enough to ensure $e^{-2N\delta^2} \leq \epsilon$:

$$P(Y > N/2) \leq e^{-2N\delta^2} \leq \epsilon \implies N \geq \frac{1}{2\delta^2} \ln\left(\frac{1}{\epsilon}\right)$$

■

Problem 2

We have $X = (X_1, \dots, X_n)$ such that $X_i \sim \mathcal{N}(0,1)$ and are independent and we need to estimate $\mathbb{E}[\|X\|_2]$ and $\mathrm{Var}(\|X\|_2)$. Notice that for a random variable $Z \geq 0$, we have:

$$\mathbb{E}[Z^p] = \int_0^\infty pt^{p-1} \cdot P(Z > t)dt \tag{1}$$

Proof. [1] For a non-negative random variable Z,

$$Z^p = \int_0^Z pt^{t-1}dt = \int_0^\infty pt^{t-1} \cdot \mathbf{1}(Z > t)dt$$

Taking expectations on both sides,

$$\begin{aligned}\mathbb{E}\left[Z^p\right] &= \mathbb{E}\left[\int_0^\infty pt^{t-1} \cdot \mathbf{1}(Z > t)dt\right] \\ &= \int_0^\infty pt^{t-1} \cdot \mathbb{E}[\mathbf{1}(Z > t)]dt \\ &= \int_0^\infty pt^{p-1} \cdot P(Z > t)dt\end{aligned}$$

□

The following inequality was proved in the class:

$$P(\left|\|X\|_2 - \sqrt{n}\right| > \epsilon) \leq 2e^{-\epsilon^2/8}, \epsilon > 0 \tag{2}$$

(a) Let $Z = |\|X\|_2 - \sqrt{n}|$ and $p = 1$,

$$\begin{aligned}\left|\mathbb{E}\left[\|X\|_2 - \sqrt{n}\right]\right| &\leq \mathbb{E}\left[\left|\|X\|_2 - \sqrt{n}\right|\right] && \text{Jensen's Inequality, i.e., } |\mathbb{E}[X]| \leq \mathbb{E}[|X|] \\ &\overset{(1)}{=} \int_0^\infty P(\left|\|X\|_2 - \sqrt{n}\right| > t)dt \\ &\overset{(2)}{\leq} \int_0^\infty 2e^{-t^2/8}dt \\ &= \int_0^\infty 4\sqrt{2} \cdot e^{-r^2}dr \\ &= 2\sqrt{2\pi} && \text{Using } \int_0^\infty e^{-t^2}dt = \frac{1}{2}\int_{-\infty}^\infty e^{-t^2}dt = \frac{1}{2}\sqrt{\pi}\end{aligned}$$

$$\text{So, } \left|\mathbb{E}\left[\|X\|_2 - \sqrt{n}\right]\right| \leq 2\sqrt{2\pi} \implies \sqrt{n} - 2\sqrt{2\pi} \leq \mathbb{E}\left[\|X\|_2\right] \leq \sqrt{n} + 2\sqrt{2\pi}$$

■

[1] https://math.stackexchange.com/questions/172841

(b) Let $Z = |\|X\|_2 - \sqrt{n}|$ and $p = 2$,

$$
\begin{aligned}
\mathrm{Var}(\|X\|_2) &= \mathrm{Var}\left(\|X\|_2 - \sqrt{n}\right) \\
&\leq \mathbb{E}[(\|X\|_2 - \sqrt{n})^2] && \text{As } \mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 \leq \mathbb{E}[X^2] \\
&= \mathbb{E}\left[\left|\|X\|_2 - \sqrt{n}\right|^2\right] \\
&\overset{(1)}{=} \int_0^\infty 2t \cdot P(\left|\|X\|_2 - \sqrt{n}\right| > t)dt \\
&\overset{(2)}{\leq} \int_0^\infty 2te^{-t^2/8}dt = 2 \cdot \frac{e^{-t^2/8}}{-1/4}\Big|_0^\infty = 8
\end{aligned}
$$

$$\text{So, } \mathrm{Var}(\|X\|_2) \leq 8$$

$\blacksquare$

Problem 3

Given $F_d(\boldsymbol{\theta}) : \mathbb{R}^{\binom{d}{2}} \to \mathbb{R}$,

$$F_d(\boldsymbol{\theta}) = \log\left(\sum_{x\in\{\pm 1\}^d} \exp\left(\frac{1}{\sqrt{d}}\sum_{i<j}\theta_{ij}x_i x_j\right)\right)$$

Let $\boldsymbol{\theta}, \boldsymbol{x}$ be $\binom{d}{2}$-length vectors such that $\boldsymbol{\theta} = (\theta_{ij})_{1\le i<j\le d}$, $\boldsymbol{x} = f(x) = (x_i x_j)_{1\le i<j\le d}$ and $\mathcal{X} = \{\boldsymbol{x} | \boldsymbol{x} = f(x), \forall x \in \{\pm 1\}^d\}$. Note that $\mathcal{X}$ is a multiset, i.e., there could be multiple copies of the same element. So, $F_d(\boldsymbol{\theta})$ can be re-written as follows:

$$F_d(\boldsymbol{\theta}) = \log\left(\sum_{\boldsymbol{x}\in\mathcal{X}} \exp\left(\frac{1}{\sqrt{d}}\boldsymbol{\theta}^\top \boldsymbol{x}\right)\right)$$

