Lecture-27: Martingale Concentration Inequalities

1 Introduction

Consider a probability space $(\Omega, \mathcal{F}, P)$ and a discrete filtration $\mathcal{F}_\bullet = (\mathcal{F}_n \subseteq \mathcal{F} : n \in \mathbb{N})$. Let $X : \Omega \to \mathbb{R}^{\mathbb{N}}$ be discrete random process and stopping time $\tau : \Omega \to \mathbb{N}$, both adapted to the filtration $\mathcal{F}_\bullet$.

Remark 1. Recall that for a submartingale X and a stopping time τ bounded above by n, both adapted to the same filtration $\mathcal{F}_\bullet$, we have $\mathbb{E}X_1 \leqslant \mathbb{E}X_\tau \leqslant \mathbb{E}X_n$.

Theorem 1.1 (Kolmogorov's inequality for submartingales). *For a non-negative submartingale $X : \Omega \to \mathbb{R}_+^{\mathbb{N}}$ and $a > 0$,*

$$P\left\{\max_{i \in [n]} X_i > a\right\} \leqslant \frac{\mathbb{E}[X_n]}{a}.$$

Proof. We define a random time $\tau_a \triangleq \inf\{i \in \mathbb{N} : X_i > a\}$ and stopping time $\tau \triangleq \tau_a \wedge n$. It follows that,

$$\left\{\max_{i \in [n]} X_i > a\right\} = \cup_{i \in [n]} \{X_i > a\} = \{X_\tau > a\}.$$

Using this fact and Markov inequality, we get $P\left\{\max_{i \in [n]} X_i > a\right\} = P\{X_\tau > a\} \leqslant \frac{\mathbb{E}[X_\tau]}{a}$. Since $\tau \leqslant n$ is a bounded stopping time, result follows from the Remark 1. □

Corollary 1.2. *For a martingale X and positive constant a,*

$$P\left\{\max_{i \in [n]} |X_i| > a\right\} \leqslant \frac{\mathbb{E}|X_n|}{a}, \qquad P\left\{\max_{i \in [n]} |X_i| > a\right\} \leqslant \frac{\mathbb{E}X_n^2}{a^2}.$$

Proof. The proof the above statements follow from and Kolmogorov's inequality for submartingales, and by considering the convex functions $f(x) = |x|$ and $f(x) = x^2$. □

Theorem 1.3 (Strong Law of Large Numbers). *Let $S : \Omega \to \mathbb{R}^{\mathbb{N}}$ be a random walk with* i.i.d. *step size X having finite mean μ. If the moment generating function $t \mapsto M(t) \triangleq \mathbb{E}[e^{tX_1}]$ exists for all $t \in \mathbb{R}_+$, then*

$$P\left\{\lim_{n \in \mathbb{N}} \frac{S_n}{n} = \mu\right\} = 1.$$

Proof. For a given $\epsilon > 0$, we define the following map $t \mapsto g(t) \triangleq \frac{e^{t(\mu+\epsilon)}}{M(t)}$. for all $t \in \mathbb{R}_+$. Then, it is clear that $g(0) = 1$ and from the fact that $M(0) = 1$ and $M'(0) = \mu = \mathbb{E}X_1$, we obtain

$$g'(0) = \frac{M(0)(\mu+\epsilon) - M'(0)}{M^2(0)} = \epsilon > 0.$$

Hence, there exists a value $t_0 > 0$ such that $g(t_0) > 1$. We now show that $\frac{S_n}{n}$ can be as large as $\mu + \epsilon$ only finitely often. To this end, note that

$$\left\{\frac{S_n}{n} \geqslant \mu + \epsilon\right\} \subseteq \left\{\frac{e^{t_0 S_n}}{M(t_0)^n} \geqslant g(t_0)^n\right\} \tag{1}$$

However, $Y_n \triangleq \frac{e^{t_0 S_n}}{M^n(t_0)} = \prod_{i=1}^n \frac{e^{t_0 X_i}}{M(t_0)}$ is a product of independent non negative random variables with unit mean, and hence is a non-negative martingale with $\sup_n \mathbb{E}Y_n = 1$. By martingale convergence theorem, the limit $\lim_{n \in \mathbb{N}} Y_n$ exists and is finite.

