Fundamentals of Probability, Hyunseung Kang

Definitions

Let set $\Omega$ be a set that contains all possible outcomes of an experiment. A subset of $\Omega$ is an event while a single point of $\Omega$ is considered a sample point. Next, we define a field, which is a collection of events such that it is closed under complements, and it is closed under countable addition.

**Definition**: A Sigma Field of a set $\Omega$ is a collection of sets $F \in 2^\Omega$ where (i) $\emptyset \in F$, (ii) if $A \in F$ then $A^c \in F$, and (iii) $A_i \in F$ for $i = 1, 2, \ldots \Rightarrow \bigcup_{i=1}^\infty A_i \in F$.

$F$ does not have to be unique. For example, for any set $\Omega$, $2^\Omega$ and $\{\emptyset, \Omega\}$ are both sigma fields. Also, a sigma field generated by a collection $A$, denoted as $\sigma(A)$, is the intersection of all sigma fields that contain $A$. For example, let $A$ be a collection of open sets in $\Omega = (0,1]$. Each element of $\sigma(A_{(0,1)}) = B(0,1)$ is a Borel set and the field is aptly named the Borel sigma field; it is also known as the field that contains all the open sets in $(0,1]$. Another example, consider $\sigma(X_1)$ where $X_1$ is a random variable (to be defined later). $\sigma(X_1)$ is the smallest $F$ s.t. $\forall (a,b), X_1^{-1}(a,b) \in F$; or intuitively speaking, we call it the information of $X_1$.

Finally, we define a measure which assigns probability to each element in the field we constructed.

**Definition**: A function $P : F \to [0,1]$ is a probability measure iff (i) $P(\Omega) = 1$ and (ii) $A_i \in F$ and are disjoint $\Rightarrow P\left(\bigcup_{i=1}^\infty A_i\right) = \sum_{i=1}^\infty P(A_i)$

There are numerous properties we can derive using these definitions. We highlight the important ones.

**Property**: If $A_1 \supset A_2 \supset \cdots \text{ and } \bigcap_{i=1}^\infty A_i = A$, then $P(A_n) \downarrow P(A)$

**Property**: If $A_1 \subset A_2 \subset \cdots \text{ and } \bigcup_{i=1}^\infty A_i = A$, then $P(A_n) \uparrow P(A)$

Next, we define random variables (and naturally their expectations) by first introducing simple random variables.

**Definition**: A function $X : \Omega \to R$ is a simple random variable if $X(\omega) = \sum_{i=1}^n a_i 1_{A_i}(\omega)$ where $A_i$ are $n$ disjoint partitions of $\Omega$. A random variable is $X : \Omega \to R$ on $(\Omega, F, \mathbb{P})$ where $\forall B \in B(\Omega), X^{-1}(B) \in F$.

**Definition**: The expected value of a simple random variable, $E(X) = \sum_{i=1}^n a_i P(A_i)$. The expected value of a random variable is $E(X) = \int_{\Omega} X(\omega)d\mathbb{P}(\omega) = \sup\{\int Yd\mathbb{P} : Y \leq X, Y \text{ bounded r.v.}\}$. We approximate this integral as a limit of the expectation of bounded r.v.s, which is a limit of simple random variables.

There are couple properties of random variables that come up. The most important one is uniform boundedness of random variables and independence.

**Definition**: A sequence of random variables $X_n$ is uniformly bounded when, for some $M$, $\|X_n\| \leq M \ \forall n$

**Definition**: A sequence of random variables $X_n$ is uniformly integrable if $\lim_{M \to \infty} \left(\sup_{i \in I} E(|X_i| \mathbb{1}_{|X_i| > M})\right) = 0$

**Definition**: $X$ and $Y$ are independent iff $\forall A, B \in B(\Omega), P(\{X \in A\} \cap \{Y \in B\}) = P(X \in A)P(Y \in B)$

Properties
An important concept to remember is whether a sequence of events occurs infinitely often. There are also famous theorems associated with these called the Borel Cantelli Lemmas.

**Definition**: A sequence of events, \( A_i \), occurs infinitely often, if \( \{ A_i \} = \{ \omega \in \Omega | \sum_{i=1}^{\infty} I_{A_i} = \infty \} = \cap_{i=1}^{\infty} \cup_{k=n}^{\infty} A_k \)

**Theorem (First Borel Cantelli Lemma)**: \( \sum_{i=1}^{\infty} P(A_i) < \infty \Rightarrow P(A_i \ i.o.) = 0 \)

*Proof*: Define \( B = \{ \omega \mid A_i \ i.o. \} = \{ \omega \mid \sum_{i=1}^{\infty} I_{A_i}(\omega) = \infty \} \). Then, \( \forall m, \omega \in B, mI_B(\omega) \leq \sum_{i=1}^{\infty} I_{A_i}(\omega) \). Therefore, \( E(mI_B(\omega)) \leq \sum_{i=1}^{\infty} E(I_A(\omega)) \Rightarrow P(B) \leq \sum_{i=1}^{\infty} P(A_i) \). Therefore, \( P(B) = 0 \).

**Theorem (Second Borel Cantelli Lemma)**: \( \sum_{i=1}^{\infty} P(A_i) = \infty \) and \( A_i \) indep. \( \Rightarrow P(A_i \ i.o.) = 1 \)

*Proof*: \( A_i \ i.o. \} = \{ \cup_{i=1}^{\infty} \cap_{k=n}^{\infty} A_k^c \}. Thus, we only have to show \( \forall n, P(\cap_{k=n}^{\infty} A_k^c) = 0 \). Using \( 1 - x \leq e^{-x} \), \( P(\cap_{k=n}^{\infty} A_k^c) \leq \prod_{k=n}^{\infty} (1 - P(A_k)) \leq \exp(-\sum_{k=n}^{\infty} A_k) = 0 \).

In addition to i.o., we can establish inequalities of random variables.

**Theorem (Markov Inequality)**: If \( X \geq 0 \), then \( \forall \epsilon, P(X \geq \epsilon) \leq \frac{1}{\epsilon} E(X) \). When \( X \) is not non.neg, we can use \( |X| \)

*Proof*: Note that \( \forall \epsilon, \epsilon I_{X \geq \epsilon} \leq X \). Take expectation on both sides to complete the proof.

**Theorem (Chebyshev’s Inequality)**: For any random variable \( X \) and \( \epsilon > 0 \), \( P(|X - E(X)| \geq \epsilon) \leq \frac{1}{\epsilon^2} Var(X) \)

*Proof*: Note that \( \forall \epsilon, \epsilon |I_{|X - E(X)| \geq \epsilon} \leq |X - E(X)| \). Square both sides and take expectation to complete the proof.

