Probability Theory

September 27, 2012

1 Basic Probability

**Definition: Algebra**
If \( S \) is a set then a collection \( \mathcal{A} \) of subsets of \( S \) is an algebra if

- \( S \in \mathcal{A} \)
- \( A \in \mathcal{A} \implies A^c \in \mathcal{A} \)
- \( \{ A_i \}_{i=1}^n \in \mathcal{A} \implies \bigcup_{i=1}^n A_i \in \mathcal{A} \)

**Definition: \( \sigma \)-Algebra**
If \( S \) is a set then a collection \( \mathcal{F} \) of subsets of \( S \) is an algebra if

- \( S \in \mathcal{F} \)
- \( A \in \mathcal{F} \implies A^c \in \mathcal{F} \)
- \( \{ A_i \}_{i=1}^\infty \in \mathcal{F} \implies \bigcup_{i=1}^\infty A_i \in \mathcal{F} \)

**Lemma:**
The following are basic properties of \( \sigma \)-algebras

- \( \emptyset \in \mathcal{F} \)
- \( \{ A_i \}_{i=1}^\infty \in \mathcal{F} \implies \bigcap_{i=1}^\infty A_i \in \mathcal{F} \)

**Lemma:**
If \( S \) is a non-empty set and \( \mathcal{G} \in \mathcal{P}(S) \) then \( \exists \mathcal{F} := \sigma(\mathcal{G}) \) the smallest \( \sigma \)-algebra containing \( \mathcal{G} \)

**Proof:**
We know that \( \mathcal{P}(S) \) is a \( \sigma \)-algebra containing \( \mathcal{G} \)
Define \( \Gamma := \{ \mathcal{A} \text{ \( \sigma \)-algebra} : \mathcal{G} \in \mathcal{A} \} \)
Claim \( \mathcal{F} = \bigcap_{\mathcal{A} \in \Gamma} \mathcal{A} \)
If this is a \( \sigma \)-algebra it is clearly the smallest one containing \( \mathcal{G} \) by definition of intersection.
Remark:
From the above lemma we say that \( F \) is the \( \sigma \)-algebra generated by \( G \) and that \( G \) is the generator of \( F \)

Definition: Borel \( \sigma \)-Algebra
Let \( S = \mathbb{R}^n, G := \{ \prod_{i=1}^{n} [a_i, b_i] : a_i \leq b_i \in \mathbb{Q} \} \)
Then the Borel \( \sigma \)-algebra is \( B(\mathbb{R}^n) = \sigma(G) \)

Remark:
If \( S \) is a topological space then \( B(S) \) is the \( \sigma \)-algebra generated by open sets of \( S \)

Definition: Measurable Space
\((S, \mathcal{F})\) is a measurable space where
- \( S \) is a set
- \( \mathcal{F} \) is a \( \sigma \)-algebra

Definition: Measurable Set
\( A \) is measurable if \( A \in \mathcal{F} \) where \( \mathcal{F} \) is a \( \sigma \)-algebra

Definition: Additive
If \( S \) is a set, \( A \) is an algebra then
\( \mu : A \rightarrow \mathbb{R}_{[0,\infty)} \)
is additive if \( F, G \in A, F \cap G = \phi \implies \mu(F \cup G) = \mu(F) + \mu(G) \)

Definition: \( \sigma \)-Additive
If \( S \) is a set, \( A \) is an algebra then
\( \mu : A \rightarrow \mathbb{R}_{[0,\infty)} \)
is \( \sigma \)-additive if \( \{A_i\}_{i=1}^{\infty} \in \mathcal{F}, A_i \cap A_j = \phi \forall i \neq j \implies \mu(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i) \)

Definition: Measure
If \((S, \mathcal{F})\) is a measurable space and \( \mu : \mathcal{F} \rightarrow \mathbb{R}_{[0,\infty]} \) is a measure if \( \mu \) is \( \sigma \)-additive

Definition: Finite Measure
A measure \( \mu \) on measurable space \((S, \mathcal{F})\) is finite if \( \mu(S) < \infty \)

Definition: \( \sigma \)-Finite Measure
A measure \( \mu \) on measurable space \((S, \mathcal{F})\) is \( \sigma \)-finite if
\( \exists \{A_i\}_{i=1}^{\infty} \in \mathcal{F} \) s.t.
- \( \mu(A_i) < \infty \ \forall i \)
- \( \bigcup_{i=1}^{\infty} A_i = S \)

Definition: Probability Measure
A measure \( \mu \) on \( S \) is a probability measure if \( \mu(S) = 1 \)

Definition: Measure Space
A measure space is \((S, \mathcal{F}, \mu)\) where \( S \) is a set, \( \mathcal{F} \) is a \( \sigma \)-algebra on \( S \) and \( \mu \) is a measure on \((S, \mathcal{F})\)

Definition: Probability Space
A measure space \((S, \mathcal{F}, \mu)\) is a probability space if \( \mu \) is a probability measure

Definition: \( \pi \)-System
A family \( \mathcal{Y} \) of subsets of set \( S \) is a \( \pi \)-system if \( Y \) is closed under intersection
i.e. \( A, B \in Y \implies A \cap B \in Y \)

Definition: Dynkin System
If \( S \) is a set and \( \mathcal{D} \) is a set of subsets of \( S \) then \( \mathcal{D} \) is a Dynkin system of \( S \) if
- \( S \in \mathcal{D} \)
- \( A, B \in \mathcal{D}, A \subseteq B \implies B \setminus A \in \mathcal{D} \)
- \( \{A_i\}_{i=1}^{\infty} \in \mathcal{D}, A_i \cap A_j = \phi \forall i \neq j \implies \bigcup_{i=1}^{\infty} A_i \in \mathcal{D} \)

Remark:
We denote \( d(\mathcal{G}) \) to be the smallest Dynkin system containing \( \mathcal{G} \) which exists

Lemma:
If \( \mathcal{G} \) is a \( \pi \)-system then \( \sigma(\mathcal{G}) = d(\mathcal{G}) \)
Proof:
Clearly every \( \sigma \)-algebra is also a Dynkin system hence \( d(\mathcal{G}) \subseteq \sigma(\mathcal{G}) \) is trivial hence it remains to show that \( d(\mathcal{G}) \) is a \( \sigma \)-algebra

- Clearly \( \mathcal{S} \in d(\mathcal{G}) \) by definition
- Suppose \( A \in d(\mathcal{G}) \) then \( B \setminus A \in d(\mathcal{G}) \forall B \in d(\mathcal{G}) \) hence \( A^c = \mathcal{S} \setminus A \in d(\mathcal{G}) \)
- Want to show that \( d(\mathcal{G}) \) is a \( \pi \)-system
  - Define \( D_1 := \{ A \subseteq \mathcal{S} : A \cap B \in d(\mathcal{G}) \forall B \in \mathcal{G} \} \)
    - Is a Dynkin system and \( \mathcal{G} \subseteq D_1 \)
    - Define \( D_2 := \{ A \subseteq \mathcal{S} : A \cap B \in d(\mathcal{G}) \forall B \in d(\mathcal{G}) \} \subseteq d(\mathcal{G}) \)
      - Is a Dynkin system and \( \mathcal{G} \subseteq D_2 \)
      - Hence \( d(\mathcal{G}) \) is a Dynkin system hence is a \( \pi \)-system
- Let \( \{ A_i \}_{i=1}^\infty \in d(\mathcal{G}) \), we need to show that \( \bigcup_{i=1}^\infty A_i \in d(\mathcal{G}) \)
  - Define \( B_i = A_i \setminus \bigcup_{j=1}^{i-1} A_j \in d(\mathcal{G}) \forall i \)
  - Hence \( \bigcup_{i=1}^\infty A_i = \bigcup_{i=1}^\infty B_i \in d(\mathcal{G}) \)

since this is a disjoint union

hence indeed \( d(\mathcal{G}) \) is a \( \sigma \)-algebra

Theorem:
Let \((\mathcal{S}, \mathcal{F}, \mu)\) be a probability space and suppose \( \mathcal{F} = \sigma(\mathcal{G}) \) for some \( \pi \)-system \( \mathcal{G} \in \mathcal{P}(\mathcal{S}) \)
If \( \nu \) is another probability measure on \((\mathcal{S}, \mathcal{F})\) s.t.
\( \mu(A) = \nu(A) \) \( \forall A \in \mathcal{G} \)
then \( \mu = \nu \)

Proof:
write \( \mathcal{D} = \{ A \in \mathcal{F} : \mu(A) = \nu(A) \} \)

is a Dynkin system with \( \mathcal{G} \subseteq \mathcal{D} \) hence
\( d(\mathcal{G}) \subseteq \mathcal{D} \)

hence by the previous lemma \( \mathcal{F} = \sigma(\mathcal{G}) = d(\mathcal{G}) = \mathcal{D} \)

Definition: Content
\( \mu \) is a content on \((\mathcal{S}, \mathcal{A})\) if
- \( \mu(\emptyset) = 0 \)
- \( \mu(A \cup B) = \mu(A) + \mu(B) \) \( \forall A, B \in \mathcal{A} : A \cap B = \emptyset \)

Theorem Caratheodory’s Theorem
If \( \mu \) is a \( \sigma \)-additive content on \((\mathcal{S}, \mathcal{A})\) where \( \mathcal{A} \) is an algebra of \( \mathcal{S} \) then \( \mu \) extends to a measure on \((\mathcal{S}, \sigma(\mathcal{A}))\)

Lemma: Cantor
If \( \{ K_n \}_{n=1}^\infty \) are compact sets in a metric space s.t.
\( K_n \neq \emptyset \) \( \forall n \) and
\( K_{n+1} \subseteq K_n \) \( \forall n \)
then
\( \bigcap_{n=1}^\infty K_n \neq \emptyset \)

Proof:
Choose \( x_n \in K_n \) for each \( n \in \mathbb{N} \) since \( K_n \neq \emptyset \)
Notice that \( \{x_n\}_{n=r}^{\infty} \subseteq K_r \) since \( K_n \) are decreasing.

\[ \exists \{x_{kn}\}_{k=1}^{\infty}, x_0 \in K_1 \text{ s.t.} \]

\[ \lim_{k \to \infty} x_{nk} = x_0 \]

by compactness

\[ \{x_n\}_{k=r}^{\infty} \subseteq K_r \]

hence \( x_0 \in K_n \) \( \forall n \) hence \( x_0 \in \bigcap_{n=1}^{\infty} K_n \)

**Lemma:**

If \( \{A_n\}_{n=1}^{\infty} \subseteq A \) and \( \mu \) is a \( \sigma \)-additive content s.t.

- \( A = a \{Int(a, b) : -\infty \leq a \leq b \leq \infty\} \) where
  \[
  Int(a, b) = \begin{cases} 
  (a, b) & b < \infty \\
  (a, \infty) & b = \infty 
  \end{cases}
  \]

- \( A_{n+1} \subseteq A_n \)
- \( \lim_{n \to \infty} A_n = \phi \)
- \( \mu(A_n) < \infty \) \( \forall n \)

then

\[ \lim_{n \to \infty} \mu(A_n) = 0 \]

**Proof:**

By contradiction assume that \( \lim_{n \to \infty} \mu(A_n) = \delta > 0 \)

then \( \mu(A_n) \geq \delta \) \( \forall n \in \mathbb{N} \)

We can choose sets \( \{F_n\}_{n=1}^{\infty} \in A \) s.t.

- \( F_n \subseteq A_n \) for each \( n \)
- \( \mu(A_n) - \mu(F_n) \leq 2^{-n}\delta \)
- \( F_{n+1} \subseteq F_n \)

then we have that

\[ \mu(F_n) \geq \mu(A_n) - \delta 2^{-n} \geq \delta - \delta 2^{-n} \geq \delta/2 \]

hence \( F_n \neq \phi \)

moreover \( F_n \neq \phi \)

\( F_n \) is compact so by the previous lemma

\[ \bigcap_{n=1}^{\infty} F_n \neq \phi \]

but

\[ \bigcap_{n=1}^{\infty} F_n \subseteq \bigcap_{n=1}^{\infty} A_n \neq \phi \]

which is a contradiction to

\[ \bigcap_{n=1}^{\infty} A_n = \lim_{n \to \infty} A_n = \phi \]

**Proposition:**

The Lebesgue measure \( \mathcal{L} \) on \( (\mathbb{R}, A) \) where

\[ A = a \{Int(a, b) : -\infty \leq a \leq b \leq \infty\} \]

extends to a measure on \( (\mathbb{R}, \sigma(A) = (\mathbb{R}, \mathcal{B}(\mathbb{R})) \)

**Proof:**

By Carathéodory’s theorem it suffices to show that \( \mathcal{L} \) is \( \sigma \)-additive
i.e. $\forall\{A_i\}_{i=1}^{\infty} \in \mathcal{A}$ pairwise disjoint s.t. $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$ we have that

$$L\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} L(A_i)$$

For $\mathcal{A}_{\text{finite}} := \{ A \in \mathcal{A} : L(A) < \infty \}$

suppose $\{A_i\}_{i=1}^{\infty} \in \mathcal{A}_{\text{finite}}$ be pairwise disjoint s.t. $A = \bigcup_{i=1}^{\infty} A_i \in \mathcal{A}_{\text{finite}}$

Define $B_n := \bigcup_{i=1}^{n} A_i$

$L(A) = L(A \setminus B_n) + \sum_{i=1}^{n} L(A_i)$

by the previous lemma $\lim_{n \to \infty} L(A \setminus B_n) = 0$

hence

$$L\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} L(A_i)$$

so it remains to show this holds for $A \in \mathcal{A} \setminus \mathcal{A}_{\text{finite}}$

suppose $A = (-\infty, a]$ let $\{A_i\}_{i=1}^{\infty} \in \mathcal{A}$ be pairwise disjoint s.t. $\bigcup_{i=1}^{\infty} A_i = A$

Since $\{A_i\}_{i=1}^{\infty}$ are pairwise disjoint $\exists 1 A_k \in \{A_i\}_{i=1}^{\infty}$ s.t. $A_k = (-\infty, a_k]$ for some $a_k$

$L(A) \geq 0$ since it is a content hence

$$\infty = L(A_k) \leq \sum_{i=1}^{\infty} L(A_i)$$

$$\infty = L(A_k) \leq L\left(\bigcup_{i=1}^{\infty} A_i\right)$$

hence indeed the proposition holds.

**Definition: Measurable Function**

If $(\mathcal{S}, \mathcal{F}), (\mathcal{S}', \mathcal{F}')$ are measurable spaces then

$f : \mathcal{S} \to \mathcal{S}'$ is $(\mathcal{S}, \mathcal{S'})$-measurable if

$\forall A' \in \mathcal{F}'$ $f^{-1}(A') \in \mathcal{F}$

**Lemma:**

If $(\mathcal{S}, \mathcal{F}), (\mathcal{S}', \mathcal{F}')$ are measurable spaces and $\mathcal{F}' = \sigma(\mathcal{G})$, then it is sufficient that

$f^{-1}(A) \in \mathcal{F}$ $\forall A \in \mathcal{G}$

for $f$ to be measurable.

**Corollary:**

Let $(\mathcal{S}, \mathcal{F})$ is measurable and $f : \mathcal{S} \to \mathbb{R}$

then if $\{f \leq a\} \in \mathcal{F}$ $\forall a \in \mathbb{R}$ then $f$ is $(\mathcal{S}, \mathbb{R})$ measurable.

**Proof:**

By the previous lemma take $\mathcal{G} = \{(-\infty, a] : a \in \mathbb{R}\}$ which generates $\mathcal{B}(\mathbb{R})$

**Lemma:**

If $\{f_i\}_{i=1}^{\infty}$ are measurable then

- $h_1 h_2$
- $\alpha h_1 + \beta h_2$
- $h_1 \circ h_2$
- $\inf_i h_i$
- $\liminf_i h_i$
- $\limsup_i h_i$
are measurable when well defined.

**Corollary:**
Let $S$ be a topological space with $\sigma$-algebra $\mathcal{B}(S)$
If $f : S \to \mathbb{R}$ is continuous then it is Borel measurable.