Since $g(t_0) > 1$, it follows from (1) that

$$P\left\{\frac{S_n}{n} \geqslant \mu + \epsilon \text{ for an infinite number of n}\right\} = 0.$$

Similarly, defining the function $f(t) \triangleq \frac{e^{t(\mu-\epsilon)}}{M(t)}$ and noting that since $f(0) = 1$ and $f'(0) = -\epsilon$, there exists a value $t_0 < 0$ such that $f(t_0) > 1$, we can prove in the same manner that

$$P\left\{\frac{S_n}{n} \leqslant \mu - \epsilon \text{ for an infinite number of n}\right\} = 0.$$

Hence, result follows from combining both these results, and taking limit of arbitrary ϵ decreasing to zero. □

Definition 1.4. A discrete random process $X : \Omega \to \mathbb{R}^{\mathbb{N}}$ with distribution function $F_n \triangleq F_{X_n}$ for each $n \in \mathbb{N}$, is said to be **uniformly integrable** if for every $\epsilon > 0$, there is a y_ϵ such that for each $n \in \mathbb{N}$

$$\mathbb{E}[|X_n| \mathbb{1}_{\{|X_n|>y_\epsilon\}}] = \int_{|x|>y_\epsilon} |x| dF_n(x) < \epsilon.$$

Lemma 1.5. *If $X : \Omega \to \mathbb{R}^{\mathbb{N}}$ is uniformly integrable then there exists finite M such that $\mathbb{E}|X_n| < M$ for all $n \in \mathbb{N}$.*

Proof. Let y_1 be as in the definition of uniform integrability. Then

$$\mathbb{E}|X_n| = \int_{|x|\leqslant y_1} |x| dF_n(x) + \int_{|x|>y_1} |x| dF_n(x) \leqslant y_1 + 1.$$

□

1.1 Generalized Azuma Inequality

Lemma 1.6. *For a zero mean random variable X with support $[-\alpha, \beta]$ and any convex function f*

$$\mathbb{E}f(X) \leqslant \frac{\beta}{\alpha+\beta} f(-\alpha) + \frac{\alpha}{\alpha+\beta} f(\beta).$$

Proof. From convexity of f, any point (X, Y) on the line joining points $(-\alpha, f(-\alpha))$ and $(\beta, f(\beta))$ is

$$Y = f(-\alpha) + (X+\alpha)\frac{f(\beta) - f(-\alpha)}{\beta + \alpha} \geqslant f(X).$$

Result follows from taking expectations on both sides. □

Lemma 1.7. *For $\theta \in [0,1]$ and $\bar{\theta} \triangleq 1 - \theta$, we have $\theta e^{\bar{\theta}x} + \bar{\theta}e^{-\theta x} \leqslant e^{x^2/8}$ for all $x \in \mathbb{R}$.*

Proof. Defining $\alpha \triangleq 2\theta - 1, \beta \triangleq \frac{x}{2}$, and $f(\alpha, \beta) \triangleq \cosh\beta + \alpha \sinh\beta - e^{\alpha\beta+\beta^2/2}$, we can write

$$\theta e^{\bar{\theta}x} + \bar{\theta}e^{-\theta x} - e^{x^2/8} = \frac{(1+\alpha)}{2} e^{(1-\alpha)\beta} + \frac{(1-\alpha)}{2} e^{-(1+\alpha)\beta} - e^{\beta^2/2} = e^{-\alpha\beta} f(\alpha, \beta).$$

Therefore, we need to show that $f(\alpha, \beta) \leqslant 0$ for all $\alpha \in [-1, 1]$ and $\beta \in \mathbb{R}$. This inequality is true for $|\alpha| = 1$ and sufficiently large β. Therefore, it suffices to show this for $\beta < M$ for some M. We take the partial derivative of $f(\alpha, \beta)$ with respect to variables α, β and equate it to zero to get the stationary point,

$$\sinh\beta + \alpha\cosh\beta = (\alpha+\beta)e^{\alpha\beta+\beta^2/2}, \qquad \sinh\beta = \beta e^{\alpha\beta+\beta^2/2}.$$

If $\beta \neq 0$, then the stationary point satisfies $1 + \alpha \coth\beta = 1 + \frac{\alpha}{\beta}$, with the only solution being $\beta = \tanh\beta$. By Taylor series expansion, it can be seen that there is no other solution to this equation other than $\beta = 0$. Since $f(\alpha, 0) = 0$, the lemma holds true. □

Proposition 1.8. *Let X be a zero-mean martingale with respect to filtration $\mathcal{F}_\bullet$, such that $-\alpha \leqslant X_n - X_{n-1} \leqslant \beta$ for each $n \in \mathbb{N}$. Then, for any positive values a and b*

$$P\{X_n \geqslant a + bn \text{ for some } n\} \leqslant \exp\left(-\frac{8ab}{(\alpha+\beta)^2}\right). \tag{2}$$