**Theorem (Cauchy-Schwartz Inequality)**: \( E(|XY|) \leq \left( E(X^2) \right)^{1/2} \left( E(Y^2) \right)^{1/2} \)

*Proof*: We know \( \forall X, Y, (X - Y)^2 \geq 0 \Rightarrow XY \leq \frac{X^2 + Y^2}{2} \). Let \( X = \frac{|X|}{\sqrt{E(X^2)}} \) and \( Y = \frac{|Y|}{\sqrt{E(Y^2)}} \). Then, \( \frac{|X||Y|}{\sqrt{E(X^2)E(Y^2)}} \leq \frac{1}{2} \left( \frac{X^2}{E(X^2)} + \frac{Y^2}{E(Y^2)} \right) \). Take expectation on both sides, multiply the square root factor to get the desired inequality.

**Theorem (Holder’s Inequality)**: If \( \frac{1}{p} + \frac{1}{q} = 1, p > 1, q > 1 \), then \( E|XY| \leq \left( \frac{1}{p} E|X|^p \right)^{1/p} \left( \frac{1}{q} E|Y|^q \right)^{1/q} \)

*Proof*: From Young’s Inequality, \( XY \leq \frac{1}{p} X^p + \frac{1}{q} Y^q \). Let \( X = \frac{|X|}{(E|X|^p)^{1/p}} \) and \( Y = \frac{|Y|}{(E|Y|^q)^{1/q}} \), do the same trick as above.

**Theorem (Kolmogorov Maximal Inequality)**: If \( X_i \) are indep. and \( E(X_i) = 0 \), then \( P(\max_{1 \leq k \leq n} |\sum_{i=1}^{k} X_i| \geq \epsilon) \leq \frac{1}{\epsilon^2} Var(\sum_{i=1}^{k} X_i) \)

*Proof*: Let \( S_k = \sum_{i=1}^{k} X_i \). Define \( T = \min\{k : |S_k| \geq \epsilon\} \). Then, \( \sum_{i=1}^{\infty} I_{T=i} \leq \frac{\sum_{i=1}^{\infty} I_{T=i} |S_k^2|}{\epsilon^2} \). Taking expectation on both sides, \( P(\max_{1 \leq k \leq n} |\sum_{i=1}^{k} X_i| \geq \epsilon) \leq \frac{1}{\epsilon^2} \sum_{i=1}^{\infty} E(I_{T=i} |S_k^2|) = \frac{1}{\epsilon^2} \sum_{i=1}^{\infty} E(I_{T=i}^2 (S_k^2 + 2S_i (S_n - S_i))) \). Note that \( E(2S_i I_{T=i}) (S_n - S_i) = 0 \) because the stopping time \( T \) is independent from future observations.

Rearranging the terms, we get \( \frac{1}{\epsilon^2} \sum_{i=1}^{\infty} E(I_{T=i} (S_i^2 + 2S_i (S_n - S_i))) = \frac{1}{\epsilon^2} \sum_{i=1}^{\infty} E(I_{T=i} (S_n^2 - (S_n - S_i)^2)) \leq \frac{1}{\epsilon^2} E(S_n^2) \sum_{i=1}^{\infty} I_{T=i} \leq \frac{1}{\epsilon^2} E(S_n^2) = \frac{1}{\epsilon^2} Var(S_n)

*Theorem (Kolmogorov Zero-One Law)**: If \( X_i \) indep., and \( T = \cap_{i=1}^{\infty} \sigma(X_i, X_{i+1}, ...) \), then \( \forall A \in T, P(A) = 0 \) or 1
Note: $T$ stands for tail events. Intuitively speaking, these are events such that it is determined by all the values of $X_i$, but is not determined by any finite subset of $X_i$ (or probabilistically independent of any finite subset). An example of a tail event is $\{ \omega : \sum_{i=1}^{\infty} X_i(x) \to 0 \}$, $\{ \omega : \sum_{i=1}^{\infty} X_i \text{ converges} \}$, or $\{ \omega : \prod_{i=1}^{\infty} X_i \}$. Generally, it is hard to determine whether it’s 0 or 1.

Proof: Since $P(A) = P(A \cap A) = P(A)^2 \Rightarrow P(A) = 0$ or 1, we prove $P(A) = P(A)^2$. First, let $F_1 = \sigma(X_1), F_2 = \sigma(X_1, X_2), \ldots$. Then, if $A \in \sigma(\bigcup_{i=1}^{\infty} \sigma(X_1, \ldots, X_i))$, by the approximation of sets, $\exists A(\epsilon) \in \sigma(X_1, \ldots, X_k_0)$ s.t.

$P(A \Delta A(\epsilon)) \leq \epsilon$. Now, restrict $A \subset \sigma(X_{k_0+1}, X_{k_0+2}, \ldots)$. Since the sigma fields do not overlap, $A$ and $A(\epsilon)$ are independent. Now, $I_A = I_{A(A(\epsilon))} = I_A(I_{A(\epsilon)} + I_{A \Delta A(\epsilon)})$. Taking expectation on both sides and using independence, we get $P(A) = P(A)P(A(\epsilon)) + P(A \Delta A(\epsilon)) \Rightarrow |P(A) - P(A)P(A(\epsilon))| \leq \epsilon$. As $\epsilon \to 0$, $P(A(\epsilon)) \to P(A)$, which imply $P(A) = P(A)^2$.

**Theorem (Berry-Essen Theorem):** If $X_i$ i.i.d., $E(X_i) = 0, E(X_i^2) = 1, E(X_i^3) = \rho < \infty$, then $\forall x \left| P\left(\frac{\sum_{i=1}^{n} X_i}{\sqrt{n}} \leq x\right) - \Phi(x)\right| \leq \frac{Cp}{\sqrt{n}}$ where $C \leq 2.4$.

**Theorem (Berry-Essen Inequality):** For any two random variables, $X, Y$, with cdfs $F(x), G(x)$, and chfs $\phi(t), \psi(t)$, $\exists A, B$, which are constants s.t. $\sup_x |F(x) - G(x)| \leq 2 \int_{-T}^{T} \frac{|\phi(t) - \psi(t)|}{t} dt + \frac{B}{T} |\int G(x)|_\infty$.

We state other properties, mostly having to do with inequalities and expectation.

**Theorem:** If $X$ is a bounded r.v., then $|\int X dP| \leq \int |X| dP$

**Theorem:** If $X \leq Y$, then $\int X dP \leq \int Y dP$ (monotonicity of the integral)

**Theorem (Jensen’s Inequality):** If $f$ is a proper, convex function and $X$ is r.v., then $f\left(E(X)\right) \leq E\left(f(X)\right)$

**Theorem (Hoeffding’s Inequality, Limited):** If $X_i, i.i.d., |X_i| \leq 1$, and $E(X_i) = 0$, then $\forall \epsilon > 0, P\left(\frac{\sum_{i=1}^{n} X_i}{n} \geq \epsilon\right) \leq \exp\left(-\frac{1}{2} \epsilon^2 n\right)$

**Theorem (Tail Bound for Normal(0,1)):** If $X \sim N(0,1), \forall \epsilon > 0, P(X \geq \epsilon) \leq e^{-\frac{\epsilon^2}{2}}$

**Theorem:** For any sequence of events $A_i$, $\lim_{n \to \infty} \frac{\sum_{i=1}^{n} I_{A_i}}{\sum_{i=1}^{n} P(A_i)} = 1$ a.s.

