Problem 1

Solution

(a) Because $\forall i$, f_i is postive non-decreasing, $\prod_{i=1}^{k'} f_i$, $\forall 1 \le k' \le k$ is also non-decreasing. Thus,

$$\mathbb{E}\left[\prod_{i=1}^{k} f_i(\mathbf{X})\right] \le \mathbb{E}\left[\prod_{i=1}^{k-1} f_i(\mathbf{X})\right] \mathbb{E}\left[f_k(\mathbf{X})\right] \le \cdots \le \prod_{i=1}^{k} \mathbb{E}\left[f_i(\mathbf{X})\right].$$

(b) NA r.v.s $\{X_1, \ldots, X_n\}$ may not be independent, but if we can prove that Chernoff's trick also holds for NA r.v.s, then Hoeffding's inequality holds too. Using the results in Problem 2.(a).ii, $\{X_1 - \mathbb{E}[X_1], \ldots, X_n - \mathbb{E}[X_n]\}$ are also NA. For any $a > 0$ and $\theta > 0$, we have

$$P(\sum_{i=1}^{n} X_i - \mathbb{E}[X_i] \ge a) \le \frac{\mathbb{E}\left[\prod_{i=1}^{n} e^{\theta(X_i - \mathbb{E}[X_i])}\right]}{e^{\theta a}} \le \frac{\prod_{i=1}^{n} \mathbb{E}\left[e^{\theta(X_i - \mathbb{E}[X_i])}\right]}{e^{\theta a}},$$

where the first derivation uses Markov's inequality, and the last derivation uses the result in (a). Suppose $X_i \in [a_i, b_i]$, then we have

$$P(\sum_{i=1}^{n} X_i - \mathbb{E}[X_i] \ge a) \le \exp\left(\prod_{i=1}^{n} \frac{(b_i - a_i)^2}{8}\theta^2 - \theta a\right) \le \exp\left(\frac{-2x^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right)$$

Problem 2

Solution

(a) (i) Here, we only consider increasing functions. Proof for decreasing functions is similar. If $f(X, Y)$ is an increasing function of (X, Y), then given $Y = y$, $f(X, y)$ is an increasing function of X. Also, for any $y_1 \le y_2$, we have $\forall X$, $f(X, y_1) \le f(X, y_2)$ and thus $\mathbb{E}[f(X, y_1)] \le \mathbb{E}[f(X, y_2)]$. Therefore, as a function of Y, $\mathbb{E}[f(X, Y) \mid Y]$ is also increasing. Thus, for any increasing function f and g,

$$\begin{aligned}
&\mathbb{E}[f(X, Y) \cdot g(X, Y)] = \mathbb{E}[\mathbb{E}[f(X, Y) \cdot g(X, Y) \mid Y]] \\
&\quad \le \mathbb{E}[\mathbb{E}[f(X, Y) \mid Y] \cdot \mathbb{E}[g(X, Y) \mid Y]] \le \mathbb{E}[\mathbb{E}[f(X, Y) \mid Y]] \cdot \mathbb{E}[\mathbb{E}[g(X, Y) \mid Y]] \\
&\quad = \mathbb{E}[f(X, Y)] \cdot \mathbb{E}[g(X, Y)]
\end{aligned}$$

(ii) For any two monotone functions g and h depending on disjoint subsets of Y, $g(Y) = g(\{Y_i,\ i \in \mathcal{I}\}) = g(\{f_i(X),\ i \in \mathcal{I}\})$ and $h(Y) = h(\{f_i(X),\ i \in \mathcal{J}\})$. Since $\{f_i(X),\ i \in [k]\}$ depends on disjoint subsets of X, g and h also depend on disjoint subsets of X. Also, since monotone functions are closed under composition, both g and h are still monotone (either increasing or decreasing) functions of X depending disjoint subsets of X. Thus, $\mathbb{E}[g(Y) \cdot h(Y)] \le \mathbb{E}[g(Y)] \cdot \mathbb{E}[h(Y)]$, and thus Y^k are NA.