Proof. Let $X_0 = 0$ and $c > 0$, then we define a random sequence $W : \Omega \to \mathbb{R}^{\mathbb{N}}$ adapted to filtration $\mathcal{F}_\bullet$, such that

$$W_n \triangleq e^{c(X_n - a - bn)} = W_{n-1}e^{-cb}e^{c(X_n - X_{n-1})}, \quad n \in \mathbb{Z}_+.$$

We will show that W is a supermartingale with respect to the filtration $\mathcal{F}_\bullet$. It is easy to see that $\sigma(W_n) \subseteq \mathcal{F}_n$ for each $n \in \mathbb{N}$. We can also see that $\mathbb{E}|W_n| < \infty$ for all n. Further, we observe

$$\mathbb{E}[W_n|\mathcal{F}_{n-1}] = W_{n-1}e^{-cb}\mathbb{E}[e^{c(X_n - X_{n-1})}|\mathcal{F}_{n-1}].$$

Applying Lemma 1.6 to the convex function $f(x) = e^{cx}$, replacing expectation with conditional expectation, the fact that $\mathbb{E}[X_n - X_{n-1}|\mathcal{F}_{n-1}] = 0$, and setting $\theta = \frac{\alpha}{(\alpha+\beta)} \in [0,1]$, we obtain that

$$\mathbb{E}[e^{c(X_n - X_{n-1})}|\mathcal{F}_{n-1}] \leqslant \frac{\beta e^{-c\alpha} + \alpha e^{c\beta}}{\alpha + \beta} = \bar{\theta}e^{-c(\alpha+\beta)\theta} + \theta e^{c(\alpha+\beta)\bar{\theta}} \leqslant e^{c^2(\alpha+\beta)^2/8}.$$

The second inequality follows from Lemma 1.7 with $x = c(\alpha + \beta)$ and $\theta = \frac{\alpha}{(\alpha+\beta)} \in [0,1]$. Fixing the value $c = \frac{8b}{(\alpha+\beta)^2}$, we obtain

$$\mathbb{E}[W_n|\mathcal{F}_{n-1}] \leqslant W_{n-1}e^{-cb + \frac{c^2(\alpha+\beta)^2}{8}} = W_{n-1}.$$

Thus, W is a supermartingale. For a fixed positive integer k, define the bounded stopping time τ by

$$\tau \triangleq \inf\{n \in \mathbb{N} : X_n \geqslant a + bn\} \wedge k.$$

Now, using Markov inequality and optional stopping theorem, we get

$$P\{X_\tau \geqslant a + b\tau\} = P\{W_\tau \geqslant 1\} \leqslant \mathbb{E}[W_\tau] \leqslant \mathbb{E}[W_0] = e^{-ca} = e^{-\frac{8ab}{(\alpha+\beta)^2}}.$$

The above inequality is equivalent to $P\{X_n \geqslant a + bn \text{ for some } n \leqslant k\} \leqslant e^{-\frac{8ab}{(\alpha+\beta)^2}}$. Since, the choice of k was arbitrary, the result follow from letting $k \to \infty$. ☐

Theorem 1.9 (Generalized Azuma inequality). *Let X be a zero-mean martingale, such that $-\alpha \leqslant X_n - X_{n-1} \leqslant \beta$ for all $n \in \mathbb{N}$. Then, for any positive constant c and integer m*

$$P\{X_n \geqslant nc \text{ for some } n \geqslant m\} \leqslant e^{-\frac{2mc^2}{(\alpha+\beta)^2}}, \qquad P\{X_n \leqslant -nc \text{ for some } n \geqslant m\} \leqslant e^{-\frac{2mc^2}{(\alpha+\beta)^2}}.$$

Proof. Observe that if there is an n such that $n \geqslant m$ and $X_n \geqslant nc$ then for that n, we have $X_n \geqslant nc \geqslant \frac{mc}{2} + \frac{nc}{2}$. Using this fact and previous proposition for $a = \frac{mc}{2}$ and $b = \frac{c}{2}$, we get

$$P\{X_n \geqslant nc \text{ for some } n \geqslant m\} \leqslant P\left\{X_n \geqslant \frac{mc}{2} + \frac{c}{2}n \text{ for some } n\right\} \leqslant e^{-\frac{8\frac{mc}{2}\frac{c}{2}}{(\alpha+\beta)^2}}.$$

This proves first inequality, and second inequality follows by considering the martingale $-X$. ☐