**Theorem (Approximation of Sets/Approximation Lemma):** If $F_1 \subset F_2 \subset \ldots$ is a sequence of $\sigma$-fields and $A \in \sigma(\bigcup_{i=1}^{\infty} F_i)$, then $\forall \epsilon > 0, \exists A(\epsilon) \in \bigcup_{i=1}^{\infty} F_i$ s.t. $P(A \Delta A(\epsilon)) \leq \epsilon$ where $A \Delta A(\epsilon) = (A \cup A(\epsilon)) \setminus (A \cap A(\epsilon))$ (aka the stuff not shared by two sets $A$ and $A(\epsilon)$).

**Theorem (Kronecker’s Lemma):** For $a_i \in R, \sum_{i=1}^{\infty} \frac{a_i}{n}$ converges imply $\sum_{i=1}^{n} \frac{a_i}{n} \to 0$

**Proof:** $\sum_{i=1}^{\infty} \frac{a_i}{n} \to a_1 + \ldots + a_n + n \sum_{i=1+n}^{\infty} \frac{a_i}{n} \Rightarrow \left(\frac{n a_1}{n} + \frac{n a_2}{n} + \ldots \frac{n a_n}{n}\right)^2 \to a_1 + \ldots + a_n$. Note that $\sum_{i=n+1}^{\infty} \frac{a_i}{n} \to 0$ as $n \to \infty$ by our hypothesis. Taking limits on both sides, you get the result.

**Theorem (Sterling’s Formula):** $n! \approx \sqrt{2\pi n} n^n e^{-n}$
Theorem (Le Cam’s Theorem): If \( X_i \) are Bernoulli with parameters \( p_i \), independent, \( \lambda_n = \sum_{i=1}^{n} p_i \), and \( S_n = \sum_{i=1}^{n} X_i \), then \( \sum_{k=0}^{\infty} \frac{\lambda_{k} e^{-\lambda_n}}{k!} | < 2 \sum_{i=1}^{n} p_i^2 \). Note: This is saying that the sum of Bernoullis roughly has a Poisson distribution.

Convergence of Random Variables

First, we define the three different types of convergence encountered in

**Definition (convergence in dist):** A sequence of probability spaces \((\Omega_k, F_k, P_k)\) with an associated random variable \( X_k \), converges in distribution of \( X \) if for all continuous points \( x \), in \( P(X \leq x) \), \( P(X_n \leq x) \rightarrow P(X \leq x) \)

**Definition (convergence in prob):** A sequence of functions \( X_n: \Omega \rightarrow R \) converges in probability if \( \forall \epsilon > 0, \lim_{n \rightarrow \infty} P(|X_n - X| \geq \epsilon) = 0 \)

**Definition (convergence a.s.):** A sequence of functions \( X_n: \Omega \rightarrow R \) converges in probability if \( \forall \epsilon > 0, P(|X_n - X| \geq \epsilon \text{ i.o.}) = 0 \).

Next, there are properties along with theorems related to these convergence.

**Proposition:** \( X_n \xrightarrow{a.s.} X \Rightarrow X_n \xrightarrow{prob} X \)

**Proof:** Let \( P(A_n) = P(\cup_{n \geq 2N} \{ \omega : |X_n(\omega) - X(\omega)| \geq \epsilon \}) \geq P(\{ \omega : |X_n(\omega) - X(\omega)| \geq \epsilon \}) \). Note that \( A_n \supset A_{n+1} \supset A_{n+2} \supset \ldots \) and \( A = \cap_{n=1}^{\infty} A_n \). Thus, \( P(A_n) \rightarrow P(A) = 0 \Rightarrow \lim_{n \rightarrow \infty} P(\{ \omega : |X_n(\omega) - X(\omega)| \geq \epsilon \}) = 0 \)

**Proposition:** If \( X_n \xrightarrow{prob} X \Rightarrow X_n \xrightarrow{dist} X \)

**Proposition:** If \( X_n \xrightarrow{prob} X \Rightarrow \) These are identical statements: (i) \( X_n \) is uniformly integrable, (ii) \( X_n \rightarrow X \) in \( L^1 \), and (iii) \( E|X_n| \rightarrow E|X| < \infty \)

**Proposition:** If \( X_n \xrightarrow{prob} X \), then there is a subsequence, \( X_{n_k} \), s.t. \( X_{n_k} \xrightarrow{a.s.} X \)

**Proof:** Since \( \forall \epsilon > 0, P(|X_n - X| \geq \epsilon) \rightarrow 0 \). Then, by defn. of limits, for each \( k \), we can find \( X_{n_k} \) s.t. \( P(A_k) \leq \frac{1}{2^k} \) where \( A_k = \{ |X_{n_k} - X| \geq \epsilon \} \). Since \( \sum P(A_k) < \infty \), by first B.C. \( P(A_k \text{ i.o.}) = 0 \), as desired.

**Proposition:** If \( X_n \xrightarrow{prob} X \Leftrightarrow \forall X_{n_k}, \) there is a further subsequence s.t. \( X_{n_{k_l}} \xrightarrow{a.s.} X \)

**Proposition:** If \( X_n \xrightarrow{prob} X \) and \( X_n \xrightarrow{prob} Y \), then \( P(X = Y) = 1 \)

**Theorem (Bounded Convergence Theorem):** If \( X_n \) is a sequence of uniformly bounded random variables defined on the same probability space and if \( X_n \xrightarrow{a.s./prob} X \), then \( \int X_n dP \rightarrow \int X dP \)

**Proof:** \(|\int (X_n - X) dP| \leq \int |X_n - X| dP = \int_{\omega:|X_n - X| \geq \epsilon} |X_n - X| dP + \int_{\omega:|X_n - X| < \epsilon} |X_n - X| dP \leq \int_{\omega:|X_n - X| \geq \epsilon} 2MdP + \int_{\omega:|X_n - X| < \epsilon} \epsilon dP = 2M P(|X_n - X| \geq \epsilon) + \epsilon \leq K \epsilon' \), where \( K \) is some constant. This is the defn. of limit.

