AI Institute for Advances in Optimization **Georgia Institute of Technology**

Lecture 2: Probabilistic 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 11 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.

Additional resources are available through my colleague, Professor Ioannis Karatzas. He has designed a valuable series of lecture notes on martingale-related theory, which is comprehensive and accessible to new learners. The link is https://github.com/HeyuanYAO2704/Lecture-Notes-MATH-GU-4156-Advanced-Probability-Theory-Ioannis-Karatzas

Theorem 2.4 (Bernstein's inequality) *Let $\{x_i\}$ be a sequence of independent variables. Assume that $\mathbb{E}(x_i) = 0$ and $\mathbb{E}(x_i^2) = \sigma_i^2 < \infty$ and set $x_n^2 = \sum_{i=1}^n \sigma_i^2$. Furthermore, assume that there exists a constant $0 < c < \infty$ such that $\mathbb{E}(|x_i|^k) \le (k!/2)\sigma_i^2 c^{k-2}$ for all $k > 2$ (satisfied by subexponential random variables). Then for all $x > 0$.*

$$\mathbb{P}\left(\sum_{i=1}^n x_i > x\right) \le \exp\left(-\frac{x^2}{2(v_n^2 + cx)}\right).$$

Theorem 2.5 (Hoeffing Inequality for Sampling Without Replacement) *Let the population C consist of N values $c_1, ..., c_N$ (e.g. a deck of cards, N=52), and let $d_1, ..., d_n$ ($n \le N$) denote a random sample* ***without*** *replacement drawn from C, and $y_1, ..., y_n$ denote a random sample* ***with*** *replacement from C. The random variables $y_1, ..., y_n$ are i.i.d. and $d_i \overset{\mathscr{L}}{=} y_i$ for all i. And we have*

$$\mathbb{E}\left[\Phi(\sum_{i=1}^n d_i)\right] \le \mathbb{E}\left[\Phi(\sum_{i=1}^n y_i)\right]. \tag{2.4}$$

Theorem 2.6 (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 $t > 0$,*

$$P\left(\max_{1\le k\le n}\left|\sum_{i=1}^k X_i\right| \ge t\right) \le \frac{1}{t^2}\sum_{i=1}^n \sigma_i^2. \tag{2.5}$$

Theorem 2.7 (Doob's Maximal Inequalities) *If $X_1, X_2, \ldots, X_n$ are martingale differences (i.e., $M_n := \sum\limits_{i=1}^n X_n$ is a martingale with $M_0 = 0$), then for any $t > 0$,*

$$\mathbb{P}\left(\max_{1\le k\le n} |M_k| \ge t\right) \le \frac{1}{t^2}\mathbb{E}[(M_n)^2]. \tag{2.6}$$

And for all $p > 1$,

$$\mathbb{E}\left(\max_{1\le k\le n} |M_k|^p\right) \le \left(\frac{p}{p-1}\right)^p \mathbb{E}\,|M_n|^p \tag{2.7}$$

Theorem 2.8 (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*

$$\limsup_{n\to\infty} \frac{|S_n|}{\sqrt{2n\sigma^2 loglogn}} = 1 \; a.s.. \tag{2.8}$$

We call it a day with the following important inequalities and one definition, which we will give further discussion tomorrow.

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

$$c_p E\left[\left(\sum_{i=1}^n X_i^2\right)^{\frac{p}{2}}\right] \leq E\left[\left(\max_{1\leq k\leq n}\left|\sum_{i=1}^k X_i\right|\right)^p\right] \leq C_p E\left[\left(\sum_{i=1}^n X_i^2\right)^{\frac{p}{2}}\right], \tag{2.9}$$

where c_p and C_p are positive constants depending only on p.

And generally, For any $1 \leq p < \infty$ there exist positive constants c_p, C_p such that, for all local martingales M with $M_0 = 0$ and stopping times τ, the following inequality holds.

$$c_p \mathbb{E}\left[[M]_\tau^{p/2}\right] \leq \mathbb{E}\left[(M_\tau^*)^p\right] \leq C_p \mathbb{E}\left[[M]_\tau^{p/2}\right],$$

where $[M]$ is the quadratic variation process of M and $M_t^* \equiv \sup_{s\leq t}|M_s|$. Furthermore, for continuous local martingales, this statement holds for all $0 < p < \infty$.

Theorem 2.10 (Khintchine inequality) *Let $\epsilon_i, i = 1, ..., m$ be i.i.d. Rademacher variables with $\mathbb{P}(\epsilon_i = \pm 1) = 1/2$. Let $0 < p < \infty$ and let $x_1, ..., x_n \in \mathbb{C}$. Then*

$$A_p \left(\sum_{i=1}^n |x_i|^2\right)^{1/2} \leq \left(\mathbb{E}|\sum_{i=1}^n \epsilon_i x_i|^p\right)^{1/p} \leq B_p \left(\sum_{i=1}^n |x_i|^2\right)^{1/2} \tag{2.10}$$

for some constants $A_p, B_p > 0$ depending only on p.

Definition 2.11 (K-function) *Consider a nontrivial centered random variable Y (i.e., $\mathbb{E}Y = 0$). Then the K-function, $K_Y(x)$, is implicitly defined by the inverse of*

$$g(x) = \frac{x^2}{\int\limits_0^x \mathbb{E}|Y|I_{|Y|>u}du}. \tag{2.11}$$

Exercise: Please verify the equivalent definition of K-function, which is the unique solution of

$$K_Y(x)^2 = x\mathbb{E}[Y^2 \wedge (|Y|K_Y(x))] = x\mathbb{E}Y^2 I_{|Y|\leq K_Y(x)} + xK_Y(x)\mathbb{E}|Y|I_{|Y|>K_Y(x)}. \tag{2.12}$$

You can find the proof in Section 1.4.3 of the book [1].

Theorem 2.12 (K-function Bounds) *Consider $\{y_i\}_{i=1}^n$ a sequence of i.i.d. random centered variables such that $y_1 \sim Y$. Then we have*

$$0.67K_Y(n) \leq \mathbb{E}|\sum_{i=1}^n y_i| \leq 2K_Y(n). \tag{2.13}$$

References

[1] Victor de la Peña and Evarist Giné. *Decoupling: from dependence to independence.* Springer Science & Business Media, 2012.