Lecture 23: Martingale Concentration Inequalities

1 Introduction

**Lemma 1.1.** If \( \{X_n : n \in \mathbb{N}\} \) is a submartingale and \( N \) is a stopping time such that \( \Pr\{N \leq n\} = 1 \) then

\[
\mathbb{E}X_1 \leq \mathbb{E}X_N \leq \mathbb{E}X_n.
\]

**Proof.** It follows from optional stopping theorem that since \( N \) is bounded, \( \mathbb{E}[X_N] \geq \mathbb{E}[X_1] \). Now, since \( N \) is a stopping time, we see that for \( \{N = k\} \)

\[
\mathbb{E}[X_n|X_1,\ldots,X_n,N = k] = \mathbb{E}[X_n|X_1,\ldots,X_k,N = k] = \mathbb{E}[X_n|X_1,\ldots,X_k] \geq X_k = X_N.
\]

Result follows by taking expectation on both sides. \( \square \)

**Theorem 1.2 (Kolmogorov’s inequality for submartingales).** If \( \{X_n : n \in \mathbb{N}\} \) is a submartingale, then

\[
\Pr\{\max\{X_1,X_2,\ldots,X_n\} > a\} \leq \frac{\mathbb{E}[X_n]}{a}, \text{ for } a > 0.
\]

**Proof.** We define a stopping time

\[
N = \min\{i \in [n] : X_i > a\} \land n \leq n.
\]

It follows that, \( \{\max\{X_1,\ldots,X_n\} > a\} = \{X_N > a\} \). Using this fact and Markov inequality, we get

\[
\Pr\{\max\{X_1,\ldots,X_n\} > a\} = \Pr\{X_N > a\} \leq \frac{\mathbb{E}[X_N]}{a}.
\]

Since \( N \leq n \) is a bounded stopping time, result follows from the previous Lemma \( \square \)

**Corollary 1.3.** Let \( \{X_n : n \in \mathbb{N}\} \) be a martingale. Then, for \( a > 0 \) the following hold.

\[
\Pr\{\max\{|X_1|,\ldots,|X_n|\} > a\} \leq \frac{\mathbb{E}[|X_n|]}{a},
\]

\[
\Pr\{\max\{|X_1|,\ldots,|X_n|\} > a\} \leq \frac{\mathbb{E}[X_n^2]}{a^2}.
\]

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

**Theorem 1.4 (Strong Law of Large Numbers).** Let \( S_n \) be a random walk with iid step size \( \{X_i : i \in \mathbb{N}\} \) with finite mean \( \mu \). Then

\[
\Pr\left\{\lim_{n \in \mathbb{N}} \frac{S_n}{n} = \mu\right\} = 1.
\]
Proof. We will prove the theorem under the assumption that the moment generating function
\(\psi(t) = E[e^{tX}]\) for random variable \(X\) exists. For a given \(\epsilon > 0\), we define

\[ g(t) \triangleq e^{t(\mu + \epsilon)}/\psi(t). \]

Then, it is clear that \(g(0) = 1\) and

\[ g'(0) = \frac{\psi(0)(\mu + \epsilon) - \psi'(0)}{\psi^2(0)} = \epsilon > 0. \]

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

\[ \left\{ \frac{S_n}{n} \geq \mu + \epsilon \right\} \subseteq \left\{ \frac{e^{t_0 S_n}}{\psi(t_0)^n} \geq g(t_0)^n \right\} \tag{1} \]

However, \(\frac{e^{t_0 S_n}}{\psi(t_0)^n}\) is a product of independent non negative random variables with unit mean, and
hence is a martingale. By martingale convergence theorem, we have

\[ \lim_{n \in \mathbb{N}} \frac{e^{t_0 S_n}}{\psi(t_0)^n} \text{ exists and is finite.} \]

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

\[ \Pr \left\{ \frac{S_n}{n} \geq \mu + \epsilon \text{ for an infinite number of } n \right\} = 0. \]

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

\[ \Pr \left\{ \frac{S_n}{n} \leq \mu - \epsilon \text{ for an infinite number of } n \right\} = 0. \]

Hence, result follows from combining both these results, and taking limit of arbitrary \(\epsilon\) decreasing
to zero.

Definition 1.5. A sequence of random variables \(\{X_n : n \in \mathbb{N}\}\) with distribution functions
\(\{F_n : n \in \mathbb{N}\}\), is said to be uniformly integrable if for every \(\epsilon > 0\), there is a \(y_\epsilon\) such that

\[ \int_{|x| > y_\epsilon} |x| dF_n(x) < \epsilon \ \forall n \in \mathbb{N}. \]

Lemma 1.6. If \(\{X_n : n \in \mathbb{N}\}\) is uniformly integrable then there exists finite \(M\) such that
\(E[|X_n|] < M\) for all \(n \in \mathbb{N}\).