(b) Let $I_{j,i}$ be an indicator random variable to show if the j-th ball falls into the i-th bin. Then, for any $j \in [m]$, $\sum_{i=1}^{n} I_{j,i} = 1$. Next, we will show $I = \{I_{j,i}, j \in [m]\}$ are NA. For any increasing functions f and g, denoting $\tilde{f}(I)$ and $\tilde{g}(I)$ as $f(I) - f(\mathbf{0})$ and $g(I) - g(\mathbf{0})$ respectively, then we have $\tilde{f}(I) \ge 0$ and $\tilde{g}(I) \ge 0$. Since only one element of I can be 1, and f and g depend on disjoint subsets of I, at least one of $\tilde{f}(I)$ and $\tilde{g}(I)$ is zero, so $\mathbb{E}\left[\tilde{f}(I) \cdot \tilde{g}(I)\right] = 0$. Thus, $LHS =$

$\mathbb{E}[f(I)\cdot g(I)] = \mathbb{E}\left[(\tilde{f}(I)+f(\mathbf{0}))\cdot(\tilde{g}(I)+g(\mathbf{0}))\right] = g(\mathbf{0})\mathbb{E}\left[\tilde{f}(I)\right] + f(\mathbf{0})\mathbb{E}[\tilde{g}(I)] + f(\mathbf{0})g(\mathbf{0})$, and $RHS = (\mathbb{E}\left[\tilde{f}(I)\right] + f(\mathbf{0}))(\mathbb{E}[\tilde{g}(I)] + g(\mathbf{0})) \geq LHS$. Thus, $\{I_{j,i}, j\in[m]\}$ are NA.

Using the results in (a.i), we have $\{I_{j,i},$ for any $j\in[m]$ and $i\in[n]\}$ are NA. Using the results in (a.ii), we have $X_i = \sum_{j=1}^m I_{j,i}$ are NA.

(c) Let Y_i be an indicator random variable to indicate if the i-th bin is non-empty. We have $Y_i \sim Ber(1-(\frac{n-1}{n})^m)$, and $\mathbb{E}[Y_i] = 1-(\frac{n-1}{n})^m$. Using the results in (a.ii), $Y_i = \min\{1, X_i\}$ is NA. Using the results in Problem 1 (b), we have $\forall o > 0$,

$$P\left(O - \mathbb{E}[O] = \sum_{i=1}^n (Y_i - \mathbb{E}[Y_i]) \geq o\right) \leq \exp\left(\sum_{i=1}^n \frac{(b_i-a_i)^2}{8}\theta^2 - \theta o\right) = \exp\left(\frac{n}{8}\theta^2 - \theta o\right) \leq e^{\frac{-2o^2}{n}}$$

Problem 3

Solution

Let $Z_i \in \{1,2,3,\ldots,n\}$ be the index of the bin into which the i-th ball is put. Let f be a function such that $f(Z_1,Z_2,Z_3,\ldots,Z_m)$ is the number of non-empty bins. Obviously, Z_i's are independent, and $|f(Z_1,\ldots,Z_i,\ldots,Z_m) - f(Z_1,\ldots,Z_i',\ldots,Z_m)| < 1$. Thus, McDiarmid's inequality applys: $\forall o > 0$,

$$P(O - \mathbb{E}[O] = f(Z_1,Z_2,Z_3,\ldots,Z_m) - \mathbb{E}[f(Z_1,Z_2,Z_3,\ldots,Z_m)] \geq o) \leq e^{\frac{-2o^2}{n}},$$

which is the same as the Hoeffding bound.

