7.1 Introduction

In this lecture, we discuss the McDiarmid’s inequality and its applications. First, we recall Azuma-Hoeffding’s inequality and McDiarmid’s inequality, then we go through a number of examples where these inequalities find applications such as bin packing, longest common subsequence, longest increasing subsequence, log partition function in SK spin glass model, Glauber dynamics.

7.1.1 Recap

**Theorem 7.1** (Azuma-Hoeffding’s inequality). Let $(M_n, F_n)_{n \geq 1}$ be a mean-zero martingale and $A_K \leq M_K - M_{K-1} \leq B_K$ almost surely $\forall k \geq 1$ where $A_K, B_K$ are $F_{K-1}$ measurable, then

$$M_n \in \mathcal{G} \left( \frac{1}{4} \sum_{k=1}^{n} C_K^2 \right),$$

where $C_K = \|B_K - A_K\|_{\infty}$.

**Corollary 7.2** (McDiarmid’s inequality). Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be such that,

$$C_k = \sup_{x_1, \ldots, x_n, x'_k} \left( f(x_1, \ldots, x_n) - f(x_1, \ldots, x'_k, \ldots, x_n) \right) < \infty,$$

$\forall k$, and $X_1, \ldots, X_n$ be independent random variables. Then,

$$f(X_1, \ldots, X_n) - \mathbb{E}[f(X_1, \ldots, X_n)] \in \mathcal{G} \left( \frac{1}{4} \sum_{k=1}^{n} C_K^2 \right)$$

7.1.2 Examples

7.1.2.1 Bin packing

**Example 7.3.** Let $[x_1, \ldots, x_n] \in (0, 1)$ represent the size of $n$ elements $[1, \ldots, n]$ respectively, and let $[B_1, \ldots, B_k]$ represent $k$ bins each of size 1. We want to find the minimum number of bins required to fill in all the $n$ elements in the $k$ bins. More formally, we want to partition the elements
\[ [i_1, \ldots, i_n] \text{ using the minimum number of bins such that the sum of size of elements in each bin is less than or equal to 1. To that end, define } f(x_1, \ldots, x_n) \text{ as} \]
\[ f(x_1, \ldots, x_n) = \min \{ k \mid \exists \text{ a partition } [B_1, \ldots, B_k] \text{ of } [1, \ldots, n] \text{ such that } \sum_{i \in B_j} x_i \leq 1 \forall j \in \{1, \ldots, k\} \} \]
\[ (7.4) \]

Note that the \( C_k \) corresponding to (7.2) for the function \( f \) in (7.4) is 1, i.e. \( C_k = 1 \). Thus, from Corollary. 7.2 it follows that if \( X_1, \ldots, X_n \) are independent random variables and \( 0 \leq X_i \leq 1 \) almost surely for all \( i \in [1, \ldots, n] \), then
\[ f(X_1, \ldots, X_n) \in \mathcal{G}(\frac{1}{4} \times 4n) = \mathcal{G}(n), \]
\[ (7.5) \]
where \( f(X_1, \ldots, X_n) = f(X_1, \ldots, X_n) - \mathbb{E}[f(X_1, \ldots, X_n)] \).

### 7.1.2.2 Longest common subsequence

Let \( x \in \{0, 1\} \) and \( y \in \{0, 1\} \). Let \( f(x, y) \) represent the longest common subsequence between the sequences \( x \) and \( y \) respectively. More formally,
\[ f(x, y) = \max \{ k \mid \exists i_1 < \cdots < i_k \text{ and } j_1 < \cdots < j_k \text{ such that } x_{i_1} = y_{j_1}, \ldots, x_{i_k} = y_{j_k} \} \]
\[ (7.6) \]
Here we have \( C_k = 1 \) where \( C_k \) is as defined in (7.2) with respect to the function in (7.6). Thus, from Corollary. 7.2 we have that if \( X_1, \ldots, X_n, Y_1, \ldots, Y_n \) are independent random variables, then
\[ f(X, Y) \in \mathcal{G} \left( \frac{1}{4} \times 1 \times 2n \right) = \mathcal{G} \left( \frac{n}{2} \right), \]
\[ (7.7) \]
where \( f(X, Y) = f(X, Y) - \mathbb{E}[f(X, Y)] \).