**Proof:**
Open sets generate $\mathcal{B}(\mathbb{R})$
Since the preimage of open sets by a continuous function are open we have that $\forall A$ open $f^{-1}(A)$ is open hence is
in the $\sigma$-algebra generated by open sets in $S$

**Definition: Random Variable**
If $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space and $(S, F')$ is measurable then
$X : \Omega \to S$ is a random variable if it is a measurable mapping
i.e. $\forall A' \in F' \quad X^{-1}(A') \in F$

**Definition: Push Forward Measure**
If $X : \Omega \to S$ is a random variable on $(\Omega, \mathcal{F}, \mathbb{P})$ then the distribution of $X$ is the probability measure

$$\mathbb{P}_X(A) = \mathbb{P}(\{X \in A\}) \quad \forall A \in F'$$

**Definition: Cumulative Distribution Function**
Define

$$F_X(x) = \mathbb{P}_X((-\infty, x]) = \mathbb{P}(\{X \leq x\})$$

**Lemma:**
Suppose $F = F_X$ for some random variable $X$ then
- $F : \mathbb{R} \to [0, 1]$ is non decreasing
- $F$ is right-continuous
- $\lim_{x \to -\infty} F(x) = 0$
- $\lim_{x \to \infty} F(x) = 1$

**Lemma:**
If $F : \mathbb{R} \to [0, 1]$ s.t.
- $F : \mathbb{R} \to [0, 1]$ is non decreasing
- $F$ is right-continuous
- $\lim_{x \to -\infty} F(x) = 0$
- $\lim_{x \to \infty} F(x) = 1$

then $\exists (\Omega, \mathcal{F}, \mathbb{P})$ and random variable $X$ s.t. $F_X = F$

**Definition: Product $\sigma$-Algebra**
If $(S_i, \mathcal{F}_i)_{i=1}^n$ are measurable spaces then the product $\sigma$-algebra on these is

$$\bigotimes_{i=1}^n \mathcal{F}_i = \sigma(\{A_1 \times ... \times A_n : A_i \in \mathcal{F}_i\})$$

**Lemma:**
If $(\Omega, \mathcal{F}), (S_i, \mathcal{F}_i)_{i=1}^n$ are measurable spaces then
$\{X_i : \Omega \to S_i\}_{i=1}^n$ are all random variables iff
$Z := (X_1, ... X_n) : \Omega \to S_1 \times ... \times S_n$ is $(\mathcal{F}, \bigotimes_{i=1}^n \mathcal{F}_i)$-measurable

**Proof:**
Suppose $Z$ is a random variable
let $\pi_i : S_1 \times ... \times S_n \to S_i$ be the $i$th canonical projection
then $X_i = \pi_i \circ Z$
Compositions of measurable functions are measurable so it is sufficient that the projections \( \pi_i \) are measurable which is true since
\[
\pi^{-1}_i(A) = \mathbb{R} \times \ldots \times \mathbb{R} \times A \times \mathbb{R} \times \ldots \times \mathbb{R}
\]
Suppose \( \{X_i\}_{i=1}^n \) are random variables and let \( A \in \bigotimes_{i=1}^n \mathcal{F}_i \)
\[
Z^{-1}(A) = \{\omega : z(\omega) \in A\} = \{\omega : X_i(\omega) \in A_i \ \forall i\} = \bigcap_{i=1}^n X_i^{-1}(A_i) \in \mathcal{F}
\]

**Definition: Joint Distribution**
If \( \{X_i\}_{i=1}^n : (\Omega, \mathcal{F}) \to (\mathcal{S}_i, \mathcal{F}_i) \) are random variables for measurable spaces \((\Omega, \mathcal{F}), (\mathcal{S}_i, \mathcal{F}_i)_{i=1}^n\) then the distribution of \( Z = (X_1, \ldots, X_n) : (\Omega, \mathcal{F}) \to (\mathcal{S}_n \times \ldots \times \mathcal{S}_n, \bigotimes_{i=1}^n \mathcal{F}_i) \) is called the joint distribution of \((X_1, \ldots, X_n)\):
\[
\mathbb{P}_Z(A) = \mathbb{P}(\{Z \in A\}) = \mathbb{P}\left(\bigcap_{i=1}^n \{X_i \in A_i\}\right)
\]

2 Independence

**Definition: Independence**
Let \((\Omega, \mathcal{F}, \mathbb{P})\) be a probability space then
- Sub-\(\sigma\) algebras \(\{\mathcal{F}_i\}_{i=1}^n\) are independent if whenever \(\{i_k\}_{k=1}^n\) are distinct and \(A_{i_k} \in \mathcal{F}_{i_k}\) then
  \[
  \mathbb{P}\left(\bigcap_{i=1}^n A_{i_k}\right) = \prod_{i=1}^k \mathbb{P}(A_{i_k})
  \]
- Random variables \(\{X_i\}_{i=1}^n\) are independent if the sub-\(\sigma\) algebras \(\{\sigma(X_i)\}_{i=1}^n\) are independent where \(\sigma(X_i) = \sigma(\{X_i^{-1}(A) : A \in \mathcal{F}_i\})\)
- Events \(\{E_i\}_{i=1}^n \in \mathcal{F}\) are independent if the sub-\(\sigma\) algebras \(\{\sigma(E_i)\}_{i=1}^n\) are independent where \(\sigma(E_i) = \{\phi, \Omega, E_i, \Omega \setminus E_i\}\)

**Lemma:**
Suppose \(\mathcal{G}, \mathcal{H}\) are sub-\(\sigma\) algebras of \(\mathcal{F}\) on the probability space \((\Omega, \mathcal{F}, \mathbb{P})\) s.t. \(\mathcal{G} = \sigma(I), \mathcal{H} = \sigma(J)\) for \(\pi\)-systems \(I, J\) then \(\mathcal{G}, \mathcal{H}\) are independent iff
\[
\mathbb{P}(I \cap J) = \mathbb{P}(I)\mathbb{P}(J) \quad \forall I \in I, J \in J
\]

**Proof:**
Independence \(\implies\) \(\mathbb{P}(I \cap J) = \mathbb{P}(I)\mathbb{P}(J) \quad \forall I \in I, J \in J\)
is trivial since independence \(\implies\) \(\mathbb{P}(I \cap J) = \mathbb{P}(I)\mathbb{P}(J) \quad \forall I \in I \supseteq J, J \in \mathcal{H} \supseteq J\)
Define \(\mu(A) = \mathbb{P}(A \cap I), \nu(A) = \mathbb{P}(A)\mathbb{P}(I) \quad \forall I \in I, A \in \mathcal{H}\)
clearly these are measures and by the property we have that they coincide on \(\mathcal{H}\)
Define \(\eta(A) = \mathbb{P}(A \cap H), \kappa(A) = \mathbb{P}(A)\mathbb{P}(H) \quad \forall A \in \mathcal{G}, H \in \mathcal{H}\)
clearly these are measures and by the property we have that they coincide on \(\mathcal{G}\)
So indeed we have independence.

**Corollary:**
Let \((\Omega, \mathcal{F}, P)\) be a probability space and \(X, Y : \Omega \to \mathbb{R}\) be random variables. If
\[
\mathbb{P}(\{X \leq a\} \cap \{Y \leq b\}) = \mathbb{P}(X \leq a)\mathbb{P}(Y \leq b) \quad \forall a, b \in \mathbb{R}
\]
then \(X, Y\) are independent.

**Proof:**
\(\sigma(\{-\infty, a\} : a \in \mathbb{R}) = \mathcal{B}(\mathbb{R})\) hence this holds by the previous lemma.

**Corollary:**
\(X, Y\) are independent iff
\[
F_X(a)F_Y(b) = \mathbb{P}(X \leq a)\mathbb{P}(Y \leq b) = \mathbb{P}(\{X \leq a\} \cap \{Y \leq b\}) = F_{X,Y}(a,b)
\]
\(\forall a, b \in \mathbb{R}\)

**Lemma:**
Let \((\Omega_i, \mathcal{F}_i, \mu_i)_{i=1,2}\) be \(\sigma\)-finite measure spaces then
\[
\exists \mu = \mu_1 \bigotimes \mu_2\text{ which is }\sigma\text{-finite and }
\mu(A_1 \times A_2) = \mu(A_1)\mu_2(A_2) \quad \forall A_i \in \mathcal{F}_i \ i = 1,2
\]
\(\mu\) is a measure on \((\Omega_1 \times \Omega_2, \mathcal{F}_1 \bigotimes \mathcal{F}_2)\)

**Corollary:**
Random variables \(X, Y\) with joint distribution \(\mathbb{P}_{X,Y}\) are independent iff
\[
\mathbb{P}_{X,Y} = \mathbb{P}_X\mathbb{P}_Y
\]

**Definition: Tail-\(\sigma\) Algebras**
If \(\{\mathcal{F}_i\}_{i=1}^\infty\) are a sequence of \(\sigma\)-algebras then the tail-\(\sigma\) algebra is defined as
\[
\mathcal{T} = \bigcap_{n=1}^{\infty} \sigma\left(\bigcup_{k=n}^{\infty} \mathcal{F}_k\right)
\]
If \(\{X_i\}_{i=1}^\infty\) are a sequence of random variables then the tail-\(\sigma\) algebra is defined as
\[
\mathcal{T} = \bigcap_{n=1}^{\infty} \sigma\left(\bigcup_{k=n}^{\infty} \{\sigma(X_k)\}\right)
\]

**Definition: Deterministic**
A random variable \(X\) is deterministic if \(\exists c \in [-\infty, \infty] \text{ s.t. } \mathbb{P}(X = c) = 1\)

**Definition: P-Trivial**
\(\mathcal{T}\) is \(P\)-trivial if
- \(A \in \mathcal{T} \implies \mathbb{P}(A) \in \{0, 1\}\)
- If \(X\) is a \(\mathcal{T}\)-measurable random variable then \(X\) is deterministic.

**Lemma:**
Let \(\mu\) be a measure on measurable space \((\Omega, \mathcal{F})\) and \(\{A_i\}_{i=1}^\infty \in \mathcal{F}\)
- If \(A_i \subseteq A_{i+1} \forall i\) then \(\mu(\lim_{i \to \infty} A_i) = \lim_{i \to \infty} \mu(A_i)\)
- If \(A_i \supseteq A_{i+1} \forall i\) and \(\mu(A_i) < \infty\) then \(\mu(\lim_{i \to \infty} A_i) = \lim_{i \to \infty} \mu(A_i)\)
Theorem: Kolmogorov

Suppose that \( \{F_i\}_{i=1}^{\infty} \) are an independent sequence of \( \sigma \)-algebras then the associated \( \sigma \)-algebra is \( P \)-trivial

Proof:

\[ P(A) \in \{0, 1\} \iff P(A)^2 = P(A) \]

\[ \iff P(A)P(A) = P(A \cap A) \]

\[ \iff A \perp A \]

Let \( T_n = \sigma\left( \bigcup_{k=n+1}^{\infty} F_k \right) \)

and \( T = \bigcap_{n=1}^{\infty} T_n \)

\( H_n := \sigma\left( \bigcup_{k=1}^{n} F_k \right) \)

\( \{F_k\}_{k=1}^{\infty} \) are independent hence

\( H_n \perp T_n \) hence

\( H_n \perp T \)

therefore \( A \in T \) is independent from every \( H_n \)

\( \bigcup_{n=1}^{\infty} H_n \) is a \( \pi \)-system hence it is sufficient that

\[ A \perp \sigma\left( \bigcup_{n=1}^{\infty} H_n \right) = \sigma\left( \bigcup_{k=1}^{\infty} F_k \right) = T_0 \supset T \]

hence \( A \perp A \)

Let \( c \in \mathbb{R} \) by the first part of the theorem we have that \( P(X \leq c) \in \{0, 1\} \)

define \( c := \sup\{x : P(X \leq x) = 0\} \)

If \( c = -\infty \) then \( P(X = -\infty) = 1 \)

If \( c = \infty \) then \( P(X \leq \infty) = 1 \)

If \( |c| < \infty \) then

\[ P(X \leq c - \frac{1}{n}) = 0 \quad \forall n \in \mathbb{N} \]

\[ P(X \leq c + \frac{1}{n}) = 1 \quad \forall n \in \mathbb{N} \]

hence by the previous lemma

\[ P\left( \bigcup_{n=1}^{\infty} \{X \leq c - \frac{1}{n}\} \right) = P(X < c) \]

\[ = 0 \]

\[ P\left( \bigcap_{n=1}^{\infty} \{X \leq c + \frac{1}{n}\} \right) = P(X \leq c) \]

\[ = 1 \]

\[ P(X = c) = P(X \leq c) - P(X < c) \]

\[ = 1 \]

Definition: Limsup

Let \((\Omega, \mathcal{F}, \mathbb{P})\) be a probability space with \( \{A_i\}_{i=1}^{\infty} \in \mathcal{F} \) then
Definition: Liminf

Let \((Ω, F, P)\) be a probability space with \(\{A_i\}_{i=1}^{∞} \in F\) then

\[
\{A_n: \text{i.o.}\} = \bigcap_{m=1}^{∞} \bigcup_{n \geq m} A_n
\]
\[
= \{ω: \forall m; \exists n \geq m \text{ s.t. } ω \in A_n\}
\]
\[
= \liminf_n A_n
\]

Lemma: Fatou

Let \((Ω, F, µ)\) be a measure space and \(\{A_n\}_{n=1}^{∞} \in F\) then

- \(µ(\liminf_n A_n) \leq \liminf_n µ(A_n)\)
- If \(µ(Ω) < ∞\) then \(µ(\limsup_n A_n) ≥ \limsup_n µ(A_n)\)

Proof:

- Define \(Z_m := \bigcap_{n \geq m} A_n\) is monotonic and \(µ(Z_m) < µ(A_n) \quad ∀n \geq m\)

\[
µ(\liminf_n A_n) = µ\left( \bigcup_{m=1}^{∞} Z_m \right)
\]
\[
= \lim_{m \to ∞} µ(Z_m)
\]
\[
≤ \lim_{m \to ∞} \inf_{n \geq m} µ(A_n)
\]
\[
= \liminf_n µ(A_n)
\]

- \(µ(Ω) - µ(\limsup_n A_n) = µ(Ω \setminus \limsup_n A_n)\)

\[
= µ(\liminf_n (Ω \setminus A_n))
\]
\[
≤ \liminf_n µ(Ω \setminus A_n)
\]
\[
= µ(Ω) - \limsup_n µ(A_n)
\]

Since \(µ(Ω) < ∞\) we have that \(µ(\limsup_n A_n) ≥ \limsup_n µ(A_n)\)

Lemma: Borel-Cantelli

Let \((Ω, F, P)\) be a probability space and \(\{A_i\}_{i=1}^{∞} \in F\) then
If \( \sum_{n=1}^{\infty} P(A_n) < \infty \) then \( P(A_n : \text{i.o.}) = 0 \)

If \( \{A_n\}_{n=1}^{\infty} \) are independent and \( \sum_{n=1}^{\infty} P(A_n) = \infty \) then \( P(A_n : \text{i.o.}) = 1 \)

**Proof:**

- define \( G_m : \bigcup_{n \geq m} A_n \)
  then for \( k \in \mathbb{N} \) we have \( \bigcap_{m=1}^{\infty} G_m \subset G_k \)

\[
\mathbb{P}\left( \bigcap_{m=1}^{\infty} G_m \right) = \mathbb{P}(G_k) \leq \mathbb{P}\left( \bigcup A_n \right) \leq \sum_{n=k}^{\infty} \mathbb{P}(A_n)
\]

\[
\lim_{k \to \infty} \sum_{n=k}^{\infty} \mathbb{P}(A_n) = 0
\]

- \( \{A_i\}_{i=1}^{\infty} \) independent \( \implies \{A_i^c\}_{i=1}^{\infty} \) are independent.

\[
\mathbb{P}\left( \bigcap_{n=m}^{r} A_n^c \right) = \prod_{n=m}^{r} \mathbb{P}(A_n^c) \quad \text{by independence}
\]

\[
= \prod_{n=m}^{r} (1 - \mathbb{P}(A_n)) \\
\leq \prod_{n=m}^{r} e^{-\mathbb{P}(A_n)} \\
= e^{-\sum_{n=m}^{r} \mathbb{P}(A_n)}
\]

\[
\lim_{r \to \infty} e^{-\sum_{n=m}^{r} \mathbb{P}(A_n)} = 0
\]

hence we have that \( \mathbb{P}\left( \bigcap_{n=m}^{\infty} A_n^c \right) = 0 \)

\[
(limsup_n A_n)^c = \bigcup_{m=1}^{\infty} \bigcap_{n \geq m} A_n^c
\]

is the countable union of null sets hence it itself a null set hence

\[
\mathbb{P}(limsup_n A_n) = 1
\]

**Corollary:**
If \( \{A_i\}_{i=1}^{\infty} \) are independent then \( \mathbb{P}(A_i : \text{i.o.}) \in \{0, 1\} \)

**Corollary:**
If we write \( \mathcal{F}_i = \sigma(A_n) \) then \( \{A_n : \text{i.o.}\} \) is a tail event of \( \mathcal{T} = \bigcap_{i=1}^{\infty} \sigma(\{\mathcal{F}_i\}_{i=1}^{\infty}) \)

**Lemma: Existence Of Independent Random Variables**
Let \( (\Omega_i, \mathcal{F}_i, \mathbb{P}_i) \) \( i=1, 2 \) be probability spaces with respective random variables \( X_1, X_2 \) which have distributions \( \mathbb{P}_{X_1}, \mathbb{P}_{X_2} \)
If we set $\Omega = \Omega_1 \times \Omega_2$, $\mathcal{F} = \mathcal{F}_1 \otimes \mathcal{F}_2$, $\mathbb{P} = \mathbb{P}_1 \times \mathbb{P}_2$ then for $\omega = (\omega_1, \omega_2) \in \Omega$ define

$X_1(\omega) = X_1(\omega_1)$

$X_2(\omega) = X_2(\omega_2)$

are defined on $(\Omega, \mathcal{F}, \mathbb{P})$ and are independent.