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

\[ E[|X_n|] = \int_{|x| \leq y_1} |x| dF_n(x) + \int_{|x| > y_1} |x| dF_n(x) \leq y_1 + 1. \]
1.1 Generalized Azuma Inequality

Proposition 1.7. Let \( \{X_n : n \in \mathbb{N}\} \) be a martingale with mean \( X_0 = 0 \), such that
\[-\alpha \leq X_n - X_{n-1} \leq \beta \ \forall \ n \in \mathbb{N}.\]

Then, for any positive values \( a \) and \( b \)
\[
Pr\{X_n \geq a + bn \text{ for some } n\} \leq \exp\left( -\frac{8ab}{(\alpha + \beta)^2} \right).
\]

Proof. For \( n \geq 0 \), we define
\[ W_n = \exp\{c(X_n - a - bn)\} = W_{n-1}e^{-cb}\exp\{c(X_n - X_{n-1})\} \]
Since exponential is an invertible function, \( \sigma(W_n, i \in [n]) = \sigma(X_i, i \in [n]) \), and hence
\[
\mathbb{E}[W_n|W_1 \ldots W_{n-1}] = W_{n-1}e^{-cb}\mathbb{E}[\exp\{c(X_n - X_{n-1})\}|X_1 \ldots X_{n-1}] = \mathbb{E}[W_{n-1}e^{-cb}]\exp\{c(X_n - X_{n-1})\}\]
Using Jensen’s inequality for convex function \( f(x) = e^x \), we obtain
\[
\mathbb{E}[\exp\{c(X_n - X_{n-1})\}|X_1 \ldots X_{n-1}] \leq W_{n-1}e^{-cb}\exp\left( \frac{\beta e^{-c\alpha} + \alpha e^\beta}{\alpha + \beta} \right) \leq W_{n-1}e^{-cb}\exp\left( \frac{\alpha + \beta}{2}\right)^2.
\]
where second inequality follows from with \( \theta = \alpha/(\alpha + \beta) \), \( x = c(\alpha + \beta) \). Hence, fixing the value of \( c \) as \( c = 8b/(\alpha + \beta)^2 \) yields
\[
\mathbb{E}[W_n|W_1 \ldots W_{n-1}] \leq W_{n-1},
\]
and so \( \{W_n : n \in \mathbb{N}_0\} \) is a supermartingale. For a fixed positive integer \( k \), define the bounded stopping time \( N \) by
\[
N = \min\{n : \text{ either } X_n \geq a + bn \text{ or } n = k\}.
\]
Now, using Markov inequality and optional stopping theorem, we get
\[
Pr\{X_N \geq a + bN\} = Pr\{W_N \geq 1\} \leq \mathbb{E}[W_N] \leq \mathbb{E}[W_0].
\]
But the above inequality is equivalent to
\[
Pr\{X_n \geq a + bn \text{ for some } n \leq k\} \leq e^{-8ab/(\alpha + \beta)^2}.
\]
Letting \( k \to \infty \) gives the result.

\[ \square \]

Theorem 1.8 (Generalized Azuma Inequality). Let \( \{X_n : n \in \mathbb{N}_0\} \) be a martingale with mean \( X_0 = 0 \), such that \(-\alpha \leq X_n - X_{n-1} \leq \beta \) for all \( n \in \mathbb{N} \). Then, for any positive constant \( c \) and integer \( m \):
\[
Pr\{X_n \geq nc \text{ for some } n \geq m\} \leq \exp\left( -\frac{2mc^2}{(\alpha + \beta)^2} \right),
\]
\[
Pr\{X_n \leq -nc \text{ for some } n \geq m\} \leq \exp\left( -\frac{2mc^2}{(\alpha + \beta)^2} \right).
\]

Proof. Observe that if there is an \( n \) such that \( n \geq m \) and \( X_n \geq nc \) then for that \( n, X_n \geq nc \geq mc/2 + nc/2 \). Using this fact and previous proposition for \( a = mc/2 \) and \( b = c/2 \), we get
\[
Pr\{X_n \geq nc \text{ for some } n \geq m\} \leq Pr\{X_n \geq mc/2 + (c/2)n \text{ for some } n\} \leq \exp\left\{ -\frac{8(mc^2)(c/2)}{(\alpha + \beta)^2} \right\}.
\]
This proves first inequality, and second inequality follows by considering the martingale \( \{-X_n : n \in \mathbb{N}_0\} \).

\[ \square \]