**Theorem (Fatou’s Lemma):** If \( X_n \geq 0 \) and \( X_n \xrightarrow{a.s./prob} X \), then \( \int X dP \leq \liminf_{n \rightarrow \infty} \int X_n dP \)
Proof: Define \( Y_n = X_n \wedge Z \) where \( Z \geq 0 \) and is a bounded r.v. Note that \( Y_n \leq Z \leq M \) and \( Y_n \xrightarrow{a.s/prob} X \wedge Z \). Then,
\[
\int X_n \wedge Z \, dP \leq \int X_n \, dP \leq \liminf \int X_n \wedge Z \, dP \leq \liminf \int X_n \, dP.
\]
Then, by BCT, we get \( \int X \wedge Z \, dP \leq \liminf \int X_n \, dP \). Finally, take the sup over all \( Z \) and we end up with \( \int X \, dP \leq \liminf \int X_n \, dP \).

Note that Fatou’s Lemma is often used to show that the expectation of a random variable is bounded.

**Theorem (Dominated Convergence Theorem):** If \( |X_n| \leq Y \), \( E(Y) < \infty \), and \( X_n \xrightarrow{a.s/prob} X \), then \( \int X_n \, dP \to \int X \, dP \).

Proof: Let \( Z'_n = X_n + Y \geq 0 \) and \( Z_n = Y - X_n \geq 0 \), where both \( Z'_n \xrightarrow{a.s/prob} X + Y \) and \( Z_n \xrightarrow{a.s/prob} Y - X \). By Fatou’s Lemma, we get \( \int X + Y \, dP \leq \liminf \int Z'_n \, dP \Rightarrow \int X \, dP \leq \liminf \int X_n \, dP \) and \( \int Y - X \, dP \leq \liminf \int Z_n \, dP \Rightarrow \int X \, dP \geq \limsup \int X_n \, dP \). Combining them together, we get our result.

**Theorem (Monotone Convergence Theorem):** If \( 0 \leq X_n \leq X_{n+1} \) and \( X_n \uparrow X \) (a.s. or in prob), then \( \int X_n \, dP \to \int X \, dP \).

Proof: Using Fatou, we get \( \int X \, dP \leq \liminf_{n \to \infty} \int X_n \, dP \). Also, taking the integral and taking \( \limsup \) on both sides, we get \( \limsup_{n \to \infty} \int X_n \, dP \leq \int X \, dP \). Thus, we get our desired result.

There are several results on convergence. Here, we state what’s relevant for this section.

**Theorem (Kolmogorov One-Series Theorem):** If \( X_i \) independent, \( E(X_i) = 0, \sum_{i=1}^{\infty} \frac{1}{i^2} \text{Var}(X_i) < \infty \), then \( \sum_{i=1}^{\infty} \frac{X_i}{i} \) converges a.s.

Proof: Use Kolmogorov Maximal Inequality

**Theorem (Kolmogorov Condensation Test):** Suppose \( X_i \) indep and \( P(|X_i| \leq B) = 1 \forall i \). If \( \sum_{i=1}^{\infty} X_i \) converges (a.s. or in prob), \( \sum_{i=1}^{\infty} \text{Var}(X_i) \) converges.

Now, there are two very important types of law of large numbers, the strong law and the weak law. The weak law is basically convergence in probability while strong law is convergence a.s. For here on out, \( S_n = \sum_{i=1}^{n} X_i \)

**Theorem (Weak Law of Large Numbers):** If \( X_i \) are independent with \( E(X_i) = 0, \text{Var}(X_i) = \sigma^2 \), then \( \frac{S_n}{n} \xrightarrow{prob} 0 \)

Proof: Using Chebyshev, \( \forall \varepsilon > 0, P(|\sum X_i| \geq n\varepsilon) \leq \frac{\sigma^2}{\varepsilon^2} \). Take the limit on both sides to obtain the result.

There are several flavors of the strong law of large numbers, each relaxing assumptions from the previous ones

**Theorem (Strong Law of Large Numbers, v1):** If \( X_i \text{ i.i.d. and bounded, } E(X_i) = 0 \), then \( P\left( \lim_{n \to \infty} \frac{S_n}{n} = 0 \right) = 1 \)

Proof: Pick quadratic subsequence terms, \( n_k = k^2 \). Then, we know from Chebyshev and first BC,

\[
\forall \varepsilon > 0, \sum_{k=1}^{\infty} P\left( \frac{S_{n_k}}{n_k} > \varepsilon \right) \leq \frac{\sigma^2}{\varepsilon^2} \sum_{k=1}^{\infty} \frac{1}{k^2} < \infty \Rightarrow P\left( \frac{S_{n_k}}{n_k} > \varepsilon \text{ i.o.} \right) = 0 \Rightarrow P\left( \lim_{k \to \infty} \frac{S_{n_k}}{n_k} = 0 \right) = 1.
\]

Now, let \( n_k \leq n \leq n_{k+1} \). We can write \( \frac{S_n}{n} = \frac{S_n + X_{n+1} + \ldots + X_{n+k}}{n} \leq \frac{S_{n_k} + M(2k+1)}{n} \leq \frac{S_{n_k}}{n_k} + \frac{M(2k+1)}{n} \). As \( n \to \infty \), we get \( \lim_{n \to \infty} \frac{S_n}{n} \leq \lim_{n \to \infty} \frac{S_{n_k}}{n_k} \). Since the event that \( \lim_{n \to \infty} \frac{S_{n_k}}{n_k} = 0 \) w.p. 1, \( \lim_{n \to \infty} \frac{S_n}{n} = 0 \) w.p. 1

**Theorem (Strong Law of Large numbers, v2):** If \( X_i \text{ i.i.d. } E(X_i) = 0, \text{Var}(X_i) = \sigma^2 \), then \( P\left( \lim_{n \to \infty} \frac{S_n}{n} = 0 \right) = 1 \)
Proof: By Kolmogorov One-Series Theorem, \( \sum_{i=1}^{\infty} \frac{\text{Var}(X_i)}{i^2} < \infty \Rightarrow \sum_{i=1}^{\infty} \frac{X_i}{i} \) converges a.s. By Kronecker, this implies \( \sum_{i=1}^{\infty} \frac{X_i}{n} \) converges a.s.

