ESE 6180: Learning, Dynamics and Control	**Fall 2024**

Homework 3

Assigned: 11/4/2024 — *Due: 11/20/2024*

Homework must be LaTeX'd or it will not be graded.

Grading: Each problem will be graded on a scale of 0-4. If you get 80% of the problem or more correct, and make an honest attempt at the rest, you will get 4/4. If you get 60% of the problem or more correct, you will get 3/4, etc.

Canvas: Please submit your HW as a single pdf file, with pages correctly tagged to go with each problem.

Working in groups: You are allowed to, and in fact encouraged, to discuss and work on problems with your classmates. However, each student must write up their own homework independently. Further, please make note of your collaborators in the designated spot in the homework template.

Citing references: If you referred to solutions found in published material (papers, textbooks, websites, etc.), you must cite these in your homework solutions. It is ok to use proofs that you find online for guidance, but you should indicate where and how you did so, and you should always make a first attempt at the answer on your own. Importantly, even if you are following the guidance of a proof from a paper, you must be sure to fully explain all steps, as well as fill in any missing steps.

Useful inequalities: This cheat sheet may come in handy throughout the course.

Please note that all four exercises this time are meant to be solved in order. Results from the previous exercise **will be useful for subsequent exercises.**

The overarching theme of this sheet is to study a method for analyzing more general (than linear) dependent random variables.

1 Martingales

Fix an increasing family of sigma-fields $\mathcal{F}_1 \subset \cdots \subset \mathcal{F}_n$. A sequence of random variables $X_{1:n}$, adapted to $\mathcal{F}_{1:n}$,[1] is called a martingale if for any time $t \in [n]$ 1) $\mathbf{E}[X_{t+1}|\mathcal{F}_t] = X_t$ and, 2) $\mathbf{E}|X_t| < \infty$. In this question, we consider the canonical example of a martingale, namely a random walk.

(a) Let $W_{1:n} \in N(0, I_n)$ and define $S_k \triangleq \sum_{i=1}^k W_i$. Show that $S_{1:n}$ is a martingale. Hint: condition on $W_{1:k-1}$.

(b) Fix $\sigma > 0$ and suppose now instead that $W_{1:n}$ is a sequence of random variables satisfying $\mathbf{E}[W_k|W_{1:k-1}] = 0$ and $\mathbf{E}\left[\exp\left(\lambda W_k\right) \middle| W_{1:k-1}\right] \le \exp\left(\frac{\lambda^2\sigma^2}{2}\right)$ for every $\lambda \in \mathbb{R}$. Such variables are called sub-Gaussian martingale difference sequences and can be thought of as a general class of increment process for a random walk. Let $S_k \triangleq \sum_{i=1}^k W_i$ for $k \in [n]$ and show that

$$\mathbf{E}\left[\exp\left(\lambda n^{-1/2} S_n\right)\right] \le \exp\left(\frac{\lambda^2\sigma^2}{2}\right), \quad \forall \lambda \in \mathbb{R}. \tag{1}$$

[1] By adapted we mean that each $X_{1:k}$ is $\mathcal{F}_k$-measurable.

2 The Bounded Differences Inequality

In this exercise we continue our study of martingales and apply them to understand random functions. Let $f : \mathbb{R}^n \to \mathbb{R}$ and let $X_{1:n}$ be a sequence of iid random variables in $\mathbb{R}$. The *Doob-decomposition* of $f(X_{1:n})$ is constructed as

$$\Delta_k^f \triangleq \mathbf{E}[f(X_{1:n})|X_{1:k}] - \mathbf{E}[f(X_{1:n})|X_{1:k-1}] \tag{2}$$

where $\mathbf{E}[f(X_{1:n})|X_{1:0}] \triangleq \mathbf{E}[f(X_{1:n})]$. Remark: the idea behind this decomposition is to break the dependence in f into its smallest constituent parts.

(a) Show that $f(X_{1:n}) - \mathbf{E}f(X_{1:n}) = \sum_{k=1}^n \Delta_k^f$.

(b) Suppose that f is bounded. Show that the partial sums process $S_t \triangleq \sum_{k=1}^t \Delta_k^f$ is martingale with respect to the increasing family of sigma-fields $\sigma(X_1) \subset \sigma(X_{1:2}), \dots, \sigma(X_{1:n})$.[2]

(c) Suppose that the function f satisfies the so-called *bounded differences property.* Namely, for fixed $b > 0$ and any two vectors $x_{1:n}, x'_{1:n}$ differing in at most one position (i.e., $x_i \neq x'_i$ for at most one index i), we assume that $|f(x_{1:k}) - f(x'_{1:k})| \leq b$. Show that

$$\mathbf{P}\left(f(X_{1:n}) - \mathbf{E}f(X_{1:n}) \geq t\right) \leq \exp\left(-\frac{2t^2}{nb^2}\right) \tag{3}$$

Hint: what variance proxy can you establish for the Δ_k^f?

(d) Apply (3) to the function $f(x_{1:n}) \triangleq \frac{1}{n}\sum_{i=1}^n x_i$ evaluated at $x_{1:n} = W_{1:n}$ where the W_i are iid and bounded: $|W_i| \leq 1$—is the scaling of the bound (3) as expected?

3 Concentration Inequalities for Lipschitz Stable Dynamics

Let $W_{1:n+1}$ be iid real random variables and bounded: $|W_i| \leq b_W$ for some $b_W > 0$. Let $\phi : \mathbb{R} \to \mathbb{R}$ be ρ-Lipschitz for $\rho \in (0, 1)$ and bounded: for $x, y \in \mathbb{R}$, $|\phi(x) - \phi(y)| \leq \rho|x - y|$ and $|\phi(\cdot)| \leq b_\phi$. We consider a nonlinear dynamical system generating samples $X_{1:n+1}$ as:

$$X_{k+1} = \phi(X_k) + W_{k+1}, \qquad k = 1, \dots, n, \qquad X_1 = W_1. \tag{4}$$

(a) Show that $|X_i| \leq \frac{C}{1-\rho}$ independently of the index i for some constant C depending only on b_W and b_ϕ.

(b) Prove a concentration inequality for

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n X_i^2 - \mathbf{E}X_i^2. \tag{5}$$

In particular, show that this random variable has a sub-Gaussian tail with variance proxy independent of n. Hint: Use the second problem for an appropriate function f.

4 Linear Regression with Nonlinear Dynamics

Let $X_{1:n}$ be constructed as in the third exercise (4). Fix a further sequence $V_{1:n}$ that is iid (and independent of $W_{1:n}$) and b_V-bounded. Suppose that you observe $(X, Y)_{1:n}$ where

$$Y_i = \theta_\star X_i + V_i, i \in [n]. \tag{6}$$

(a) Show that $\sum_{i=1}^n X_i V_i$ is a martingale difference sequence

[2] For a random variable X, $\sigma(X)$ denotes the smallest sigma-field to which X is measurable.

(b) Show that $n^{-1/2}\sum_{i=1}^{n} X_i V_i$ is sub-Gaussian and compute an upper bound for its variance proxy of that does not depend on n.

(c) Show that $\widehat{\theta} = \frac{\sum_{i=1}^{n} X_i Y_i}{\sum_{i=1}^{n} X_i^2}$ converges to $\theta_\star$ at a rate proportional to $n^{-1/2}$.