**Fact 7.4** (Sub-additive theorem). If \( X_1, \ldots, X_n \) are i.i.d. random variables distributed as \( \mu \), and \( Y_1, \ldots, Y_n \) are i.i.d. random variables distributed as \( \nu \), then,
\[ \frac{1}{n} \mathbb{E}[f(X_1, \ldots, X_n, Y_1, \ldots, Y_n)] \rightarrow C(\mu, \nu). \]
\[ (7.8) \]

### 7.1.2.3 Longest increasing subsequence

Let \( \pi_n \sim U(S_n) \), i.e. \( \pi_n \) is a random variable uniformly taking values from \( S_n \) where \( S_n \) is the finite symmetric group on \( n \) elements consisting of all permutations that can be performed on the \( n \) elements. Let \( \bar{\pi} = [i_1, \ldots, i_n] \) be \( n \) elements and \( \pi_n(i) = [\pi_n(i_1), \ldots, \pi_n(i_n)] \) be their corresponding positions. Then the longest increasing subsequence of \( \bar{\pi}, W \), can be written as
\[ W(\pi_n) = \sup \{ k \mid \exists i_1 < \cdots < i_k \text{ such that } \pi_n(i_1) < \cdots < \pi_n(i_k) \}. \]
\[ (7.9) \]
Take \( X_1, \ldots, X_n \sim U(0, 1) \) and define
\[ f(x_1, \ldots, x_n) = \sup \{ k \mid \exists \ i_1 < \cdots < i_k \text{ such that } x_{i_1} < \cdots < x_{i_k} \}. \quad (7.10) \]

We can show that \( W(\pi_n) = f(X_1, \ldots, X_n) \) from the following exercise.

**Exercise 7.1.** Let \( X_{(1)} < \cdots < X_{(n)} \) be the ordered values of \( X_1, \ldots, X_n \). Define \( X_{\pi_k} = X_{(k)} \) for \( k \in \{1, \ldots, n\} \). Then one can show that \( \pi_n \sim U(S_n) \).

Observe that for \( W(\pi_n) \), we have \( C_k = 1 \forall k \in \{1, \ldots, n\} \). Thus, from Corollary. 7.2 it follows that

\[ W(\pi_n) - \mathbb{E}[W(\pi_n)] \in \mathcal{O}(\frac{1}{4} \times n \times 1 = \frac{n}{4}). \]

### 7.1.2.4 Log partition function in SK spin glass model

Let \( \sigma_1, \ldots, \sigma_n \in \{-1, +1\} \) and let \( J_{ij} \) be independent random variables such that \( \mathbb{E}[J_{ij}] = 0 \). The goal here is to obtain \( \sup_{\sigma} \sum_{i<\j} \sigma_i \sigma_j J_{ij} \).

**Finite temperature model**

For finite \( \beta \), define \( F(J) \) as

\[ F(J) = \frac{1}{\beta} \log \sum_{\sigma \in \{-1, +1\}^n} \exp \left( \frac{\beta}{\sqrt{n}} \sum_{i<\j} \sigma_i \sigma_j J_{ij} \right). \]

Then taking partial derivative of \( F(J) \) with respect to \( J_{ij} \), we have

\[ \left| \frac{\partial F(\beta)}{J_{ij}} \right| = \frac{\sum_{\sigma} \exp \left( \frac{\beta}{\sqrt{n}} \sum_{i<\j} \sigma_i \sigma_j J_{ij} \right) \sigma_i \sigma_j}{\sum_{\sigma} \exp \left( \frac{\beta}{\sqrt{n}} \sum_{i<\j} \sigma_i \sigma_j J_{ij} \right)} = \left| \frac{\sigma_i \sigma_j}{\sqrt{n}} \right| \approx \left| \frac{1}{\sqrt{n}} \right|. \]

Thus, we have, if \( |J_{ij}| \leq 1 \), then \( C_{ij} = \frac{2}{\sqrt{n}} \) using Mean Value Theorem, which states \( F_\beta(J) - F_\beta(J') = \frac{\partial F_\beta}{\partial J_{ij}} \mid_{J_{ij}} \times (J_{ij} - J'_{ij}) \) for some \( J_{ij}^* \in [J_{ij}, J'_{ij}] \). Then,

\[ F_\beta(J) - \mathbb{E}[F_\beta(J)] \in \mathcal{O}(\frac{1}{4} \times \left( \frac{n}{2} \right) \times \frac{4}{n} = \frac{n-1}{2}). \]