Lemma 1. *The log-sum-exp function ,i.e., $f(x) = \log(e^{x_1} + \cdots + e^{x_n})$ is convex in $\mathbb{R}^n$.*

Proof. We prove $f(x)$ is convex by showing that $\nabla^2 f(x) \succeq 0$. Let $y = [e^{x_1}, \ldots, e^{x_n}]^\top$

$$\begin{aligned}
\nabla f(x) &= \frac{1}{e^{x_1} + \cdots + e^{x_n}}[e^{x_1}, \ldots, e^{x_n}]^\top = \frac{y}{\mathbf{1}^\top y} \\
\left[\nabla^2 f(x)\right]_{ij} &= \frac{(\mathbf{1}^\top y)e^{x_i} \cdot \mathbf{1}(i=j) - e^{x_i}e^{x_j}}{(\mathbf{1}^\top y)^2} \\
\implies \nabla^2 f(x) &= \frac{1}{(\mathbf{1}^\top y)^2}\left(\left(\mathbf{1}^\top y\right)\mathbf{diag}(y) - yy^\top\right) \\
\text{For any } v \in \mathbb{R}^n, v^\top \nabla^2 f(x) v &= \frac{1}{(\mathbf{1}^\top y)^2}\left(\left(\sum_{i=1}^n y_i\right)\left(\sum_{i=1}^n y_i v_i^2\right) - \left(\sum_{i=1}^n v_i y_i\right)^2\right) \ge 0
\end{aligned}$$

The last line follows by applying Cauchy-Schwarz inequality for the vectors $a = (\sqrt{y_1}, \ldots, \sqrt{y_n})$ and $b = (v_1\sqrt{y_1}, \ldots, v_n\sqrt{y_n})$. So, $\nabla^2 f(x) \succeq 0 \implies f(x)$ is convex. ☐

(a) Clearly $F_d(\boldsymbol{\theta})$ is a log-sum-exp function of the affine transformation of $\boldsymbol{\theta}$. So, using Lemma 1 $F_d(\boldsymbol{\theta})$ is convex.

(b) We will first compute the gradient of $F_d(\boldsymbol{\theta})$ w.r.t $\boldsymbol{\theta}$.

$$\begin{aligned}
\nabla_{\boldsymbol{\theta}} F_d(\boldsymbol{\theta}) &= \frac{d}{d\boldsymbol{\theta}}\log\left(\sum_{\boldsymbol{x}\in\mathcal{X}} \exp\left(\frac{1}{\sqrt{d}}\boldsymbol{\theta}^\top \boldsymbol{x}\right)\right) \\
&= \frac{1}{\left(\sum_{\boldsymbol{x}\in\mathcal{X}} \exp\left(\frac{1}{\sqrt{d}}\boldsymbol{\theta}^\top \boldsymbol{x}\right)\right)} \cdot \sum_{\boldsymbol{x}\in\mathcal{X}} \frac{\boldsymbol{x}}{\sqrt{d}} \cdot \exp\left(\frac{1}{\sqrt{d}}\boldsymbol{\theta}^\top \boldsymbol{x}\right) \\
\text{Rearranging, } \nabla_{\boldsymbol{\theta}} F_d(\boldsymbol{\theta}) &= \frac{1}{\sqrt{d}}\sum_{\boldsymbol{x}\in\mathcal{X}} \boldsymbol{x} \cdot \frac{\exp\left(\frac{1}{\sqrt{d}}\boldsymbol{\theta}^\top \boldsymbol{x}\right)}{\left(\sum_{\boldsymbol{x}\in\mathcal{X}} \exp\left(\frac{1}{\sqrt{d}}\boldsymbol{\theta}^\top \boldsymbol{x}\right)\right)} = \frac{1}{\sqrt{d}}\mathbb{E}_p[\boldsymbol{x}]
\end{aligned}$$

where, p is a probability distribution of $\boldsymbol{x}$ such that $p(\boldsymbol{x}) = \frac{\exp\left(\frac{1}{\sqrt{d}}\boldsymbol{\theta}^\top \boldsymbol{x}\right)}{\left(\sum_{\boldsymbol{x}\in\mathcal{X}} \exp\left(\frac{1}{\sqrt{d}}\boldsymbol{\theta}^\top \boldsymbol{x}\right)\right)}$

Now, we will estimate the norm of the gradient of $F_d(\boldsymbol{\theta})$ w.r.t $\boldsymbol{\theta}$.