**Theorem (Strong Law of Large Numbers, v3):** If \( X_i \) i.i.d. \( E(X_i) = 0 \), and \( E(|X_i|) < \infty \), \( P \left( \lim_{n \to \infty} \frac{S_n}{n} = 0 \right) = 1 \)

Proof: Let \( \frac{X_1 + \ldots + X_n}{n} = \frac{\sum_{i=1}^{n} X_i I_{|X_i| \geq i}}{n} + \frac{\sum_{i=1}^{n} Y_i}{n} \) where \( Y_i = X_i I_{|X_i| \leq i} \). For the first sum, \( \sum_{i=1}^{\infty} P(|X_i| \geq i) \leq E(|X_1|) < \infty \), by first BC \( P(|X_i| \geq i, \ i.o.) = 0 \) and thus, it goes to zero a.s. For the second sum, using Kolmogorov’s One-Series Theorem, \( \sum_{i=1}^{\infty} \frac{\text{Var}(Y_i)}{i^2} \leq E(Y_i^2) \sum_{i=1}^{\infty} \frac{I_{|X_i| \leq i}}{i^2} \leq E\left( \frac{Y_i^2}{|Y_i|} \right) = E(|Y_i|) < \infty \Rightarrow \sum_{i=1}^{\infty} \frac{Y_i}{i} \) converges a.s. By Kronecker’s Lemma, this implies that \( \sum_{i=1}^{\infty} \frac{Y_i}{n} \to 0 \) a.s.

*Theorem (Strong Law of Large Numbers, v4):* If \( X_i \) are pairwise independent, \( E(X_i) = 0 \), and \( E(|X_i|) < \infty \), \( P \left( \lim_{n \to \infty} \frac{S_n}{n} = 0 \right) = 1 \)

Proof: Requires Edimonti’s Theorem and a whole lot of crap.

**Distribution Theory**

We first define a characteristic function and state some properties associated with it. Note that for any r.v., the characteristic function always exists.

**Definition:** For any random variable \( X \), the characteristic function, \( \phi: R \to C \), of \( X \) is defined as \( \phi(t) = E(e^{itX}) \)

**Properties:** If \( \phi(t) \) is a characteristic function, then the following statements are true

(i) \( \phi(0) = 1 \)
(ii) \( \phi(t) \) is uniformly continuous (and hence, continuous)
(iii) \( \limsup \phi(t) > 0 \Rightarrow \) there is an atom in some place
(iv) \( \phi(t) \) is real \( \Leftrightarrow X = -X \) in distribution (aka symmetric)
(v) \( \int_{-\infty}^{\infty} |\phi(t)|dt < \infty \Rightarrow \) We have a pdf
(vi) \( E(|X|^n) < \infty \Rightarrow \phi(t) \) is \( n \) times continuous differentiable with \( \phi^{(k)}(0) = i^k E(X^k), \ 1 \leq k \leq n \)
(vii) \( \phi(t) \) is \( n \) derivatives at zero, then \( E(|X|^k) < \infty \) \( \forall 1 \leq k \leq n \) if \( n \) is even and up to \( n-1 \) if \( n \) is odd.
(viii) (Levy’s Uniqueness Theorem) If \( X \) and \( Y \) have the same chfs, then \( X \) and \( Y \) have the same distribution
(ix) (Levy’s Inversion Formula): \( \frac{1}{2\pi i} \int_{-\infty}^{\infty} \frac{(e^{-iat}-e^{-ibt})\phi(t)}{t} \ dt = \frac{1}{2} P(X = a) + P(a < X < b) + \frac{1}{2} P(X = b) \) and thus, the chf determines the cumulative distribution of \( X \).

**Theorem (Polya’s Theorem):** If \( \phi(t) \) is (i) symmetric, (ii) convex for \( t > 0 \), (iii) uniformly continuous, and (iv) \( \phi(0) = 1 \), then \( \phi(t) \) is a characteristic function of Polya type

In addition to characteristic functions, we define the concept of tightness and how it can be used to prove various properties of convergence for characteristic functions and its cdfs.

**Definition:** A sequence of cdfs, \( F_n \), is tight if \( \forall \epsilon > 0, \exists T \ s.t. F_n(-T) + 1 - F_n(T) \leq \epsilon \ \forall n \)

**Definition:** A sequence of r.v.s, \( X_n \), is tight if \( \forall \epsilon > 0, \exists T \ s.t. P(|X_n| \geq T) \leq \epsilon \ \forall n \)
**Theorem (Helly Selection Theorem):** For a sequence of cdfs, $F_n$, there is a subsequence $n_j$ and a right-continuous non-decreasing function $F$ s.t. $F_{n_j}(x) \to F(x)$ at all continuity points in $F$. (Note: Steele states $F$ is a distribution function only if $F_n$ is tight)

**Proof:** By compactness of $[0,1]$ and $F_n \in [0,1]$, $\exists F_{nk}$ that converges in $[0,1]$. Let $q_1, q_2, \ldots$ be a sequence of rationals.

Then, $F_n(q_1)$ has a subsequence, $F_{nk}(q_1)$, that converges to $F(q_1)$. Move on to $q_2$. $F_{nk}(q_2)$ has a further subsequence that converges to $F(q_2)$. Repeating this diagonal argument through $q_1$, we end up with a super-small subsequence s.t. $\lim_{n \to \infty} F_{nk}(x)$.

Next, since $F_{nk}$ is a non-decreasing function, $F_{nk}$ must also be a non-increasing function. Furthermore, $F(x) = \lim_{n \to \infty} F_{nk}(x)$ (right continuous). Finally, if $x$ is a continuous point in $F$, we can pick $r' < x < r''$ so that $F(r') \leq F(x) \leq F(r'')$. Since $F_{nk}(r') \to F(r')$ and $F_{nk}(r'') \to F(r'')$, then $F_{nk}(r') \leq F_{nk}(x) \leq F_{nk}(r'')$, as desired.

**Theorem (Tail Bound):** For any random variable $X$, $P(|X|) \geq \frac{2}{\epsilon} \leq 2(1 - \frac{1}{2\epsilon}) \phi(t)dt$.

**Proof:** $P(|X| \leq T) + \frac{1}{\epsilon} P(|X| > T) = 1 - P(|X| > T) + \frac{1}{T} P(|X| > T)$. Rearrange the equations to obtain $1 - \frac{1}{\epsilon} P(|X| \leq T) = 1 - \frac{1}{2\epsilon} \phi(t)dt$. Pick $T = 2$ and we’re done.

**Theorem (Tightness Lemma):** If a sequence of chfs, $\phi_n(t) \to \phi(t)$ for $t \in (-\epsilon, \epsilon)$ and some $\epsilon > 0$ and $\phi(t)$ is continuous at zero, then the cdfs associated with $\phi_n(t)$ is tight. (Note: $\phi(t)$ does not have to be a chf)

**Proof:** ($F_n$ denotes the cdfs for each chf). By Tail Bound, we get $F_n \left( \left[ -\frac{2}{\epsilon}, \frac{2}{\epsilon} \right] \right) \leq 2 \left( 1 - \frac{1}{2\epsilon} \phi(t)dt \right)$. Taking limsup on both sides, we get $F_n \left( \left[ -\frac{2}{\epsilon}, \frac{2}{\epsilon} \right] \right) \leq \limsup_{n \to \infty} F_n \left( \left[ -\frac{2}{\epsilon}, \frac{2}{\epsilon} \right] \right) \leq \lim_{n \to \infty} 2 \left( 1 - \frac{1}{2\epsilon} \phi(t)dt \right)$. We can control the right hand to be a very small number because $\lim_{\epsilon \to 0} 2 \left( 1 - \frac{1}{2\epsilon} \phi(t)dt \right) = 0$ where we use that $\phi(t)$ is continuous at zero. Thus, $F_n$ are tight.