Problem 4

Solution

(a) We have $\mathbb{E}[\exp(-\theta X)] \leq \exp(\nu^2\theta^2/2)$. Thus for any $t > 0$ and $\theta > 0$,

$$\begin{aligned} P(|X|>t) &= P(X>t) + P(-X>t) = P(e^{\theta X} > e^{\theta t}) + P(e^{-\theta X} > e^{\theta t}) \\ &\leq \frac{\mathbb{E}\left[e^{\theta X}\right]}{e^{\theta t}} + \frac{\mathbb{E}\left[e^{-\theta X}\right]}{e^{\theta t}} \leq 2\exp(\nu^2\theta^2/2 - \theta t) \leq 2\exp(-\frac{t^2}{2\nu^2}) \end{aligned}$$

(b) Let the CDF of $|X|$ be F. Then, $F(x) = 0$, $\forall x \leq 0$. Because $g(x) = x^k$ is an increasing function, using the Integration-by-parts Theorem [1], we have

$$\mathbb{E}\left[|X|^k\right] = \int_0^\infty g'(x)P(|X|>x)dx \leq 2\int_0^\infty kx^{k-1}\exp(-\frac{x^2}{2\nu^2})dx.$$

Subtitute $\frac{x^2}{2\nu^2}$ with y, and we have

$$\mathbb{E}\left[|X|^k\right] \leq 2k\int_0^\infty \nu^2(2\nu^2 y)^{k/2-1}\exp(-y)dy = k(2\nu^2)^{k/2}\cdot\Gamma(\frac{k}{2}).$$

[1] See Example 7 in https://www.math.arizona.edu/~jwatkins/g-expectation.pdf

Problem 5

Solution

First, we partation the indices into K index group such that $G_k = \{j : X_j \in A_k\}$. We also have for any $f \in F$ and $j_1, j_2 \in G_i$, $f(X_{j_1}) = f(X_{j_2})$. Thus,

$$R_n(F(X^n)) = \mathbb{E}\left[\sup_{f\in F}\left|\frac{1}{n}\sum_{i=1}^{n}\sigma_i f(X_i)\right|\right] = \mathbb{E}\left[\sup_{f\in F}\left|\frac{1}{n}\sum_{k=1}^{K}\sum_{i\in G_k}\sigma_i f(X_i)\right|\right] = \mathbb{E}\left[\frac{1}{n}\sum_{k=1}^{K}\left|\sum_{i\in G_k}\sigma_i\right|\right]$$
$$\le \mathbb{E}\left[\frac{1}{n}\sum_{k=1}^{K}\sqrt{\sum_{i\in G_k}\sigma_i^2}\right] = \mathbb{E}\left[\frac{1}{n}\sum_{k=1}^{K}\sqrt{n_k}\right] = \frac{1}{n}\sum_{k=1}^{K}\sqrt{n_k}$$

Problem 6

Solution

For $n = 3$, we can choose $z_2 = z_1 + 0.55$ and $z_3 = z_2 + 0.55$. It is easy to verify z_1, z_2, and z_3 can be shattered by F. When $n = 4$, no matter what $z_1 < z_2 < z_3 < z_4$ are, they cannot be assigned values 1, 0, 1, 0. Thus $VC(F) = 3$.

Problem 7

Solution

For n points, we can choose $z_i = 2^i\pi$, $i = 1, 2, 3, \ldots, n$. Then z^n can be shattered by F. Given a valuation $b_1, \ldots, b_n$, we can choose $\theta = 2^{-n-1} + \sum_{i=1}^{n} 2^{-i}b_i$. Then, $\theta z_i = \pi(b_i + \sum_{j<i} 2^{i-j}b_j + \sum_{j>i}^{n} 2^{i-j}b_j + 2^{i-n-1})$, where $\sum_{j<i} 2^{i-j}b_j$ is an even number, and $\sum_{j>i}^{n} 2^{i-j}b_j + 2^{i-n-1} \in (0, 1)$. Thus, $\mathbf{1}_{\{\sin(\theta z_i > 0)\}} = b_i$.