$$\begin{aligned}
\|\nabla_{\boldsymbol{\theta}} F_d(\boldsymbol{\theta})\|_2 &= \frac{1}{\sqrt{d}} \|\mathbb{E}_p[\boldsymbol{x}]\|_2 \\
&\leq \frac{1}{\sqrt{d}} \mathbb{E}_p[\|\boldsymbol{x}\|_2] && \text{Jensen's inequality for } \ell_2\text{-norm, i.e.,convex function} \\
&= \frac{1}{\sqrt{d}} \cdot \sqrt{\frac{d(d-1)}{2}} && \text{As each coordinate of } \boldsymbol{x} \in \{\pm 1\} \text{ and } \boldsymbol{x} \text{ is } \binom{d}{2}\text{-dimensional} \\
&\leq \sqrt{d/2} && (3)
\end{aligned}$$

In the rest of the proof, norm refers to ℓ_2-norm unless stated otherwise. As $F_d(\boldsymbol{\theta})$ is a convex function, by using the first order condition for convexity, we get:

$$\begin{aligned}
F_d(\boldsymbol{\theta}') &\geq F_d(\boldsymbol{\theta}) + \nabla_{\boldsymbol{\theta}} F_d(\boldsymbol{\theta})^\top (\boldsymbol{\theta}' - \boldsymbol{\theta}) \\
\implies F_d(\boldsymbol{\theta}') - F_d(\boldsymbol{\theta}) &\geq \nabla_{\boldsymbol{\theta}} F_d(\boldsymbol{\theta})^\top (\boldsymbol{\theta}' - \boldsymbol{\theta}) && (4) \\
F_d(\boldsymbol{\theta}) &\geq F_d(\boldsymbol{\theta}') + \nabla_{\boldsymbol{\theta}'} F_d(\boldsymbol{\theta}')^\top (\boldsymbol{\theta} - \boldsymbol{\theta}') \\
\implies F_d(\boldsymbol{\theta}') - F_d(\boldsymbol{\theta}) &\leq \nabla_{\boldsymbol{\theta}'} F_d(\boldsymbol{\theta}')^\top (\boldsymbol{\theta}' - \boldsymbol{\theta}) && (5) \\
\implies \nabla_{\boldsymbol{\theta}} F_d(\boldsymbol{\theta})^\top (\boldsymbol{\theta}' - \boldsymbol{\theta}) &\overset{(4)}{\leq} F_d(\boldsymbol{\theta}') - F_d(\boldsymbol{\theta}) \overset{(5)}{\leq} \nabla_{\boldsymbol{\theta}'} F_d(\boldsymbol{\theta}')^\top (\boldsymbol{\theta}' - \boldsymbol{\theta}) \\
\text{So, } |F_d(\boldsymbol{\theta}') - F_d(\boldsymbol{\theta})| &\leq \max(|\nabla_{\boldsymbol{\theta}} F_d(\boldsymbol{\theta})^\top (\boldsymbol{\theta}' - \boldsymbol{\theta})|, |\nabla_{\boldsymbol{\theta}'} F_d(\boldsymbol{\theta}')^\top (\boldsymbol{\theta}' - \boldsymbol{\theta})|) \\
&\leq \max_{\boldsymbol{\alpha}\in\{\boldsymbol{\theta},\boldsymbol{\theta}'\}} (\|\nabla_{\boldsymbol{\alpha}} F_d(\boldsymbol{\alpha})\| \left\|\boldsymbol{\theta} - \boldsymbol{\theta}'\right\|) && \text{As } |a^\top b| \leq \|a\| \|b\| \\
&= \left\|\boldsymbol{\theta} - \boldsymbol{\theta}'\right\| \max_{\boldsymbol{\alpha}\in\{\boldsymbol{\theta},\boldsymbol{\theta}'\}} (\|\nabla_{\boldsymbol{\alpha}} F_d(\boldsymbol{\alpha})\|) \\
&\overset{(3)}{\leq} \sqrt{d/2} \left\|\boldsymbol{\theta} - \boldsymbol{\theta}'\right\|
\end{aligned}$$

(c) We state an extension to a theorem discussed in the class (without detailed proof) that which will be pivotal in proving the required inequality.

Theorem 2 (Extension to Theorem 2.26 of MJW[2])**.** *Let $(X_1, \cdots, X_n)$ be a vector of i.i.d. Gaussian random variables such that $X_i \sim \mathcal{N}(0, \sigma^2)$ and $f : \mathbb{R}^n \to \mathbb{R}$ be L-Lipschitz w.r.t. ℓ_2-norm, then $f(X) - \mathbb{E}[f(X)]$ is sub-Gaussian with parameter at most $L\sigma$.*

Proof. The proof can be obtained by following the proof schema as mentioned in MJW for a scaled version of X. Specifically, if $X \sim \mathcal{N}(0, \sigma^2)$, then $Y = \frac{X}{\sigma} \sim \mathcal{N}(0, 1)$. □

[2] Wainwright, Martin J. High-dimensional statistics: A non-asymptotic viewpoint. Vol. 48. Cambridge University Press, 2019.