3 Expectation And Convergence

Definition: Simple Function

A function $f : S \to \mathbb{R}$ on measure space $(S, \mathcal{F}, \mu)$ is simple if

$\exists \{\alpha_i\}_{i=1}^n \in \mathbb{R}$ and partition $\{A_i\}_{i=1}^n \in \mathcal{F}$ s.t.

$f = \sum_{i=1}^n \alpha_i 1_{A_i}$

Definition: Integration Of Simple Functions

If $f : S \to \mathbb{R}$ is simple with $f = \sum_{i=1}^n \alpha_i 1_{A_i}$ then

$\int f d\mu = \sum_{i=1}^n \alpha_i \mu(A_i)$

Definition: Integration Of Positive Measurable Functions

If $f : S \to \mathbb{R}_{[0, \infty]}$ is measurable then

$\int f d\mu = \sup \{ \int g d\mu : g \leq f, g \text{ simple} \}$

Definition: $\mu$-Integrable

$f : S \to \mathbb{R}$ is $\mu$-integrable if

$\int |f| d\mu < \infty$

Definition: Integration Of $\mu$-Integrable Functions

If $f : S \to \mathbb{R}$ is $\mu$-integrable then

$\int f d\mu = \int f^+ d\mu - \int f^- d\mu$

Remark

We denote the space of $\mu$-integrable functions as $L^1(\mu) = \mathcal{L}(S, \mathcal{F}, \mu)$

Lemma:

The following properties hold for $L^1(\mu)$

- $f, g \in L^1(\mu)$, $\alpha, \beta \in \mathbb{R}$ $\implies$

  $\int \alpha f + \beta g d\mu = \alpha \int f d\mu + \beta \int g d\mu$

- $0 \leq f \in L^1(\mu)$ $\implies$

  $\int f d\mu \geq 0$

- $\{f_n\}_{n=1}^\infty$, measurable s.t. $\lim_{n \to \infty} f_n = f$ and $f_n \leq f_{n+1}$ $\forall n$ $\implies$

  $\lim_{n \to \infty} \int f_n d\mu = \int f d\mu$

- $f = 0 \text{ a.e.}$ $\iff$

  $\int f d\mu = 0$
Theorem: Fubini’s Theorem
Let \((S_1 \times S_2, \mathcal{F}_1 \otimes \mathcal{F}_2, \mu_1 \otimes \mu_2)\) be a measure space and \(f \in \mathcal{L}^1(\mu_1 \otimes \mu_2)\) then let \(
abla = \mu_1 \otimes \mu_2\)
\[
\int \int f(s_1, s_2) d\mu_1(s_1) d\mu_2(s_2) = \int f(s) d\mu(s) = \int \int f(s_1, s_2) d\mu_2(s_2) d\mu_1(s_1)
\]
Moreover it is sufficient that \(f \geq 0\) measurable for the above integral inequality to hold.

Definition: Expectation
If \(X\) is a r.v. on \((\Omega, \mathcal{F}, \mathbb{P})\) probability space \(X \in \mathcal{L}^1(\mathbb{P})\) then
\[
\mathbb{E}[X] := \int X d\mathbb{P}
\]

Lemma:
If \(X\) is a r.v. on \((\Omega, \mathcal{F}, \mathbb{P})\) with distribution \(\mathbb{P}_X\) and \(h : \mathbb{R} \rightarrow \mathbb{R}\) is Borel measurable then
- \(h \circ X \in \mathcal{L}^1(\mathbb{P}) \iff h \in \mathcal{L}^1(\mathbb{P}_X)\)
- \(\mathbb{E}[h \circ X] = \int h \circ X d\mathbb{P} = \int h d\mathbb{P}_X\)

Corollary:
\[
\mathbb{E}[X] = \int z d\mathbb{P}_X(z) = \int X d\mathbb{P}
\]

Corollary:
Take \(h = 1_B\) for \(B \in \mathcal{B}(\mathbb{R})\) then
\[
\mathbb{E}[1_B \circ X d\mathbb{P}] = \int 1_B \circ X d\mathbb{P}
\]
\[
= \int 1_{X^{-1}(B)} d\mathbb{P}
\]
\[
= \mathbb{P}(X^{-1}(B))
\]
\[
= \mathbb{P}(X \in B)
\]
\[
= \mathbb{P}_X(B)
\]
\[
= \int 1_B d\mathbb{P}_X
\]
This holds for all indicators hence by linearity holds for all simple functions therefore by monotone convergence theorem holds for all positive measurable functions and all integrable functions.

Lemma:
If \(X, Y\) are independent r.v.s on probability space \((\Omega, \mathcal{F}, \mathbb{P})\) then
\[
\mathbb{E}[XY] = \mathbb{E}[X] \mathbb{E}[Y]
\]

Proof:
\[
\mathbb{E}[XY] = \int xy d\mathbb{P}_{X,Y}
\]
\[
= \int \int XY d\mathbb{P}_X d\mathbb{P}_Y
\]
\[
= \mathbb{E}[Y] \int X d\mathbb{P}_X d\mathbb{P}_Y
\]
\[
= \int X d\mathbb{P}_X \int Y d\mathbb{P}_Y
\]
\[
= \mathbb{E}[X] \mathbb{E}[Y]
\]

by Fubini
Remark:
If $X$ is a r.v on $(\Omega, \mathcal{F}, P)$ and $A \in \mathcal{F}$ then we use the following notation:

$$E[X; A] := E[X1_A] = \int X1_A dP$$

Lemma: Markov’s Inequality
If $X$ is a random variable, $\epsilon > 0$ and $g : \mathbb{R} \to \mathbb{R}_{[0, \infty]}$ is Borel measurable and non-decreasing then

- $P(g(X) > \epsilon) \leq E[g(X)]$
- $P(X > \epsilon) g(\epsilon) \leq E[g(X)]$

Proof:
$g \circ X$ is measurable and non-negative by properties of measurable functions hence

- \[
E[g(X)] = \int g(X) dP \\
\geq \int g(X)1_{\{g(X) > \epsilon\}} dP \\
\geq \int \epsilon 1_{\{g(X) > \epsilon\}} dP \\
\geq \epsilon \int 1_{\{g(X) > \epsilon\}} dP \\
= \epsilon P(g(X) > \epsilon)
\]

- \[
E[g(X)] = \int g(X) dP \\
\geq \int g(X)1_{\{X > \epsilon\}} dP \\
\geq \int g(\epsilon) 1_{\{X > \epsilon\}} dP \\
\geq g(\epsilon) \int 1_{\{X > \epsilon\}} dP \\
= g(\epsilon) P(X > \epsilon)
\]

Corollary: Chebyshev’s Inequality
If $X$ is a r.v. with finite variance then

$$P(|X - E[X]| > \epsilon) \leq \frac{Var[X]}{\epsilon^2}$$

$\forall \epsilon > 0$

Proof:
$$P(|X - E[X]| > \epsilon) = P((X - E[X])^2 > \epsilon^2)$$
$$\leq \frac{E[(X - E[X])^2]}{\epsilon^2}$$
$$= \frac{Var[X]}{\epsilon^2}$$
**Theorem:**
If \( \{X_i\}_{i=1}^n \) are independent random variables each with mean \( \mu \) and variance \( \sigma^2 \) then for \( \varepsilon > 0 \)

\[
P\left( \left| \frac{1}{n} \sum_{i=1}^n X_i - \mu \right| > \varepsilon \right) \leq \frac{\sigma^2}{n\varepsilon^2}
\]

**Proof:**
define \( Z = \sum_{i=1}^n X_i \) then
\[
E[Z] = \mu \quad \text{Var}[Z] = \sigma^2/n
\]
hence by Chebyshev’s inequality this clearly holds.

**Lemma: Jensen’s Inequality**
If \( X \) is a random variable with \( E[|X|] < \infty \) s.t. \( f \) is convex on an interval containing the range of \( X \) and \( E[|f \circ X|] < \infty \) then

\[
f(E[X]) \leq E[f(X)]
\]

**Proof:**
Suppose that \( X \) is simple so let \( X = \sum_{i=1}^n a_i 1_{A_i} \)

\[
f(E[X]) = f \left( \sum_{i=1}^\infty a_i P(A_i) \right)
\]

\[
\leq \sum_{i=1}^\infty P(A_i) f(a_i) \quad \text{by convexity}
\]

\[
= E[f(X)]
\]

Suppose that \( X \) is a positive r.v. then \( \exists \{X_n\}_{n=1}^\infty \) r.vs which are increasing and converge to \( X \) a.s.
then

\[
f(E[X]) = \lim_{n \to \infty} f(E[X_n])
\]

\[
\leq \lim_{n \to \infty} E[f(X_n)]
\]

\[
= E[f(X)]
\]

### 4 Convergence

**Definition: Almost Sure Convergance**
If \( \{X_i\}_{i=1}^n \) are a sequence of r.vs on \((\Omega, \mathcal{F}, P)\) then \( \{X_i\}_{i=1}^n \) converge to r.v. \( X \) on \((\Omega, \mathcal{F}, P)\) if

\[
P(\{\omega \in \Omega : \lim_{n \to \infty} X_n(\omega) = X(\omega)\}) = 1
\]

**Remark:**
Almost sure limits are unique only up to equality a.s.

**Corollary:**
\( X = Y \) a.s. iff

\[
P(\{\omega \in \Omega : X(\omega) = Y(\omega)\}) = 1
\]

**Lemma:**
\( \{X_n\}_{n=1}^\infty \) converge to \( X \) a.s. iff

\[
P(|X_n - X| > \varepsilon \ i.o.) = 0 \quad \forall \varepsilon > 0, \ n \in \mathbb{N}
\]

**Corollary: Stability Properties**
- If \( \{X_n\}_{n=1}^\infty \) are r.vs s.t. \( X_n \to X \) a.s. and \( g : \mathbb{R} \to \mathbb{R} \) is continuous then \( g(X_n) \to g(X) \) a.s.
Definition: Convergence In Probability
If \( \{X_n\}_{n=1}^{\infty} \), \( \{Y_n\}_{n=1}^{\infty} \) are r.vs s.t. \( X_n \rightarrow X \) a.s. and \( Y_n \rightarrow Y \) a.s. with \( \alpha, \beta \in \mathbb{R} \) then \( \alpha X_n + \beta Y_n \rightarrow \alpha X + \beta Y \).

Definition: Cauchy Convergence In Probability
If \( \{X_n\}_{n=1}^{\infty} \) are r.vs on \((\Omega, \mathcal{F}, \mathbb{P})\) then \( X_n \rightarrow X \) in probability if
\[
\lim_{n \to \infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0
\]

Lemma:
Let \( \{X_n\}_{n=1}^{\infty} \), \( \{Y_n\}_{n=1}^{\infty} \) be r.vs on \((\Omega, \mathcal{F}, \mathbb{P})\)
If \( X_n \rightarrow \infty \) a.s. then \( X_n \rightarrow X \) in probability.

Proof:
\( X_n \rightarrow \infty \) a.s. implies
\[
\forall \varepsilon > 0 \quad \mathbb{P}(|X_n - X| > \varepsilon) \leq \mathbb{P}(\sup_{k \geq n}|X_k - X| > \varepsilon)
\]
hence we have
\[
\lim_{n \to \infty} \mathbb{P}(|X_n - X| > \varepsilon) \leq \lim_{n \to \infty} \mathbb{P}(\sup_{k \geq n}|X_k - X| > \varepsilon)
\]
\[
\leq \mathbb{P}(lim_{n \to \infty}\sup_{k \geq n}|X_k - X| > \varepsilon)
\]
\[
= \mathbb{P}(|X_n - X| > \varepsilon \text{ i.o.})
\]
\[
= 0
\]

Lemma:
Let \( \{X_n\}_{n=1}^{\infty} \) be r.vs on \((\Omega, \mathcal{F}, \mathbb{P})\)
\( X_n \rightarrow X \) in probability iff \( X_n \rightarrow X \) Cauchy in probability.

Proof:
Suppose \( X_n \rightarrow X \) in probability.
then
\[
|X_n - X_m| \leq |X_n - X| + |X - X_m|
\]
\[
\{|X_n - X_m| > \varepsilon\} \subseteq \{|X_n - X| > \varepsilon/2\} \cup \{|X - X_m| > \varepsilon/2\}
\]
\[
\mathbb{P}(|X_n - X_m| > \varepsilon) \leq \mathbb{P}(|X_n - X| > \varepsilon/2) + \mathbb{P}(|X - X_m| > \varepsilon/2)
\]
\[
\lim_{n,m \to \infty} \mathbb{P}(|X_n - X_m| > \varepsilon) \leq \lim_{n,m \to \infty} \mathbb{P}(|X_n - X| > \varepsilon/2) + \mathbb{P}(|X - X_m| > \varepsilon/2)
\]
\[
= 0
\]

Now suppose \( X_n \rightarrow X \) Cauchy in probability
for any \( k \geq 1 \) choose \( \varepsilon = 2^{-k} \)
\( \exists m_k \text{ s.t. } \forall n > m \geq m_k \) we have that
\[
\mathbb{P}(|X_n - X_m| > 2^{-k}) < 2^{-k}
\]
set \( n_1 = m_1 \) and recursively define \( n_{k+1} = \max\{n_k + 1, m_{k+1}\} \)
define \( X'_k := X_{n_k} \)
define \( A_k := \{|X'_k - X'_k| > 2^{-k}\} \)
then \( \sum_{k=1}^{\infty} \mathbb{P}(A_k) < \infty \)
then by Borel Cantelli
\[
|X'_{k+1}(\omega) - X'_k(\omega)| \leq 2^{-k} \quad \forall \omega \in \Omega \setminus A, \quad \forall k \geq K_0
\]
where \( \mathbb{P}(A) = 0 \)
hence \( \forall \omega \in \Omega \setminus A \)
\[ \lim_{n \to \infty} \sup_{m > n} |X'_m - X'_n| \leq \lim_{n \to \infty} \sum_{k=n}^{\infty} |X'_m - X'_n| \]
\[ \leq \lim_{n \to \infty} \sum_{k=n}^{\infty} 2^{-k} \]
\[ = 0 \]

hence
\[ \mathbb{P}( \limsup_{n} X'_n = \liminf_{n} X'_n < \infty ) = 1 \]

define \( X := \liminf_{n} X'_n \) since \( X_n \to X \) a.s.
by the previous lemma we have that \( X_n \to X \) in probability
hence \( \forall \varepsilon > 0 \)
\[ \lim_{k \to \infty} \mathbb{P}(|X_k - X| > \varepsilon) \leq \lim_{k \to \infty} (\mathbb{P}(|X_k - X_{n_k}| > \varepsilon/2) + \mathbb{P}(|X_{n_k} - X| > \varepsilon/2)) \]
\[ = 0 \]

**Lemma:**
Let \( \{X_n\}_{n=1}^{\infty}, X \) be random variables on \((\Omega, \mathcal{F}, \mathbb{P})\)
If \( \sum_{n=1}^{\infty} \mathbb{P}(|X_n - X| > \varepsilon) < \infty \quad \forall \varepsilon > 0 \) then
\( X_n \to X \) a.s.

**Proof:**
This follows directly from Borel Cantelli.

**Lemma:**
If \( \{X_n\}_{n=1}^{\infty}, X \) are r.vs then
\( X_n \to X \) in probability iff
\( \forall \{X_{n_{k_r}}\}_{k_r=1}^{\infty} \exists \{X_{n_{kr}}\}_{r=1}^{\infty} \) s.t. \( X_{n_{kr}} \to X \) a.s.

