Topics in Probability Theory	**Fall 2023 (Georgia Institute of Technology)**

Lecture 2: Introduction to Inequalities

Lecturer: ***Víctor H. de la Peña*** *Scribes: Heyuan Yao*

In this lecture, we shall immerse ourselves in the realm of inequalities, embarking on a journey that traverses from rudimentary probabilistic inequalities to the intricacies of exponential and decoupling inequalities, culminating in the domain of martingale inequalities [1]. Furthermore, I will introduce an interesting limit theorem: the Law of the Iterated Logarithm, poised for an unveiling in the days to come when we start to learning self-normalization.

Theorem 2.1 (Markov's Inequality) *For any integrable random variable X (i.e. $\mathbb{E}X < \infty$), we have that, for any $a > 0$,*

$$\mathbb{P}(|X| \geq a) \leq \frac{\mathbb{E}(|X|)}{a}. \tag{2.1}$$

Theorem 2.2 (Hőlder's inequality/ Cauchy-Schwarz Inequality) *For any square-integrable random variable X, Y (i.e. $\mathbb{E}X^2, \mathbb{E}Y^2 < \infty$), we have that*

$$\mathbb{E}|XY| \leq \sqrt{\mathbb{E}X^2\mathbb{E}Y^2}. \tag{2.2}$$

And I leave the random vector's version as an exercise.

Theorem 2.3 (Jensen's Inequality) *We assume X an integrable real-valued random variable and Φ a convex function. Then*

$$\Phi(\mathbb{E}(X)) \leq \mathbb{E}(\Phi(X)). \tag{2.3}$$

[1] For those unacquainted with martingale theory, you can conveniently skip to Lecture Note 3 to get familiar with it, which is not measure-theory based. This concept will frequently resurface throughout our forthcoming sessions. Its wondrous properties find applications spanning mathematical finance, fluid mechanics, statistics, and, of course, machine learning.

Theorem 2.4 (Bernstein's Inequality) *If $X_1, X_2, \dots, X_n$ are independent random variables with $\mathbb{E}(X_i) = 0$ and $\sigma_i^2 := \mathbb{E}(X_i^2) < \infty$, and*

$$\mathbb{E}|X_i|^k = \|X_i\|_k^k \leq \frac{k!}{2}\sigma_i^2 M^{k-2} \tag{2.4}$$

holds for any $k \geq 2$ and for some constant $M \in (0, \infty)$. Then letting $\sigma^2 = \sum_{i=1}^n \sigma_i^2$ and $S = \sum_{i=1}^n X_i$, we have

$$\mathbb{E}exp(tS) \leq \mathbb{E}exp\left(\frac{\sigma^2 t^2}{2 - 2M|t|}\right) \tag{2.5}$$

for $M|t| < 1$, and

$$\mathbb{P}(S > u) \leq exp\left(-\frac{u^2}{2\sigma^2 + 2Mu}\right) \tag{2.6}$$

for all $u \geq 0$.

Theorem 2.5 (Chernoff's Exponential Inequality) *Assume that $X_1, X_2, \dots, X_n$ are independent random variables with $0 \leq X_i \leq 1$, $E[X_i] = \mu_i$ and $Var[X_i] = \sigma_i^2$. Let $\mu = \sum_{i=1}^n \mu_i$, $\sigma^2 = \sum_{i=1}^n \sigma_i^2$ and $X = \sum_{i=1}^n X_i$, then for any $0 < t < \frac{2\sigma^2}{\max_i max(\mu_i, 1-\mu_i)}$,*

$$\mathbb{P}(X - \mu > t) \leq exp(\frac{-t^2}{4\sigma^2}) \tag{2.7}$$

Theorem 2.6 (Hoeffing Inequality for Sampling Without Replacement) *If $Y_1, Y_2, \ldots, Y_n$ are independent nonnegative random variables, and $d_1, d_2, \ldots, d_n$ are possibly dependent random variables. Assume each Y_i has the same law as d_i. Then for every continuous convex function Φ,*

$$\mathbb{E}\left[\Phi(\sum_{i=1}^{n} d_i)\right] \leq \mathbb{E}\left[\Phi(\sum_{i=1}^{n} Y_i)\right]. \tag{2.8}$$

Theorem 2.7 (Kolmogorov's Maximal Inequality) *If $X_1, X_2, \ldots, X_n$ are independent random variables with $E[X_i] = 0$ and $Var(X_i) = \sigma_i^2$, then for any $t > 0$,*

$$P\left(\max_{1\leq k\leq n}\left|\sum_{i=1}^{k} X_i\right| \geq t\right) \leq \frac{1}{t^2}\sum_{i=1}^{n}\sigma_i^2. \tag{2.9}$$

Theorem 2.8 (Doob's Maximal Inequalities) *If $X_1, X_2, \ldots, X_n$ are martingale differences, then for any $t > 0$,*

$$\mathbb{P}\left(\max_{1\leq k\leq n}\left|\sum_{i=1}^{k} X_i\right| \geq t\right) \leq \frac{1}{t^2}E\left[\left(\sum_{i=1}^{n} X_i\right)^2\right]. \tag{2.10}$$

And for all $p > 1$,

$$\mathbb{E}\left(\max_{1\leq k\leq n}\left|\sum_{i=1}^{k} X_i\right|^p\right) \leq \left(\frac{p}{p-1}\right)^p \mathbb{E}\left|\sum_{i=1}^{k} X_i\right|^p \tag{2.11}$$

Theorem 2.9 (Burkholder-Davis-Gundy's Square Function Inequality) *If $X_1, X_2, \ldots, X_n$ are martingale differences, then for any $p \geq 1$,*

$$c_p E\left[\left(\sum_{i=1}^{n} X_i^2\right)^{\frac{p}{2}}\right] \leq E\left[\left(\max_{1\leq k\leq n}\left|\sum_{i=1}^{k} X_i\right|\right)^{p}\right] \leq C_p E\left[\left(\sum_{i=1}^{n} X_i^2\right)^{\frac{p}{2}}\right], \tag{2.12}$$

where c_p and C_p are positive constants depending on p.

Theorem 2.10 (Law of the Iterated Logarithm) *Let Y_n be independent, identically distributed random variables with means zero and variances σ^2. Let $S_n = Y_1 + ... + Y_n$. Then*

$$\limsup_{n\to\infty} \frac{|S_n|}{\sqrt{2n\sigma^2 loglogn}} = 1 \ a.s.. \tag{2.13}$$