Given all $\theta_{ij} \sim \mathcal{N}(0, \sigma^2)$ for $1 \leq i < j \leq d$ and are independent, we have,

$$
\begin{aligned}
F_d(\boldsymbol{\theta}) &= \log\left(\sum_{x\in\{\pm 1\}^d} \exp\left(\frac{1}{\sqrt{d}}\sum_{i<j}\theta_{ij}x_ix_j\right)\right) \\
\implies \mathbb{E}[F_d(\boldsymbol{\theta})] &= \mathbb{E}\left[\log\left(\sum_{x\in\{\pm 1\}^d} \exp\left(\frac{1}{\sqrt{d}}\sum_{i<j}\theta_{ij}x_ix_j\right)\right)\right] \\
&\leq \log\left(\mathbb{E}\left[\sum_{x\in\{\pm 1\}^d} \exp\left(\frac{1}{\sqrt{d}}\sum_{i<j}\theta_{ij}x_ix_j\right)\right]\right) && \text{As } \mathbb{E}[\log X] \leq \log \mathbb{E}[X] \\
&= \log\left(\sum_{x\in\{\pm 1\}^d} \mathbb{E}\left[\exp\left(\frac{1}{\sqrt{d}}\sum_{i<j}\theta_{ij}x_ix_j\right)\right]\right) \\
&= \log\left(\sum_{x\in\{\pm 1\}^d} \mathbb{E}\left[\exp\left(\frac{1}{\sqrt{d}}\theta_{12}x_1x_2\right)\right]^{\binom{d}{2}}\right) && \text{As } \theta_{ij}\text{' s are i.i.d} \\
&= \log\left(\sum_{x\in\{\pm 1\}^d} \exp\left(\frac{x_1^2x_2^2\sigma^2}{2d}\cdot\binom{d}{2}\right)\right) && \text{MGF for } X \sim \mathcal{N}(0,\sigma^2) \\
&= \log\left(\sum_{x\in\{\pm 1\}^d} \exp\left(\frac{\sigma^2(d-1)}{4}\right)\right) \\
&= \log\left(2^d \cdot \exp\left(\frac{\sigma^2(d-1)}{4}\right)\right) \\
&= d\ \log 2 + \frac{\sigma^2(d-1)}{4} \leq d\ \log 2 + \frac{\sigma^2 d}{4}
\end{aligned}
$$

$$
\implies F_d(\boldsymbol{\theta}) - \mathbb{E}[F_d(\boldsymbol{\theta})] \geq F_d(\boldsymbol{\theta}) - d\ \log 2 - \frac{\sigma^2 d}{4} \tag{6}
$$

As $(X_1, \ldots, X_n)$ is a Gaussian vector such that $X_i \sim \mathcal{N}(0, \sigma^2)$ and $F_d(\boldsymbol{\theta})$ is $\sqrt{d/2}$-Lipschitz, from Theorem 2, we know that $f(X) - \mathbb{E}[f(X)]$ is sub-Gaussian random variable with parameter atmost $\sigma\sqrt{d/2}$. So, using one sided tail bound for sub-Gaussian random variable, we have,

$$
\begin{aligned}
P(F_d(\boldsymbol{\theta}) - \mathbb{E}[F_d(\boldsymbol{\theta})] > t) &\leq \exp\left(\frac{-t^2}{2\cdot\frac{d}{2}\sigma^2}\right) \\
\implies P(F_d(\boldsymbol{\theta}) - \mathbb{E}[F_d(\boldsymbol{\theta})] > dt) &\leq \exp\left(\frac{-dt^2}{\sigma^2}\right) \\
P(F_d(\boldsymbol{\theta}) - d\ \log 2 - \frac{\sigma^2 d}{4} > dt) &\overset{(6)}{\leq} \exp\left(\frac{-dt^2}{\sigma^2}\right) \\
\implies P\left(\frac{F_d(\boldsymbol{\theta})}{d} > \log 2 + \frac{\sigma^2}{4} + t\right) &\leq \exp\left(\frac{-dt^2}{\sigma^2}\right)
\end{aligned}
$$

■

Problem 4

(a) Let $X = (X_1, \ldots, X_{2C})$ be a random vector such that X_i is the number of balls in bin i after all the T balls are thrown randomly. As each of the throws is independent of the other and each of the bins is equally probable to be obtained, X follows a *multinomial distribution* with parameters $T, p_i = 1/2C, i \in [2C]$.

Let $\mathcal{S}$ be the set of $2C$-length binary vectors such that each vector has equal number of 0s and 1s. So, $|\mathcal{S}| = \binom{2C}{C}$. Let $v \in \mathcal{S}$. Then $v^\top X$ is the total number of balls in exactly in C bins. So, the number of balls in the top C bins, i.e., $M_C(T)$ equals the maximum of $v^\top X$, for $v \in \mathcal{S}$.