**Proof:**
Suppose \( X_n \to X \) in probability
the \( \forall \varepsilon > 0 \) find \( n_{k_r} \in \mathbb{N} \) s.t. \( \mathbb{P}(|X_{n_{k_r}} - X| > 1/r) < 2^{-r} \)
then
\[ \sum_{r=1}^{\infty} \mathbb{P}(|X_{n_{k_r}} - X| > 1/r) < \sum_{r=1}^{\infty} 2^{-r} < \infty \]
moreover \( \forall \varepsilon > 0 \)
\[ \sum_{r > 1/\varepsilon} \mathbb{P}(|X_{n_{k_r}} - X| > \varepsilon) < \sum_{r > 1/\varepsilon} \mathbb{P}(|X_{n_{k_r}} - X| > 1/r) \]
\[ \leq \sum_{r > 1/\varepsilon} 2^{-k} \]
\[ < \infty \]

hence we simply take the sequence \( \{X_{n_{k_r}}\}_{r > 1/\varepsilon} \)
which converges to \( X \) a.s.
Assume that every subsequence has a subsequence converging a.s.
we need to show that \( X \) is unique.
fix \( \varepsilon > 0 \)
\( b_n := \mathbb{P}(|X_n - X| > \varepsilon) \in [0, 1] \) is compact hence
∃b_{n_k} converging subsequence
by our assumption X_{n_k} has a subsequence X_{n_{k_r}} → X a.s.

$$\lim_{r \to \infty} b_{n_{k_r}} = \lim_{r \to \infty} P(|X_{n_{k_r}} - X| > \varepsilon) = 0$$

since a.s. convergence implies convergence in probability.

hence since b_{n_k} is convergent we must have that \( \lim_{k \to \infty} b_{n_k} = 0 \)
hence every subsequence of b_k which converges does so to 0

**Theorem: The Weak Law Of Large Numbers**
If \( \{X_n\}_{n=1}^\infty \) are i.i.d r.v.s with \( X_i \in L^1(\Omega, \mathcal{F}, \mathbb{P}) \) then

$$\lim_{n \to \infty} \frac{\sum_{k=1}^n X_k}{n} = \mathbb{E}[X_1]$$

**Proof:**
Write

$$X_k^* := X_k \mathbb{1}_{|X_k| \leq k^{1/4}}$$
$$X_k' := X_k \mathbb{1}_{|X_k| > k^{1/4}}$$
$$Y_n^* := \frac{\sum_{k=1}^n X_k^* - \mathbb{E}[X_k^*]}{n}$$
$$Y_n' := \frac{\sum_{k=1}^n X_k' - \mathbb{E}[X_k']}{n}$$

It suffices to show that \( Y_n^*, Y_n' \to 0 \) in probability.

$$\text{Var}[X_k^*] \leq \mathbb{E}[X_k^*]^2 \leq k^{1/2}$$
$$\text{Var}[Y_n^*] \leq \sum_{k=1}^n \frac{\text{Var}[X_k^*]}{n^2} \leq \sum_{k=1}^n \frac{k^{1/2}}{n^2} \leq \sum_{k=1}^n \frac{n^{1/2}}{n^2} = n^{-1/2}$$

by Chebyshev’s inequality

$$\mathbb{P}(|Y_n^*| > \varepsilon) \leq \frac{\text{Var}[Y_n^*]}{\varepsilon^2}$$

but

$$\lim_{n \to \infty} \frac{\text{Var}[Y_n^*]}{\varepsilon^2} = 0$$

hence \( Y_n^* \to 0 \) in probability

$$\mathbb{E}[|X_k'|] = \mathbb{E}[|X_1'| 1_{|X_1'| > k^{1/4}}] \to 0$$

by MCT
hence
\[ P(|Y'_n| \geq \varepsilon) \leq \frac{E[|Y'_n|]}{\varepsilon} \leq \frac{2}{\varepsilon} \sum_{k=1}^{n} E[|X'_k|] \]
\[ \lim_{n \to \infty} P(|Y'_n| \geq \varepsilon) = 0 \]

hence \( Y'_n \to 0 \) in probability.

**Theorem: The Strong Law Of Large Numbers**

Let \( \{X_n\}_{n=1}^{\infty} \) be independent r.vs s.t.

- \( E[|X_n|^2] < \infty \quad \forall n \)
- \( v := \sup_{X_n} \{Var[X_n]\} < \infty \)

then
\[ \sum_{k=1}^{n} \frac{X_k - E[X_k]}{n} \to 0 \quad a.s. \]

**Proof:**

WLOG assume \( E[X_k] = 0 \)

Define \( S_n = \sum_{i=1}^{n} X_i \)

Notice that \( S_n \to 0 \) in probability so we want a subsequence \( S_{n_k} \to 0 \) a.s.

\[ P(|S_{n_k}| > \varepsilon) \leq \frac{Var[S_{n_k}]}{n^2 \varepsilon^2} \]

hence
\[ \sum_{n=1}^{\infty} P(|S_n| > \varepsilon) < \infty \quad \forall \varepsilon > 0 \]

so we have that \( S_{n_k} \to 0 \) a.s.

For \( m \in \mathbb{N} \) let \( n = n_m \) s.t.

\( n^2 \leq m \leq (n+1)^2 \)

write \( y_k = ks_k = \sum_{i=1}^{k} X_i \)

\[ P(|y_m - y_{n_k^2}| > \varepsilon n^2) \leq \varepsilon^{-2} n^{-4} Var\left[ \sum_{i=n^2+1}^{m} X_i \right] \]
\[ \leq \varepsilon^{-2} n^{-4} v(m-n^2) \]
\[ \sum_{m=1}^{\infty} P(|y_m - y_{n_k^2}| > \varepsilon n^2) \leq \frac{v}{\varepsilon^2} \sum_{n=1}^{\infty} \sum_{m=n^2+1}^{(n+1)^2-1} \frac{m-n^2}{n^4} \]
\[ \leq \frac{v}{\varepsilon^2} \sum_{n=1}^{\infty} n = 1^\infty \sum_{k=1}^{\infty} \frac{2n}{n^4} \]
\[ = \frac{v}{\varepsilon^2} \sum_{n=1}^{\infty} \frac{2n(2n+1)}{2n^4} \]
\[ \leq \frac{v}{\varepsilon^2} \sum_{n=1}^{\infty} \frac{2}{n^2} \]
\[ < \infty \]

hence since the intersection of two sets of full measure is also a set of full measure (where finite)
we get that
$$|S_m| = |y_m/m| \leq |y_m/n^2| \to 0$$
a.s.

**Definition: Absolute Moment**
Let $X$ be a random variable on probability space $(\Omega, \mathcal{F}, \mathbb{P})$
then for $p \in [1, \infty)$ the $p$th moment of $X$ is defined as
$$E[|X|^p] = \int |X|^p d\mathbb{P}$$

**Definition: $L^p$ Space**
$$L^p(\Omega, \mathcal{F}, \mathbb{P}) := \{ X : \Omega \to \mathbb{R} : E[|X|^p] < \infty, \ X \text{ measurable} \}$$

**Definition: $L^p$ Norm**
The $L^p$ norm is defined as
$$||X||_p := E[|X|^p]^{1/p}$$

**Lemma: Holder’s Inequality**
For $p, q \in [1, \infty)$ s.t. $\frac{1}{p} + \frac{1}{q} = 1$
we have that for $X \in L^p$, $Y \in L^q$
$$E[|XY|] \leq ||X||_p ||Y||_q$$
m moreover this holds for $p = 1, q = \infty$ where
$$||Y||_\infty := \inf \{ s \geq 0 : \mathbb{P}(|Y| > s) = 0 \}$$

**Lemma:**
For $p \in [1, \infty]$, $X, Y \in L^p$ we have
$$||X + Y||_p \leq ||X||_p + ||Y||_p$$

**Corollary:**
$$||X||_p = 0 \implies X = 0 \ a.s.$$
for $p \in [1, \infty]$ let $\{X_n\}_{n=1}^{\infty} \in L^p$ be a Cauchy sequence in $L^p$ then

$\exists X \in L^p$ s.t. $X_n \to X$ in $L^p$

**Lemma:**
If $1 \leq s \leq r$ then

$$X_n \to X \text{ in } L^s \implies X_n \to X \text{ in } L^r$$

**Proof:**
For $s = 1, r = 2$

$$E[|X_n - X|^s] = \int |X_n - X|d\mathbb{P}$$

$$= \int |X_n - X|d\mathbb{P}$$

$$\leq \left( \int |X_n - X|^2d\mathbb{P} \right)^{1/2} \left( \int 1^2d\mathbb{P} \right)^{1/2}$$

$$= E[|X_n - X|^2]^{1/2}$$

$$\lim_{n \to \infty} E[|X_n - X|^2] = 0 \implies \lim_{n \to \infty} E[|X_n - X|^2]^{1/2} = 0$$

hence indeed the lemma holds for $s = 1, r = 2$

**Lemma:**
Let $\{X_n\}_{n=1}^{\infty}, X$ be r.v.s s.t. $X_n \to X$ in $L^p$ for $p \geq 1$ then

$X_n \to X$ in probability.

**Proof:**
let $\varepsilon > 0$

$$P(|X_n - X| > \varepsilon) = P(|X_n - X|^p > \varepsilon^p)$$

$$\leq \frac{E[|X_n - X|^p]}{\varepsilon^p}$$

$$\lim_{n \to \infty} \frac{E[|X_n - X|^p]}{\varepsilon^p} = 0$$

by Markov’s inequality

by convergence in $L^p$

**Lemma:**
$X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ iff

$$\lim_{k \to \infty} \int_{\{|X| > k\}} |X|d\mathbb{P} = 0$$

**Definition: Uniformly Integrable**
A set of random variables $C$ on $(\Omega, \mathcal{F}, \mathbb{P})$ is uniformly integrable if

$$\lim_{k \to \infty} \sup_{X \in C} \int_{\{|X| > k\}} |X|d\mathbb{P} = 0$$

**Theorem:**
If $X, \{X_n\}_{n=1}^{\infty} \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ then

$X_n \to X$ in $L^1$ iff

- $X_n \to X$ in probability
- $\{X_n\}_{n=1}^{\infty}$ is uniformly integrable.
**Proof:**

$X_n \to X$ in $\mathcal{L}^1 \implies X_n \to X$ in probability by a previous lemma.

Moreover if $X_n \to X$ in $\mathcal{L}^1$ we have finite integrability since otherwise $\mathbb{P}(X = \infty) > 0$ hence the expectation cannot be finite.

Assume that $X_n \to X$ in probability and $\{X_n\}_{n=1}^\infty$ is uniformly integrable.

For $k > 0$ define

$$\varphi_k(x) = \begin{cases} k & x > k \\ x & x \in [-k, k] \\ -k & x < -k \end{cases}$$

Then we have that

$$\mathbb{E}[|X_n - X|] \leq \mathbb{E}[|X_n - \varphi_k(X_n)|] + \mathbb{E}[|\varphi_k(X_n) - \varphi_k(X)|] + \mathbb{E}[|\varphi_k(X) - X|]$$

Notice that

$$|\varphi_k(X_n) - X_n| = 1_{\{|X_n| > k\}}|k - |X_n|| \leq 2|X_n|1_{\{|X_n| > k\}}$$

Hence let $\varepsilon > 0$ then by uniform integrability we have that

$$\exists k_1 \in [0, \infty) \text{ s.t. } \forall k \geq k_1 \quad \sup_n \int_{\{|X_n| > k\}} |X_n|d\mathbb{P} \leq \varepsilon/6$$

Hence $\forall n$ we have that $k > k_1 \implies$

$$\mathbb{E}[|X_n - \varphi_k(X_n)|] \leq \sup_n 2\int_{\{|X_n| > k\}} |X_n|d\mathbb{P} \leq \varepsilon/3$$

Since $X \in \mathcal{L}^1$ we have that

$$\lim_{k \to \infty} \int_{\{|X| > k\}} |X|d\mathbb{P} = 0$$

Hence $\exists k_2 \in [0, \infty) \text{ s.t. } \forall k > k_2 \quad \mathbb{E}[|X - \varphi_k(X)|] \leq \varepsilon/3$

Set $k_0 = \min\{k_1, k_2\}$

$$\mathbb{P}(\varphi_{k_0}(X_n) - \varphi_{k_0}(X) > \varepsilon) \leq \mathbb{P}(X_n - X > \varepsilon)$$

Hence since $X_n \to X$ in probability we have that $\varphi_k(X_n) \to \varphi_k(X)$ in probability so by the Dominated Convergence Theorem $|\varphi_{k_0}| \leq k_0$ so

$\exists N \in \mathbb{N}$ s.t. $\forall n \geq N$ we have that

$$\mathbb{E}[|\varphi_k(X_n) - \varphi(X)|] < \varepsilon/3$$

Hence indeed we have that for $k = k_0$ and $n \geq N$ we have that

$$\mathbb{E}[|X_n - X|] \leq \mathbb{E}[|X_n - \varphi_k(X_n)|] + \mathbb{E}[|\varphi_k(X_n) - \varphi_k(X)|] + \mathbb{E}[|\varphi_k(X) - X|] \leq \varepsilon$$

**Definition: Probability Set**

We define $\text{Prob}(\mathbb{R})$ to be the set of probability measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$

**Definition: Continuous Bounded Functions**

We define $C_b(\mathbb{R})$ to be the set of continuous bounded functions $f : \mathbb{R} \to \mathbb{R}$

**Definition: Weak Convergence**

- If $\mu, \{\mu_n\}_{n=1}^\infty \in \text{Prob}(\mathbb{R})$ then $\mu_n \to \mu$ weakly if
  $$\lim_{n \to \infty} \int f d\mu_n = \int f d\mu \quad \forall f \in C_b(\mathbb{R})$$

- If $F, \{F_n\}_{n=1}^\infty$ are distribution functions then $F_n \to F$ weakly if $\mu_n \to \mu$ weakly
  
  where $\mu, \{\mu_n\}_{n=1}^\infty$ are the associated probability measures
  $$F(x) = \mu((-\infty, x])$$
• If $X, \{X_n\}_{n=1}^\infty$ are random variables then
  $X_n \to X$ weakly if $\mathbb{P}_{X_n} \to \mathbb{P}_X$ weakly
  where $\mathbb{P}_X(A) = \mathbb{P}(X \in A)$

Definition: Convergence In Distribution

• If $F, \{F_n\}_{n=1}^\infty$ are distribution functions then
  $F_n \to F$ in distribution if for all continuous points $x$ of $F$
  $$\lim_{n \to \infty} F_n(x) = F(x)$$

• If $X, \{X_n\}_{n=1}^\infty$ are random variables then
  $X_n \to X$ in distribution if
  $$\mathbb{P}_{F_n} \to \mathbb{P}_F$$
  where $\mathbb{P}_F(x) = \mathbb{P}(X \leq x)$

Theorem:
If $\{F_n\}_{n=1}^\infty, F$ are distribution functions of probability measures $\{\mu_n\}_{n=1}^\infty \mu$ then
$F_n \to F$ weakly iff $F_n \to F$ in distribution.

Proof:
Suppose $F_n \to F$ weakly.

fix $x \in \mathbb{R}$ s.t. $F$ is continuous at $x$ and let $\delta > 0$ define
$$h(y) = \begin{cases} 1 & y \leq x \\ 1 - \frac{y - x}{\delta} & y \in (x, x + \delta) \\ 0 & y \geq x + \delta \end{cases}$$
and
$$g(y) = \begin{cases} 1 & y \leq x - \delta \\ 1 - \frac{y - x}{\delta} & y \in (x - \delta, x) \\ 0 & y \geq x \end{cases}$$
then since $F_n(x) = \int 1_{(-\infty, x]} d\mu_n$ we have that
$$\int g d\mu_n \leq F(x) \leq \int h d\mu_n$$
$$\limsup_n F_n(x) \leq \limsup_n \int h d\mu_n = \int h d\mu \leq F(x + \delta)$$
$$\liminf_n F_n(x) \geq \liminf_n \int g d\mu_n = \int g d\mu \geq F(x - \delta)$$
by continuity of $h, g$

By continuity of $F$ at $x$ we have that
$$\lim_{\delta \to 0} F(x + \delta) = F(x) = \lim_{\delta \to 0} F(x - \delta)$$
hence
$$\lim_{n \to \infty} F_n(x) = F(x)$$
hence $F_n \to F$ in distribution.

Suppose $F_n \to F$ in distribution.

then
$$\lim_{n \to \infty} \int 1_{(-\infty, x]} d\mu_n = \int 1_{(-\infty, x]} d\mu$$
$\forall x \in \mathbb{R}$

Since any continuous bounded function can be approximated by sums of indicator functions and linearity of the integral we indeed have that $F_n \to F$ weakly

Lemma:
Let $\{X_i\}_{i=1}^\infty, X$ be random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ then if $X_n \to X$ in probability then $X_n \to X$ weakly

Proof:
There are two different proofs of this
• Since $X_n \rightarrow X$ in probability and $f$ is continuous we have that $f(X_n) \rightarrow f(X)$ in probability. $f(X_n)$ is bounded by the properties of $f$ hence by the dominated convergence theorem

$$
\lim_{n \rightarrow \infty} \int f(X_n) dP = \int f(X) dP
$$

hence indeed $X_n \rightarrow X$ weakly.

• Since $f$ is bounded we know that $\exists k > 0$ s.t.

$$
l_k = \sup_{x \in \mathbb{R}} |f(x)| 
$$

hence by the dominated convergance theorem

$$
\lim_{n \rightarrow \infty} \int f(X_n) dP = \int f(X) dP
$$

so $f(X_n)$ is uniformly integrable.