**Theorem (Levy Continuity Theorem):** Given a sequence of chfs, $\phi_{nk}(t) \to \phi(t)$ and its corresponding cdfs $F_{nk}$, where $\phi_{nk}(t) \to \phi(t)$ for all $t$ and $\phi(t)$ is continuous at zero, then (i) $\phi(t)$ is a chf and (ii) $F_{nk} \to F$

**Proof:** By Tightness Lemma, $F_{nk}$ must be tight. By Helly, $\exists F_{nk} \to F$ and thus, $\phi(t)$ has to be a chf. Now, by contradiction, suppose $F_{nk}$ does not converge to $F$. This implies that there is a subsequence $F_{nk}$ does not converge to $F$. By Helly, we can choose a further subsequence of $F_{nk}$ s.t. $F_{nk} \to G$, where $G(x) \neq F(x)$ where $x$ is a continuity point of $F$. But, since the chfs converge, $\phi_{nk}(t) \to \phi_{n}\phi(t)$ and $\phi_{nk}(t) \to \phi_{n}\phi(t) \Rightarrow \phi_{nk}(t) = \phi_{n}(t) = \phi_{n}(t)$. Hence, $F_{nk} = G(x)$, which is a contradiction.

With these criterions, we can now prove the central limit theorem. There are many flavors of this as well, each relaxing the assumptions of the other.

**Theorem (Central Limit Theorem, v1):** If $X_i$ i.i.d. $E(X_i) = 0, E(X_i^2) = 1$, $\frac{S_n}{\sqrt{n}} \to N(0,1)$

**Proof:** We need three key ingredients:

1. **From Taylor Series:** $e^{it} = \sum_{n=0}^{\infty} \frac{(it)^n}{n!} = \sum_{n=0}^{\infty} \frac{(it)^n}{n!} + \frac{it^{k+1}}{k!} \int_0^u e^{it(u-t)}k dt \Rightarrow \left| \phi(t) - \sum_{n=0}^{\infty} \frac{(it)^k}{k!} E(X^k) \right| \leq E \left( \min \left( \frac{|X|^{k+1}}{(k+1)!}, \frac{|tX|^k}{k!} \right) \right)$. Take $k = 2$ and we get $|\phi(t) - \left( 1 - \frac{1}{2} t^2 \right)| \leq E \left( \min \left( \frac{|X|^3}{6}, |tX|^2 \right) \right)$
2. By induction, we get $|\prod _{i=1}^{n} a_i - \prod _{i=1}^{n} b_i| \leq \sum _{i=1}^{n} |a_i - b_i|$ where $|a_i| \leq 1$ and $|b_i| \leq 1$ (lemma 1, pg 358)

3. From the definition of natural $e$, $\lim _{n \to \infty} \left( 1 + \frac{t}{n} \right)^n = e^t$

Since $\frac{S_n}{\sqrt{n}}$ is represented by a characteristic function $\phi \left( \frac{t}{\sqrt{n}} \right)$ and using (1) and (2), we get $\left| \phi \left( \frac{t}{\sqrt{n}} \right) - \left( 1 - \frac{1}{2n} t^2 \right)^n \right| \leq t^2 E \left( \min \left( \frac{|t||X|^3}{\sqrt{n}}, |X|^2 \right) \right)$. By DCT, $\min \left( \frac{|t||X|^3}{\sqrt{n}}, |X|^2 \right) \leq |X|^2$ and $t^2 E \left( \min \left( \frac{|t||X|^3}{\sqrt{n}}, |X|^2 \right) \right) \to 0$. By (3), we get $\left( 1 - \frac{1}{2n} t^2 \right)^n \to e^{\frac{1}{2} t^2}$. Since this is the chf of a normal, by Levy Continuity Theorem, we're done.

**Theorem (Central Limit Theorem, v2):** For each $n$, we have $X_{n,1}, ..., X_{n,n}$, where each $X_{n,k}$ are independent and each set of sequence may live on different probability spaces. If $E(X_{n,k}) = 0$, $E(X_{n,k}^2) = \sigma_{n,k}^2$, $s_n^2 = \sum _{k=1}^{n} \sigma_{n,k}^2$, and $\forall \varepsilon > 0, \sum _{k=1}^{n} E(X_{n,k}^2 | |X_{n,k}| \geq \varepsilon s_n) \to 0$, then $\frac{\sum _{k=1}^{n} X_{n,k}}{s_n} \to N(0,1)$. Note: The tail bound condition is called the Lindenberg Condition (LC)

**Proof:** WLOG, let $Y_{n,k} = \frac{X_{n,k}}{s_n}$ to get $\sum _{k=1}^{n} E(Y_{n,k}^2) = 1$, $E(Y_{n,k}^2) = \sigma_{n,k}^2$, and $\forall \varepsilon > 0, \sum _{k=1}^{n} E(Y_{n,k}^2 | |Y_{n,k}| \geq \varepsilon) \to 0$.

Then, using (1) and (2) from v1, we get $|\Pi _{k=1}^{n} \phi_{Y_{n,k}}(t) - \Pi _{k=1}^{n} \left( 1 - \frac{1}{2} t^2 \sigma_{n,k}^2 \right) - \Pi _{k=1}^{n} e^{\frac{1}{2} t^2 \sigma_{n,k}^2}| \leq |\Pi _{k=1}^{n} \phi_{Y_{n,k}}(t) - \Pi _{k=1}^{n} \left( 1 - \frac{1}{2} t^2 \sigma_{n,k}^2 \right) - \Pi _{k=1}^{n} e^{\frac{1}{2} t^2 \sigma_{n,k}^2}|$. The first and the second absolute values can be bounded by the following

- Using (1) and (2) from v1
  $|\Pi _{k=1}^{n} \phi_{Y_{n,k}}(t) - \Pi _{k=1}^{n} \left( 1 - \frac{1}{2} t^2 \sigma_{n,k}^2 \right)| \leq \sum _{k=1}^{n} E \left( t^2 Y_{n,k}^2 \min (|t Y_{n,k}|, 1) (|Y_{n,k}| < \varepsilon) + (|Y_{n,k}| \geq \varepsilon) \right)$

- Again, using the same trick again where we Taylor expand around $\phi_{Y_{n,k}}(t) \sim N(0, \sigma_{n,k}^2)$, $|\Pi _{k=1}^{n} \left( 1 - \frac{1}{2} t^2 \sigma_{n,k}^2 \right) - \Pi _{k=1}^{n} e^{\frac{1}{2} t^2 \sigma_{n,k}^2}| \leq E \left( t^2 N_{n,k}^2 \min (|t N_{n,k}|, |N_{n,k}| \geq \varepsilon) \right)$

Finally, we have a pretty cool theorem showing that two random variables have identical distribution.