$$
\begin{aligned}
\mathbb{E}[M_C(T)] &= \mathbb{E}[\max_{v \in \mathcal{S}} v^\top X] \\
&\leq \mathbb{E}\left[\frac{1}{\lambda}\log\left(\sum_{i=1}^{|\mathcal{S}|} \exp\left(\lambda v_i^\top X\right)\right)\right] && \text{summation} \geq \text{maximum} \\
&\leq \frac{1}{\lambda}\log\left(\mathbb{E}\left[\sum_{i=1}^{|\mathcal{S}|} \exp\left(\lambda v_i^\top X\right)\right]\right) && \text{Jensen's inequality} \\
&= \frac{1}{\lambda}\log\left(|\mathcal{S}| \cdot \mathbb{E}\left[\exp\left(\lambda v_i^\top X\right)\right]\right) && v_i\text{s are identical} \\
&= \frac{1}{\lambda}\log\left(|\mathcal{S}| \cdot \left(\frac{Ce^\lambda + C}{2C}\right)^T\right) && \text{for multinomial } X, \mathbb{E}\left[e^{t^\top X}\right] = \left(\sum_{i=1}^{2C} p_i e^{t_i}\right)^T \\
&= \frac{1}{\lambda}\log\left(|\mathcal{S}| \cdot e^{\lambda T/2} \cdot \left(\frac{e^{\lambda/2} + e^{-\lambda/2}}{2}\right)^T\right) \\
&\leq \frac{1}{\lambda}\log\left(|\mathcal{S}| \cdot e^{\lambda T/2} \cdot e^{\lambda^2 T/2}\right) && \frac{e^{\lambda/2} + e^{-\lambda/2}}{2} \leq e^{\lambda^2/2}, \lambda \in \mathbb{R} \\
&= \frac{T}{2} + \frac{\log|\mathcal{S}|}{\lambda} + \frac{\lambda T}{2}
\end{aligned}
$$

Choosing $\lambda = \sqrt{\frac{2\log|\mathcal{S}|}{T}}$, we get the tightest bound. Subsituting the value of λ, we get:

$$
\begin{aligned}
\mathbb{E}[M_C(T)] &\leq \frac{T}{2} + \sqrt{2T\log|\mathcal{S}|} \\
&= \frac{T}{2} + \sqrt{2T\log\binom{2C}{C}} \\
&\leq \frac{T}{2} + \sqrt{2T\log(2e)^C} && \binom{n}{k} \leq \left(\frac{ne}{k}\right)^k \\
&= \frac{T}{2} + \sqrt{2CT(1 + \log 2)} \\
&\leq \frac{T}{2} + 2\sqrt{CT}
\end{aligned}
$$

■

(b) Let us consider the case when $C = 1$. Let the random variable X denote the number of balls in the first bin. Note that $X = \sum_{i=1}^{T} Y_i$ where Y_i is a Bernoulli random variable with $p = 0.5$. So,

$$\begin{aligned}
\mathbb{E}[M_C(T)] &= \mathbb{E}[\max(X, T - X)] \\
&= \frac{T}{2} + \mathbb{E}\left[\max\left(X - \frac{T}{2}, \frac{T}{2} - X\right)\right] \\
&= \frac{T}{2} + \mathbb{E}\left[\left|X - \frac{T}{2}\right|\right] \\
&= \frac{T}{2} + \mathbb{E}\left[\left|\sum_{i=1}^{T} Y_i - \frac{T}{2}\right|\right] \\
&= \frac{T}{2} + \frac{1}{2}\mathbb{E}\left[\left|\sum_{i=1}^{T} (2Y_i - 1)\right|\right] \\
&= \frac{T}{2} + \frac{1}{2}\mathbb{E}\left[\left|\sum_{i=1}^{T} \epsilon_i\right|\right] && \text{where } \epsilon_i\text{s are i.i.d. Radamacher r.v} \\
&\overset{(a)}{=} \frac{T}{2} + \frac{1}{2} \cdot 2^{1-T} \left\lceil \frac{T}{2} \right\rceil \binom{T}{\lceil \frac{T}{2} \rceil} \\
&= \frac{T}{2} + \frac{1}{\sqrt{\pi}} \frac{\Gamma(S + 1/2)}{\Gamma(S)} && \text{where } S = \left\lceil \frac{T}{2} \right\rceil, \Gamma(n) = \int_0^\infty x^{n-1} e^{-x} dx \\
&\overset{(b)}{=} \frac{T}{2} + \frac{1}{\sqrt{\pi}} \cdot \sqrt{S}\left(1 - \frac{1}{8S} + \frac{1}{128S^2} + \dots\right) \\
&\overset{(c)}{=} \frac{T}{2} + \sqrt{\frac{T}{2\pi}}\left(1 \mp \frac{1}{4T} + \frac{1}{32T^2} \pm \dots\right) && (7) \\
&\geq \frac{T}{2} + \sqrt{\frac{T}{2\pi}} - \frac{1}{4\sqrt{2\pi T}} && \text{higher order terms are small for } T > 1 \\
&= \frac{T}{2} + \sqrt{\frac{T}{2\pi}} - O\left(\frac{1}{\sqrt{T}}\right) && (8)
\end{aligned}$$

The expression (a) follows from the expression of expected distance covered in a T-step 1-D random walk[3]. The step (b) follows from Graham *et al.*[4]. The expression (c) is obtained by subsituting $T = 2S$ or $T = 2S + 1$ depending upon the parity of T and the top signs are considered for T even and bottom signs for T odd.