Since $f$ is continuous and $X_n \rightarrow X$ in probability we have that $f(X_n) \rightarrow f(X)$ in probability. This along with uniform integrability implies that $f(X_n) \rightarrow f(X)$ in $L^1$ i.e. $\lim_{n \rightarrow \infty} E[|f(X_n) - f(X)|] = 0$

hence indeed we have weak convergance.

Corollary:
Let $\{X_i\}_{i=1}^{\infty}, X$ be random variables on $(\Omega, F, P)$ then if $X_n \rightarrow X$ a.s. then $X_n \rightarrow X$ weakly.

**Definition: Conditional Probability**

If $(\Omega, F, P)$ is a probability space with $A, B \in F$ then the conditional probability of $B$ given $A$ has occurred (where $P(A) > 0$ is

$$
P(B|A) = \frac{P(A \cap B)}{P(A)}
$$

**Definition: Conditional Expectation**

If $(\Omega, F, P)$ is a probability space with $A \in F$ with strictly positive probability then the conditional expectation of r.v $X$ given $A$ has occurred is

$$
E[X|A] = \frac{1}{P(A)} \int_A X dP
$$

Moreover if $F_0 = \sigma(G)$ for a countable partition $G = \{A_i\}_{i=1}^{\infty}$ of measurable sets on $\Omega$ then

$$
E[X|F_0] = \sum_{i: P(A_i) > 0} \frac{1_{A_i}}{P(A_i)} \int_{A_i} X dP
$$

is a random variable.

**Theorem:**
Let $X \in L^1(\Omega, F, P), F_0 = \sigma(G)$ s.t. $G$ is a countable partition $\{A_i\}_{i=1}^{\infty} \in \Omega$ Then $E[X|F_0]$ has the following properties:

• $E[X|F_0]$ is $F_0$ measurable.

• $\forall A \in F_0$ we have that

$$
E[X1_A] = E[1_A E[X|F_0]]
$$

**Proof:**

• This is trivial since $\frac{E[X|A]}{P(A)}$ is constant, $1_{A_i}$ are measurable since $A_i \in F_0$ and the countable sum of measurable functions is measurable.

• Let $A \in G$ with $P(A) > 0$
\[
\begin{align*}
\mathbb{E}[1_A \mathbb{E}[X | \mathcal{F}_0]] &= \mathbb{E} \left[ 1_A \sum_{i : \mathbb{P}(A_i) > 0} \frac{1_{A_i}}{\mathbb{P}(A_i)} \mathbb{E}[X; A_i] \right] \\
&= \mathbb{E} \left[ \frac{1_A}{\mathbb{P}(A)} \mathbb{E}[X; A] \right] \\
&= \frac{\mathbb{E}[1_A \mathbb{E}[X; A]]}{\mathbb{P}(A)} \\
&= \frac{\mathbb{E}[X; A]}{\mathbb{P}(A)} \\
&= \mathbb{E}[X 1_A]
\end{align*}
\]

This extends to \( A \in \mathcal{F}_0 \) by standard operations of elements of \( \sigma \)-algebras. If \( \mathbb{P}(A) = 0 \) then \( X 1_A, \mathbb{E}[X | \mathcal{F}_0] 1_A = 0 \) a.s. hence both sides of the equation are null.

**Corollary:**
Taking \( y = 1 \) in the second property we have that
\[
\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X | \mathcal{F}_0]]
\]

**Definition: Version Of Conditional Expectation**
A random variable \( X_0 \in L^1(\Omega, \mathcal{F}, \mathbb{P}) \) is called a version \( \mathbb{E}[X | \mathcal{F}_0] \) if

- \( X_0 \) is \( \mathcal{F}_0 \) measurable
- \( \forall A \in \mathcal{F}_0 \) we have
  \[ \mathbb{E}[1_A X] = \mathbb{E}[1_A X_0] \]

**Theorem:**
Let \( X_1, X_2 \in L^1(\Omega, \mathcal{F}, \mathbb{P}) \) then

- \( \mathbb{P} \) a.s.
  \[
  \mathbb{E}[X_1 + X_2 | \mathcal{F}_0] = \mathbb{E}[X_1 | \mathcal{F}_0] + \mathbb{E}[X_2 | \mathcal{F}_0]
  \]

- \( \forall c \in \mathbb{R} \)
  \[
  \mathbb{E}[cX | \mathcal{F}_0] = c\mathbb{E}[X | \mathcal{F}_0]
  \]

\( \mathbb{P} \) a.s.

**Proof:**
For any choice of \( \mathbb{E}[X_i | \mathcal{F}_0] \) we have that \( \mathbb{E}[X_1 | \mathcal{F}_0] + \mathbb{E}[X_2 | \mathcal{F}_0] \) is measurable hence
\[
\begin{align*}
\mathbb{E}[1_A (\mathbb{E}[X_1 | \mathcal{F}_0] + \mathbb{E}[X_2 | \mathcal{F}_0])] &= \mathbb{E}[1_A \mathbb{E}[X_1 | \mathcal{F}_0] + 1_A \mathbb{E}[X_2 | \mathcal{F}_0]] \\
&= \mathbb{E}[1_A \mathbb{E}[X_1 | \mathcal{F}_0]] + \mathbb{E}[1_A \mathbb{E}[X_2 | \mathcal{F}_0]] \\
&= \mathbb{E}[1_A X_1] + \mathbb{E}[1_A X_2] \\
&= \mathbb{E}[1_A (X_1 + X_2)] \\
&= \mathbb{E}[1_A \mathbb{E}[X_1 + X_2 | \mathcal{F}_0]]
\end{align*}
\]

by linearity of expectation

**Theorem:**
Let \( X_1, X_2 \in L^1(\Omega, \mathcal{F}, \mathbb{P}) \) s.t. \( X_1 \leq X_2 \)
then
\[ \mathbb{E}[X_1 | \mathcal{F}_0] \leq \mathbb{E}[X_2 | \mathcal{F}_0] \]

\( \mathbb{P} \) a.s.
Proof:
Let \( B_0 := \{ \omega \in \Omega : E[X_1|\mathcal{F}_0] > E[X_2|\mathcal{F}_0] \} \in \mathcal{F}_0 \) we want to show that \( P(B_0) = 0 \)

\[
0 \leq E[1_{B_0}(E[X_1|\mathcal{F}_0] - E[X_2|\mathcal{F}_0])]
= E[1_{B_0}(X_1 - X_2)] \\
\leq 0
\]

hence \( P(B_0) = 0 \) as required.

Theorem:
If \( Z, W \) are conditional expectations of \( X|\mathcal{F}_0 \) for some \( X \in L^1 \) then \( Z = W \) \( P \) a.s.

Proof:
\( Z, W \) are \( \mathcal{F}_0 \)-measurable so write
\[
A_0 := \{ Z > W \} \in \mathcal{F}_0
\]
then we have

\[
E[1_{A_0}(Z - W)] = E[1_{A_0}Z] - E[1_{A_0}W] \\
= E[1_{A_0}X] - E[1_{A_0}X] \\
= 0
\]

since \( 1_{A_0}(Z - W) \geq 0 \) we have that \( 1_{A_0}(Z - W) = 0 \) a.s.
but since \( Z > W \) on \( A_0 \) it must be that \( 1_{A_0} = 0 \) a.s.
hence \( P(A_0) = 0 \)
this holds similarly for \( A_1 := \{ W > Z \} \)

\[
P(Z \neq W) = P(\{ Z > W \} \cup \{ Z < W \}) \\
\leq P(A_0) + P(A_1) \\
= 0
\]

Definition: Respective Density
If \( \mu, \nu \) are measures on measure space \( (\Omega, \mathcal{F}) \) then \( \nu \) has density with respect to \( \mu \) if
\[
\exists f : \Omega \to [0, \infty) \text{ that is } \mathcal{F}\text{-measurable s.t.} \\
\forall A \in \mathcal{F} \text{ we have}
\nu(A) = \int_A f \, d\mu
\]

Theorem: Radon-Nikodym
If \( \mu, \nu \) are finite measures on measurable space \( (\Omega, \mathcal{F}) \) then the following are equivalent:

- \( \mu(A) = 0 \implies \nu(A) = 0 \)
- \( \nu \) has density w.r.t. \( \mu \)

Lemma:
If \( 0 \leq X \in L^1 \) and \( \mathcal{F}_0 \subseteq \mathcal{F} \) is a sub-\( \sigma \)-algebra then take
\[
\mu := P_{|\mathcal{F}_0} \text{ and} \\
\nu : \mathcal{F}_0 \to [0, \infty) \text{ s.t.} \\
\forall A \in \mathcal{F}_0 \quad \nu(A) = \int_A X \, dP
\]
then if \( A \) is \( \mu \)-null then by Radon-Nikodym \( \nu \) has density \( g \) w.r.t. \( \mu \)
If $g$ is the density of $\nu$ w.r.t. $\mu$ where 
\[ \nu(A) := \int_A X \, d\mu \]
then $g$ is a conditional expectation $E[X | \mathcal{F}_0]$.

**Proof:**
let $A_0 \in \mathcal{F}_0$

\[
E[g_{1A_0}] = \int_{A_0} g \, d\mu \\
= \int_{A_0} g \, d\mu \\
= \int 1_{A_0} \, d\nu \\
= \int 1_{A_0} X \, d\mu \\
= E[1_{A_0} X]
\]

**Corollary:**
If $X \in \mathcal{L}^1$ then $X = X_+ - X_-$ and if $g_+, g_-$ are versions of $E[X_+ | \mathcal{F}_0], E[X_- | \mathcal{F}_0]$ then by linearity

\[ E[X | \mathcal{F}_0] = g_+ - g_- \]

**Theorem:**
If $X \in \mathcal{L}^1(\Omega, \mathcal{F}, P)$ and $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \mathcal{F}$ are sub-$\sigma$-algebras then 

\[ X_0 := E[E[X | \mathcal{F}_1] | \mathcal{F}_0] = E[X | \mathcal{F}_0] \]

**Proof:**
We want to show that $E[X_0 1_{A_0}] = E[X 1_{A_0}]$

\[
E[X_0 1_{A_0}] = E[1_{A_0} E[E[X | \mathcal{F}_1] | \mathcal{F}_0]] \\
= E[1_{A_0} E[X | \mathcal{F}_1] | \mathcal{F}_0] \\
= E[E[X 1_{A_0}] | \mathcal{F}_0] \quad \text{since } A_0 \in \mathcal{F}_0 \\
= E[X 1_{A_0}] \quad \text{since } A_0 \in \mathcal{F}_1
\]

**Theorem:**
If $X \in \mathcal{L}^1(\Omega, \mathcal{F}, P)$ then 

\[ E[X | \mathcal{F}_0] \in \mathcal{L}^1(\Omega, \mathcal{F}, P) \]

**Proof:**
Define $A_0 := \{E[X | \mathcal{F}_0] \geq 0\} \in \mathcal{F}_0$
then 

\[
E[E[X | \mathcal{F}_0]^+] = E[E[X | \mathcal{F}_0] 1_{A_0}] \\
= E[X 1_{A_0}] \\
< \infty \quad \text{since } X \in \mathcal{L}^1
\]

this holds similarly for $E[E[X | \mathcal{F}_0]^+]$

hence this holds for $E[X | \mathcal{F}_0]$ by linearity.
Let \((\Omega, \mathcal{F}, P)\) be a probability space with \(\mathcal{F}_0 \subset \mathcal{F}\) sub-\(\sigma\)-algebra, \(X \in L^1(\Omega, \mathcal{F}, P)\) where \(\mathcal{F}_0 \perp \sigma(X)\) then

\[ E[X|\mathcal{F}_0] = E[X] \quad P\text{-a.s.} \]

**Proof:**

\(E[X]\) is \(\mathcal{F}_0\) measurable since constant so let \(A \in \mathcal{F}_0\)

\[
E[1_A E[X]] = E[X] E[1_A]
\]

\[
= E[1_A X]
\]

\[
= E[1_A E[X|\mathcal{F}_0]]
\]

**Theorem: MCT For Conditional Expectations**

Let \(\{X_n\}_{n=1}^\infty \in L^1\) be an increasing sequence s.t. \(X_n \to X\ \mathbb{P}\text{-a.s.}\) for some \(X \in L^1\) then

\[ E[X_n|\mathcal{F}_0] \to E[X|\mathcal{F}] \]

both \(\mathbb{P}\text{-a.s.}\) and in \(L^1\)

**Proof:**

Since \(E[X_n|\mathcal{F}_0]\) is an increasing sequence of random variables \(\exists Z\) which is an \(\mathcal{F}_0\) measurable r.v. s.t.

\[ E[X_n|\mathcal{F}_0] \to Z \quad \mathbb{P}\text{-a.s.} \]

For \(A \in \mathcal{F}_0\) we have that

\[
E[Z 1_A] = E[\lim_{n \to \infty} E[X_n|\mathcal{F}_0] 1_A]
\]

\[
= \lim_{n \to \infty} E[E[X_n|\mathcal{F}_0] 1_A]
\]

\[
= \lim_{n \to \infty} E[X_n 1_A]
\]

\[
= E[X 1_A]
\]

hence \(Z = E[X|\mathcal{F}_0] \quad \mathbb{P}\text{-a.s.}\)

\[
0 \leq E[X|\mathcal{F}_0] - E[X_n|\mathcal{F}_0]
\]

\[
\leq E[X|\mathcal{F}_0] - E[X_1|\mathcal{F}_0] \in L^1
\]

hence by dominated convergence theorem

\[
\lim_{n \to \infty} E[|E[X|\mathcal{F}_0] - E[X_n|\mathcal{F}_0]|] = 0
\]

**Theorem: DCT For Conditional Expectations**

Let \(\{X_n\}_{n=1}^\infty \in L^1\) converge to \(X\ \mathbb{P}\text{-a.s.}\)

If \(\exists Z \in L^1\) s.t. \(\forall n \in \mathbb{N}\) we have that \(|X_n| \leq Z\ \mathbb{P}\text{-a.s.}\) then

- \(X \in L^1\)
- \(E[X_n|\mathcal{F}_0] \to E[X|\mathcal{F}_0]\ \mathbb{P}\text{-a.s.}\)
- \(E[X_n|\mathcal{F}_0] \to E[X|\mathcal{F}_0] \in L^1\)

**Proof:**

- \(X \in L^1\) since \(|X_n| \leq Z\ \mathbb{P}\text{-a.s.}\) \(\implies |X| \leq Z\ \mathbb{P}\text{-a.s.}\)
- define $U_n := \inf_{m \geq n} X_m$ is an increasing sequence
  and $V_n := \sup_{m \geq n} X_m$ is a decreasing sequence
  then $U_n, V_n \to X$ \( \mathbb{P} \)-a.s.
  moreover

  $$-Z \leq U_n \leq X_n \leq V_n \leq Z$$

  by MCT for conditional expectations we have that

  $$\mathbb{E}[U_n | \mathcal{F}_0] \to \mathbb{E}[X | \mathcal{F}_0]$$
  $$\mathbb{E}[V_n | \mathcal{F}_0] \to \mathbb{E}[X | \mathcal{F}_0]$$

  both \( \mathbb{P} \)-a.s. and in \( L^1 \)
  and by monotonicity of conditional expectations we have that

  $$\mathbb{E}[U_n | \mathcal{F}_0] \leq \mathbb{E}[X_n | \mathcal{F}_0] \leq \mathbb{E}[V_n | \mathcal{F}_0]$$

  so indeed \( \mathbb{E}[X_n | \mathcal{F}_0] \to \mathbb{E}[X | \mathcal{F}_0] \) in \( L^1 \) and \( \mathbb{P} \)-a.s.

