Topics in Random Processes and Concentrations: EE 6112
Indian Institute of Technology Madras April 27, 2021
Instructor: Abhishek Sinha

Problem Set 4

- This problem set is due on **May 12, 2021** in the class.
- Each problem carries 10 points.
- No collaboration allowed. Each student must write his/her own solution.

---

1. **(Multiplicative System)** In the class, we have proved Hoeffding's inequality that gives an exponential bound on the deviation probability $\mathbb{P}(|X1 + ... + Xn) \geq t]$ for a sum of independent random variables that are bounded and have zero mean. In this problem, you will develop a generalization of Hoeffding's inequality to sums of dependent random variables that satisfy a certain weak orthogonality condition.

   (a) In preparation for the rest of the problem, derive the inequality

   $$\cosh(x) \leq \exp(x^2/2), \quad \forall x \in \mathbb{R}.$$

   (b) We say that a collection $X_1, ..., X_n$ of random variables is a multiplicative system if, for any $1 \leq k \leq n$ and any set of k indices $1 \leq i_1 < i_2 < \ldots < i_k \leq n$,

   $$\mathbb{E}[X_{i_1} X_{i_2} \ldots X_{i_k}] = 0.$$

   Prove that if $X_1, \ldots, X_n$ are a multiplicative system, then

   $$\mathbb{E}\Big[\prod_{i=1}^{n}(a_i X_i + b_i)\Big] = \prod_{i=1}^{n} b_i.$$

   for any choice of real constants $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$.

   (c) Let $U_1, \ldots, U_n$ be n possibly dependent random variables, and let Z be any real-valued random variable jointly distributed with them. For each i , let $X_i = \mathbb{E}[Z|U^i] - \mathbb{E}[Z|U^{i-1}]$ (where $\mathbb{E}[Z|U^0] \equiv \mathbb{E}Z$). Prove that $X_1, \ldots, X_n$ are a multiplicative system.

   (d) Consider a multiplicative system $X_1, \ldots, X_n$, such that $|X_i| \leq c_i$ for each i , where $c_i > 0$ are some finite constants. Prove that, for any $t > 0$:

   $$\mathbb{E}\Big[\exp(t\sum_{i=1}^{n} X_i)\Big] \leq \prod_{i=1}^{n} \cosh(tc_i).$$

   (e) Now for the final step: prove that if $X_1, \ldots, X_n$ are a multiplicative system of random variables satisfying the boundedness condition of part (c), then

   $$\mathbb{P}\Big(|\sum_{i=1}^{n} X_i| \geq t\Big) \leq 2\exp\big(-\frac{t^2}{2\sum_{i=1}^{n} c_i^2}\big).$$

2. **(Sudakov-Fernique inequality)** Recall the **Gaussian interpolation lemma** proved in the class: for two independent Gaussian random vectors $X \sim \mathcal{N}(0, \Sigma^X)$ and $Y \sim \mathcal{N}(0, \Sigma^Y)$, define the interpolated Gaussian vector

$$Z(u) := \sqrt{u}X + \sqrt{1-u}Y, \quad u \in [0,1].$$

Then for any twice-differentiable function $f : \mathbb{R}^n \to \mathbb{R}$, we have

$$\frac{d}{du}\mathbb{E}f(Z(u)) = \frac{1}{2}\sum_{i,j=1}^{n}(\Sigma_{ij}^X - \Sigma_{ij}^Y)\mathbb{E}\left(\frac{\partial^2 f}{\partial x_i \partial x_j}(Z(u))\right) \tag{1}$$

Use (1) to prove the Sudakov-Fernique inequality stated below:

(Sudakov-Fernique) Let $(X_t)_{t\in T}$ and $(Y_t)_{t\in T}$ be two mean-zero Gaussian processes. For simplicity, assume $|T| < \infty$. Assume that, for all $t, s \in T$, we have

$$\mathbb{E}(X_t - X_s)^2 \le \mathbb{E}(Y_t - Y_s)^2.$$

Then

$$\mathbb{E}\sup_{t\in T} X_t \le \mathbb{E}\sup_{t\in T} Y_t.$$

`Hint:` Take $f(x) = f_\beta(x) = \frac{1}{\beta}\ln\sum_{i=1}^n \exp(\beta x_i)$. Note that $f_\beta(x) \nearrow \max_i x_i$ as $\beta \to \infty$.