Now, let us combine bins $(2k - 1, 2k), k \in [C]$ and call it a *superbin*[5]. Let M_i, T_i respectively denote the maximum and the total number of balls in the ith superbin. From (8), for $T_i > 0$,

[3] https://mathworld.wolfram.com/RandomWalk1-Dimensional.html

[4] Graham, R. L.; Knuth, D. E.; and Patashnik, O. Answer to problem 9.60 in Concrete Mathematics: A Foundation for Computer Science, 2nd ed. Reading, MA: Addison-Wesley, 1994.

[5] Parts of the subsequent proof closely follows a similar proof in: Bhattacharjee, Rajarshi, Subhankar Banerjee, and Abhishek Sinha. "Fundamental Limits on the Regret of Online Network-Caching." Abstracts of the 2020 SIGMETRICS/Performance Joint International Conference on Measurement and Modeling of Computer Systems. 2020.

we get:

$$\mathbb{E}[M_i|T_i] \geq \frac{T_i}{2} + \sqrt{\frac{T_i}{2\pi}} - \frac{1}{4\sqrt{2\pi T_i}} \tag{9}$$

Note that $M_C(T) \geq \sum_{i=1}^{C} M_i = CM_1$. The last equality follows as the M_i's are identical. Furthermore, notice that T_i follows a Binomial distribution with parameters T and $1/C$. So, $\mathbb{E}[T_i] = T/C$ and $\mathrm{Var}(T_i) = T\frac{1}{C}\left(1 - \frac{1}{C}\right)$.

$$\begin{aligned}
\mathbb{E}[M_C(T)] &\geq C\mathbb{E}[M_1] \\
&\stackrel{(a)}{=} C\mathbb{E}[M_1 \cdot \mathbf{1}(T_1 > 0)] \\
&\stackrel{(b)}{=} C\mathbb{E}[\mathbb{E}[M_1 \cdot \mathbf{1}(T_1 > 0)|T_1]] \\
&\stackrel{(9)}{\geq} C\mathbb{E}\left[\frac{T_1 \cdot \mathbf{1}(T_1 > 0)}{2} + \sqrt{\frac{T_1 \cdot \mathbf{1}(T_1 > 0)}{2\pi}} - \frac{\mathbf{1}(T_1 > 0)}{4\sqrt{2\pi T_1}}\right] \\
&= \underbrace{\frac{C}{2}\mathbb{E}\left[T_1 \cdot \mathbf{1}(T_1 > 0)\right]}_{\text{(I)}} + \underbrace{\frac{C}{\sqrt{2\pi}}\mathbb{E}\left[\sqrt{T_1 \cdot \mathbf{1}(T_1 > 0)}\right]}_{\text{(II)}} - \underbrace{\frac{C}{4\sqrt{2\pi}}\mathbb{E}\left[\frac{\mathbf{1}(T_1 > 0)}{\sqrt{T_1}}\right]}_{\text{(III)}}
\end{aligned}$$

The expression (a) follows because if $T_1 = 0 \implies M_1 = 0$. Step (b) describes the law of iterated expectations. As $T_1 \geq 0$,

$$\begin{aligned}
T_1 &= T_1 \cdot \mathbf{1}(T_1 = 0) + T_1 \cdot \mathbf{1}(T_1 > 0) \\
\sqrt{T_1} &= \sqrt{T_1} \cdot \mathbf{1}(T_1 = 0) + \sqrt{T_1} \cdot \mathbf{1}(T_1 > 0)
\end{aligned} \tag{10}$$

From (10), we obtain $\mathbb{E}[T_1] = \mathbb{E}[T_1 \cdot \mathbf{1}(T_1 > 0)]$ and $\mathbb{E}[\sqrt{T_1}] = \mathbb{E}[\sqrt{T_1} \cdot \mathbf{1}(T_1 > 0)]$. So, (I) becomes:

$$\frac{C}{2}\mathbb{E}\left[T_1 \cdot \mathbf{1}(T_1 > 0)\right] = \frac{C}{2}\mathbb{E}\left[T_1\right] = \frac{T}{2} \tag{11}$$

Consider,

$$\begin{aligned}
\sqrt{x} - \left(1 + \frac{x-1}{2} - \frac{(x-1)^2}{2}\right) &= \frac{1}{2}\sqrt{x}(\sqrt{x} - 1)^2(\sqrt{x} + 2) \geq 0 \\
\implies \sqrt{x} \geq 1 + \frac{x-1}{2} - \frac{(x-1)^2}{2}
\end{aligned}$$