**Theorem:** If $\sum _{k=0}^{n} \frac{1}{2k!} \sigma_{2k}^2 m_{2k} < \infty$ and $E(X^k) = m_k = E(Y^k)$ for $k = 0, 1, 2, ...$, then $X = Y$ in distribution.

**Martingales**

First, note that we work in $L^2$ (aka Hilbert spaces) where all Cauchy sequence converges. A related work to Hilbert spaces is the theory behind conditional expectation.

**Definition:** If $G \subseteq F, x \in F, F$ is a field, and $E|X| < \infty$, then $Y$ is a **conditional expectation** of $X$ given $G$ if (a) $Y \in G$ and (b) $E(I_A Y) = E(I_A X) \forall A \in G$. Written another way, $Y = E(X|G)$
Properties: Take the notation above. Then, the following are true

(i) (Existence) $Y$ exists
(ii) (Uniqueness) If $Y$ above exists, then $Y$ is a.s. unique.
(iii) (Linearity of Conditional Expectation). If $Y_1, Y_2$ is the conditional expectation of $X_1, X_2$ given $G$, then $Y_1 + Y_2$ is the conditional expectation of $X_1 + X_2$ given $G$
(iv) (Non-negativity): If $X \geq 0$, then $Y$ is non-negative.
(v) (Contraction): $|Y| \leq |X|$ (or $E(|Y|) \leq E(|X|)$)
(vi) (Tower Property): Given $G_1 \subset G_2 \subset F$, $E(E(X|G_2)|G_1) = E(E(X|G_1)|G_2) = E(X|G_1)$

We highlight a convergence of monotone martingales.

Theorem (Monotone Convergence of Conditional Expectation): If $X_n \geq 0$ and $X_n \leq X_{n+1}$, $X_n \to X$ w.p.1, and $E(X) < \infty$, then $E(X_n|G) \to E(X|G)$ w.p.1.

Proof: Since $Y_n = E(X_n|G) \leq E(X_{n+1}|G) = Y_{n+1}$, and $E(|G)$ is some positive operator on the monotone sequence, $\exists Y$ s.t. $E(X_n|G) \to Y$ w.p.1. (Note $Y$ can be infinity). Next, using MCT, $E(Y_n) \to E(Y)$ and since $E(Y_n) = E(X_n) < \infty$, we get $E(Y) < \infty$. Finally, to show that $Y = E(X|G)$, again, using MCT, we see that $\forall A \in G$, $E(Y_n|A) \to E(Y|A)$ and $E(X_n|A) \to E(X|A)$ and $E(Y_n|A) = E(X_n|A)$. Since limits are unique, $E(Y|A) = E(X|A)$ and we’re done.

Now, onto the definition of a martingale

Definition: Suppose we have $(X_n, F_n)$ s.t. $X_n \in F_n, F_1 \subset F_2 \subset \ldots$, and $E|X_n| < \infty$. Then $X_n$ is a martingale iff $E(X_{n+1}|F_n) = X_n$ $\forall n$

Definition: Suppose we have $F_1 \subset F_2 \subset \ldots$, and a random variable $Z$ where $E(|Z|) < \infty$. Then, $(Y_n, F_n)$, where $Y_n = E(Z|F_n)$, is a martingale (called Levy’s Martingale)

Definition: Suppose we have $(X_n, F_n)$ s.t. $X_n \in F_n, F_1 \subset F_2 \subset \ldots$, and $E|X_n| < \infty$. Then $X_n$ is a submartingale iff $E(X_{n+1}|F_n) \geq X_n$. Note that any convex function of a MG, $X_n$, is subMG. If $X_n$ is a subMG and monotonic, any convex function on it will be a subMG.

Now, there are a lot of ways in which martingales can converge. We highlight a couple along with the proofs

Theorem: If $(X_n, F_n)$ is a martingale and $d_k = X_k - X_{k-1}$, then for any $j \neq k, E(d_j d_k) = 0$.

Proof: WLOG, $j < k$. Then, $E(d_j d_k) = E((d_j d_k|F_j)) = E(d_j E(d_k|F_j)) = E(d_j E(X_k - X_{k-1}|F_{k-1} - F_j)) = 0$

Theorem (Doob $L^2$ Maximal Inequality): For any martingale $(X_n, F_n)$ in $L^2$ and $\lambda$, $P(\max_{1 \leq i \leq n}|X_i| \geq \lambda) \leq \frac{1}{\lambda^2} E(X_n^2)$

Proof: Let $T = \min\{i: |X_i| \geq \lambda\}$ and $A_i = i$. Then, $\max_{1 \leq i \leq n}|X_i| \geq \lambda \leq \frac{1}{\lambda^2} \sum_{i=1}^{n} E(X_i^2 A_i)$. Take expectation on both sides, we get $P(\max_{1 \leq i \leq n}|X_i| \geq \lambda) \leq \frac{1}{\lambda^2} \sum_{i=1}^{n} E((X_n - X_i + X_i)^2 A_i)$ where the second inequality works because $E((X_n - X_i)X_i A_i) = E(E((X_n - X_i X_i A_i|F_i)) = E(X_i A_i E(X_n - X_i|F_i)) = 0$. Therefore, $\frac{1}{\lambda^2} \sum_{i=1}^{n} E((X_n - X_i + X_i)^2 A_i) \leq \frac{1}{\lambda^2} E(X_n^2 \sum_{i=1}^{n} A_i) \leq \frac{1}{\lambda^2} E(X_n^2)$, as desired. Note: This also works if MG is in $L^p$.

Theorem ($L^2$ Convergence): If $(X_n, F_n)$ is a martingale in $L^2$, then $X_n$ converges in $L^2$, a.s.

Proof: First, some facts about expectation and martingales. For any $m < n$, $E((X_n - X_m)^2) = E(\sum_{k=m+1}^{n} d_k)^2 = \sum_{k=m+1}^{n} E(d_k)^2$, thanks to the theorem above. Now, $E(X_n^2) = E(\sum_{k=1}^{n} d_k)^2 = \sum_{k=1}^{n} E(d_k)^2 < B$ for any $n \Rightarrow$
\[ \sum_{k=1}^{\infty} E(d_k^2) < B. \] Therefore, \( \sup_{n,m \geq N} E(X_n - X_m)^2 \leq \sum_{k=N+1}^{\infty} E(d_k^2) \). Notice that the right hand sum shrinks as \( N \) grows (thanks to the bound on the sum).