3. **(Exponentially many mutually almost orthogonal points)** From linear algebra, we know that any set of orthonormal vectors in $\mathbb{R}^n$ contains at most n vectors. However, if we allow the vectors to be almost orthogonal, there can be *exponentially many* of them! Prove this counterintuitive fact as follows. Fix $\epsilon \in (0,1)$. Show that there exists a set $\{x_1, x_2, \dots, x_N\}$ of unit vectors in $\mathbb{R}^n$ which are mutually almost orthogonal:

$$|\langle x_i, x_j\rangle| \le \epsilon, \forall i \ne j,$$

and the set is exponentially large in n:

$$N \ge \exp(c(\epsilon)n).$$

4. **(Non-parametric Least-Square Estimation)** Consider the function class $S_{\alpha,\gamma}(C_{\max}, L)$ which we introduced in the notes. Recall that,

$$\begin{aligned}S_{\alpha,\gamma}(C_{\max}, L) = \{f : [0,1] \to \mathbb{R} : |f^{(j)}|_\infty \le C_{\max}, \forall 0 \le j \le \alpha, \text{ and} \\ |f^\alpha(x) - f^\alpha(y)| \le L|x-y|^\gamma, \forall x, y \in [0,1].\}\end{aligned}$$

It can be shown that for some C (which depends on the parameters), the δ-covering number of $S_{\alpha,\gamma}(C_{\max}, L)$ in the sup-norm may be bounded as follows:

$$\log N(\delta, S_{\alpha,\gamma}(C_{\max}, L), ||\cdot||_\infty) \le C\left(\frac{1}{\delta}\right)^{1/(\alpha+\gamma)}.$$

Suppose we observe

$$Y_i = f^*(x_i) + \epsilon_i, \quad 1 \le i \le n,$$

where $f^* \in S_{\alpha,\gamma}(C_{\max}, L)$, and ϵ_i are i.i.d. standard Gaussians and the x_i's are deterministic points in $[0,1]$. Consider the non-parametric least-square estimator

$$\hat{f} \in \arg \min_{f \in S_{\alpha,\gamma}(C_{\max},L)} \frac{1}{n} \sum_{i=1}^{n} \big(Y_i - f(x_i)\big)^2.$$

Using the notion of Gaussian complexity for the function class $S_{\alpha,\gamma}(C_{\max}, L)$ and **Dudley's entropy integral**, prove an upper-bound for the mean-squared estimation error:

$$\mathsf{MSE} \equiv \mathbb{E}\left(\frac{1}{n} \sum_{i=1}^{n} \big(\hat{f}(x_i) - f^*(x_i)\big)^2\right).$$

5. **(Fundamental Limits of Sign Identification in Sparse Linear Regression)** In sparse linear regression, we have n observations $Y_i = \langle X_i, \theta^* \rangle + \epsilon_i$, where $X_i \in \mathbb{R}^d$ are known (fixed) matrices and the vector θ^* has a small number $k \ll d$ of non-zero entries, and $\epsilon_i \sim \mathsf{N}(0, \sigma^2)$. In this problem, we investigate the problem of *sign recovery*, that is, identifying the vector of signs $\mathsf{sign}(\theta^*_j), \forall j$, where $\mathsf{sign}(0) = 0$.
Assume we have the following process: fix a signal threshold $\theta_{\min} > 0$. First, a vector $S \in \{-1, 0, +1\}^d$ is chosen uniformly at random from the set of vectors $\mathcal{S}_k \equiv \{s \in \{-1, 0, +1\}^d : ||s||_1 = k\}, k \ge 2$. Then we define the vectors θ^s so that $\theta^s_j = \theta_{\min} s_j$, and conditional on $S = s$, we observe

$$Y = X\theta^s + \epsilon, \quad \epsilon \sim \mathsf{N}(0, \sigma^2 I_{n\times n}).$$

(Here $X \in \mathbb{R}^{n\times d}$ is a known fixed matrix.)

(a) Use Fano's inequality to show that for any estimator $\hat{S}$ of S, we have

$$\mathbb{P}(\hat{S} \ne S) \ge \frac{1}{2} \quad \text{unless} \quad n \ge \frac{\frac{d}{k} \ln \binom{d}{k}}{||n^{-1/2} X||^2_{\mathsf{Fr}}} \frac{\sigma^2}{\theta^2_{\min}}.$$

(b) Assume that $X \in \{-1, +1\}^{n\times d}$. Give a lower bound on how large n must be for sign recovery. Give a one line interpretation of the quantity $\frac{\theta^2_{\min}}{\sigma^2}$.