Let $x = \frac{X}{\mathbb{E}[X]}$, we have almost surely:

$$\begin{aligned}
\sqrt{\frac{X}{\mathbb{E}[X]}} &\geq 1 + \frac{\frac{X}{\mathbb{E}[X]} - 1}{2} - \frac{(\frac{X}{\mathbb{E}[X]} - 1)^2}{2} \\
\implies \mathbb{E}\left[\sqrt{\frac{X}{\mathbb{E}[X]}}\right] &\geq \mathbb{E}\left[1 + \frac{\frac{X}{\mathbb{E}[X]} - 1}{2} - \frac{(\frac{X}{\mathbb{E}[X]} - 1)^2}{2}\right] \\
\implies \frac{\mathbb{E}[\sqrt{X}]}{\sqrt{\mathbb{E}[X]}} &\geq 1 - \frac{\mathrm{Var}(X)}{2(\mathbb{E}[X])^2}
\end{aligned}$$

$$\begin{aligned}
\implies \mathbb{E}[\sqrt{T_1}] &\geq \sqrt{\mathbb{E}[T_1]}\left(1 - \frac{\mathrm{Var}(T_1)}{2(\mathbb{E}[T_1])^2}\right) \\
&= \sqrt{\frac{T}{C}}\left(1 - \frac{T/C(1-1/C)}{2(T/C)^2}\right) \\
&= \sqrt{\frac{T}{C}}\left(1 - \frac{C-1}{2T}\right) \\
&\geq \sqrt{\frac{T}{C}} - \frac{1}{2}\sqrt{\frac{C}{T}}
\end{aligned}$$

So, (II) becomes:

$$\frac{C}{\sqrt{2\pi}}\mathbb{E}\left[\sqrt{T_1 \cdot \mathbf{1}(T_1 > 0)}\right] = \frac{C}{\sqrt{2\pi}}\mathbb{E}\left[\sqrt{T_1}\right] \geq \sqrt{\frac{CT}{2\pi}} - \frac{C^{3/2}}{2\sqrt{2\pi T}} \tag{12}$$

$$\begin{aligned}
\mathbb{E}\left[\frac{\mathbf{1}(T_1 > 0)}{\sqrt{T_1}}\right] &\overset{(1)}{=} \int_0^\infty P\left(\frac{\mathbf{1}(T_1 > 0)}{\sqrt{T_1}} > t\right) dt \\
&= \int_0^1 P\left(\frac{\mathbf{1}(T_1 > 0)}{\sqrt{T_1}} > t\right) dt + \int_{1^+}^\infty P\left(\frac{\mathbf{1}(T_1 > 0)}{\sqrt{T_1}} > t\right) dt
\end{aligned} \tag{13}$$

Consider the second term of the RHS. Since, $T_1 \sim Binom(T, 1/C) \implies T_1 \in \{0, 1, \ldots, T\}$. If $T_1 = 0 \implies \frac{\mathbf{1}(T_1>0)}{\sqrt{T_1}} = 0 \implies P(0 > t) = 0$ almost surely. Now, if $T_1 \geq 1$, then $\frac{1}{\sqrt{T_1}} \leq 1 \implies P\left(\frac{\mathbf{1}(T_1>0)}{\sqrt{T_1}} > t\right) = 0$ almost surely for $t \in (1, \infty)$. So, (13) becomes,

$$\begin{aligned}
\mathbb{E}\left[\frac{\mathbf{1}(T_1 > 0)}{\sqrt{T_1}}\right] &= \int_0^1 P\left(\frac{\mathbf{1}(T_1 > 0)}{\sqrt{T_1}} > t\right) dt \\
&\leq \int_0^1 dt = 1
\end{aligned}$$

So, (III) becomes,

$$\frac{C}{4\sqrt{2\pi}}\mathbb{E}\left[\frac{\mathbf{1}(T_1 > 0)}{\sqrt{T_1}}\right] \leq \frac{C}{4\sqrt{2\pi}} \implies -\frac{C}{4\sqrt{2\pi}}\mathbb{E}\left[\frac{\mathbf{1}(T_1 > 0)}{\sqrt{T_1}}\right] \geq -\frac{C}{4\sqrt{2\pi}} \tag{14}$$

Combining (11),(12) and (14), we get:

$$\begin{aligned}
\mathbb{E}[M_C(T)] &\geq \frac{T}{2} + \sqrt{\frac{CT}{2\pi}} - \frac{C^{3/2}}{2\sqrt{2\pi T}} - \frac{C}{4\sqrt{2\pi}} \\
\implies \mathbb{E}[M_C(T)] &\geq \frac{T}{2} + \sqrt{\frac{CT}{2\pi}} - O\left(\frac{1}{\sqrt{T}}\right)
\end{aligned}$$

■

Problem 5

Let $s_n = \max\left(\frac{z_n}{n}, 1 - \frac{z_n}{n}\right)$, where z_n is the number of zeros in the sequence be score assigned to a length-n binary sequence. Clearly, $0.5 \leq s_n \leq 1$, and lower bound equality holds only when the sequence is uninformative, i.e., number of zeros and ones are the same. So, we design our loss function as:

$$\phi_n = 1 - s_n = 1 - \max\left(\frac{z_n}{n}, 1 - \frac{z_n}{n}\right) \tag{15}$$

Consider a sequence $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)$ and $\boldsymbol{y'} = (y'_1, y'_2, \ldots, y'_n)$ obtained by flipping exactly one bit. So,

$$\left|\phi_n(\boldsymbol{y}) - \phi_n(\boldsymbol{y'})\right| = \left|\max\left(\frac{z_n(\boldsymbol{y})}{n}, 1 - \frac{z_n(\boldsymbol{y})}{n}\right) - \max\left(\frac{z_n(\boldsymbol{y'})}{n}, 1 - \frac{z_n(\boldsymbol{y'})}{n}\right)\right| \leq \frac{1}{n}$$

