Homework 2 *Martingale Theory, S. S. Mukherjee, Fall 2025*

Topics: Applications of martingales Due on Nov 7, 2025

Name of student:
Roll number:

1. The Hoeffding decomposition and CLT for U-statistics [5]

Let $X_1, X_2, \ldots$ be i.i.d. random variables and h be an order-m symmetric kernel, with $\mathbb{E}[|h(X_1, \ldots, X_m)|] < \infty$. For $n \geq m$, consider the U-statistic

$$U_n(h) = \frac{1}{\binom{n}{m}} \sum_{1 \leq i_1 < \cdots < i_m \leq n} h(X_{i_1}, \ldots, X_{i_m}).$$

Prove that

$$U_n(h) = \sum_{k=0}^{m} \binom{m}{k} U_n(h_k),$$

where

$$h_k(x_1, \ldots, x_k) = \sum_{\ell=0}^{k} (-1)^{k-\ell} \sum_{1 \leq i_1 < \cdots < i_\ell \leq k} g_\ell(x_{i_1}, \ldots, x_{i_\ell}),$$

with

$$g_\ell(x_1, \ldots, x_\ell) := \mathbb{E}[h(X_1, \ldots, X_m) \mid X_1 = x_1, \ldots, X_\ell = x_\ell], \quad 1 \leq \ell \leq k,$$

and $g_0 \equiv \mathbb{E}[h(X_1, \ldots, X_m)]$. This representation is called the *Hoeffding decomposition.*

Show that the $U_n(h_k)$'s are pairwise orthogonal in L^2 (i.e. uncorrelated) across k and $\{\binom{n}{k} U_n(h_k)\}_{n \geq m}$ is a martingale for each $0 \leq k \leq n$.

Assuming that $\mathbb{E}[g_1(X_1)^2] < \infty$, show that

$$\sqrt{n}\,(U_n(h) - \mathbb{E}[h(X_1, \ldots, X_m)]) \xrightarrow{d} \mathcal{N}(0, m^2 \cdot \mathrm{Var}(g_1(X_1))).$$

2. Concentration in the binary hypercube [5]

Let $\mathcal{C} = \{0,1\}^n$ denote the binary hypercube. For $x, y \in \mathcal{C}$, let $d_H(x, y) := \sum_{i=1}^n \mathbb{I}(x_i \neq y_i)$ denote the *Hamming distance* between x and y. For $A \subset \mathcal{C}$, we define $d_H(x, A) := \inf_{y \in A} d_H(x, y)$. Suppose Z is a uniform random point in $\mathcal{C}$. Prove that for any $A \subset \mathcal{C}$, with $|A| > 2^{n-1}$, one has

$$\mathbb{E} d_H(Z, A) < \sqrt{2n \log 2}.$$

Deduce that for any $\delta > 0$, one has

$$\left|\{x : d_H(x, A) \leq (\delta + \sqrt{2 \log 2})\sqrt{n}\}\right| \geq (1 - e^{-\frac{\delta^2}{2}}) 2^n,$$

i.e. a non-trivial fraction of the points of $\mathcal{C}$ are within $O(\sqrt{n})$ distance from the set A.

Hint: Apply Azuma on an appropriate Doob martingale!