**Theorem:**
Let $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ and $Y$ be \( \mathcal{F}_0 \) measurable s.t. $XY \in L^1$ then

$$\mathbb{E}[XY | \mathcal{F}_0] = Y \mathbb{E}[X | \mathcal{F}_0]$$

**Proof:**
Suppose $Y = 1_A$ for some $A \in \mathcal{F}_0$
then take $B \in \mathcal{F}_0$

$$\mathbb{E}[1_B Y \mathbb{E}[X | \mathcal{F}_0]] = \mathbb{E}[1_B \cap A \mathbb{E}[X | \mathcal{F}_0]]$$
$$= \mathbb{E}[1_B \cap A X]$$
$$= \mathbb{E}[1_B Y X]$$
$$= \mathbb{E}[1_B \mathbb{E}[XY | \mathcal{F}_0]]$$

hence the theorem holds for indicator functions
Suppose $Y = \sum_{n=1}^{\infty} \alpha_n 1_{A_n}$

$$\mathbb{E}[XY | \mathcal{F}_0] = \mathbb{E}\left[ \sum_{n=1}^{\infty} \alpha_n 1_{A_n} X | \mathcal{F}_0 \right]$$
$$= \sum_{n=1}^{\infty} \alpha_n \mathbb{E}[1_{A_n} X | \mathcal{F}_0]$$
$$= \sum_{n=1}^{\infty} \alpha_n \mathbb{E}[1_{A_n} X | \mathcal{F}_0]$$
$$= \sum_{n=1}^{\infty} \alpha_n \mathbb{E}[X | \mathcal{F}_0]$$
$$= \mathbb{E}[XY | \mathcal{F}_0]$$

hence the theorem holds for simple functions
Suppose $Y$ is positive
let \{ $Y_n$ \}_{n=1}^{\infty} be an increasing sequence of simple functions s.t. $Y_n \to Y$
$|Y_n X| \leq |Y X| \in L^1$
and $Y_n X \to Y X$

hence by DCT for conditional expectations we have that
\[ Y_n \mathbb{E}[X | \mathcal{F}_0] = \mathbb{E}[Y_n X | \mathcal{F}_0] \rightarrow \mathbb{E}[Y X | \mathcal{F}_0] \quad \text{P} - \text{a.s.} \]

moreover
\[ Y_n \mathbb{E}[X | \mathcal{F}_0] \rightarrow Y \mathbb{E}[X | \mathcal{F}_0] \quad \text{P} - \text{a.s.} \]

hence the theorem holds for positive functions.

For the general case write \( Y = Y_+ - Y_- \)

\[ \mathbb{E}[XY | \mathcal{F}_0] = \mathbb{E}[X (Y_+ - Y_-) | \mathcal{F}_0] \]
\[ = \mathbb{E}[XY_+ | \mathcal{F}_0] - \mathbb{E}[XY_- | \mathcal{F}_0] \]
\[ = Y_+ \mathbb{E}[X | \mathcal{F}_0] - Y_- \mathbb{E}[X | \mathcal{F}_0] \]
\[ = Y \mathbb{E}[X | \mathcal{F}_0] \]

Lemma:
If \( X \in \mathcal{L}^1 \), \( \mathcal{F}_0 \subseteq \mathcal{F} \) is a sub-\( \sigma \) algebra then
\[ ||\mathbb{E}[X | \mathcal{F}_0]||_1 \leq ||X||_1 \]

Proof:
\[ ||\mathbb{E}[X | \mathcal{F}_0]||_1 = \int |\mathbb{E}[X | \mathcal{F}_0]| d\mathbb{P} \]
\[ = \int \mathbb{E}[X | \mathcal{F}_0]_+ d\mathbb{P} + \int \mathbb{E}[X | \mathcal{F}_0]_- d\mathbb{P} \]
\[ = \int \mathbb{E}[X | \mathcal{F}_0] I_{\mathbb{E}[X | \mathcal{F}_0] > 0} d\mathbb{P} + \int \mathbb{E}[X | \mathcal{F}_0] I_{\mathbb{E}[X | \mathcal{F}_0] < 0} d\mathbb{P} \]
\[ = \mathbb{E}[X I_{\mathbb{E}[X | \mathcal{F}_0] > 0}] + \mathbb{E}[X I_{\mathbb{E}[X | \mathcal{F}_0] < 0}] \]
\[ \leq ||X||_1 \]
\[ = ||X||_1 \]

Theorem: Conditional Jensen’s Inequality
Let \( I \subseteq \mathbb{R} \) be open, \( \varphi : I \rightarrow \mathbb{R} \) be convex and \( X \in \mathcal{L}^1 \) a r.v. with \( X : \Omega \rightarrow I \) hence \( \varphi \circ X \in \mathcal{L}^1 \) then
\[ \mathbb{E}[\varphi \circ X | \mathcal{F}_0] \geq \varphi(\mathbb{E}[X | \mathcal{F}_0]) \]

Proof:
From Jensen’s inequality we have that
\[ \varphi(X) \geq \varphi(\mathbb{E}[X | \mathcal{F}_0]) + D_- \varphi(\mathbb{E}[X | \mathcal{F}_0])(X - \mathbb{E}[X | \mathcal{F}_0]) \]

let \( A_n = \{ \omega \in \Omega : |D_- \varphi(\mathbb{E}[X | \mathcal{F}_0])| \leq n \} \in \mathcal{F}_0 \)

then
\[ 1_{A_n} \mathbb{E}[\varphi(X) | \mathcal{F}_0] = \mathbb{E}[1_{A_n} \varphi(X) | \mathcal{F}_0] \]
\[ \geq \mathbb{E}[1_{A_n} \varphi(\mathbb{E}[X | \mathcal{F}_0])] + \mathbb{E}[1_{A_n} D_- \varphi(\mathbb{E}[X | \mathcal{F}_0])(X - \mathbb{E}[X | \mathcal{F}_0])] \]
\[ = 1_{A_n} \varphi(\mathbb{E}[X | \mathcal{F}_0]) + 1_{A_n} D_- \varphi(\mathbb{E}[X | \mathcal{F}_0]) \mathbb{E}[X - \mathbb{E}[X | \mathcal{F}_0]] \]
\[ = 1_{A_n} \varphi(\mathbb{E}[X | \mathcal{F}_0]) + 1_{A_n} D_- \varphi(\mathbb{E}[X | \mathcal{F}_0]) (\mathbb{E}[X | \mathcal{F}_0] - \mathbb{E}[X | \mathcal{F}_0]) \]
\[ = 1_{A_n} \varphi(\mathbb{E}[X | \mathcal{F}_0]) \]

since \( 1_{A_n} - \mathcal{F}_0 \) measurable
this holds for all \( n \) hence also for \( \lim_{n \to \infty} A_n = \Omega \)

**Theorem:**
Let \( p \geq 1, X \in L^p \) then \( \|\mathbb{E}[X|F_0]\|_p \leq \|X\|_p \)

**Proof:**
Write \( \varphi(x) = x^p \) is convex hence by Jensen’s inequality
\[
\|\mathbb{E}[X|F_0]\|_p = \int \varphi(\mathbb{E}[X|F_0])dP \\
\leq \int \mathbb{E}[\varphi(X)|F_0]dP \\
= \|X\|_p
\]

**Corollary:**
If \((\Omega, \mathcal{F}), (\Omega', \mathcal{F}')\) are measurable spaces,
\( Y : (\Omega, \mathcal{F}) \to (\Omega', \mathcal{F}') \) is \( \mathcal{F} - \mathcal{F}' \) measurable and
\( Z : \Omega \to \mathbb{R} \) then \( Z \) is \( \sigma(Y) \) measurable iff
\( \exists g : (\Omega', \mathcal{F}') \to (\mathbb{R}, \mathcal{B}(\mathbb{R})) \) s.t. \( Z = g \circ Y \)

**Proof:**
Clearly if \( \exists g : (\Omega', \mathcal{F}') \to (\mathbb{R}, \mathcal{B}(\mathbb{R})) \) s.t. \( Z = g \circ Y \) then \( Z \) is \( \sigma(Y) \) measurable since the composition of measurable functions is measurable.
Suppose
\[
Z = \sum_{k=1}^{n} \alpha_k 1_{A_k}
\]
for \( \{A_k\}_{k=1}^{n} \in \sigma(Y) \)
notice that \( \sigma(Y) = \{Y^{-1}(A'_k) : A'_k \in \mathcal{F}'\} \)
hence \( A_k = Y^{-1}(A'_k) \) for some \( A'_k \in \mathcal{F}' \)
so
\[
Z = \sum_{k=1}^{n} \alpha_k 1_{A_k} \\
= \sum_{k=1}^{n} \alpha_k 1_{Y^{-1}(A'_k)} \\
= \sum_{k=1}^{n} \alpha_k (1_{A'_k} \circ Y) \\
= \left( \sum_{k=1}^{n} \alpha_k 1_{A'_k} \right) \circ Y \\
= g \circ Y
\]
where clearly \( g \) is \( \mathcal{F}' - \mathcal{B}(\mathbb{R}) \) measurable since \( \{A'_k\}_{k=1}^{n} \in \mathcal{F}' \)
Suppose \( Z \) is positive.
then \( \exists \{Z_n\}_{n=1}^{\infty} \) simple increasing functions s.t. \( \lim_{n \to \infty} Z_n = Z \)
then \( \exists \{g_n\}_{n=1}^{\infty} \) s.t. \( Z_n = g_n \circ Y \)
then define \( g = \sup_n g_n \) is clearly \( \mathcal{F}' - \mathcal{B}(\mathbb{R}) \) measurable and \( Z = g \circ Y \)
For general $Z$ write $Z = Z_+ - Z_-$
then $\exists g_+, g_- : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ s.t. $Z_+ = g_+ \circ Y$, $Z_- = g_- \circ Y$
so write $g = g_+ - g_-$

hence

$$Z = (g_+ - g_-) \circ Y = g \circ Y$$

**Definition: Factorised Conditional Expectation**

$f : (\mathbb{R}, \mathcal{B}(\mathbb{R})) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is a FCE of the r.v. $X$ given the r.v. $Y : \Omega \to \mathbb{R}$ if

$f \circ Y$ is a version of the conditional expectation of $X$ given $Y$

i.e.

$$\mathbb{E}[X|Y] = \mathbb{E}[X|\sigma(Y)] = f(Y)$$

**Corollary:**

Let $Y : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ and $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$

then

$\exists f$ FCE of $X$ given $Y$ and

$$\int_A f d\mathbb{P}_Y = \mathbb{E}[1_{Y^{-1}(A)} X]$$

**Proof:**

By the previous lemma

write $Z = \mathbb{E}[X|Y]$ which is $\sigma(Y)$ measurable hence indeed

$\exists f$ FCE of $X$ given $Y$

moreover

$$\int_A f d\mathbb{P}_Y = \int A 1_{Y^{-1}(A)} X d\mathbb{P}$$

$$= \mathbb{E}[1_{Y^{-1}(A)} X]$$

**Remark:**

$\mathbb{E}[X|Y = y]$ is unique $\mathbb{P}_Y$-a.s.

**Theorem:**

Let $Y : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ and $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$

If $\lambda, \mu$ are $\sigma$-finite measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ s.t. $(X, Y)$ has density $g$ w.r.t. $\lambda \otimes \mu$

then

$$f(Y) := \frac{\int_{\mathbb{R}} xg(x, y)d\lambda(x)}{\int_{\mathbb{R}} g(x, y)d\lambda(x)}$$

is an FCE of $X$ given $Y$

**Proof:**

recall that

$$\mathbb{P}((X, Y) \in A_1 \times A_2) = \int_{A_1 \times A_2} g(x, y)d(\lambda \otimes \mu)(x, y)$$

$$\mathbb{E}[h(X, Y)] = \int_{\mathbb{R}^2} h(x, y)g(x, y)d(\lambda \otimes \mu)(x, y)$$

Suppose $f$ takes the required form, hence clearly it is Borel measurable.
Let \( A \in \mathcal{B}(\mathbb{R}) \)

\[
\mathbb{E}[1_A Y - 1_{(A)}] = \int x 1_A(y) g(x, y) d(\lambda \otimes \mu)(x, y) \\
= \int 1_A(y) \int x g(x, y) d\lambda(x) d\mu(y) \\
= \int 1_A(y) f(y) \int g(x, y) d\lambda(x) d\mu(y) \\
= \int 1_A(y) f(y) g(x, y) d(\lambda \otimes \mu)(x, y) \\
= \mathbb{E}[1_A(y) f(y)]
\]

hence indeed \( f(y) = \mathbb{E}[1_A Y] \)

## 5 Martingales

**Definition: Filtration**

If \((\Omega, \mathcal{F})\) is a measurable space then \( \{\mathcal{F}_i\}_{i=1}^{\infty} \) sub-\( \sigma \)-algebras of \( \mathcal{F} \) are a filtration if \( \mathcal{F}_i \subseteq \mathcal{F}_{i+1} \subseteq \mathcal{F} \quad \forall i \)

**Definition: Adapted**

A sequence of random variables \( \{X_i\}_{i=1}^{\infty} \) is called adapted to \( \{\mathcal{F}_i\}_{i=1}^{\infty} \) if \( X_n \) is \( \mathcal{F}_n \)-measurable \( \forall n \)

**Definition: Martingale**

If \((\Omega, \mathcal{F}, \mathbb{P})\) is a probability space and \( \{\mathcal{F}_i\}_{i=1}^{\infty} \) is a filtration of \((\Omega, \mathcal{F})\) then \( \{X_n\}_{n=1}^{\infty} \) is a martingale if:

- \( \mathbb{E}[|X_n|] < \infty \quad \forall n \)
- \( \{X_n\}_{n=1}^{\infty} \) are adapted to \( \{\mathcal{F}_n\}_{n=1}^{\infty} \)
- \( \mathbb{E}[X_{n+1}|\mathcal{F}_n] = X_n \)

**Remark:**

If \( \{X_n\}_{n=1}^{\infty} \) has the first three properties of a martingale but

- \( \mathbb{E}[X_{n+1}|\mathcal{F}_n] \geq X_n \) then \( \{X_n\}_{n=1}^{\infty} \) is a sub-martingale
- \( \mathbb{E}[X_{n+1}|\mathcal{F}_n] \leq X_n \) then \( \{X_n\}_{n=1}^{\infty} \) is a super-martingale

**Lemma:**

If \( \{X_n\}_{n=1}^{\infty} \) is a martingale w.r.t. \( \{\mathcal{F}_n\}_{n=1}^{\infty} \) then

\( \mathbb{E}[X_{n+m}|\mathcal{F}_n] = 0 \)

**Proof:**

\[
\mathbb{E}[X_{n+m}|\mathcal{F}_n] = \mathbb{E}[\mathbb{E}[X_{n+m}|\mathcal{F}_n \cap \mathcal{F}_{n+m-1}]] \\
= \mathbb{E}[\mathbb{E}[X_{n+m}|\mathcal{F}_{n+m-1}]|\mathcal{F}_n] \\
= \mathbb{E}[X_{n+m-1}|\mathcal{F}_n] \\
= \mathbb{E}[X_{n+1}|\mathcal{F}_n] \\
= X_n
\]

**Theorem:**

If \( \{X_n\}_{n=1}^{\infty} \in \mathcal{L}^2 \) is a martingale w.r.t. \( \{\mathcal{F}_n\}_{n=1}^{\infty} \) and \( r \leq s \leq t \) then

- \( \mathbb{E}[X_r(X_s - X_r)|\mathcal{F}_r] = 0 \)
- \( \mathbb{E}[X_r(X_t - X_s)|\mathcal{F}_s] = 0 \)
\( \mathbb{E}[(X_t - X_s)(X_s - X_r)|\mathcal{F}_s] = 0 \)

\( \mathbb{E}[X_r (X_s - X_r)] = 0 \)

\( \mathbb{E}[X_r (X_t - X_s)] = 0 \)

\( \mathbb{E}[(X_t - X_s)(X_s - X_r)] = 0 \)

**Proof:**

\[
\mathbb{E}[X_r (X_s - X_r)|\mathcal{F}_r] = \mathbb{E}[X_r X_s - X_r^2|\mathcal{F}_r] \\
= \mathbb{E}[X_r X_s|\mathcal{F}_r] - \mathbb{E}[X_r^2|\mathcal{F}_r] \\
= X_r \mathbb{E}[X_s|\mathcal{F}_r] - X_r^2 \\
= 0
\]

\[
\mathbb{E}[X_r (X_t - X_s)|\mathcal{F}_s] = \mathbb{E}[X_r X_t - X_r X_s|\mathcal{F}_s] \\
= \mathbb{E}[X_r X_t|\mathcal{F}_s] - \mathbb{E}[X_r X_s|\mathcal{F}_s] \\
= X_r \mathbb{E}[X_t|\mathcal{F}_s] - X_r \mathbb{E}[X_s|\mathcal{F}_s] \\
= X_r X_t - X_r X_s \\
= 0
\]

\[
\mathbb{E}[(X_t - X_s)(X_s - X_r)|\mathcal{F}_s] = \mathbb{E}[X_t X_s - X_t X_s + X_s X_r + X_s^2|\mathcal{F}_s] \\
= \mathbb{E}[X_t X_s|\mathcal{F}_s] - \mathbb{E}[X_s^2|\mathcal{F}_s] - \mathbb{E}[X_t X_s|\mathcal{F}_s] + \mathbb{E}[X_s X_r|\mathcal{F}_s] \\
= X_t \mathbb{E}[X_s|\mathcal{F}_s] - X_s^2 - X_r \mathbb{E}[X_t|\mathcal{F}_s] + X_s X_r \\
= X_s^2 - X_s^2 - X_r X_s + X_s X_r \\
= 0
\]

\[
\mathbb{E}[X_r (X_s - X_r)] = \mathbb{E}[\mathbb{E}[X_r (X_s - X_r)|\mathcal{F}_r]] \\
= \mathbb{E}[0] \\
= 0
\]

\[
\mathbb{E}[X_r (X_t - X_s)] = \mathbb{E}[\mathbb{E}[X_r (X_t - X_s)|\mathcal{F}_s]] \\
= \mathbb{E}[0] \\
= 0
\]

\[
\mathbb{E}[(X_t - X_s)(X_s - X_r)] = \mathbb{E}[\mathbb{E}[(X_t - X_s)(X_s - X_r)|\mathcal{F}_s]] \\
= \mathbb{E}[0] \\
= 0
\]
Definition: $\mathcal{L}^n$ Bounded
A set of r.vs $C$ is $\mathcal{L}^n$ bounded if
$$\sup_{X \in C} \mathbb{E}[|X|^n] < \infty$$