Second, to prove that \( X_n \) converges a.s., we just have to show that the sequence is Cauchy OR
\[ \lim_{N \to \infty} P(\max_{n,m \geq N}|X_n - X_m| \geq \varepsilon) = 0 \quad \forall \varepsilon > 0. \] Note that \( P(\max_{M \geq n,M \geq N}|X_n - X_M + X_N - X_m| \geq 2\varepsilon) \leq P(\max_{M \geq n,M \geq N}|X_n - X_N| \geq \varepsilon) + P(\max_{M \geq n,M \geq N}|X_N - X_m| \geq \varepsilon). \) Applying Doob \( L^2 \) on each of the terms, we get
\[ P(\max_{M \geq n,M \geq N}|X_n - X_N| \geq \varepsilon) + P(\max_{M \geq n,M \geq N}|X_N - X_m| \geq \varepsilon) \leq \frac{2}{\varepsilon^2} E((X_M - X_N)^2) \leq \frac{2}{\varepsilon^2} \sum_{k=N+1}^{\infty} E(d_k^2). \] Take \( M \to \infty \), we end up with \( P(\max_{n,m \geq N}|X_n - X_m| \geq \varepsilon) \leq \frac{2}{\varepsilon^2} \sum_{k=N+1}^{\infty} E(d_k^2). \) Using the first fact, we get the RHS to 0.

**Theorem (Doob \( L^1 \) Maximal Inequality):** For any subMG\([X_n, F_n]\) that is non-negative and for any \( \lambda \),
\[ P(\max_{1 \leq k \leq n} X_k \geq \lambda) \leq \frac{1}{\lambda} E(X_n) \]

**Proof:** Define \( T = \min\{k: X_k \geq \lambda\} \). Then, \( I_{\max_{1 \leq k \leq n} X_k \geq \lambda} \leq \sum_{k=1}^{n} \frac{1}{\lambda} X_k I_{T=k} \). Take expectation on both sides,
\[ P(\max_{1 \leq k \leq n} X_k \geq \lambda) \leq \frac{1}{\lambda} \sum_{k=1}^{n} E(X_k I_{T=k}) \leq \frac{1}{\lambda} \sum_{k=1}^{n} E(X_n I_{T=k}) = \frac{1}{\lambda} E(X_n \sum_{k=1}^{n} I_{T=k}) \leq \frac{1}{\lambda} E(X_n) \] where the second inequality uses \( E(X_{n+1}) \geq E(X_n) \). Moving out the expectation and using the same reasoning as \( L^2 \), we get the above inequality.

**Theorem (\( L^1 \) Convergence):** If \( [X_n, F_n] \) are subMG in \( L^1 \), then \( X_n \) converges a.s.

**Proof:** \( E(N_n) \leq E(X_n^+ + |a|) \). By MCT and the fact that \( E(X_n^+) < \infty, E(N_\infty) < \infty \Rightarrow P(N_\infty < \infty) = 1. \) Furthermore, \( P(U_{(a,b)} \in \mathbb{Q} \liminf X_n < a < b < \limsup X_n) \) \( \leq P(\liminf X_n < a < b < \limsup X_n) = 0 \Rightarrow \liminf X_n = \limsup X_n \) a.s. and thus, \( X_n \) exists. By Fatou’s lemma, \( E(X^+) \leq \liminf E(X_n^+) < \infty \) and thus \( X < \infty \) a.s.
Similarly, for the case \( E(X_n^+) = E(X_n^+) - E(X_n) \leq E(X_n^+) - E(X_0) \). Using Fatou again, \( E(X^-) \leq \liminf E(X_n^-) \leq \sup E(X_n^-) - E(X_0) < \infty \). Hence, \( X > -\infty \) a.s. Therefore, \( X_n \to X \) a.s.

**Theorem (\( L^p \) Maximal Inequality):** If \( (X_n, F_n) \) is a non-negative martingale, \( E(\max_{1 \leq k \leq n} X_k^p) \leq \frac{p-1}{p} E(|X_n|^p) \) for \( p > 1 \).

**Proof:** First, looking at Doob \( L^1 \) proof, we know \( \sum_{k=1}^{n} I_{T=k} = I_{\max_{1 \leq k \leq n} X_k \geq \lambda} \). Replacing that sum with the equality,
\[ \lambda P(\max_{1 \leq k \leq n} X_k \geq \lambda) \leq E(X_n I_{\max_{1 \leq k \leq n} X_k \geq \lambda}) \]
Manipulating this and using the fact that
\[ E(|Z|^p) = p \int_{0}^{\infty} \lambda^{p-1} P(Z \geq \lambda) d\lambda \]
for any non-negative \( Z \), \( p \lambda^{p-1} P(max_{1 \leq k \leq n} X_k \geq \lambda) \leq \frac{p}{p-1} E(X_n(p-1) \int_{0}^{\infty} \lambda^{p-2} I_{\max_{1 \leq k \leq n} X_k \geq \lambda} d\lambda) \leq \frac{p}{p-1} E(X_n(p-1) \int_{0}^{\max_{1 \leq k \leq n} X_k} \lambda^{p-2} d\lambda) = \frac{p}{p-1} E(X_n \left( \frac{\max_{1 \leq k \leq n} X_k}{p-1} \right)^{p-1}) \]
Then, applying the Holder’s Inequality on the RHS,
\[ \frac{p}{p-1} E(X_n \left( \frac{\max_{1 \leq k \leq n} X_k}{p-1} \right)^{p-1}) \leq \frac{p}{p-1} E(|X_n|^p) \frac{1}{p} E\left( \left( \max_{1 \leq k \leq n} X_k \right)^\frac{p}{p-1} \right)^{p-1} \leq \frac{p}{p-1} E(|X_n|^p)^\frac{p}{p-1}. \] Take the power off and we’re done.

**Theorem (Up-Crossing Inequality):** If \( X_n, n \geq 0 \) is a subMG, \( (b-a)E(N_n) \leq E((X_n - a)^+) - E((X_0 - a)^+) \) where \( N_n \) is the number of times the process up-crosses \( b \) and \( a \). Note: The idea is to show that the MG goes between \([a, b]\) finitely many times for \( L^1 \) convergence

**Proof:** In Durret. Very nasty

**Theorem (Doob Stopping Time):** If \( [X_n, F_n] \) is a MG, and \( T_n \) is a stopping time, then \( \{\min(X_n, T_n), F_n\} \) is a martingale.
**Theorem (Wald Identity):** If $X_i$ iid $E(|X_i|) < \infty$, and $N$ is a stopping time, $E(N) < \infty \Rightarrow E(S_N) = E(X_1)E(N)$

**Theorem (Optional Sampling Theorem):** If $\{X_n, F_n\}$ is MG, $E(N) < \infty$, is a bounded stopping time, $E(X_N) = E(X_1)$