So, the candidate ϕ_n is smooth. Consider the case when $k = 1$. As it is known apriori that there is an imbalance of zeros and ones in the sequence, $s_n > 0.5 \implies \phi_n < 0.5$.

$$\begin{aligned}
\phi_n &= 1 - \max\left(\frac{z_n}{n}, 1 - \frac{z_n}{n}\right) \\
\implies \mathbb{E}\left[\phi_n\right] &= \mathbb{E}\left[1 - \max\left(\frac{z_n}{n}, 1 - \frac{z_n}{n}\right)\right] \\
&= 1 - \mathbb{E}\left[\max\left(\frac{z_n}{n}, 1 - \frac{z_n}{n}\right)\right] \\
&= 1 - \mathbb{E}\left[\frac{1}{2} + \max\left(\frac{z_n}{n} - \frac{1}{2}, \frac{1}{2} - \frac{z_n}{n}\right)\right] \\
&= \frac{1}{2} - \mathbb{E}\left[\left|\frac{z_n}{n} - \frac{1}{2}\right|\right] < \frac{1}{2}
\end{aligned}$$

As $\mathbb{E}[\phi_n] < 1/2$, by Cover's argument, we can say that there does not exist an algorithm that realises this loss function. So, if add an residual term and design a new loss function ϕ'_n as described in (16), we can prove that exists an algorithm that serves our purpose. Furthermore, it can be easily verified that $\mathbb{E}[\phi'_n] = 1/2$.

$$\phi'_n = \phi_n + \mathbb{E}\left[\left|\frac{z_n}{n} - \frac{1}{2}\right|\right] \tag{16}$$

Now, we extend for the case $k > 1$. Let $x_1, x_2, \ldots, x_k$ be the k-partitions, $L(x_1), L(x_2), \ldots, L(x_n)$ be the respective lengths of the partitions and $z(x_1), z(x_2), \ldots, z(x_n)$ be the number of zeros in the respective partitions. Let Φ_n be defined as follows:

$$\Phi_n = \max_{x_1, \ldots, x_n} \frac{1}{k} \sum_{i=1}^{k} \phi_{L(x_i)} \tag{17}$$

The max refers to maximum over all possible k-partitions of the sequence. Following the steps as earlier, we can show that Φ_n is also smooth.

$$
\begin{aligned}
\Phi_n &= \max_{x_1,\dots,x_n} \frac{1}{k}\sum_{i=1}^{k} \phi_{L(x_i)} \\
\implies \mathbb{E}[\Phi_n] &= \frac{1}{k}\mathbb{E}\left[\max_{x_1,\dots,x_n} \sum_{i=1}^{k} \phi_{L(x_i)}\right] \\
&= \frac{1}{k}\mathbb{E}\left[\max_{x_1,\dots,x_n} \sum_{i=1}^{k} \left(1 - \max\left(\frac{z(x_i)}{L(x_i)}, 1 - \frac{z(x_i)}{L(x_i)}\right)\right)\right] \\
&= \frac{1}{k}\mathbb{E}\left[\max_{x_1,\dots,x_n} \sum_{i=1}^{k} \left(\frac{1}{2} - \max\left(\frac{z(x_i)}{L(x_i)} - \frac{1}{2}, \frac{1}{2} - \frac{z(x_i)}{L(x_i)}\right)\right)\right] \\
&= \frac{1}{k}\mathbb{E}\left[\max_{x_1,\dots,x_n} \sum_{i=1}^{k} \left(\frac{1}{2} - \left|\frac{z(x_i)}{L(x_i)} - \frac{1}{2}\right|\right)\right] \\
&= \frac{1}{2} + \frac{1}{k}\mathbb{E}\left[\max_{x_1,\dots,x_n} \sum_{i=1}^{k} -\left|\frac{z(x_i)}{L(x_i)} - \frac{1}{2}\right|\right] \\
&= \frac{1}{2} - \frac{1}{k}\mathbb{E}\left[\min_{x_1,\dots,x_n} \sum_{i=1}^{k} \left|\frac{z(x_i)}{L(x_i)} - \frac{1}{2}\right|\right] < \frac{1}{2}
\end{aligned}
$$

The last inequality follows because one can always choose a k-partition such that there is an imbalance in at least one of the partitions, so the second term > 0. As $\mathbb{E}[\Phi_n] < 1/2$, there is no algorithm that achieves the said loss function. So, we devise a modified loss function Φ'_n in (18) which satisfies $\mathbb{E}[\Phi'_n] = 1/2$.

$$
\Phi'_n = \Phi_n + \frac{1}{k}\mathbb{E}\left[\min_{x_1,\dots,x_n} \sum_{i=1}^{k} \left|\frac{z(x_i)}{L(x_i)} - \frac{1}{2}\right|\right] \tag{18}
$$

■