Theorem:
Every $\mathcal{L}^2$ bounded martingale $\{X_n\}_{n=1}^\infty$ w.r.t. $\{\mathcal{F}_n\}_{n=1}^\infty$ converges in $\mathcal{L}^2$
i.e. $\exists Y \in \mathcal{L}^2$ s.t. $\lim_{n \to \infty} \mathbb{E}[|X_n - Y|^2] = 0$

Proof:
$\mathcal{L}^2$ is complete hence every cauchy sequence converges

$$\sum_{k=1}^{n-1} \mathbb{E}[(X_{k+1} - X_k)^2] = \sum_{k=1}^{n-1} \mathbb{E}[(X_{k+1}^2 + X_k^2 - 2X_kX_{k+1})]$$
$$= \sum_{k=1}^{n-1} \mathbb{E}[(X_{k+1}^2 + X_k^2 - 2X_kX_{k+1})]$$
$$= \mathbb{E}[(X_n - X_1)^2]$$
by a telescoping sum, orthogonality of increments
$$\leq 2c$$
since $X_n$ is $\mathcal{L}^2$ bounded

$$\lim_{k \to \infty} \sup_{m \geq k} \mathbb{E}[(X_m - X_n)^2] = \lim_{n \to \infty} \sup_{m \geq k} \sum_{l=n}^{m-1} \mathbb{E}[(X_{l+1} - X_l)^2]$$
$$\leq \lim_{k \to \infty} \sup_{m \geq k} \sum_{l=k}^{\infty} \mathbb{E}[(X_{l+1} - X_l)^2]$$
$$= 0$$

Remark:
The previous theorem also gives us that $Y \in \mathcal{L}^2(\Omega, \mathcal{F}_\infty, \mathbb{P})$
where $\mathcal{F}_\infty := \sigma(\bigcup_{k=1}^\infty \mathcal{F}_k)$

Theorem: Levy
Let $X \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ and $\{\mathcal{F}_k\}_{k=1}^\infty$ a filtration
define $X_n := \mathbb{E}[X|\mathcal{F}_n]$
then $X_n \to \mathbb{E}[X|\mathcal{F}_\infty]$ in $\mathcal{L}^2$

Proof:
We have that $X_n$ converges in $\mathcal{L}^2$ so let $X_n \to Y$
We want to show that $\mathbb{E}[Y1_A] = \mathbb{E}[X1_A] \quad \forall A \in \mathcal{F}_n$
let $A \in \mathcal{F}_n$ and $m \geq n$ then notice that
$$\mathbb{E}[X_m1_A] = \mathbb{E}[\mathbb{E}[X|\mathcal{F}_m]1_A]$$
$$= \mathbb{E}[X1_A]$$
since $X_m \to Y$ in $\mathcal{L}^2$ we have that

$$\lim_{m \to \infty} |\mathbb{E}[X_m1_A - Y1_A]| \leq \lim_{m \to \infty} (\mathbb{E}[X_m - Y]^21_A)^{1/2}$$
by the Cauchy-Schwarz inequality
$$= 0$$

hence indeed we have that $\mathbb{E}[X1_A] = \mathbb{E}[Y1_A] \quad \forall A \in \mathcal{F}_n$
We need this $\forall A \in \mathcal{F}_\infty$
5 MARTINGALES

define

\[ \mathcal{D} := \{ A \in \mathcal{F}_\infty : E[X_1|A] = E[Y_1|A] \} \]

We want to show that \( \mathcal{D} \) is a dynkin system.
\( \Omega \in \mathcal{D} \) holds trivially since \( \Omega \in \mathcal{F}_n \) \( \forall n \)

Let \( A \in \mathcal{D} \)

\[
E[Y_1|A^c] = E[Y_1|\Omega] - E[Y_1|A]
= E[X_1|\Omega] - E[X_1|A]
= E[X_1|A^c]
\]

as required.

Suppose \( \{A_i\}_{i=1}^\infty \in \mathcal{D} \) are disjoint and \( A := \bigcup_{i=1}^\infty A_i \)

\[
E[Y_1|A] = E \left[ Y \lim_{n \to \infty} \sum_{k=1}^n 1_{A_k} \right]
= \lim_{n \to \infty} E \left[ Y \sum_{k=1}^n 1_{A_k} \right] \quad \text{by DCT}
= \lim_{n \to \infty} \sum_{k=1}^n E[Y_1|A_k]
= \lim_{n \to \infty} \sum_{k=1}^n E[X_1|A_k]
= \lim_{n \to \infty} E \left[ X \sum_{k=1}^n 1_{A_k} \right]
= E \left[ X \lim_{n \to \infty} \sum_{k=1}^n 1_{A_k} \right]
= E[X_1|A]
\]

Since \( \mathcal{D} \) is a Dynkin system if \( I \subset \mathcal{D} \) is a \( \pi \)-system then \( \sigma(I) \subset \mathcal{D} \)
\( I := \bigcup_{n=1}^\infty \mathcal{F}_n \subset \mathcal{D} \) is a \( \pi \)-system since
\( \{A_i\}_{i=1}^\infty \in I \) then \( \forall i \exists k_i \text{ s.t. } A_i \in \mathcal{F}_{k_i} \)
hence since finite \( \exists K = \max\{k_i\} \)
then since \( \{\mathcal{F}_n\}_{n=1}^\infty \) is a filtration \( A_i \in \mathcal{F}_K \) \( \forall i \)
hence since \( \mathcal{F}_K \) is a \( \sigma \)-algebra \( \bigcap_{i=1}^n A_i \in \mathcal{F}_K \)
thus \( \mathcal{F}_\infty = \sigma(I) \subset \mathcal{D} \subset \mathcal{F}_\infty \)

**Lemma: DOOB**

Let \( \{\mathcal{F}_n\}_{n=1}^\infty \) be a filtration of \((\Omega, \mathcal{F}, \mathbb{P})\) and \( \{X_n\}_{n=1}^\infty \in \mathcal{L}^2 \) an adapted sequence of \( \{\mathcal{F}_n\}_{n=1}^\infty \) then

- \( \exists \{A_n\}_{n=1}^\infty \subset \mathcal{L}^1 \) and martingale \( \{M_n\}_{n=1}^\infty \subset \mathcal{L}^1 \) with
  - \( A_1 = 0 \)
  - \( A_{n+1} \) \( \mathcal{F}_n \)-measurable
  - \( X_n = M_n + A_n \)
which is a \( \mathbb{P} \)-a.s. unique decomposition
- For \( X_n = M_n + A_n \) above we have that \( \{X_n\}_{n=1}^\infty \) is a sub-martingale iff \( \{A_n\}_{n=1}^\infty \) is increasing \( \mathbb{P} \)-a.s.
Proof:

- If we construct the sequence recursively then the sequence will be \( \mathbb{P} \)-a.s unique
  \[ A_1 = 0 \implies M_1 = X_1 \]
  \[ X_{n+1} = M_{n+1} + A_{n+1} \]
  \[ \mathbb{E}[X_{n+1} | \mathcal{F}_n] = \mathbb{E}[M_{n+1} | \mathcal{F}_n] + \mathbb{E}[A_{n+1} | \mathcal{F}_n] \]
  \[ = M_n + A_{n+1} \]
  \[ A_{n+1} = \mathbb{E}[X_{n+1} | \mathcal{F}_n] - M_n \]

  is clearly \( \mathcal{F}_n \) measurable.

  \[ \mathbb{E}[M_{n+1} | \mathcal{F}_n] = \mathbb{E}[X_{n+1} - A_{n+1} | \mathcal{F}_n] \]
  \[ = \mathbb{E}[X_{n+1} | \mathcal{F}_n] - \mathbb{E}[A_{n+1} | \mathcal{F}_n] \]
  \[ = A_{n+1} + M_n - A_{n+1} \]
  \[ = M_n \]

  hence \( \{M_n\}_{n=1}^\infty \) is a martingale.

- \[ A_{n+1} - A_n = \mathbb{E}[X_{n+1} | \mathcal{F}_n] - M_n - A_n \]
  \[ = \mathbb{E}[X_{n+1} | \mathcal{F}_n] - X_n \]

  which is positive iff \( X_n \leq \mathbb{E}[X_{n+1} | \mathcal{F}_n] \)
  i.e. \( \{X_n\}_{n=1}^\infty \) is a sub-martingale.

**Definition: Stopping Time**

Let \( \{\mathcal{F}_n\}_{n=1}^\infty \) be a filtration on \((\Omega, \mathcal{F}, \mathbb{P})\) and let \( Y : \Omega \to \mathbb{N} \cup \{\infty\} \) be a random variable

then \( T \) is a stopping time if

\[ \{T \leq n\} \in \mathcal{F}_n \quad \forall n \]

**Lemma:**

Let \((\Omega, \mathcal{F}, \mathbb{P})\) be a probability space with stopping time \( T \) on filtration \( \{\mathcal{F}_n\}_{n=1}^\infty \)

then \( T \cap S := \min(T, S), T \cup S := \max(T, S) \) are stopping times.

**Proof:**

Clearly \( T \cap S, T \cup S \) are r.v.s

\[ \{T \cap S \leq n\} = \{T \leq n\} \cup \{S \leq n\} \]
\[ \{T \cup S \leq n\} = \{T \leq n\} \cap \{S \leq n\} \]

Since \( T, S \) are stopping times \( \{T \leq n\}, \{S \leq n\} \in \mathcal{F}_n \)

hence \( \{T \cap S \leq n\}, \{T \cup S \leq n\} \in \mathcal{F}_n \)

**Lemma:**

Let \((\Omega, \mathcal{F}, \mathbb{P})\) be a probability space with stopping time \( T \) on filtration \( \{\mathcal{F}_n\}_{n=1}^\infty \)

then the set of events which are observable up top time \( T \)

\[ \mathcal{F}_T := \{A \in \mathcal{F} : A \cap \{T \leq n\} \in \mathcal{F}_n \quad \forall n\} \]

is a \( \sigma \)-algebra.

**Lemma:**

Let \((\Omega, \mathcal{F}, \mathbb{P})\) be a probability space with stopping time \( T \) on filtration \( \{\mathcal{F}_n\}_{n=1}^\infty \)

If \( \{X_n\}_{n=1}^\infty \) is adapted to \( \{\mathcal{F}_n\}_{n=1}^\infty \) then
\[ X_T(\omega) := \begin{cases} X_{T(\omega)}(\omega) & T(\omega) < \infty \\ 0 & \text{o/w} \end{cases} \]

is \( F_T \) measurable.

**Proof:**

Let \( X_n : (\Omega, F) \to (\Omega', F') \)

and let \( A \in F' \)

\[ \{ X_T \in A \} = \{ 0 \in A, T = \infty \} \cup \bigcup_{n=1}^{\infty} \{ X_n \in A \} \cap \{ T = n \} \in F \]

hence \( X_T \) is a random variable.

Moreover

\[ \{ X_T \in A \} \cap \{ T = n \} = \{ X_n \in A \} \cap \{ T = n \} \in F_n \]

since \( X_n \) adapted and \( T \) is a stopping time.

hence \( \{ X_T \in A \} \in F_T \)

**Lemma:**

Let \( \{ X_n \}_{n=1}^{\infty} \) be a martingale on \( \{ F_n \}_{n=1}^{\infty} \) and \( T \) stopping time s.t. \( T \leq m \) a.s.

then \( \mathbb{E}[X_m | T] = X_T \) \( \mathbb{P} \)-a.s.

**Proof:**

\( X_T \) is \( F_T \) measurable.

for \( k \in \mathbb{N}_{[1,m]} \) and \( A \in F_T \)

\[
\mathbb{E}[X_T 1_{A T=k}] = \mathbb{E}[X_k 1_{A \cap (T=k)}] \\
= \mathbb{E}[\mathbb{E}[X_m | F_k] 1_{A \cap (T=k)}] \\
= \mathbb{E}[\mathbb{E}[X_m 1_{A \cap (T=k)} | F_k]] \\
= \mathbb{E}[X_m 1_{A \cap (T=k)}] \\
\]

\[
\mathbb{E}[X_T 1_A] = \sum_{k=1}^{m} \mathbb{E}[X_T 1_{A T=k}] \\
= \sum_{k=1}^{m} \mathbb{E}[X_m 1_{A T=k}] \\
= \mathbb{E}[X_m 1_A] \\
\]

hence \( X_T = \mathbb{E}[X_m | F_T] \)

**Lemma:**

Let \( S \leq T \) be stopping times on \( \{ F_n \}_{n=1}^{\infty} \) then

\( F_S \subseteq F_T \)

**Proposition: Optimal Stopping**

Let \( S \leq T \) be two bounded stopping times on \( \{ F_n \}_{n=1}^{\infty} \)

and \( \{ X_n \}_{n=1}^{\infty} \) be a martingale on the same filtration.

then \( \mathbb{E}[X_T | F_S] = X_S \) \( \mathbb{P} \)-a.s.

**Proof:**

\( T \leq m \) hence

\[
\mathbb{E}[X_T | F_S] = \mathbb{E}[\mathbb{E}[X_m | F_T] | F_S] \\
= \mathbb{E}[X_m | F_S] \\
= X_S \\
\]
Proposition:
Let $S \leq T$ be $\mathbb{P}$-a.s. finite stopping times and $\{X_n\}_{n=1}^{\infty}$ a sub-martingale then
\[
\mathbb{E}[X_T | \mathcal{F}_S] \geq X_S
\]

Proposition:
Let $S \leq T$ be $\mathbb{P}$-a.s. finite stopping times and $\{X_n\}_{n=1}^{\infty}$ a super-martingale then
\[
\mathbb{E}[X_T | \mathcal{F}_S] \leq X_S
\]

Theorem: First DOOB Inequality
Let $\{X_n\}_{n=1}^{\infty}$ be a sub-martingale, $X_n := \max_{i \leq n} X_i$ and $a > 0$ then
\[
a \mathbb{P}(X_n \geq a) \leq \mathbb{E}[X_n 1_{\{X_n \geq a\}}]
\]

Proof:
Let $T = \inf \{n \in \mathbb{N} : X_n \geq a\}$
which is a stopping time.
Then
\[
\{X_n \geq a\} = \{T \leq n\} = \{X_{T \land n} \geq a\}
\]

hence
\[
a \mathbb{P}(X_n \geq a) = a \mathbb{P}(X_{T \land n} \geq a)
\]
\[= a \mathbb{E}[1_{T \land n}]
\[= \mathbb{E}[a 1_{T \land n}]
\[\leq \mathbb{E}[X_{T \land n} 1_{T \land n}]
\[\leq \mathbb{E}[\mathbb{E}[X_n | \mathcal{F}_{T \land n}] 1_{T \land n}]
\[= \mathbb{E}[\mathbb{E}[X_n 1_{T \land n} | \mathcal{F}_{T \land n}]]
\[= \mathbb{E}[X_n 1_{T \land n}]
\]

Corollary:
If $\{X_n\}_{n=1}^{\infty}$ is a non-negative sub-martingale then $\forall a \geq 0, p \geq 1$
\[
a^p \mathbb{P}(X_n \geq a) \leq \mathbb{E}[X_n^p]
\]

Theorem: Second DOOB Inequality
If $\{X_n\}_{n=1}^{\infty}$ is a non-negative sub-martingale and $p > 1$ then
\[
\mathbb{E}[X_n^p] \leq \left( \frac{p}{p-1} \right)^p \mathbb{E}[X_n]
\]

Definition: Crosses
The sequence $\{X_n\}_{n=1}^{\infty}$ crosses the interval $[a, b]$ if $\exists i, j$ s.t. $X_i < a < b < X_j$

Lemma:
Let $\{X_n\}_{n=1}^{\infty}$ which crosses any rational interval only finitely often.
Then $\{X_n\}_{n=1}^{\infty}$ converges in $\mathbb{R} \cup \{\pm \infty\}$

Proof:
Suppose $\liminf_n X_n < \limsup_n X_n$
then $\exists a, b \in \mathbb{Q}$ s.t. $\liminf_n X_n < a < b < \limsup_n X_n$
then $\{X_n\}_{n=1}^{\infty}$ crosses $[a, b]$ i.o.