### 7.1.2.5 2- D directed random polymer problems

Consider a \( n \times n \) grid made of \( n^2 \) points \((i, j)\) for \( 0 \leq i \leq n - 1, 0 \leq j \leq n - 1 \). We consider all **valid** paths from \((0, 0)\) to \((n - 1, n - 1)\) where valid path simply means that the only moves that can be made on the grid are to the right or up. Further, every unit path (i.e. 'right' or 'up') has a
weight $w_e$ associated with it. The goal here is to find $\sup P : (0,0) \rightarrow (n-1,n-1) \sum_{e \in P} w_e$. Now define $F_\beta(w)$ as follows

$$F_\beta(w) = \frac{1}{\beta} \log \sum_{P:(0,0)\rightarrow(n-1,n-1)} \exp \beta \sum_{e \in P} w_e$$

$$\frac{\partial F_\beta(w)}{\partial w_e} = \frac{1}{\beta} \frac{\sum_{P:(0,0)\rightarrow(n-1,n-1)} \exp \beta \sum_{e \in P} w_e \mathbb{1}_{e \in P}}{\sum_{P:(0,0)\rightarrow(n-1,n-1)} \exp \beta \sum_{e \in P} w_e} = \frac{\mathbb{1}_{e \in P}}{\sum_{P'} \exp \beta \sum_{e \in P'} w_e}.$$

Thus, if $|w_e| \leq 1$ for all $e$ in the grid, then

$$\frac{\partial F_\beta(w)}{\partial w_e} \in \mathcal{G}(\frac{1}{4} \times 1 \times 4n^2 = n^2). \quad (7.11)$$

This example shows that the mean is $\mathcal{O}(n)$ while the variance is $\mathcal{O}(n^2)$, hence there is no concentration here.

### 7.1.2.6 Hoeffding’s statistics

Let $A_{n \times n}$ be a $n \times n$ matrix with entries $a_{ij}$, $1 \leq i, j \leq n$, let $\pi_n \sim U(S_n)$ and $W(\pi_n) = \sum_{i=1}^{n} a_{i\pi_i}$. Now we discuss Hoeffding’s combinatorial central limit theorem. Let $a_{ij} = |i - j|$, then $W(\pi_n) = \sum_{i=1}^{n} |i - \pi_i|$.

Without loss of generality, update the matrix entries as $a_{ij} \rightarrow a_{ij} - \bar{a}_{i+} - \bar{a}_{+j} + \bar{a}_{++}$, where $\bar{a}_{i+} = \frac{1}{n} \sum_j a_{ij}$, $\bar{a}_{+j} = \frac{1}{n} \sum_i a_{ij}$, and $\bar{a}_{++} = \frac{1}{n^2} \sum_i \sum_j a_{ij}$. Further, without loss of generality, assume $\sum_{i,j} a_{i,j}^2 = 1$. If $\sum_{i,j} |a_{i,j}|^3 \rightarrow 0$ as $n \rightarrow \infty$, then, $W(\pi_n) \rightarrow \mathcal{N}(0,1)$.

### 7.1.2.7 Glauber dynamics

If $S_n$ represents the symmetry group of $n$ elements and $(I,J)$ be a random transposition chosen uniformly from the set of all transpositions on $n$ elements. Then, $\pi_n \rightarrow \pi_n \circ (I,J)$ forms a reversible Markov chain.

Now we discuss Glauber dynamics where we find similar reversible Markov chains. Define

$$\mu_\beta(\sigma_1, \ldots, \sigma_n) = \frac{\exp \beta H(\sigma_1, \ldots, \sigma_n)}{\sum_{T} \exp \beta H(T)},$$

where $H$ is the Hamiltonian. The goal is to maximize $H(\sigma_1, \ldots, \sigma_n)$ over the values of $\sigma_i$'s. We choose $I \sim U(\{1, \ldots, n\})$ and generate $\sigma_J = \sigma_I | \sigma_J$, in distribution for $J \neq I$. This forms a reversible Markov chain. The simplest case would be when all the $\sigma_i$'s are independent, in that case it is easy to see the reversible Markov chain since in that case, we are simply choosing a random index and generating an independent copy of $X_I$. 
7.1.2.8 Ising model on complete graph (Curie-Weiss model)

The Curie-Weiss model has the Hamiltonian as

$$H(\sigma_1, \ldots, \sigma_n) = \frac{1}{n} \sum_{i < j} \sigma_i \sigma_j,$$

where $\sigma_i \in \{+1, -1\}$ for all $i \in \{1, \ldots, n\}$. One can prove the following exercise:

**Exercise 7.2.** \(\mathbb{E}_\beta(\sigma_i | \sigma_j, j \neq i) = \tanh(\frac{\beta}{n} \sum_{j \neq i} \sigma_j).\)