Definition: Up-Crossing
If $\{X_n\}_{n=1}^{\infty}$ is an adapted sequence of $\{\mathcal{F}_n\}_{n=1}^{\infty}$ and $a < b$ then
define $T_0 := 0$ and recursively
$S_k := \inf \{n \geq T_{k-1} : X_n \leq a\}$
$T_k := \inf \{n \geq S_k : X_n \geq b\}$
Then \( \{X_n\}_{n=S_k}^T \) is the \( k \)th up-crossing of \([a, b]\)

**Remark:**
We write the number of up-crossings of \([a, b]\) by time \( n \) as

\[
U_{a,b}^n := \sum_{k=1}^{n} 1_{T_k \leq n}
\]

**Lemma:**
Let \( \{X_n\}_{n=1}^\infty \) be a martingale then

\[
\mathbb{E}[U_{a,b}^n] \leq \frac{\mathbb{E}[(X_n - a)_-]}{b - a}
\]

**Proof:**
Let

\[
Z := \sum_{k=1}^{\infty} (X_{T_k \cap n} - X_{S_k \cap n})
\]

then

\[
\mathbb{E}[(X_{T_k \cap n} - X_{S_k \cap n})] = \mathbb{E}[X_{T_k \cap n}] - \mathbb{E}[X_{S_k \cap n}]
\]

\[
= \mathbb{E}[\mathbb{E}[X_{T_k \cap n} | \mathcal{F}_{S_k \cap n}]] - \mathbb{E}[X_{S_k \cap n}]
\]

\[
= \mathbb{E}[X_{S_k \cap n}] - \mathbb{E}[X_{S_k \cap n}]
\]

\[
= 0
\]

\[
\mathbb{E}[Z] = 0
\]

Suppose \( U_{a,b}^n = m \) then

\[
Z \geq m(b - a) + (X_n - X_{S_m+1 \cap n})
\]

since \( Z \) crosses \([a, b] \) \( m \) times and there could be one uncompleted crossing, \( X_n - X_{S_m+1 \cap n} \geq X_n - a \) hence

\[
Z \geq \sum_{m=0}^{\infty} (1_{U_{a,b}^n = m} m(b - a) + 1_{U_{a,b}^n = m} (X_n - a))
\]

\[
0 = \mathbb{E}[Z]
\]

\[
\geq \sum_{m=0}^{\infty} \mathbb{P}(U_{a,b}^n = m) m(b - a) - \mathbb{E}[(X_n - a)_-]
\]

\[
= \mathbb{E}[U_{a,b}^n](b - a) - \mathbb{E}[(X_n - a)_-]
\]

hence we have the required result.

**Theorem: Martingale Convergence Theorem**
Let \( \{X_n\}_{n=1}^\infty \) be a martingale s.t. \( \sup_n \mathbb{E}[(X_n)_-] < \infty \)
then \( X_\infty := \lim_{n \to \infty} X_n \) exists \( \mathbb{P} \)-a.s.
moreover \( X_\infty \) is \( \mathcal{F}_\infty \) measurable and \( X_\infty \in \mathcal{L}^1 \)

**Proof:**
By the previous lemma

\[
\mathbb{E}[U_{a,b}^n] \leq \frac{\mathbb{E}[(X_n - a)_-]}{b - a} < \infty
\]

by monotone convergence of expectations we have

\[
\mathbb{E}[U_{a,b}^\infty] < \infty
\]

but \( \mathbb{E}[U_{a,b}^\infty] \geq 0 \) hence \( \mathbb{E}[U_{a,b}^\infty] < \infty \) \( \mathbb{P} \)-a.s.
Since $a, b \in \mathbb{Q}$ there are only a countable number of intervals hence
\[ \exists N \in \mathcal{F} \text{ s.t. } \mathbb{P}(N) = 0 \]
and $\forall a < b \in \mathbb{Q}, \forall \omega \in \Omega \setminus N$
we have that $\mathbb{E}[U_{a,b}] < \infty$
moreover $\lim_{n \to \infty} X_n(\omega)$ exists on $\mathbb{R} \cup \{\pm \infty\}$ $\forall \omega \in \Omega \setminus N$
$Y := \lim_{n \to \infty} X_n$ is $\mathcal{F}_\infty$ since $\{X_n\}_{n=1}^\infty$ is adapted to $\{\mathcal{F}_n\}_{n=1}^\infty$

\[
\mathbb{E}[Y_+] = \mathbb{E}[\liminf (X_n)_+] \\
\leq \liminf \mathbb{E}[(X_n)_+] \\
\leq \mathbb{E}[(X_0)_+] \\
< \infty \quad \text{by Fatou’s lemma}
\]

\[
\mathbb{E}[Y_-] = \mathbb{E}[\liminf (X_n)_-] \\
\leq \liminf \mathbb{E}[(X_n)_-] \\
\leq \mathbb{E}[(X_0)_-] \\
< \infty \quad \text{by Jensen’s inequality}
\]

therefore $\mathbb{E}[Y] < \infty$ therefore $Y \in L^1$

Remark:

- The upcrossing lemma and martingale convergence theorem remain true for super martingales
- In general we do not have $L^1$ convergence

**Theorem: Convergence For Uniformly Integrable Martingales**

Suppose that $\{X_n\}_{n=1}^\infty$ is a uniformly integrable super-martingale.
Then $X_n$ converges $\mathbb{P}$-a.s. and in $L^1(\Omega, \mathcal{F}, \mathbb{P})$ to some r.v. $X_\infty \in L^1(\Omega, \mathcal{F}_\infty, \mathbb{P})$
and $\mathbb{E}[X_\infty|\mathcal{F}_n] \leq X_n$

**Proof:**

For a super-martingale $\{X_n\}_{n=1}^\infty$ by uniform integrability

\[
\lim_{k \to \infty} \sup_n \int_{\{|X_n| > k\}} |X_n|d\mathbb{P} = 0 \quad \text{and} \quad \lim_{k \to \infty} \int_{\{|X_n| > k\}} (X_n)_-d\mathbb{P} \leq \int |X_n|d\mathbb{P}
\]

\[
= \int_{\{|X_n| > k\}} |X_n|d\mathbb{P} + \int_{\{|X_n| \leq k\}} |X_n|d\mathbb{P}
\leq \int_{\{|X_n| > k\}} |X_n|d\mathbb{P} + k \int_{\{|X_n| \leq k\}} d\mathbb{P}
\leq \sup_n \int_{\{|X_n| > k\}} |X_n|d\mathbb{P} + k \mathbb{P}(|X_n| \leq k)
\leq \infty
\]

Then by the martingale convergence theorem $\lim_{n \to \infty} X_n = X_\infty$ $\mathbb{P}$-a.s.
and since a.s. convergence and uniform integrability ensure $L^1$ convergence we have convergence in $L^1(\Omega, \mathcal{F}, \mathbb{P})$
i.e. $\lim_{n \to \infty} \|X_\infty - X_n\| = 0$
\[ ||E[X_\infty|F_m] - E[X_n|F_m]|| = ||E[X_\infty - X_n|F_m]|| \leq ||X_\infty - X_n|| \]
\[ \lim_{m \to \infty} ||E[X_\infty|F_m] - E[X_n|F_m]|| = 0 \]

**Corollary:**
Suppose that \( \{X_n\}_{n=1}^\infty \) is a uniformly integrable sub-martingale.
Then \( X_n \) converges \( \mathbb{P} \)-a.s. and in \( L^1(\Omega, \mathcal{F}, \mathbb{P}) \) to some r.v. \( X_\infty \in L^1(\Omega, \mathcal{F}_\infty, \mathbb{P}) \) and \( E[X_\infty|F_n] \geq X_n \).

**Corollary:**
Suppose that \( \{X_n\}_{n=1}^\infty \) is a uniformly integrable martingale.
Then \( X_n \) converges \( \mathbb{P} \)-a.s. and in \( L^1(\Omega, \mathcal{F}, \mathbb{P}) \) to some r.v. \( X_\infty \in L^1(\Omega, \mathcal{F}_\infty, \mathbb{P}) \) and \( E[X_\infty|F_n] = X_n \).

**Theorem: Backwards Martingale Theorem**
Let \( (\Omega, \mathcal{F}, \mathbb{P}) \) be a probability space and \( \{G_{-n} : n \in \mathbb{N}\} \) sub-\( \sigma \)-algebras of \( \mathcal{F} \) s.t.
\[ G_\infty := \bigcap_{k \in \mathbb{N}} G_{-k} \subseteq G_{-(n+1)} \subseteq G_n \subseteq G_{-1} \]
For \( X \in L^1(\Omega, \mathcal{F}, \mathbb{P}) \) define
\[ X_{-n} := E[X|G_{-n}] \]
then \( \lim_{n \to \infty} X_{-n} \) exists \( \mathbb{P} \)-a.s. and in \( L^1 \).

**Theorem: Strong Law Of Large Numbers**
Let \( \{X_i\}_{i=1}^\infty \) be i.i.d r.v.s with \( E[|X_i|] < \infty, \mu = E[X_i] \) and \( S_n = \sum_{i=1}^n X_i \) then \( S_n/n \to \mu \) \( \mathbb{P} \)-a.s.

**Proof:**
Define \( G_{-n} := \sigma(\{S_i\}_{i=n}^\infty) \) and \( G_{-\infty} = \bigcap_{n=1}^\infty G_{-n} \)
then \( E[X_1|G_{-n}] = E[X_1|S_n] = S_n/n \)
since \( G_{-n} = \sigma(S_n,\{X_i\}_{i=n+1}^\infty) \) and independence of \( \{X_i\}_{i=1}^\infty \)
hence
\[ L := \lim_{n \to \infty} \frac{S_n}{n} \]
exists \( \mathbb{P} \)-a.s. and in \( L^1 \)
moreover
\[ L = \limsup_n \frac{\sum_{i=1}^n X_i}{n} = \limsup_n \frac{\sum_{i=k}^{n+k} X_i}{n} \]
is measurable w.r.t. \( \tau_k = \sigma(\{X_i\}_{i=k}^\infty) \)
hence is measurable w.r.t. \( \tau = \bigcap_{k=1}^n \tau_k \)
hence by Komogorov’s zero-one law \( \exists c \) s.t. \( \mathbb{P}(L = c) = 1 \)
moreover \( \mu := E[S_n/n] \) and \( \lim_{n \to \infty} E[S_n/n] = E[L] = c \)
hence indeed \( \mu \) must be the limit.

**Theorem: Central Limit Theorem**
Let \( \{X_k\}_{k=1}^\infty \in L^2 \) be i.i.d random variables
denote \( \mu := E[X_i], \nu := Var[X_i] \)
then
\[ S_n := \sum_{k=1}^n \frac{X_k - \mu}{\sqrt{\nu n}} \to N(0,1) \]
weakly

**Proof:**
WLOG let \( \mu = 0, \nu = 1 \) since otherwise we can use the transformation \( X = X + \sqrt{\nu}Z \) where \( Z \sim N(0,1) \) then
\[ \sum_{k=1}^n \frac{X_k - \mu}{\sqrt{n}} = \sqrt{n} \sum_{k=1}^n \frac{Z_k}{\sqrt{n}} \]
hence indeed it suffices to show for $\mu = 0, \nu = 1$

We need to show that $\forall f \in C_b(\mathbb{R})$ that

$$
\lim_{n \to \infty} E[f(S_n)] = \int \frac{e^{x^2/2} f(x)}{\sqrt{2\pi}} dx
$$

however any $f \in C_b(\mathbb{R})$ can be approximated by some $f \in C_b(\mathbb{R})$ which has bounded and uniformly continuous first and second derivatives.

Let $\{Y_k\}_{k=1}^\infty \sim i.i.d N(0, 1)$ be independent of $\{X_k\}_{k=1}^\infty$

then it suffices to show that for $T_n := \sum_{k=1}^n Y_k / \sqrt{n}$

we have that

$$
\lim_{n \to \infty} \left| E[f(S_n)] - E[f(T_n)] \right| = 0
$$

denote $X_{k,n} = X_k / \sqrt{n}, Y_{k,n} = Y_k / \sqrt{n}$

and define

$$
W_{k,n} = \sum_{j=1}^{k-1} Y_{j,n} + \sum_{j=k+1}^n X_{j,n}
$$

Then

$$
f(S_n) - f(T_n) = \sum_{k=1}^n f(W_{k,n} + X_{k,n}) - f(W_{k,n} + Y_{k,n})
$$

by a telescoping sum, hence by Taylor’s theorem we have

$$
\begin{align*}
&f(W_{k,n} + X_{k,n}) = f(W_{k,n}) + f'(W_{k,n})X_{k,n} + \frac{1}{2} f''(W_{k,n})X_{k,n}^2 + R_{X_{k,n}} \\
&f(W_{k,n} + Y_{k,n}) = f(W_{k,n}) + f'(W_{k,n})Y_{k,n} + \frac{1}{2} f''(W_{k,n})Y_{k,n}^2 + R_{Y_{k,n}} \\
&f(W_{k,n} + X_{k,n}) - f(W_{k,n} + Y_{k,n}) = f'(W_{k,n})(X_{k,n} - Y_{k,n}) + \frac{1}{2} f''(W_{k,n})(X_{k,n}^2 - Y_{k,n}^2) + R_{X_{k,n}} - R_{Y_{k,n}} \\
&\left| E[f(W_{k,n} + X_{k,n}) - f(W_{k,n} + Y_{k,n})] \right| \leq \\
&\left| E[f'(W_{k,n})(X_{k,n} - Y_{k,n})] \right| + E[|f''(W_{k,n})(X_{k,n}^2 - Y_{k,n}^2)/2|] + \left| E[R_{X_{k,n}}] + |R_{Y_{k,n}}| \right|
\end{align*}
$$

also from Taylor’s theorem we have

$$
|R_{X_{k,n}}| \leq X_{k,n}^2 ||f''||_{\infty}
$$

so by uniform continuity we have that

$\forall \varepsilon > 0 \exists \delta > 0$ s.t.

$$
|R_{X_{k,n}}| \leq X_{k,n}^2 \varepsilon \quad \forall |X_{k,n}| < \delta
$$

thus

$$
|R_{X_{k,n}}| \leq X_{k,n}^2 \varepsilon 1_{\{|X_{k,n}| < \delta\}} + X_{k,n}^2 ||f''||_{\infty} 1_{\{|X_{k,n}| \geq \delta\}}
$$
moreover

\[ E \left[ \sum_{k=1}^{n} |R_{X_{k,n}}| + |R_{Y_{k,n}}| \right] \]

\[ \leq \sum_{k=1}^{n} E[X_{k,n}^2 (\varepsilon 1_{\{X_{k,n} < \delta\}} + ||f''||_\infty 1_{\{X_{k,n} \geq \delta\}}) + Y_{k,n}^2 (\varepsilon 1_{\{Y_{k,n} < \delta\}} + ||f''||_\infty 1_{\{Y_{k,n} \geq \delta\}})] \]

\[ = \frac{1}{n} \sum_{k=1}^{n} E[X_{k,n}^2 (\varepsilon 1_{\{X_{k,n} < \delta\}} + ||f''||_\infty 1_{\{X_{k,n} \geq \delta\}}) + Y_{k,n}^2 (\varepsilon 1_{\{Y_{k,n} < \delta\}} + ||f''||_\infty 1_{\{Y_{k,n} \geq \delta\}})] \]

\[ \leq 2 \varepsilon E[X_1^2] + \frac{1}{n} \left( \sum_{k=1}^{n} E[X_{k,n}^2 ||f''||_\infty 1_{\{X_{k,n} \geq \delta\}}] + \sum_{k=1}^{n} E[Y_{k,n}^2 ||f''||_\infty 1_{\{Y_{k,n} \geq \delta\}}] \right) \]

\[ = 2 \varepsilon E[X_1^2] + 2E[X_1^2 ||f''||_\infty 1_{\{X_n \geq \delta \sqrt{n} \}}] \rightarrow_{n \to \infty} 0 \]

By the dominated convergence theorem,

it remains to show that the first two terms are null

\[ E[f'(W_{k,n})X_{k,n}] = E[f'(W_{k,n})]E[X_{k,n}] \]

by independence

\[ = 0 \]

\[ E[f'(W_{k,n})Y_{k,n}] = E[f'(W_{k,n})]E[Y_{k,n}] \]

by independence

\[ = 0 \]

\[ E[f''(W_{k,n})(X_{k,n}^2 - Y_{k,n}^2)] = E[f''(W_{k,n})]E[X_{k,n}^2 - Y_{k,n}^2] \]

\[ = E[f''(W_{k,n})] \left( \frac{1}{n} - \frac{1}{n} \right) \]

\[ = 0 \]

hence indeed the theorem holds.